Appendix E: GRADE profiles

Contents

Appendix E: GRADE profiles **1**
A.1 RQ B **4**
A.1.1 Approaches to sample collection in children unable to expectorate spontaneously 4
A.2 RQ C **16**
A.2.1 Diagnosis of active pulmonary tuberculosis in adults who are HIV-negative 16
A.2.2 Diagnosis of active pulmonary tuberculosis in adults who are HIV-positive 20
A.2.3 Diagnosis of active pulmonary tuberculosis in adults 27
A.3 RQ D **74**
A.3.1 Diagnosis of active pulmonary tuberculosis in children and young people who are HIV-negative 74
A.3.2 Diagnosis of active pulmonary tuberculosis in children and young people who are HIV-positive 76
A.3.3 Diagnosis of active pulmonary tuberculosis in children and young people 77
A.4 RQ G **89**
A.4.1 Diagnosis of active bone and joint tuberculosis 89
A.4.2 Diagnosis of active central nervous system tuberculosis 90
A.4.3 Diagnosis of active genitourinary tuberculosis 96
A.4.4 Diagnosis of active gastrointestinal tuberculosis 98
A.4.5 Diagnosis of active lymph node tuberculosis 101
A.4.6 Diagnosis of active pericardial tuberculosis 105
A.4.7 Diagnosis of active pleural tuberculosis 107
A.5 RQ I **116**
A.5.1 Dosing frequencies in children 116
A.6 RQ K **125**
A.6.1 People coinfected with tuberculosis and HIV 125
A.6.2 People with tuberculosis and liver disease 130
A.7 RQ L **131**
A.7.1 Duration of treatment in adults with respiratory tuberculosis 131
A.8 RQ M **162**
A.8.1 Duration of treatment in children with respiratory tuberculosis 162
A.9 RQs N and Q **163**
A.9.1 Use of adjunctive corticosteroids in people with active tuberculosis 163
A.10 RQ P **210**
A.10.1 Duration of treatment in people with non-respiratory tuberculosis 210
A.11 RQs O, R and X **229**
A.11.1 Adjunctive surgery in the treatment of active PULMONARY tuberculosis 229
A.11.2 Adjunctive surgery in the treatment of active ENDOBRONCHIAL tuberculosis 231

A.11.3 Adjunctive surgery in the treatment of active CHEST WALL tuberculosis 233
A.11.4 Adjunctive surgery in the treatment of active BONE AND JOINT tuberculosis ... 234
A.11.5 Adjunctive surgery in the treatment of active SPINAL tuberculosis 236
A.11.6 Adjunctive surgery in the treatment of active CENTRAL NERVOUS SYSTEM tuberculosis 254
A.11.7 Adjunctive surgery in the treatment of active GENITOURINARY tuberculosis ... 257
A.11.8 Adjunctive surgery in the treatment of active DRUG RESISTANT tuberculosis . 260
A.11.9 Adjunctive surgery in the treatment of active DRUG RESISTANT tuberculosis . 272
A.12 RQ S 272
A.12.1 Any resistance 272
A.12.2 First-line drug resistance 276
A.12.3 Isoniazid resistance 279
A.12.4 Rifampicin resistance 294
A.12.5 Multidrug resistance 297
A.12.6 International surveillance data 305
A.13 RQ U, V & W 306
A.14 RQ Z 315
A.14.1 Management of treatment interruptions 315
A.15 RQs AA and BB 319
A.15.1 Behrman 1998. Tuberculosis control in an urban emergency department 319
A.15.2 Blumerg et al. 1995. Preventing the nosocomial transmission of tuberculosis ... 320
A.15.3 Chamie et al 2013. Household ventilation and tuberculosis transmission in Kampala, Uganda 321
A.15.4 Da Costa 2009 322
A.15.5 Gonzalez-Angulo et al 2013. Knowledge and acceptability of patient specific infection control measures for TB 323
A.15.6 Hubad et al 2012. Inadequate hospital ventilation system increases the risk of nosocomial TB 324
A.15.7 Lygizos et al 2013. Natural ventilation reduces high TB transmission risk in traditional homes in rural KwaZulu-natal, SA 325
A.15.8 Nardell et al 2008. Safety of upper-room ultraviolet germicidal air disinfection for room occupants: results from the TB UV shelter study 326
A.15.9 Richardson 2014. Shared air: a renewed focus on ventilation for the prevention of tuberculosis transmission 327
A.16 RQ CC and DD 328
A.16.1 Duration of isolation to minimise risk of infection to others 328
A.16.2 Determining level of infectiousness – time to sputum smear conversion 330
A.16.3 Risk factors for continued risk of infection – time to culture conversion 333
A.17 RQ HH 336
A.18 RQ II 349

A.1 RQ B

A.1.1 Approaches to sample collection in children unable to expectorate spontaneously

Nasogastric aspiration/lavage vs induced sputum

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration / lavage	Induced sputum	Relative (95% CI)	Absolute	
Culture positivity by specimen (assessed with: number positive/total number of specimens obtained)											
4[1,2,3,4]	cross-sectional	very serious[5,6,7,8,9]	no serious inconsistency	no serious indirectness	no serious imprecision	none	286/3086 (9.3%)	224/2747 (8.2%)	OR 1.13 (0.94 to 1.36)[10,16]	1 more per 100 (from 0 fewer to 3 more)	⊙○○○ VERY LOW
Cumulative culture positivity: 2 specimens (assessed with: number of participants with 1 or more positive culture)											
2[3,11]	cross-sectional	very serious[5,7,8,12]	no serious inconsistency	no serious indirectness	serious[13]	none	142/420 (33.8%)	78/420 (18.6%)	OR 2.24 (1.63 to 3.09)[10,17]	15 more per 100 (from 9 more to 23 more)	⊙○○○ VERY LOW
Cumulative culture positivity: 3 specimens (assessed with: number of participants with 1 or more positive culture)											
2[4,11]	cross-sectional	very serious[5,7,8,12]	no serious inconsistency	no serious indirectness	serious[13]	none	46/267 (17.2%)	58/267 (21.7%)	OR 0.74 (0.48 to 1.15)[10,18]	5 fewer per 100 (from 10 fewer to 2 more)	⊙○○○ VERY LOW
Smear positivity by specimen (assessed with: number positive/total number of specimens obtained)											
3[1,3,4]	cross-sectional	very serious[5,6,7,8,9]	no serious inconsistency	no serious indirectness	serious[13]	none	53/1217 (4.4%)	42/869 (4.8%)	OR 0.99 (0.65 to 1.5)[10,19]	0 fewer per 100 (from 2 fewer to 2 more)	⊙○○○ VERY LOW
Cumulative smear positivity: 2 specimens (assessed with: number of participants with 1 or more positive smear)											
1[3]	cross-sectional	serious[5,7,8]	no serious inconsistency	no serious indirectness	serious[13]	none	42/403 (10.4%)	23/403 (5.7%)	OR 1.92 (1.13 to 3.26)[10]	5 more per 100 (from 1 more to 11 more)	⊙○○○ VERY LOW
Cumulative smear positivity: 3 specimens (assessed with: number of participants with 1 or more positive smear)											
2[4,11]	cross-sectional	very serious[5,7,8,12]	no serious inconsistency	no serious indirectness	serious[13]	none	18/267 (6.7%)	27/267 (10.1%)	OR 0.64 (0.34 to 1.2)[10,20]	3 fewer per 100 (from 6 fewer to 2 more)	⊙○○○ VERY LOW

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration / lavage	Induced sputum	Relative (95% CI)	Absolute	
Cumulative PCR positivity: 3 specimens (assessed with: number of participants with 1 or more positive PCR)											
1[11]	cross-sectional	very serious[5,7,8,12]	no serious inconsistency	no serious indirectness	very serious[13,14]	none	2/17 (11.8%)	3/17 (17.6%)	OR 0.62 (0.09 to 4.29)[10]	6 fewer per 100 (from 16 fewer to 30 more)	⊙○○○ VERY LOW
Culture positivity by specimen (subgroup: <5 years) (assessed with: number positive/total number of specimens obtained)											
2[2,4]	cross-sectional	serious[5,7,8]	no serious inconsistency	serious[15]	no serious imprecision	none	146/2119 (6.9%)	145/2119 (6.8%)	OR 1.01 (0.79 to 1.28)[10,21]	0 more per 100 (from 1 fewer to 2 more)	⊙○○○ VERY LOW
Cumulative culture positivity: 3 specimens (subgroup: <5 years) (assessed with: number of participants with 1 or more positive culture)											
1[4]	cross-sectional	serious[5,7,8]	no serious inconsistency	no serious indirectness	serious[13]	none	38/250 (15.2%)	51/250 (20.4%)	OR 0.70 (0.44 to 1.11)[10]	5 fewer per 100 (from 10 fewer to 2 more)	⊙○○○ VERY LOW
Smear positivity by specimen (subgroup: <5 years) (assessed with: number positive/total number of specimens obtained)											
1[4]	cross-sectional	serious[5,7,8]	no serious inconsistency	no serious indirectness	serious[13]	none	8/250 (3.2%)	19/250 (7.6%)	OR 0.40 (0.17 to 0.94)[10]	4 fewer per 100 (from 0 fewer to 6 fewer)	⊙○○○ VERY LOW
Cumulative smear positivity: 3 specimens (subgroup: <5 years) (assessed with: number of participants with 1 or more positive smear)											
1[4]	cross-sectional	serious[5,7,8]	no serious inconsistency	no serious indirectness	serious[13]	none	17/250 (6.8%)	25/250 (10%)	OR 0.66 (0.35 to 1.25)[10]	3 fewer per 100 (from 6 fewer to 2 more)	⊙○○○ VERY LOW

[1] *Al-Aghbari, 2009*
[2] *Hatherill, 2009*
[3] *Mukherjee, 2013*
[4] *Zar, 2005*
[5] *Unclear if a random or consecutive sample was used*
[6] *Study did not obtain samples from all included participants (Al-Aghbari, 2009)*
[7] *Blinding of individuals administering care (all studies) and investigators unclear (Hatherhill, 2009; Jimenez, 2013; Zar, 2005); blinding of participants not stated, but unlikely given the nature of the interventions*
[8] *Precise criteria for positivity is not stated*
[9] *Unclear if there was an appropriate interval between the 2 collection techniques (Al-Aghbari, 2009)*
[10] *Calculated by reviewer*
[11] *Jiménez, 2013*
[12] *Inappropriate exclusions - excluded participants positive for non-tuberculous mycobacteria (Jiménez, 2013)*
[13] *GRADE rule of thumb: <300 events*
[14] *Wide confidence interval*
[15] *Population of Hatherhill (2009) is mostly below 5 years of age, but some over 5s are likely to have been included*

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration / lavage	Induced sputum	Relative (95% CI)	Absolute	
[16] *Forest plot (culture positivity by specimen):*											
[17] *Forest plot (cumulative culture positivity: 2 specimens):*											
[18] *Forest plot (cumulative culture positivity: 3 specimens):*											
[19] *Forest plot (smear positivity by specimen):*											
[20] *Forest plot (cumulative smear positivity: 3 specimens):*											
[21] *Forest plot (culture positivity by specimen; subgroup: <5 years)*											

Nasogastric aspiration/lavage vs induced or spontaneously produced sputum

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration / lavage	Induced or spontaneously produced sputum	Relative (95% CI)	Absolute	Quality
Culture positivity (assessed with: number of participants to be considered culture-positive)											
1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	5/67 (7.5%)	7/67 (10.4%)	OR 0.69 (0.21 to 2.3)[7]	3 fewer per 100 (from 8 fewer to 11 more)	⊙○○○ VERY LOW

[1] *Thomas, 2014*
[2] *Unclear if a random or consecutive sample was used*
[3] *Use of blinding unclear*
[4] *Precise criteria for positivity is not stated*
[5] *Comparator includes spontaneously produced sputum (not the comparator of interest)*
[5] *GRADE rule of thumb: <300 events*
[6] *Calculated by reviewer*

Nasopharyngeal aspiration vs induced sputum

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasopharyngeal aspiration	Induced sputum	Relative (95% CI)	Absolute	Quality
Culture positivity by specimen (assessed with: number positive/total number of specimens obtained)											
3[1,2,3]	cross-sectional	very serious[4,5,6,7,8]	no serious inconsistency	no serious indirectness	serious[9]	none	96/823 (11.7%)	134/839 (16%)	OR 0.69 (0.52 to 0.91)[10,12]	4 fewer per 100 (from 1 fewer to 7 fewer)	⊙○○○ VERY LOW
Smear positivity by specimen (assessed with: number positive/total number of specimens obtained)											
3[1,2,3]	cross-sectional	very serious[4,5,6,7,8]	no serious inconsistency	no serious indirectness	serious[9]	none	75/829 (9%)	86/845 (10.2%)	OR 0.86 (0.62 to 1.19)[10,13]	1 fewer per 100 (from 4 fewer to 2 more)	⊙○○○ VERY LOW
Culture positivity by specimen (subgroup: <5 years) (assessed with: number positive/total number of specimens obtained)											
1[3]	cross-sectional	serious[4,6,7]	no serious inconsistency	serious[11]	serious[9]	none	61/535 (11.4%)	84/535 (15.7%)	OR 0.69 (0.49 to 0.98)[10]	4 fewer per 100 (from 0 fewer to 7 fewer)	⊙○○○ VERY LOW

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasopharyngeal aspiration	Induced sputum	Relative (95% CI)	Absolute	Quality
Smear positivity by specimen (subgroup: <5 years) (assessed with: number positive/total number of specimens obtained)											
1[3]	cross-sectional	serious[4,6,7]	no serious inconsistency	serious[11]	serious[9]	none	57/535 (10.7%)	69/535 (12.9%)	OR 0.81 (0.55 to 1.17)[10]	2 fewer per 100 (from 5 fewer to 2 more)	⊙○○○ VERY LOW

[1] *Al-Aghbari, 2009*
[2] *Owens, 2007*
[3] *Zar, 2012*
[4] *Unclear if a random or consecutive sample was used*
[5] *Study did not obtain samples from all included participants (Al-Aghbari, 2009)*
[6] *Blinding of individuals administering care (all studies) and investigators (Owens, 2007; Zar, 2012) unclear; blinding of participants not stated, but unlikely given the nature of the interventions*
[7] *Precise criteria for positivity is not stated*
[8] *Unclear if there was an appropriate interval between the 2 collection techniques (Al-Aghbari, 2009)*
[9] *GRADE rule of thumb: <300 events*
[10] *Calculated by reviewer*
[11] *Population is mostly below 5 years of age, but some over 5s are likely to have been included*
[12] *Forest plot (culture positivity by specimen):*

[13] *Forest plot (smear positivity by specimen):*

Nasopharyngeal aspiration vs nasogastric aspiration/lavage

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasopharyngeal aspiration	Nasogastric aspiration/lavage	Relative (95% CI)	Absolute	Quality
Culture positivity by specimen (assessed with: number positive/total number of specimens obtained)											
3[1,2,3]	cross-sectional	very serious[4,5,6,7,8]	no serious inconsistency	no serious indirectness	serious[9]	none	34/729 (4.7%)	82/1101 (7.4%)	OR 0.68 (0.45 to 1.04)[10,11]	2 fewer per 100 (from 4 fewer to 0 more)	⊙○○○ VERY LOW
Smear positivity by specimen (assessed with: number positive/total number of specimens obtained)											
2[1,2]	cross-sectional	very serious[4,5,6,7,8]	no serious inconsistency	no serious indirectness	serious[9]	none	14/514 (2.7%)	25/885 (2.8%)	OR 1.12 (0.58 to 2.18)[10,12]	0 more per 100 (from 1 fewer to 3 more)	⊙○○○ VERY LOW
PCR positivity by specimen (assessed with: number positive/total number of specimens obtained)											
1[3]	cross-sectional	serious[4,6,7]	no serious inconsistency	no serious indirectness	serious[9]	none	26/218 (11.9%)	35/217 (16.1%)	OR 0.70 (0.41 to	4 fewer per 100	⊙○○○ VERY

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasopharyngeal aspiration	Nasogastric aspiration/lavage	Relative (95% CI)	Absolute	Quality
									1.22)[10]	(from 9 fewer to 3 more)	LOW

[1] *Al-Aghbari, 2009*
[2] *Oberhelman, 2006*
[3] *Oberhelman, 2010*
[4] *Unclear if a random or consecutive sample was used*
[5] *Study did not obtain samples from all included participants (Al-Aghbari, 2009)*
[6] *Blinding of individuals administering care (all studies) and investigators (Oberhelman, 2006; Oberhelman, 2010) unclear; blinding of participants not stated, but unlikely given the nature of the interventions*
[7] *Precise criteria for positivity is not stated*
[8] *Unclear if there was an appropriate interval between the 2 collection techniques (Al-Aghbari, 2009)*
[9] *GRADE rule of thumb: <300 events*
[10] *Calculated by reviewer*
[11] *Forest plot (culture positivity by specimen):*

[12] *Forest plot (smear positivity by specimen):*

Nasogastric aspiration/lavage vs bronchoalveolar lavage

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Bronchoalveolar lavage	Relative (95% CI)	Absolute	Quality
Culture positivity (assessed with: number of participants with 1 or more positive culture (cumulative yield for 3 GA specimens vs 1 BAL specimen))											
3[1,2,3]	cross-sectional	serious[4,5,6,7]	no serious inconsistency	no serious indirectness	serious[8]	none	76/273 (27.8%)	59/273 (21.6%)	OR 1.41 (0.95 to 2.1)[9,12]	6 more per 100 (from 1 fewer to 15 more)	⊙○○○ VERY LOW
Smear positivity (assessed with: number of participants with a positive smear (1 GA specimen vs 1 BAL specimen))											
1[10]	cross-sectional	serious[4,6,7]	no serious inconsistency	no serious indirectness	serious[8]	none	6/52 (11.5%)	16/52 (30.8%)	OR 0.29 (0.1 to 0.83)[9]	19 fewer per 100 (from 4 fewer to 27 fewer)	⊙○○○ VERY LOW
Smear positivity (subgroup: <5 years) (assessed with: number of participants with a positive smear (cumulative yield for 3 GA specimens vs 1 BAL specimen))											
1[10]	cross-sectional	serious[4,6,7]	no serious inconsistency	no serious indirectness	serious[8]	none	0/20 (0%)	0/20 (0%)	OR 1.00 (0.02 to 52.85)[9]	-	⊙○○○ VERY LOW
Volume of specimen (subgroup: <5 years) (measured with: mean volume of specimens obtained; better indicated by higher values)											
1[10]	cross-sectional	serious[4,6,7]	no serious inconsistency	no serious indirectness	serious[11]	none	20 mean (range) = 35 (20–55) ml	20 mean (range) = 56.5 (45 to 80)	-	MD 21.5 higher[9]	⊙○○○ VERY LOW
Need for anaesthesia (subgroup: <5 years) (assessed with: number of participants that required topical anaesthesia)											

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Bronchoalveolar lavage	Relative (95% CI)	Absolute	Quality
1[10]	cross-sectional	serious[4,6,7]	no serious inconsistency	no serious indirectness	serious[8]	none	0/20 (0%)	2/20 (10%)	OR 0.18 (0.01 to 4.01)[9]	8 fewer per 100 (from 10 fewer to 21 more)	⊙○○○ VERY LOW

[1] *Cakir, 2008*
[2] *Cakir, 2013*
[3] *Chan, 1994*
[4] *Unclear if studies made inappropriate exclusions*
[5] *Unclear if if there was an appropriate interval between specimen collections (Cakir, 2008; Cakir, 2013)*
[6] *Blinding of individuals administering care and investigators unclear; blinding of participants not stated, but unlikely given the nature of the interventions*
[7] *Precise criteria for positivity is not stated*
[8] *GRADE rule of thumb: <300 events*
[9] *Calculated by reviewer*
[10] *Abadco, 1992*
[11] *Insufficient data available to appraise imprecision*
[12] *Forest plot (culture positivity):*

Nasopharyngeal aspiration vs bronchoalveolar lavage

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasopharyngeal aspiration	Bronchoalveolar lavage	Relative (95% CI)	Absolute	Quality
Culture positivity (assessed with: number of participants with a positive culture)											
1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	none	16/50 (32%)	6/50 (12%)	OR 3.45 (1.22 to 9.76)[6]	20 more per 100 (from 2 more to 45 more)	⊙○○○ VERY LOW

[1] *Somu, 1995*
[2] *Unclear if studies made inappropriate exclusions*
[3] *Blinding of individuals administering care and investigators unclear; blinding of participants not stated, but unlikely given the nature of the interventions*
[4] *Precise criteria for positivity is not stated*
[5] *GRADE rule of thumb: <300 events*
[6] *Calculated by reviewer*

Nasogastric aspiration/lavage vs laryngeal swab

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Laryngeal swab	Relative (95% CI)	Absolute	Quality
Cumulative culture positivity: 3 specimens (assessed with: number of participants with 1 or more positive culture)											
2[1,2]	cross-	serious[3,4,5,6]	no serious	no serious	serious[7]	none	20/90	42/90	OR 0.29	26 fewer	⊙○○○

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Laryngeal swab	Relative (95% CI)	Absolute	
	sectional		inconsistency	indirectness			(22.2%)	(46.7%)	(0.14 to 0.57)[8,10]	per 100 (from 13 fewer to 36 fewer)	VERY LOW
Cumulative smear positivity: 3 specimens (assessed with: number of participants with 1 or more positive smear)											
1[1]	cross-sectional	serious[3,4,5,6]	no serious inconsistency	no serious indirectness	serious[7]	none	4/30 (13.3%)	6/30 (20%)	OR 0.58 (0.14 to 2.50)[8]	7 fewer per 100 (from 17 fewer to 18 more)	⊙○○○ VERY LOW
Cumulative culture positivity: 3 specimens (subgroup: <5 years) (assessed with: number of participants with 1 or more positive culture)											
2[1,2]	cross-sectional	serious[3,4,5,6]	no serious inconsistency	serious[9]	serious[7]	none	20/77 (26%)	41/77 (53.2%)	OR 0.29 (0.15 to 0.59)[8,11]	28 fewer per 100 (from 13 fewer to 39 fewer)	⊙○○○ VERY LOW
Cumulative smear positivity: 3 specimens (<5 years) (assessed with: number of participants with 1 or more positive smear)											
1[1]	cross-sectional	serious[3,4,5,6]	no serious inconsistency	no serious indirectness	serious[7]	none	3/17 (17.6%)	4/17 (23.5%)	OR 0.70 (0.13 to 3.72)[8]	6 fewer per 100 (from 20 fewer to 30 more)	⊙○○○ VERY LOW
Cumulative culture positivity: 3 specimens (subgroup: >5 years) (assessed with: number of participants with 1 or more positive culture)											
1[1]	cross-sectional	serious[3,4,5,6]	no serious inconsistency	no serious indirectness	serious[7]	none	0/13 (0%)	1/13 (7.7%)	OR 0.31 (0.01 to 8.30)[8]	5 fewer per 100 (from 8 fewer to 33 more)	⊙○○○ VERY LOW
Cumulative smear positivity: 3 specimens (subgroup: >5 years) (assessed with: number of participants with 1 or more positive smear)											
1[1]	cross-sectional	serious[3,4,5,6]	no serious inconsistency	no serious indirectness	serious[7]	none	1/13 (7.7%)	2/13 (15.4%)	OR 0.46 (0.04 to 5.79)[8]	8 fewer per 100 (from 15 fewer to 36 more)	⊙○○○ VERY LOW

[1] *Bhandari, 1976*
[2] *Lloyd, 1968*
[3] *Unclear if a random or consecutive sample was used*
[4] *Blinding of individuals administering care and investigators unclear; blinding of participants not stated, but unlikely given the nature of the interventions*
[5] *Precise criteria for positivity is not stated*
[6] *Unclear if exclusions were appropriate*
[7] *GRADE rule of thumb: <300 events*
[8] *Calculated by reviewer*
[9] *Population of Lloyd (1968) is >6 years of age as opposed to 5*
[10] *Forest plot (cumulative culture positivity: 3 specimens):*

[11] *Forest plot (cumulative culture positivity: 3 specimens; subgroup: <5 years):*

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Laryngeal swab	Relative (95% CI)	Absolute	Quality

Nasogastric aspiration/lavage vs lung puncture aspiration

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Lung puncture aspiration	Relative (95% CI)	Absolute	Quality
Cumulative culture positivity: 3 specimens (assessed with: number of participants with 1 or more positive culture)											
1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	none	3/30 (10%)	16/30 (53.3%)	OR 0.10 (0.02 to 0.39)[7]	43 fewer per 100 (from 23 fewer to 51 fewer)	⊙○○○ VERY LOW
Cumulative smear positivity: 3 specimens (assessed with: number of participants with 1 or more positive smear)											
1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	none	4/30 (13.3%)	5/30 (16.7%)	OR 0.77 (0.19 to 3.20)[7]	3 fewer per 100 (from 13 fewer to 22 more)	⊙○○○ VERY LOW
Cumulative culture positivity: 3 specimens (subgroup: <5 years) (assessed with: number of participants with 1 or more positive culture)											
1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	none	3/17 (17.6%)	10/17 (58.8%)	OR 0.15 (0.03 to 0.73)[7]	41 fewer per 100 (from 8 fewer to 55 fewer)	⊙○○○ VERY LOW
Cumulative smear positivity: 3 specimens (subgroup: <5 years) (assessed with: number of participants with 1 or more positive smear)											
1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	none	3/17 (17.6%)	4/17 (23.5%)	OR 0.70 (0.13 to 3.72)[7]	6 fewer per 100 (from 20 fewer to 30 more)	⊙○○○ VERY LOW
Cumulative culture positivity: 3 specimens (subgroup: >5 years) (assessed with: number of participants with 1 or more positive culture)											
1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	none	0/13 (0%)	6/13 (46.2%)	OR 0.04 (0.00 to 0.87)[7]	43 fewer per 100 (from 3 fewer to 46 fewer)	⊙○○○ VERY LOW
Cumulative smear positivity: 3 specimens (subgroup: >5 years) (assessed with: number of participants with 1 or more positive smear)											
1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	none	1/13 (7.7%)	1/13 (7.7%)	OR 1.00 (0.06 to 17.90)[7]	0 fewer per 100 (from 7 fewer to 52 more)	⊙○○○ VERY LOW

[1] *Bhandari, 1976*

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Lung puncture aspiration	Relative (95% CI)	Absolute	Quality

[2] *Unclear if a random or consecutive sample was used*
[3] *Blinding of individuals administering care and investigators unclear; blinding of participants not stated, but unlikely given the nature of the interventions*
[4] *Precise criteria for positivity is not stated*
[5] *Unclear if exclusions were appropriate*
[6] *GRADE rule of thumb: <300 events*
[7] *Calculated by reviewer*

Suctioned vs coughed induced sputum

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Suctioned induced sputum	Coughed induced sputum	Relative (95% CI)	Absolute	Quality
Culture positivity by specimen (assessed with: number positive/total number of specimens obtained)											
1[1]	observational	very serious[2,3,4]	serious[5]	no serious indirectness	serious[6]	none	129/993 (13%)	62/264 (23.5%)	OR 0.49 (0.35 to 0.68)[7]	10 fewer per 100 (from 6 fewer to 14 fewer)	⊙○○○ VERY LOW
Adverse events - none (assessed with: number of procedures completed without adverse events)											
1[1]	observational	very serious[2,3,4]	serious[5]	no serious indirectness	no serious imprecision	none	744/993 (74.9%)	259/264 (98.1%)	OR 0.06 (0.02 to 0.14)[7]	22 fewer per 100 (from 10 fewer to 47 fewer)	⊙○○○ VERY LOW
Adverse events – nose bleed (assessed with: number of procedures in which nose bleed occurred)											
1[1]	observational	very serious[2,3,4]	serious[5]	no serious indirectness	no serious imprecision	none	239/993 (24.1%)	4/264 (1.5%)	OR 20.60 (7.59 to 55.90)[7]	23 more per 100 (from 9 more to 45 more)	⊙○○○ VERY LOW
Adverse events – wheeze (assessed with: number of procedures that led to wheezing)											
1[1]	observational	very serious[2,3,4]	serious[5]	no serious indirectness	no serious imprecision	none	11/993 (1.1%)	3/264 (1.1%)	OR 0.97 (0.27 to 3.52)[7]	0 fewer per 100 (from 1 fewer to 3 more)	⊙○○○ VERY LOW
Adverse events – exacerbation of cough (assessed with: number of procedures that led to exacerbation of cough)											
1[1]	observational	very serious[2,3,4]	serious[5]	no serious indirectness	no serious imprecision	none	3/993 (0.0%)	1/264 (0.0%)	OR 0.80 (0.08 to 7.69)[7]	0 fewer per 100 (from 0 fewer to 2 more)	⊙○○○ VERY LOW

[1] *Planting, 2014*
[2] *Allocation connected to a potentially confounding factor - based on child's ability to spontaneously produce sputum*
[3] *Blinding of individuals administering care and investigators unclear; blinding of participants not stated, but unlikely given the nature of the interventions*

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Suctioned induced sputum	Coughed induced sputum	Relative (95% CI)	Absolute	Quality
[4] *Precise criteria for positivity is not stated* [5] *Unclear if groups were comparable at baseline* [6] *GRADE rule of thumb: <300 events* [7] *Calculated by reviewer*											

Nasogastric aspiration/lavage with nebulisation vs nasogastric aspiration/lavage alone

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage with nebulisation	Nasogastric aspiration/lavage alone	Relative (95% CI)	Absolute	Quality
Culture positivity (assessed with: number of participants with a positive culture)											
1[1]	randomised trial	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	none	9/36 (25%)	24/68 (35.3%)	OR 1.29 (0.49 to 3.35)[7]	6 more per 100 (from 14 fewer to 29 more)	⊙○○○ VERY LOW
Volume of specimen (measured with: mean volume of specimens obtained; better indicated by higher values)											
1[1]	randomised trial	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[8]	none	36 mean = 25 ml	68 mean = 10 ml	-	MD 15 higher[7]	⊙○○○ VERY LOW
[1] *Maciel, 2010* [2] *Unclear if a random or consecutive sample was used* [3] *Blinding of individuals administering care and investigators unclear; blinding of participants not stated, but unlikely given the nature of the interventions* [4] *Precise criteria for positivity is not stated* [5] *Unclear if exclusions were appropriate* [6] *GRADE rule of thumb: <300 events* [7] *Calculated by reviewer* [8] *Insufficient data available to appraise imprecision*											

Nasogastric aspiration/lavage with sedation vs nasogastric aspiration/lavage with placebo

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Induced sputum	Relative (95% CI)	Absolute	Quality
Acceptability of the procedure to parents – usefulness of the sedation (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 10 (10–10)	Median (range) = 5 (3–7)	-	Difference in medians = 5[4]	⊙⊙○○ LOW
Acceptability of the procedure to parents – impact on child's outlook (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised	serious[2]	no serious	no serious	serious[3]	none	Median (range) =	Median (range) =	-	Difference	⊙⊙○○

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Induced sputum	Relative (95% CI)	Absolute	Quality
	trial		inconsistency	indirectness			8.9 (7–10)	5.8 (5–7)		in medians = 3.1[4]	LOW
Acceptability of the procedure to parents – impact on parents' outlook (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 9.1 (8–10)	Median (range) = 4.9 (3–7)	-	Difference in medians = 4.2[4]	⊙⊙○○ LOW
Acceptability of the procedure to parents – child's tolerance of procedures (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 8.7 (7–10)	Median (range) = 8.5 (7–10)	-	Difference in medians = 0.2[4]	⊙⊙○○ LOW
Acceptability of the procedure to parents – would recommend to other parents (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 9.3 (9–10)	Median (range) = 4 (3–6)	-	Difference in medians = 5.3[4]	⊙⊙○○ LOW
Acceptability of the procedure to parents – would like to see the mucosal atomizer device used routinely (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 9.8 (9–10)	Median (range) = 4 (3–6)	-	Difference in medians = 5.8[4]	⊙⊙○○ LOW
Acceptability of the procedure to clinicians – usefulness of the sedation (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 10 (10–10)	Median (range) = 3 (2–4)	-	Difference in medians = 7[4]	⊙⊙○○ LOW
Acceptability of the procedure to clinicians – impact on child's outlook (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 8 (7–9)	Median (range) = 3 (2–4)	-	Difference in medians = 5[4]	⊙⊙○○ LOW
Acceptability of the procedure to clinicians – impact on clinician's outlook (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 9.5 (9–10)	Median (range) = 4 (3–5)	-	Difference in medians = 5.5[4]	⊙⊙○○ LOW

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Nasogastric aspiration/lavage	Induced sputum	Relative (95% CI)	Absolute	Quality
Acceptability of the procedure to clinicians – child's tolerance of procedures (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 8.2 (7–9)	Median (range) = 8 (7–9)	-	Difference in medians = [4]	⊙⊙○○ LOW
Acceptability of the procedure to clinicians – would recommend to other clinicians (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 9.4 (9–10)	Median (range) = 3 (1–5)	-	Difference in medians = 6.4[4]	⊙⊙○○ LOW
Acceptability of the procedure to clinicians – would like to see the mucosal atomizer device used routinely (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 10 (10–10)	Median (range) = 3 (1–5)	-	Difference in medians = 7[4]	⊙⊙○○ LOW
Acceptability of the procedure to clinicians – made the procedure more acceptable (assessed with: score derived from questionnaire, answered using a visual analogue scale ('0' for worst, '10' for best); better indicated by higher scores)											
1[1]	randomised trial	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	Median (range) = 10 (10–10)	Median (range) = 3 (1–5)	-	Difference in medians = 7[4]	⊙⊙○○ LOW

[1] *Buonsenso, 2014*
[2] *blinding of participants not stated, but unlikely given the nature of the interventions*
[3] *Insufficient data available to appraise imprecision*
[4] *Calculated by reviewer*

A.2 RQ C

A.2.1 Diagnosis of active pulmonary tuberculosis in adults who are HIV-negative

Commercial nucleic acid amplification techniques compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-negative

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Sensitivity[1]										

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Xpert MTB/RIF only	18	cross-sectional	no serious risk of bias[2]	serious[3]	serious[3]	serious[5]	Limited industry involvement All except 1 study conducted in a high incidence country[6]	2555	91.4% (95% CI 87.5 to 94.2%)	VERY LOW
Specificity[1]										
Xpert MTB/RIF only	18	cross-sectional	no serious risk of bias[2]	serious[3]	serious[3]	no serious imprecision	Limited industry involvement All except 1 study conducted in a high incidence country[6]	2555	99.5% (95% CI 98.6 to 99.8%)	LOW

[1] Forest plots for sensitivity and specificity (Xpert MTB/RIF assay):

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two

[3] Reference standard varied widely from study to study, using different culture techniques and in some cases employing a number of additional reference criteria (e.g. clinical characteristics or smear status)

[4] Wide confidence intervals

[5] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap

[6] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Use of antituberculosis antibodies to detect tuberculosis in urine compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-negative

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
LAM	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	no serious imprecision	No information available on industry involvement	397	52% 95% CI (43 to 62%)	MODERATE

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
							Conducted in a high incidence country[6]			
Specificity[1]										
LAM	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	no serious imprecision	No information available on industry involvement Conducted in a high incidence country[6]	397	86% (95% CI 77 to 93%)	MODERATE

[1] Mutetwa, 2009
[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Unclear if a consecutive or random sample used
[4] Unlcear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Interferon-gamma release assays compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-negative

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Sensitivity										
Interferon-gamma release assays[7]	2	cross-sectional	serious[2,3,4]	no serious inconsistency	serious[5,6]	no serious imprecision	Test kits supplied by industry Conducted in a high incidence country[8]	275	90.6% (95% CI 84.2 to 94.6%)	LOW
QuantiFERON-TB Gold	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	serious[5,6]	no serious imprecision	Test kits supplied by industry Conducted in a	138	89.2% (95% CI 81.7 to 96.8%)	LOW

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
							high incidence country[8]			
T-SPOT.TB	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	serious[5,6]	no serious imprecision	Test kits supplied by industry Conducted in a high incidence country[8]	137	92.2% (95% CI 85.6 to 98.8%)	LOW
Specificity										
QuantiFERON-TB Gold	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	serious[5,6]	no serious imprecision	Test kits supplied by industry Conducted in a high incidence country[8]	138	49.3% (95% CI 37.9 to 60.8%)	LOW
T-SPOT.TB	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	serious[5,6]	no serious imprecision	Test kits supplied by industry Conducted in a high incidence country[8]	137	46.6% (95% CI 35.1 to 58.0%)	LOW

[1] Kang, 2007

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two

[3] Inappropriate exclusions were not avoided – excluded patients with high clinical likelihood of active TB and a negative mycobacterial culture finding but good clinical and radiographic responses to anttuberculosis treatment

[4] Unclear if test interpretation was blinded

[5] Unclear how many participants, if any, were under 18 years old; however, it is not anticipated that the results will be significantly affected by this

[6] Reference standard included histology as an alternative to culture

[7] QuantiFERON-TB Gold and T-SPOT.TB

[8] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Tuberculin skin tests compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-negative

Test	Number of	Quality assessment						Number of	Summary of	Quality

details	evaluations	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	patients/ specimens	findings	
Sensitivity										
Mantoux	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	serious[5,6]	serious[7]	Test kits for IGRA component of trial supplied by industry Conducted in a high incidence country[8]	141	68.2% (95% CI 56.9 to 79.4%)	VERY LOW
Specificity										
Mantoux	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	serious[5,6]	serious[7]	Test kits for IGRA component of trial supplied by industry Conducted in a high incidence country[8]	141	50.7% (95% CI 39.4 to 62.0%)	VERY LOW

[1] Kang, 2007

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two

[3] Inappropriate exclusions were not avoided – excluded patients with high clinical likelihood of active TB and a negative mycobacterial culture finding but good clinical and radiographic responses to anttuberculosis treatment

[4] Unclear if test interpretation was blinded

[5] Unclear how many participants, if any, were under 18 years old; however, it is not anticipated that the results will be significantly affected by this

[6] Reference standard included histology as an alternative to culture

[7] Wide confidence interval

[8] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

A.2.2 Diagnosis of active pulmonary tuberculosis in adults who are HIV-positive

Smear microscopy compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-positive

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[2]										
All techniques	3[1,2]	cross-sectional[3]	no serious risk of bias	serious[4]	no serious indirectness[7]	serious[9]	No industry involvement All except 1 study conducted in a high incidence country[10]	1094	40.8% (95% CI 18.6 to 67.6%)	MODERATE
Fluorescence microscopy	Chaidir, 2013[1]	cross-sectional[3]	serious[5,6]	serious[4]	no serious indirectness[7]	no serious imprecision	No industry involvement Conducted in a high incidence country[10]	256	65.2% (95% CI 59.4 to 71.0%)	LOW
Ziehl-Neelson microscopy	Chaidir, 2013[1]	cross-sectional[3]	serious[5,6]	serious[4]	no serious indirectness[7]	no serious imprecision	No industry involvement Conducted in a high incidence country[10]	256	58.0% (95% CI 52.0 to 64.0%)	LOW
Specificity[11]										
Fluorescence microscopy	Lawn, 2011 Lawn, 2012 Chaidir, 2013	cross-sectional[3]	serious[5,6]	serious[4]	no serious indirectness[7]	no serious imprecision	No industry involvement Conducted in high incidence countries[10]	445 516 256	100% (95% CI 100 to 100%) 99.8% (95% CI 99.3 to 100%) 90.4% (95% CI 86.8 to 94.0%)	LOW
Ziehl-Neelson microscopy	Carriquiry, 2012 Chaidir, 2013	cross-sectional[3]	serious[5,6]	serious[4]	no serious indirectness[7]	no serious imprecision	No industry involvement All except 1 study conducted in a high incidence country[10]	133 256	96.6% (95% CI 92.8 to 100%) 96.3% (95% CI 94.0 to 98.6%)	LOW

[1] Insufficient data provided to use Chaidir (2013) in the meta-analysis

[2] Forest plots for sensitivity and specificity (grouped by technique used):

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				

[3] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[4] Reference standard varies across studies: culture technique not consistent
[5] Unclear if a consecutive or random sample of participants used in Chaidir (2013)
[6] Unclear if inappropriate exclusions were avoided in Chaidir (2013)
[7] Chaidir (2013) provide no details of the age of the study population; however, it is not anticipated that the results will be significantly affected by this
[8] Wide confidence interval
[9] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap
[10] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000
[11] Meta-analysis of relevant data not possible in STATA or R

Microscopy, chest radiography and symptoms compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-positive

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
Microscopy, chest x-ray plus symptoms[1]	1[2]	cross-sectional[3]	serious[4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement Conducted in a high incidence country[8]	445	53.7% (95% CI 40.4 to 67.0%)	MODERATE
Specificity										
Microscopy, chest x-ray plus symptoms[1]	1[2]	cross-sectional[3]	serious[4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement Conducted in a high incidence country[8]	445	76.2% (95% CI 72.0 to 80.4%)	MODERATE

[1] Any 1 of the following 4 symptoms: cough, fever, weight loss and night sweats
[2] Swindells, 2013

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				

[3] Both index test and reference standard performed in every patient, with an appropriate period of time between the two

[4] Unclear if inappropriate exclusions were avoided

[5] Unclear if test interpretation was blinded in all or most of the included comparisons

[6] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons

[7] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this

[8] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Commercial nucleic acid amplification techniques compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-positive

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Xpert MTB/RIF	16	cross-sectional	no serious risk of bias[4]	serious[5]	serious[5]	serious[7]	Limited industry involvement All except 2 studies conducted in a high incidence country[8]	2990	80.9% (95% CI 72.9 to 86.9%)	VERY LOW
Specificity[1]										
Xpert MTB/RIF	16	cross-sectional	no serious risk of bias[4]	serious[5]	serious[5]	no serious imprecision	Limited industry involvement All except 2 studies conducted in a high incidence country[8]	2990	98.8% (95% CI 97.8 to 99.4%)	LOW

[1] Forest plots for sensitivity and specificity (Xpert MTB/RIF assay):

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				

[4] Both index test and reference standard performed in every patient, with an appropriate period of time between the two

[5] Reference standard varied widely from study to study, using different culture techniques and in some cases employing a number of additional reference criteria (e.g. clinical characteristics or smear status)

[6] Wide confidence interval

[7] Significant variation in the point estimates with limited overlap of wide confidence intervals

[8] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Use of antituberculosis antibodies to detect tuberculosis in urine compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-positive

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
LAM	2[6]	cross-sectional	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement Conducted in a high incidence country[8]	1032	27.7% (95% CI 21.5 to 34.8%)	HIGH
LAM	Mutetwa, 2009[6]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	no serious imprecision	Unclear if there was industry involvement Conducted in a high incidence country[8]	397	52% 95% CI (43 to 62%)	MODERATE
Specificity[1,9]										
LAM	Lawn, 2012 Lawn, 2012 Mutetwa, 2012	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[7]	Unclear if there was industry involvement in 1 study; no involvement in the other Conducted in high incidence	516 516 397	98.1% (95% CI 96.9 to 99.4%) 98.6% (95% CI 97.5 to 99.7%) 86% (95% CI	LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							countries[8]		77 to 93%)	

[1] Forest plots:

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Unclear if a consecutive or random sample used in Mutetwa (2009)
[4] Unclear if inappropriate exclusions were avoided in Mutetwa (2009)
[5] Unclear if test interpretation was blinded in Mutetwa (2009)
[6] Insufficient data to include Mutetwa (2009) in the meta-analysis
[7] Variation in the point estimates with limited overlap of wide confidence intervals
[8] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000
[9] Meta-analysis of relevant data not possible in STATA or R

Interferon-gamma release assays compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-positive

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
QuantiFERON-TB Gold In-Tube	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	No industry involvement Conducted in a high incidence country[6]	52	85.3% (95% CI 73.4 to 97.2%)	VERY LOW
Specificity										
QuantiFERON-TB Gold In-Tube	1[2]	cross-sectional	serious[2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	No industry involvement Conducted in a high incidence country[6]	52	44.4% (95% CI 21.5 to 67.4%)	LOW

[1] Kabeer, 2009
[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				

[3] Consecutive or random sample not used
[4] Unclear if test interpretation was blinded
[5] Wide confidence interval
[6] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Tuberculin skin tests compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are HIV-positive

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
Mantoux	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	No industry involvement Conducted in a high incidence country[6]	52	25.0% (95% CI 12.2 to 37.8%)	VERY LOW
Specificity										
Mantoux	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	No industry involvement Conducted in a high incidence country[6]	52	72.7% (95% CI 54.1 to 91.3%)	LOW

[1] Kabeer, 2009
[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Consecutive or random sample not used
[4] Unclear if test interpretation was blinded
[5] Wide confidence interval
[6] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

A.2.3 Diagnosis of active pulmonary tuberculosis in adults

Smear microscopy compared to culture-based reference standard in adults with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
All techniques (Ziehl-Neelson, fluorescence, cold stain)	84	cross-sectional[2]	serious[2,3,4,5,6]	serious[7,8,9]	no serious indirectness[10]	serious[11]	Degree of industry involvement unclear in many studies; amongst those for which information is given, just under half had industry involvement Approximately half of studies were conducted in a high incidence country[13]	59984	65.6% (95% CI 61.1 to 69.9%)	VERY LOW
Specificity[1]										
All techniques (Ziehl-Neelson, fluorescence, cold stain)	84	cross-sectional[2]	serious[2,3,4,5,6]	serious[7,8,9]	no serious indirectness[10]	no serious imprecision	Degree of industry involvement unclear in many studies; amongst those for which information is given, just under half had industry involvement Approximately half of studies were conducted in a high	59984	97.9% (95% CI 97.1 to 98.5%)	LOW

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
							incidence country[13]			

Appendix E: GRADE profiles

[1] Forest plots for sensitivity and specificity (grouped by technique used):

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Unclear if a consecutive or random sample of participants used in all or most of the included comparisons
[4] Unclear if inappropriate exclusions were avoided in all or most of the included comparisons
[5] Unclear if test interpretation was blinded in all or most of the included comparisons
[6] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons
[7] Index test varies across studies: microscopy technique varies across studies
[8] Reference standard varies across studies: culture technique not consistent
[10] A number of studies include a small proportion of participants who are under 18 years old or provide no details of the age of the study population; however, it is not anticipated that the results will be significantly affected by this
[11] Significant variation in the point estimates, with limited overlap in confidence intervals
[12] Wide confidence interval
[13] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Chest radiography compared to culture-based reference standard in people with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity									
Chest radiograph – CAD4TB (computer-aided detection system) Threshold for interpretation: ≥23 points[1]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	95% (95% CI 91 to 98%)	MODERATE
Chest radiograph – CAD4TB (computer-aided detection system) Threshold for	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	85% (95% CI 79 to 90%)	MODERATE

Test details	Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
interpretation: ≥56 points[1]									
Chest radiograph – CAD4TB (computer-aided detection system) Threshold for interpretation: ≥74 points[1]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	77% (95% CI 71 to 83%)	MODERATE
Chest radiograph – CAD4TB (computer-aided detection system) Threshold for interpretation: ≥95 points[1]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	47% (95% CI 40 to 54%)	MODERATE
Chest radiograph – 'expert reader' Threshold for interpretation: category 4[4]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	59% (95% CI 52 to 66%)	MODERATE
Chest radiograph – 'expert reader' Threshold for interpretation: category 3 or	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	78% (95% CI 71 to 83%)	MODERATE

Test details	Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
4[4]									
Chest radiograph – clinical officer with practical experience, but not considered 'expert' Threshold for interpretation: category 4[4]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	7% (95% CI 4% to 12%)	MODERATE
Chest radiograph – clinical officer with practical experience, but not considered 'expert' Threshold for interpretation: category 3 or 4[4]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	76% (95% CI 69 to 82%	MODERATE
Specificity									
Chest radiograph – CAD4TB (computer-aided detection system) Threshold for interpretation: ≥23 points[1]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	33% (95% CI 27 to 39%)	MODERATE

Test details	Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Chest radiograph – CAD4TB (computer-aided detection system) Threshold for interpretation: ≥56 points[1]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	69% (95% CI 62 to 75%)	MODERATE
Chest radiograph – CAD4TB (computer-aided detection system) Threshold for interpretation: ≥74 points[1]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	79% (95% CI 74 to 84%)	MODERATE
Chest radiograph – CAD4TB (computer-aided detection system) Threshold for interpretation: ≥95 points[1]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	94% (95% CI 91 to 97%)	MODERATE
Chest radiograph – 'expert reader' Threshold for interpretation:	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	98% (95% CI 95 to 99%)	MODERATE

Test details	Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
category 4[4]									
Chest radiograph – 'expert reader' Threshold for interpretation: category 3 or 4[4]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	81% (95% CI 80 to 89%)	MODERATE
Chest radiograph – clinical officer with practical experience, but not considered 'expert' Threshold for interpretation: category 4[4]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	97% (95% CI 94 to 99%)	MODERATE
Chest radiograph – clinical officer with practical experience, but not considered 'expert' Threshold for interpretation: category 3 or 4[4]	1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	861	65% (95% CI 58 to 71%)	MODERATE

[1] Out of 100 points; system uses various subsystems for the detection of textural and shape abnormalities, for symmetry and correlation analyses operate at pixel and image level – scores generated by these subsystems are combined to an overall score for each image which summarises the result of the automated analysis as an abnormality score for the presence of active disease

[2] Breuninger, 2014

Test details	Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

[3] Protocol permitted the inclusion of children

[4] Categories:

1. normal
2. abnormal, findings not suggestive for active TB (TB sequel possible)
3. abnormal, findings consistent with active TB, but TB sequel or other lung pathology possible
4. abnormal, findings highly suggestive for active TB

Chest radiography plus an algorithm of signs, symptoms and risk factors compared to culture-based reference standard in adults with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Chest radiography plus signs, symptoms and risk factors[2]	10	systematic review[3]	serious[4,5,6]	very serious[7,8]	no serious indirectness[9]	serious[10]	Industry involvement unclear Unclear TB incidence in countries in which studies were conducted	5375	94% (24–100%)	VERY LOW
Specificity[1]										
Chest radiography plus signs, symptoms and risk factors[2]	10	systematic review[3]	serious[4,5,6]	very serious[7,8]	no serious indirectness[9]	serious[10]	Industry involvement unclear Unclear TB incidence in countries in which studies were conducted	5375	56% (21–93.1%)	VERY LOW

[1] Sensitivity and specificity:

Study	*Sensitivity (95% CI)*	*Specificity (95% CI)*

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
	Bock, 1996						81% (66 to 91%)		62% (56 to 68%)	
	El-Solh, 1997						100% (78 to 100%)		50% (44 to 57%)	
	El-Solh, 1999						100% (91 to 100%)		72% (65 to 77%)	
	Lagrange-Xelot, 2010						96% (80 to 100%)		21% (14 to 30%)	
	Moran, 2009						96% (91 to 99%)		49% (47 to 51%)	
	Mylotte, 1997						88% (47 to100%)		63% (56 to 70%)	
	Solari, 2008						93% (86 to 97%)		42% (36 to 49%)	
	Soto, 2008						93%		92%	
	Soto, 2011						24% (18 to 31%)		93% (91 to 95%)	
	Wisnivesky, 2005						95% (74 to 100%)		35% (31 to 40%)	

[2] Scoring systems used:

Study	*Details of chest radiograph scoring system*
Bock, 1996	1) chest X-ray with upper lobe infiltrate, 2) chest X-ray with cavity, 3) contact with someone with active tuberculosis, 4) self-report of positive tuberculin skin test in the past, 5) self-report of isoniazid preventive therapy in the past Test-positive: any of 1 to 3 or 4 (in the absence of 5)
El-Solh, 1997	Test-negative: upper zone disease and fever absent, or upper zone disease absent and fever present, if no weight loss and CD4+ >200 Test-positive: upper zone disease and weight loss
El-Solh, 1999	Age, CD4+ counts, diabetes mellitus, HIV, tuberculin skin test positivity; chest pain, weight loss, cough, night sweats, fever, shortness of breath; upper or lower lobe infiltrate, upper or lower lobe cavity, adenopathy, unilateral or bilateral pleural effusion, pleural thickening, miliary pattern
Lagrange-Xelot, 2010	Tuberculosis risk factors or chronic symptoms – scores 4; self-report of positive tuberculin skin test in the past – scores 5; shortness of breath – scores -3; temperature <38.5°C – scores 0; temperature 38.5-39°C – scores 3; temperature >39°C – scores 6; crackles on physical examination – scores -3; upper lobe disease on chest x-ray – scores 6 Test-positive: score of 1 or above
Moran, 2009	1) apical infiltrate, 2) cavitation, 3) immigrant, 4) weight loss, 5) positive tuberculosis history, 6) homeless, 7) incarcerated Test-positive: any of 1 to 7
Mylotte, 1997	AFB-positive smear – scores 3; localised chest X-ray change – scores 2; incarcerated – scores 2; history of weight loss – scores 1 Test-positive: score of 3 or above
Solari, 2008	Age <35 years – scores 0; age 35-60 years – scores -1; age 60 or over – scores -2; weight loss – scores 5; history of pulmonary tuberculosis – scores -3; miliary pattern – scores 10; cavity – scores 5; upper lobe infiltrate – scores 9 Test-positive: score of 3 or above

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Soto, 2008	Haemoptysis – scores 2; weight loss – scores 1; age >45 years – scores -1; expectoration – scores -1; apical infiltrate – scores 3; miliary infiltrate – scores 4 Score >4 = high probability									
Soto, 2011	Haemoptysis – scores 2; weight loss – scores 1; age >45 years – scores -1; expectoration – scores -1; apical infiltrate – scores 3; miliary infiltrate – scores 4 Score ≥5 = high probability									
Wisnivesky, 2005	Tuberculosis risk factors or chronic symptoms – scores 4; self-report of positive tuberculin skin test in the past – scores 5; shortness of breath – scores -3; temperature <38.5°C – scores 0; temperature 38.5-39°C – scores 3; temperature >39°C – scores 6; crackles on physical examination – scores -3; upper lobe disease on chest x-ray – scores 6 Test-positive: score of 1 or above									

[3] Data presented only for cross-sectional studies; case-control excluded

[4] Unclear if interpretation of reference standard was blind to the results of the index test in a number of studies (5 of 10), although interpretation of the index test was always conducted blind to the reference standard

[5] Unclear if a consecutive or random sample of participants used

[6] Unclear if inappropriate exclusions were avoided

[7] Index test varies significantly across studies

[8] Reference standard was permitted by reviewers to be liquid or solid culture; consistency in the exact techniques used across studies is not clear

[9] Reviewers provide no details of the age of the study population; however, it is not anticipated that the results will be significantly affected by this

[10] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap

Chest radiography plus an algorithm of signs, symptoms and risk factors compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are smear-negative

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Chest radiography plus signs, symptoms and risk factors[2]	4	systematic review[3]	serious[4,5,6]	very serious[7,8]	no serious indirectness[9]	serious[10]	Industry involvement unclear Unclear TB incidence in countries in	1575	94% (24–96%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							which studies were conducted			
Specificity[1]										
Chest radiography plus signs, symptoms and risk factors[2]	4	systematic review[3]	serious[4,5,6]	very serious[7,8]	no serious indirectness[9]	serious[10]	Industry involvement unclear Unclear TB incidence in countries in which studies were conducted	1575	94% (24–96%)	VERY LOW

[1] Sensitivity and specificity:

Study	*Sensitivity (95% CI)*	*Specificity (95% CI)*
Lagrange-Xelot, 2010	96% (80 to 100%)	21% (14 to 30%)
Soto, 2008	93%	92%
Soto, 2011	24% (18 to 31%)	93% (91 to 95%)
Wisnivesky, 2005	95% (74 to 100%)	35% (31 to 40%)

[2] Scoring systems used:

Study	*Details of chest radiograph scoring system*
Lagrange-Xelot, 2010	Tuberculosis risk factors or chronic symptoms – scores 4; self-report of positive tuberculin skin test in the past – scores 5; shortness of breath – scores -3; temperature <38.5°C – scores 0; temperature 38.5-39°C – scores 3; temperature >39°C – scores 6; crackles on physical examination – scores -3; upper lobe disease on chest x-ray – scores 6 Test-positive: score of 1 or above
Soto, 2008	Haemoptysis – scores 2; weight loss – scores 1; age >45 years – scores -1; expectoration – scores -1; apical infiltrate – scores 3; miliary infiltrate – scores 4 Score >4 = high probability
Soto, 2011	Haemoptysis – scores 2; weight loss – scores 1; age >45 years – scores -1; expectoration – scores -1; apical infiltrate – scores 3; miliary infiltrate – scores 4 Score ≥5 = high probability
Wisnivesky, 2005	Tuberculosis risk factors or chronic symptoms – scores 4; self-report of positive tuberculin skin test in the past – scores 5; shortness of breath – scores -3; temperature <38.5°C – scores 0; temperature 38.5-39°C – scores 3; temperature >39°C – scores 6; crackles on physical examination – scores -3; upper lobe disease on chest x-ray – scores 6

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
	Test-positive: score of 1 or above									

[3] Data presented only for cross-sectional studies; case-control excluded

[4] Unclear if interpretation of reference standard was blind to the results of the index test in a number of studies (2 of 4), although interpretation of the index test was always conducted blind to the reference standard

[5] Unclear if a consecutive or random sample of participants used

[6] Unclear if inappropriate exclusions were avoided

[7] Index test varies significantly across studies

[8] Reference standard was permitted by reviewers to be liquid or solid culture; consistency in the exact techniques used across studies is not clear

[9] Reviewers provide no details of the age of the study population; however, it is not anticipated that the results will be significantly affected by this

[10] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap

Commercial nucleic acid amplification techniques compared to culture-based reference standard in adults with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
All techniques	137	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	serious[14]	Degree of industry involvement unclear in many studies; amongst those for which information is given, approximately half had industry involvement Where information available, less than half of studies were	85438	89.0% (95% CI 87.2 to 90.6%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							conducted in a high incidence country[15]			
Amplicor	31	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	serious[14]	Degree of industry involvement unclear in many studies; amongst those for which information is given, approximately two-thirds had industry involvement Where information available, a quarter of studies were conducted in a high incidence country[15]	29937	84.8% (95% CI 81.1 to 87.9%)	VERY LOW
Amplified M. Tuberculosis Direct Test	33	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	serious[14]	Degree of industry involvement unclear in many studies; amongst those for which information is given, less than half had industry involvement Where information available, a quarter of studies	17701	91.9% (95% CI 88.1 to 94.6%)	VERY LOW

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
							were conducted in a high incidence country[15]			
BDProbeTec	3	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	serious[14]	Degree of industry involvement unclear Where information available, study was conducted in a high incidence country[15]	1416	94.4% (95% CI 90.2 to 96.8%)	VERY LOW
BDProbeTec ET	11	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	serious[14]	Degree of industry involvement unclear in many studies; information was available for 1 study, which was industry sponsored Where information available, half of studies were conducted in a high incidence country[15]	6847	88.0% (95% CI 82.8 to 91.9%)	VERY LOW
Cobas Amplicor	18	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	serious[14]	Degree of industry involvement unclear in many studies;	18000	87.2% (95% CI 80.2 to 92.0%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							information was available for 3 studies, of which 1 was industry sponsored Where information available, half of studies were conducted in a high incidence country[15]			
Enhanced Amplified M. Tuberculosis Direct Test	3	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	serious[14]	Information on industry involvement available for 2 studies, of which 1 was industry sponsored 2 of 3 studies were conducted in a high incidence country[15]	1359	78.9% (95% CI 66.6 to 87.5%)	VERY LOW
MTBDRplus assay	1[3]	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	serious[13]	Conducted in a high incidence country[15]	177	76% (95% CI 64 to 85%)	VERY LOW
TB-Biochip	1[4]	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	no serious imprecision	Conducted in a high incidence country[15]	105	97.3% (95% CI 93.5 to 100%)	VERY LOW
Xpert MTB/RIF	37	cross-sectional	serious[2,7]	serious[12]	serious[12]	serious[14]	Degree of industry involvement unclear in many studies; information	10073	90.0% (95% CI 86.5 to 92.7%)	VERY LOW

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
							provided for 5 studies, none of which were industry sponsored Majority of studies were conducted in a high incidence country[15]			
Specificity[1]										
All techniques	137	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	no serious imprecision	Degree of industry involvement unclear in many studies; amongst those for which information is given, approximately half had industry involvement Where information available, less than half of studies were conducted in a high incidence country[15]	85438	98.1% (95% CI 97.6 to 98.5%)	VERY LOW
Amplicor	31	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	no serious imprecision	Degree of industry involvement unclear in many studies; amongst those for which	29937	97.5% (95% CI 96.2 to 98.3%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							information is given, approximately two-thirds had industry involvement Where information available, a quarter of studies were conducted in a high incidence country[15]			
Amplified M. Tuberculosis Direct Test	33	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	no serious imprecision	Degree of industry involvement unclear in many studies; amongst those for which information is given, less than half had industry involvement Where information available, a quarter of studies were conducted in a high incidence country[15]	17701	97.2% (95% CI 95.5 to 98.3%)	VERY LOW
BDProbeTec	3	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	no serious imprecision	Degree of industry involvement unclear Where	1416	See forest plot below[1,16]	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							information available, study was conducted in a high incidence country[15]			
BDProbeTec ET	11	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	no serious imprecision	Degree of industry involvement unclear in many studies; information was available for 1 study, which was industry sponsored Where information available, half of studies were conducted in a high incidence country[15]	6847	97.4% (95% CI 96.0 to 98.3%)	VERY LOW
Cobas Amplicor	18	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	no serious imprecision	Degree of industry involvement unclear in many studies; information was available for 3 studies, of which 1 was industry sponsored Where information available, half of studies were conducted in a	18000	99.1% (95% CI 98.2 to 99.6%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							high incidence country[15]			
Enhanced Amplified M. Tuberculosis Direct Test	3	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	no serious imprecision	Information on industry involvement available for 2 studies, of which 1 was industry sponsored 2 of 3 studies were conducted in a high incidence country[15]	1359	See forest plot below[1,16]	VERY LOW
MTBDRplus assay	1[3]	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	no serious imprecision	Conducted in a high incidence country[15]	177	97% (95% CI 92 to 99%)	VERY LOW
TB-Biochip	1[4]	cross-sectional	very serious[2,7,8,9,10]	serious[12]	serious[11,12]	serious[13]	Conducted in a high incidence country[15]	105	78.1% (95% CI 63.8 to 92.5%)	VERY LOW
Xpert MTB/RIF	37	cross-sectional	serious[2,7]	serious[12]	serious[12]	no serious imprecision	Degree of industry involvement unclear in many studies; information provided for 5 studies, none of which were industry sponsored Majority of studies were conducted in a high incidence country[15]	10073	98.9% (95% CI 98.3 to 99.3%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
[1] Forest plots for sensitivity and specificity (grouped by technique used): Amplicor Amplified M. Tuberculosis Direct Test BDProbeTec BDProbeTec ET Cobas Amplicor Enhanced Amplified M. Tuberculosis Direct Test										
Xpert MTB/RIF assay										

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Scott, 2011
[4] Kurbatova, 2013
[7] Many studies did not use a consecutive or random sample or did not report the sampling approach used
[8] Unclear if inappropriate exclusions were avoided
[9] Many studies did not blind test interpretation or did not report the degree of blinding used
[10] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons
[11] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this
[12] Reference standard varied widely from study to study, using different culture techniques and in some cases employing a number of additional reference criteria (e.g. clinical characteristics or smear status)

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				

[13] Wide confidence interval

[14] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap

[15] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England current estimates of incidence for in the UK are 13.9 per 100,000

[16] Meta-analysis of relevant data not possible in STATA or R

Commercial nucleic acid amplification techniques compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are smear-positive

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
All techniques	66	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	no serious imprecision	Degree of industry involvement unclear in many studies; amongst those for which information is given, approximately half had industry involvement Where information available, half of studies were conducted in a high incidence country[13]	5205	98.7% (95% CI 97.8 to 99.2%)	VERY LOW
Amplicor	8	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	no serious imprecision	Degree of industry involvement unclear in many	1248	95.5% (95% CI 83.7 to 98.8%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							studies; amongst 5 studies for which information is given, 1 had industry involvement 1 study was conducted in a high incidence country[13]			
Amplified M. Tuberculosis Direct Test	11	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	no serious imprecision	Degree of industry involvement unclear in many studies; amongst 4 studies for which information is given, 3 had industry involvement Where information was available, 3 studies were conducted in a high incidence country[13]	1204	99.6% (95% CI 98.1 to 99.9%)	VERY LOW
BDProbeTec	1[3]	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	no serious imprecision	Degree of industry involvement unclear	83	98.8% (95% CI 96.5 to 100%)	VERY LOW
BDProbeTec ET	4	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	no serious imprecision	Degree of industry involvement	113	97.6% (95% CI 89.6 to 99.5%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							unclear Where information was available, 1 of 2 studies were conducted in a high incidence country[13]			
Cobas Amplicor	7	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	no serious imprecision	Degree of industry involvement unclear in many studies; only study to report had no industry involvement Where information was available, 2 of 6 studies were conducted in a high incidence country[13]	492	Median (range) = 96.2% (79.2–97.0%)	VERY LOW
Enhanced Amplified M. Tuberculosis Direct Test	2	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	serious[11]	Degree of industry involvement reported in 1 study, which did not receive industry support	45	93.0% (95% CI 75.9 to 98.2%)	VERY LOW
Xpert MTB/RIF	33	cross-sectional	serious[4,5]	serious[10]	serious[9,10]	no serious imprecision	Degree of industry involvement unclear in many studies; information	2020	98.5% (95% CI 97.5 to 99.1%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							provided for 4 studies, none of which were industry sponsored Majority of studies were conducted in a high incidence country[13]			
Specificity[1]										
All techniques	66	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	serious[11]	Degree of industry involvement unclear in many studies; amongst those for which information is given, approximately half had industry involvement Where information available, half of studies were conducted in a high incidence country[13]	5205	30.1% (95% CI 10.3 to 61.8%)	VERY LOW
Amplicor	8	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	serious[11,12]	Degree of industry involvement unclear in many studies; amongst 5 studies for which	1248	78.0% (95% CI 47.3 to 93.3%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							information is given, 1 had industry involvement 1 study was conducted in a high incidence country[13]			
Amplified M. Tuberculosis Direct Test	11	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	serious[11]	Degree of industry involvement unclear in many studies; amongst 4 studies for which information is given, 3 had industry involvement Where information was available, 3 studies were conducted in a high incidence country[13]	1204	90.4% (95% CI 68.1 to 97.7%)	VERY LOW
BDProbeTec	1[3]	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	serious[11]	Degree of industry involvement unclear	83	50.0% (95% CI 0.0 to 100%)	VERY LOW
BDProbeTec ET	4	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	serious[11,12]	Degree of industry involvement unclear Where information was	113	63.8% (95% CI 6.6 to 97.8%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							available, 1 of 2 studies were conducted in a high incidence country[13]			
Cobas Amplicor	7	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	serious[12]	Degree of industry involvement unclear in many studies; only study to report had no industry involvement Where information was available, 2 of 6 studies were conducted in a high incidence country[13]	492	See forest plot below[1,14]	VERY LOW
Enhanced Amplified M. Tuberculosis Direct Test	2	cross-sectional	very serious[4,5,6,7,8]	serious[10]	serious[9,10]	no serious imprecision	Degree of industry involvement reported in 1 study, which did not receive industry support	45	See forest plot below[1,14]	VERY LOW
Xpert MTB/RIF	33	cross-sectional	serious[4,5]	serious[10]	serious[9,10]	no serious imprecision	Degree of industry involvement unclear in many studies; information provided for 4 studies, none of which were industry	2020	See forest plot below[1,14]	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							sponsored Majority of studies were conducted in a high incidence country[13]			

[1] Forest plots for sensitivity and specificity (grouped by technique used):
Amplicor

Amplified M. Tuberculosis Direct Test

BDProbeTec ET

Cobas Amplicor

Enhanced Amplified M. Tuberculosis Direct Test

Xpert MTB/RIF assay

[2] Steingart, 2014
[3] Iinuma, 2003
[4] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[5] Many studies did not use a consecutive or random sample or did not report the sampling approach used
[6] Unclear if inappropriate exclusions were avoided
[7] Many studies did not blind test interpretation or did not report the degree of blinding used
[8] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons
[9] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this
[10] Reference standard varied widely from study to study, using different culture techniques and in some cases employing a number of additional reference criteria (e.g. clinical characteristics or smear status)
[11] Wide confidence interval
[12] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap
[13] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Health England; current estimates of incidence for in the UK are 13.9 per 100,000 [14] Meta-analysis of relevant data not possible in STATA or R										

Commercial nucleic acid amplification techniques compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are smear-negative

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
All techniques	65	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	serious[11]	Degree of industry involvement unclear in many studies; amongst those for which information is given, approximately half had industry involvement Where information available, half of studies were conducted in a high incidence country[12]	24499	72.6% (95% CI 68.1 to 76.8%)	VERY LOW
Amplicor	8	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	serious[10,11]	Degree of industry involvement unclear in many studies; amongst 4 studies for which	2739	78.0% (95% CI 60.9 to 89.0%)	VERY LOW

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
							information is given, 1 had industry involvement 1 study was conducted in a high incidence country[12]			
Amplified M. Tuberculosis Direct Test	11	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	serious[10,11]	Degree of industry involvement unclear in many studies; amongst 6 studies for which information is given, 2 had industry involvement Where information was available, 3 studies were conducted in a high incidence country[12]	5922	84.6% (95% CI 71.6 to 92.3%)	VERY LOW
BDProbeTec ET	4	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	serious[10,11]	Degree of industry involvement unclear Where information was available, 1 of 2 studies were conducted in a high incidence country[12]	2391	70.4% (95% CI 54.4 to 82.5%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Cobas Amplicor	7	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	serious[11]	Degree of industry involvement unclear in many studies; only study to report had no industry involvement Where information was available, 2 of 6 studies were conducted in a high incidence country[12]	5040	56.9% (95% CI 48.3 to 65.1%)	VERY LOW
Enhanced Amplified M. Tuberculosis Direct Test	2	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	serious[10,11]	Degree of industry involvement reported in 1 study, which did not receive industry support	1233	67.7% (95% CI 48.4 to 82.4%)	VERY LOW
Xpert MTB/RIF	3	cross-sectional	serious[3,4]	serious[9]	serious[8,9]	serious[11]	Degree of industry involvement unclear in many studies; information provided for 4 studies, none of which were industry sponsored Majority of studies were conducted in a high incidence	619	71.1% (95% CI 65.5 to 76.0%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							country[12]			
Specificity[1]										
All techniques	65	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	no serious imprecision	Degree of industry involvement unclear in many studies; amongst those for which information is given, approximately half had industry involvement Where information available, half of studies were conducted in a high incidence country[12]	24499	98.6% (95% CI 97.9 to 99.0%)	VERY LOW
Amplicor	8	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	no serious imprecision	Degree of industry involvement unclear in many studies; amongst 4 studies for which information is given, 1 had industry involvement 1 study was conducted in a high incidence country[12]	2739	96.5% (95% CI 92.3 to 98.5%)	VERY LOW
Amplified M.	11	cross-	very	serious[9]	serious[8,9]	no serious	Degree of	5922	98.0% (95%	VERY

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Tuberculosis Direct Test		sectional	serious[3,4,5,6,7]			imprecision	industry involvement unclear in many studies; amongst 6 studies for which information is given, 2 had industry involvement Where information was available, 3 studies were conducted in a high incidence country[12]		CI 94.7 to 99.2%)	LOW
BDProbeTec ET	4	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	no serious imprecision	Degree of industry involvement unclear Where information was available, 1 of 2 studies were conducted in a high incidence country[12]	2391	96.4% (95% CI 94.2 to 97.8%)	VERY LOW
Cobas Amplicor	7	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	no serious imprecision	Degree of industry involvement unclear in many studies; only study to report had no industry involvement Where	5040	99.3% (95% CI 98.1 to 99.8%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							information was available, 2 of 6 studies were conducted in a high incidence country[12]			
Enhanced Amplified M. Tuberculosis Direct Test	2	cross-sectional	very serious[3,4,5,6,7]	serious[9]	serious[8,9]	no serious imprecision	Degree of industry involvement reported in 1 study, which did not receive industry support	1233	See forest plot below[1,14]	VERY LOW
Xpert MTB/RIF	33	cross-sectional	serious[3,4]	serious[9]	serious[8,9]	no serious imprecision	Degree of industry involvement unclear in many studies; information provided for 4 studies, none of which were industry sponsored Majority of studies were conducted in a high incidence country[12]	7180	99.0% (95% CI 98.3 to 99.4%)	VERY LOW

[1] Forest plots for sensitivity and specificity (grouped by technique used):
Amplicor

Amplified M. Tuberculosis Direct Test

BDProbeTec ET

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Cobas Amplicor Enhanced Amplified M. Tuberculosis Direct Test Xpert MTB/RIF assay [3] Both index test and reference standard performed in every patient, with an appropriate period of time between the two [4] Many studies did not use a consecutive or random sample or did not report the sampling approach used [5] Unclear if inappropriate exclusions were avoided [6] Many studies did not blind test interpretation or did not report the degree of blinding used [7] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons [8] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this [9] Reference standard varied widely from study to study, using different culture techniques and in some cases employing a number of additional reference criteria (e.g. clinical characteristics or smear status) [10] Wide confidence interval [11] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap [12] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000 [13] Meta-analysis of relevant data not possible in STATA or R										

Phage-based tests compared to culture-based reference standard in adults with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Phage-based tests	5[2]	cross-sectional	serious[4,5,6,7,8]	serious[9]	serious[10,11]	serious[12]	Industry involvement unclear Unclear TB incidence in	3033	69.5% (95% CI 47.5 to 85.1%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							countries in which studies were conducted			
Specificity[1]										
Phage-based tests	5[2]	cross-sectional	serious[4,5,6,7,8]	serious[9]	serious[10,11]	serious[12]	Industry involvement unclear Unclear TB incidence in countries in which studies were conducted	3033	See forest plot below[1,13]	VERY LOW

[1] Forest plots:

[2] Dinnes, 2007
[4] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[5] Unclear if a consecutive or random sample was used
[6] Unclear if inappropriate exclusions were avoided
[7] Unclear if test interpretation was blinded
[8] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons
[9] Reference standard varied widely from study to study
[10] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this
[11] Reference standard sometimes included more than just culture, for example X-ray, clinical features and treatment response
[12] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap
[13] Meta-analysis of relevant data not possible in STATA or R

Phage-based tests compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are smear-positive

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
FASTPlaque TB	2	cross-sectional	serious[2,3,4,5,6]	serious[7]	serious[8,9]	no serious imprecision	Industry involvement unclear Unclear TB incidence in countries in which studies were conducted	277	86.3% (95% CI 81.4 to 90.1%)	VERY LOW
Specificity[1]										
FASTPlaque TB	2	cross-sectional	serious[2,3,4,5,6]	serious[7]	serious[8,9]	serious[12]	Industry involvement unclear Unclear TB incidence in countries in which studies were conducted	277	See forest plot below[1,11]	VERY LOW

[1] Forest plots:

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Unclear if a consecutive or random sample was used
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons
[7] Reference standard varied widely from study to study
[8] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this
[9] Reference standard sometimes included more than just culture, for example X-ray, clinical features and treatment response
[10] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap
[11] Meta-analysis of relevant data not possible in STATA or R

Phage-based tests compared to culture-based reference standard in adults with suspected pulmonary tuberculosis who are smear-negative

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
FASTPlaque TB	2	cross-sectional	serious[2,3,4,5,6]	serious[7]	serious[8,9]	serious[10]	Industry involvement unclear Unclear TB incidence in countries in which studies were conducted	1016	58.6% (95% CI 39.6 to 75.3%)	VERY LOW
Specificity[1]										
FASTPlaque TB	2	cross-sectional	serious[2,3,4,5,6]	serious[7]	serious[8,9]	no serious imprecision	Industry involvement unclear Unclear TB incidence in countries in which studies were conducted	1016	See forest plot below[1,11]	VERY LOW

[1] Forest plots:

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Unclear if a consecutive or random sample was used
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons
[7] Reference standard varied widely from study to study
[8] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this
[9] Reference standard sometimes included more than just culture, for example X-ray, clinical features and treatment response
[10] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap
[11] Meta-analysis of relevant data not possible in STATA or R

Antituberculosis antibody detection compared to culture-based reference standard in adults with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Antituberculosis antibody detection tests	9	cross-sectional	serious[2,3,4,5,6]	serious[7]	serious[8,9]	serious[10]	Degree of industry involvement unclear in many studies; 1 of the 2 studies that provided information had industry involvement Both studies for which information available were conducted in a high incidence country[11]	2703	68.2% (95% CI 40.9 to 86.9%)	VERY LOW
Specificity[1]										
Antituberculosis antibody detection tests	9	cross-sectional	serious[2,3,4,5,6]	serious[7]	serious[8,9]	serious[10]	Degree of industry involvement unclear in many studies; 1 of the 2 studies that provided information had industry involvement Both studies for which information available were conducted in a high incidence	2703	85.3% (95% CI 76.8 to 91.0%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							country[11]			

[1] Forest plots:

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Unclear if a consecutive or random sample was used
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons
[7] Reference standard varied widely from study to study
[8] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this
[9] Reference standard sometimes included more than just culture, for example X-ray, clinical features and treatment response
[10] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap
[11] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Use of antituberculosis antibodies (LAM) to detect tuberculosis in serum

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Antituberculosis antibody detection tests	2	cross-sectional	serious[2,3,4,5,6]	serious[7]	serious[8,9]	serious[10]	Industry involvement unclear Unclear TB incidence in countries in which studies were conducted	370	54.1% (95% CI 30.4 to 76.2%)	VERY LOW
Specificity[1]										
Antituberculosis antibody	2	cross-sectional	serious[2,3,4,5,6]	serious[7]	serious[8,9]	no serious imprecision	Industry involvement unclear	370	See forest plot below[1,11]	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
detection tests							Unclear TB incidence in countries in which studies were conducted			

[1] Forest plots:

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Unclear if a consecutive or random sample was used
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons
[7] Reference standard varied widely from study to study
[8] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this
[9] Reference standard sometimes included more than just culture, for example X-ray, clinical features and treatment response
[10] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap
[11] Meta-analysis of relevant data not possible in STATA or R

Use of antituberculosis antibodies (LAM) to detect tuberculosis in urine

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Antituberculosis antibody detection tests	3	cross-sectional	serious[2,3,4,5,6]	serious[7]	serious[8,9]	serious[10]	No industry involvement Conducted in a high incidence country[11]	1429	32.9% (95% CI 22.6 to 45.2%)	VERY LOW
Specificity[1]										
Antituberculosis	3	cross-sectiona	serious[2,3,4,5,6]	serious[7]	serious[8,9]	no serious	No industry	1429	See forest plot	VERY

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
antibody detection tests		I				imprecision	involvement Conducted in a high incidence country[11]		below[1,12]	LOW

[1] Forest plots:

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[3] Unclear if a consecutive or random sample was used
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Unclear if a threshold for test interpretation was prespecified in all or most of the included comparisons
[7] Reference standard varied widely from study to study
[8] Unclear how many participants, if any, are under 18 years old; however, it is not anticipated that the results will be significantly affected by this
[9] Reference standard sometimes included more than just culture, for example X-ray, clinical features and treatment response
[10] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap
[11] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000
[12] Meta-analysis of relevant data not possible in STATA or R

Interferon-gamma release assays compared to culture-based reference standard in adults with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
IGRAs	3[2,3]	cross-sectional	serious[4,5,6,7]	serious[8]	serious[9,10]	no serious imprecision	Industry involvement in Kang (2007) Conducted in a high incidence country[12]	327	89.3% (95% CI 83.4 to 93.3%)	VERY LOW
Specificity[1]										

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
IGRAs	3[2,3]	cross-sectional	serious[4,5,6,7]	serious[8]	serious[9,10]	no serious imprecision	Industry involvement in Kang (2007) Conducted in a high incidence country[12]	327	See forest plot below[1,13]	VERY LOW

[1] Forest plots:

[2] Kabeer, 2009
[3] Kang, 2007
[4] Both index test and reference standard performed in every patient, with an appropriate period of time between the two
[5] Consecutive or random sample not used in Kabeer (2009)
[6] Inappropriate exclusions were not avoided – Kang (2007) excluded patients with high clinical likelihood of active TB and a negative mycobacterial culture finding but good clinical and radiographic responses to anttuberculosis treatment
[7] Unclear if test interpretation was blinded
[8] Reference standard varied
[9] Unclear how many participants, if any, were under 18 years old in Kang (2007); however, it is not anticipated that the results will be significantly affected by this
[10] Reference standard included histology as an alternative to culture in Kang (2007)
[11] Wide confidence interval
[12] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000
[13] Meta-analysis of relevant data not possible in STATA or R

Tuberculin skin tests compared to culture-based reference standard in adults with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Mantoux	2	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7,8]	serious[10]	Degree of industry involvement unclear 1 study;	108	46.1% (95% CI 12.1 to 84.2%)	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							amongst the 2 for which information is given, 1 had industry involvement Where information available, both studies were conducted in a high incidence country[11]			
Specificity[1]										
Mantoux	2	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7,8]	serious[10]	Degree of industry involvement unclear 1 study; amongst the 2 for which information is given, 1 had industry involvement Where information available, both studies were conducted in a high incidence country[11]	108	See forest plot below[1,12]	VERY LOW

[1] Forest plots:

[2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two

[3] Consecutive or random sample not used in Kabeer (2009)

[4] Inappropriate exclusions were not avoided – Kang (2007) excluded patients with high clinical likelihood of active TB and a negative mycobacterial culture finding but

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
good clinical and radiographic responses to anttuberculosis treatment [5] Unclear if test interpretation was blinded [6] Reference standard varied [7] Unclear how many participants, if any, were under 18 years old in Kang (2007); however, it is not anticipated that the results will be significantly affected by this [8] Reference standard included histology as an alternative to culture in Kang (2007) [9] Wide confidence interval [10] Significant variation in the point estimates, as well as wide confidence intervals with limited overlap [11] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000 [12] Meta-analysis of relevant data not possible in STATA or R										

Gas chromatography mass spectrometry for tuberculostearic acid compared to culture-based reference standard in adults with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
Gas chromatography mass spectrometry for tuberculostearic acid	1[1]	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency[7]	no serious indirectness	serious[8]	Degree of industry involvement unclear	145	55.3% (95% CI 39.5 to 71.1%)	LOW
Specificity										
Gas chromatography mass spectrometry for tuberculostearic acid	1[1]	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency[7]	no serious indirectness	serious[8]	Degree of industry involvement unclear	145	86.9% (80.5% to 93.3%)	LOW

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
[1] Savić, 1992 [2] Both index test and reference standard performed in every patient, with an appropriate period of time between the two [3] Unclear if a consecutive or random sample of patients were enrolled [4] Unclear if the study avoided inappropriate exclusions [5] Unclear if test interpretation was blinded [6] Unclear if a test threshold was prespecified [7] Unclear how many participants, if any, were under 18 years old; however, it is not anticipated that the results will be significantly affected by this [8] Wide confidence interval										

Time-to-detection

Test	Time	Reference
Time to diagnosis (median (range), unless otherwise indicated)		
Xpert MTB/RIF	0 days	Balcells, 2012
	0 (0–1) days	Boehme, 2011
	2 hours	Helb, 2010
	4 (3–6) days	Lawn, 2011
	<2 hours	Marlowe, 2011
	113 minutes	Miller, 2011
	2 hours	Moure, 2011
	within two hours	Rachow, 2011
	results available the same day	Van Rie, 2013
	3 to 24 hours	Zeka, 2011
Microscopy	1 (IQR 0–1) days	Boehme, 2011
	3 (2–5) days	Lawn, 2011
	minimum of 1 day; routinely available within 3 days	Kambashi, 2001
Liquid culture	10 (5–22) days	Balcells, 2012
	16 (13–21) days	Boehme, 2011
	smear-positive: 12 (10–14) days smear-negative: 20 (17–27) days	Lawn, 2011
	mean (range) = 19 (3–42) days	
Solid culture	30 (23–43) days	Boehme, 2011
Time to treatment initiation (median (range) or [interquartile range])		
Xpert MTB/RIF	before Xpert MTB/RIF introduced: 56 (39–81) days after Xpert MTB/RIF introduced: 5 (2–8) days	Boehme, 2011
	Xpert MTB/RIF positive patients: 0 (0–0) days patients diagnosed by other methods: 13 (10–20) days	Van Rie, 2013

A.3 RQ D

A.3.1 Diagnosis of active pulmonary tuberculosis in children and young people who are HIV-negative

Commercial nucleic acid amplification techniques compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis who are HIV-negative

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Xpert MTB/RIF	4	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6,7]	No industry involvement All studies conducted in a high incidence country[8]	1428	65.4% (95% CI 53.1 to 76.0%)	LOW
Specificity[1]										
Xpert MTB/RIF	4	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement All studies conducted in a high incidence country[8]	1428	See forest plot below[1,9]	MODERATE

[1] Forest plots for sensitivity and specificity:

[2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two

[3] Random sample of patients enrolled in Zar (2012); unclear if consecutive or random sample of patients enrolled in Bates (2013)

[4] Blinding of test interpretation employed in Zar (2012); unclear if blinding of test interpretation employed in Bates (2013)

[5] Threshold for interpretation unclear

[6] Significant variation in point estimates with little overlap in confidence intervals

[7] Wide confidence interval

[8] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

[9] Meta-analysis of relevant data not possible in STATA or R

Interferon-gamma release assays compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis who are HIV-negative

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
QuantiFERON-TB Gold In-Tube	1[1]	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7]	no serious imprecision	No industry involvement Conducted in a high incidence country[9]	362	79.7% (95% CI 72.7 to 86.7%)	VERY LOW
Specificity										
QuantiFERON-TB Gold In-Tube	1[1]	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7]	serious[8]	No industry involvement Conducted in a high incidence country[9]	362	16.7% (95% CI 11.9 to 21.4%)	VERY LOW

[1] Lodha, 2013
[2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two
[3] Unclear if consecutive or random sample of patients was enrolled
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Not all diagnoses were made with the same reference standard
[7] Reference diagnoses could be made by microscopy alone
[8] Wide confidence interval
[9] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Tuberculin skin tests compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis who are HIV-negative

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Mantoux	1[1]	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7]	no serious imprecision	No industry involvement Conducted in a high incidence country[9]	362	89.8% (95% CI 84.6 to 95.1%)	VERY LOW
Specificity										
Mantoux	1[1]	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7]	no serious imprecision	No industry involvement Conducted in a high incidence country[9]	362	5.1% (95% CI 2.3 to 8.0%)	VERY LOW

[1] Lodha, 2013
[2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two
[3] Unclear if consecutive or random sample of patients was enrolled
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Not all diagnoses were made with the same reference standard
[7] Reference diagnoses could be made by microscopy alone
[8] Wide confidence interval
[9] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

A.3.2 Diagnosis of active pulmonary tuberculosis in children and young people who are HIV-positive

Commercial nucleic acid amplification techniques compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis who are HIV-positive

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
Xpert	4	cross-	serious[2,3,4,5]	no serious	no serious	serious[6]	No industry	513	82.0% (55.2	LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
MTB/RIF		sectional		inconsistency	indirectness		involvement Conducted in a high incidence country[7]		to 94.4%)	
Specificity[1]										
Xpert MTB/RIF	4	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement Conducted in a high incidence country[7]	513	99.5% (96.2 to 99.9%)	MODERATE

[1] Forest plots for sensitivity and specificity:

[2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two

[3] Random sample of patients enrolled in Zar (2012); unclear if consecutive or random sample of patients enrolled in Bates (2013)

[4] Blinding of test interpretation employed in Zar (2012); unclear if blinding of test interpretation employed in Bates (2013)

[5] Threshold for interpretation unclear

[6] Significant variation in point estimates with little overlap/wide confidence intervals

[7] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

A.3.3 Diagnosis of active pulmonary tuberculosis in children and young people

Smear microscopy compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
All techniques	8	cross-sectional	serious[2,3,4,5,6]	serious[7]	no serious indirectness	serious[9]	Limited industry involvement,	2491	56.3% (95% CI 32.7 to	VERY LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							although 2 studies do not provide any information on this All except 1 study conducted in a high incidence country[9]		77.4%)	
Fluorescence microscopy	6	cross-sectional	serious[2,3,5,6]	no serious inconsistency	no serious indirectness	serious[9]	No industry involvement All studies conducted in high incidence countries[9]	2384	43.1% (95% CI 22.5 to 66.4%)	LOW
Ziehl-Neelson microscopy	1	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency	no serious indirectness	serious[8]	No information available on industry involvement	60	81.5% (95% CI 66.8 to 96.1%)	LOW
Specificity[1]										
All techniques	8	cross-sectional	serious[2,3,4,5,6]	serious[7]	no serious indirectness	no serious imprecision	Limited industry involvement, although 2 studies do not provide any information on this All except 1 study conducted in a high incidence country[9]	2491	99.7% (95% CI 98.8 to 99.9%)	LOW
Fluorescence	6	cross-sectional	serious[2,3,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement	2384	See forest plot below[1,10]	MODERATE

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
microscopy							All studies conducted in high incidence countries[9]			
Ziehl-Neelson microscopy	1	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	No information available on industry involvement	60	97.6% (95% CI 90.9 to 100%)	MODERATE

[1] Forest plots for sensitivity and specificity (grouped by technique used):

[2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two

[3] Random sample of patients enrolled in Zar (2012 and 2013); unclear if consecutive or random sample of patients enrolled in Bates (2013), El-Sayed Zaki (2008) and Shata (1996)

[4] Unclear if El-Sayed Zaki (2008) avoided inappropriate exclusions

[5] Blinding of test interpretation employed in Zar (2012 and 2013); unclear if blinding of test interpretation employed in Bates (2013), El-Sayed Zaki (2008) and Shata (1996)

[6] Threshold for interpretation unclear

[7] Shata (1996) uses a different culture technique as a reference standard than the other included studies (solid *vs* liquid culture)

[8] Significant variation in point estimates with little overlap in or wide confidence intervals

[9] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

[10] Meta-analysis of relevant data not possible in STATA or R

Chest radiography compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
Chest X-ray	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	No information available on industry	110	72%	LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							involvement Conducted in a high incidence country[6]			
Specificity										
Chest X-ray	1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	No information available on industry involvement Conducted in a high incidence country[6]	110	54%	LOW

[1] Iriso, 2005
[2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two
[3] Unclear if study avoided inappropriate exclusions
[4] Index test interpretation was blinded, though it was unclear if interpretation of the reference standard was also blinded
[5] Insufficient data to assess imprecision
[6] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Commercial nucleic acid amplification techniques compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity[1]										
All techniques	9	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency	no serious indirectness	serious[8]	Limited industry involvement, although 1 study did not provide any information on this	2828	71.3% (95% CI 54.3 to 83.8%)	LOW

Test details	Number of evaluations	Quality assessment						Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
							All studies except 1 conducted in a high incidence country[9]			
Amplified M. Tuberculosis Direct Test	1	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	No information available on industry involvement	60	97.5% (95% CI 92.7 to 100%)	MODERATE
Xpert MTB/RIF	8	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency	no serious indirectness	serious[8]	No industry involvement All studies conducted in a high incidence country[9]	2768	65.0% (95% CI 51.9 to 76.1%)	LOW
Specificity[1]										
All techniques	9	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	Limited industry involvement, although 1 study did not provide any information on this All studies except 1 conducted in a high incidence country[9]	2828	98.6% (95% CI 98.0 to 99.1%)	MODERATE
Amplified M. Tuberculosis Direct Test	1	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	No information available on industry involvement	60	97.6% (95% CI 90.9 to 100%)	MODERATE
Xpert MTB/RIF	8	cross-sectional	serious[2,3,4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement All studies conducted in a high incidence country[9]	2768	98.7% (95% CI 98.1 to 99.1%)	MODERATE

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
[1] Forest plots for sensitivity and specificity (grouped by technique used): [2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two [3] Random/consecutive sample of patients enrolled in Nicol (2011), Sekadde (2013) and Zar (2012 and 2013); unclear if consecutive or random sample of patients enrolled in Bates (2013) and El-Sayed Zaki (2008) [4] Unclear if El-Sayed Zaki (2008) avoided inappropriate exclusions [5] Blinding of test interpretation employed in Zar (2012 and 2013); unclear if blinding of test interpretation employed in Bates (2013), El-Sayed Zaki (2008), Nicol (2011) and Sekadde (2013) [6] Threshold for interpretation unclear [7] Wide confidence interval [8] Significant variation in point estimates with little overlap in or wide confidence intervals [9] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000										

Commercial nucleic acid amplification techniques compared to culture-based reference standard in children under 2 years old

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
Xpert MTB/RIF	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	No industry involvement Conducted in a high incidence country[7]	930	63.2% (95% CI 41.5 to 84.9%)	LOW
Specificity										
Xpert MTB/RIF	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement Conducted in a high incidence country[7]	930	99.8% (95% CI 99.3 to 100%)	MODERATE
[1] Bates, 2013 [2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two										

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
[3] Unclear if consecutive or random sample of patients was enrolled [4] Unclear if test interpretation was blinded [5] Threshold for interpretation unclear [6] Wide confidence interval [7] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000										

Commercial nucleic acid amplification techniques compared to culture-based reference standard in children between 2 and 4 years old

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
Xpert MTB/RIF	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	No industry involvement Conducted in a high incidence country[7]	201	66.7% (95% CI 44.9 to 88.4%)	LOW
Specificity										
Xpert MTB/RIF	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement Conducted in a high incidence country[7]	201	99.5% (95% CI 98.4 to 100%)	MODERATE
[1] Bates, 2013 [2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two [3] Unclear if consecutive or random sample of patients was enrolled [4] Unclear if test interpretation was blinded [5] Threshold for interpretation unclear [6] Wide confidence interval [7] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000										

Commercial nucleic acid amplification techniques compared to culture-based reference standard in children between 5 and 9 years old

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
Xpert MTB/RIF	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	No industry involvement Conducted in a high incidence country[7]	124	50.0% (95% CI 10.0 to 90.0%)	LOW
Specificity										
Xpert MTB/RIF	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement Conducted in a high incidence country[7]	124	97.5% (95% CI 94.6 to 100%)	MODERATE

[1] Bates, 2013

[2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two

[3] Unclear if consecutive or random sample of patients was enrolled

[4] Unclear if test interpretation was blinded

[5] Threshold for interpretation unclear

[6] Wide confidence interval

[7] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Commercial nucleic acid amplification techniques compared to culture-based reference standard in children and young people between 10 and 15 years old

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
All techniques	2[1,8]	cross-sectional	serious[2,3,4,5,7]	no serious inconsistency	no serious indirectness	serious[6]	Limited industry involvement, although 1 study	198	96.5% (95% CI 87.0 to 99.1%)	LOW

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
							did not provide any information on this All studies except 1 conducted in a high incidence country[9]			
Amplified M. Tuberculosis Direct Test	1[8]	cross-sectional	serious[2,3,4,5,7]	no serious inconsistency	no serious indirectness	no serious imprecision	No information available on industry involvement	60	97.5% (95% CI 92.7 to 100%)	MODERATE
Xpert MTB/RIF	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	No industry involvement Conducted in a high incidence country[9]	138	96.8% (95% CI 88.0 to 100%)	LOW
Specificity										
Amplified M. Tuberculosis Direct Test	1[8]	cross-sectional	serious[2,3,4,5,7]	no serious inconsistency	no serious indirectness	no serious imprecision	No information available on industry involvement	60	100% (95% CI 100% to 100%)	MODERATE
Xpert MTB/RIF	1[1]	cross-sectional	serious[2,3,4,5]	no serious inconsistency	no serious indirectness	no serious imprecision	No industry involvement Conducted in a high incidence country[9]	138	98.4% (95% CI 96.1 to 100%)	MODERATE

[1] Bates, 2013
[2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two
[3] Unclear if consecutive or random sample of patients was enrolled
[4] Unclear if test interpretation was blinded
[5] Threshold for interpretation unclear
[6] Wide confidence interval
[7] Unclear if El-Sayed Zaki (2008) avoided inappropriate exclusions
[8] El-Sayed Zaki (2008)

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
[9] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000										

Interferon-gamma release assays compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
QuantiFERON-TB Gold In-Tube	1[1]	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7]	no serious imprecision	No industry involvement Conducted in a high incidence country[9]	5886	79.7 (72.7 to 86.7)	VERY LOW
Specificity										
QuantiFERON-TB Gold In-Tube	1[1]	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7]	serious[8]	No industry involvement Conducted in a high incidence country[9]	5886	16.7 (11.9 to 21.4)	VERY LOW
[1] Lodha, 2013 [2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two [3] Unclear if consecutive or random sample of patients was enrolled [4] Unclear if inappropriate exclusions were avoided [5] Unclear if test interpretation was blinded [6] Not all diagnoses were made with the same reference standard [7] Reference diagnoses could be made by microscopy alone [8] Significant variation in point estimates with little overlap in confidence intervals [9] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000										

Tuberculin skin tests compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
Mantoux	2[1,9]	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7]	no serious imprecision	Limited industry involvement, although 1 study did not provide any information on this Conducted in a high incidence country[9]	5543	89.8 (84.6 to 95.1)[1] 47[9]	VERY LOW
Specificity										
Mantoux	2[1,9]	cross-sectional	serious[2,3,4,5]	serious[6]	serious[7]	serious[8]	Limited industry involvement, although 1 study did not provide any information on this Conducted in a high incidence country[9]	5543	5.1 (2.3 to 8.0)[1] 60[9]	VERY LOW

[1] Lodha, 2013
[2] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two
[3] Unclear if consecutive or random sample of patients was enrolled in Lodha (2013) and Mahomed (2013)
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Not all diagnoses were made with the same reference standard
[7] Reference diagnoses could be made by microscopy alone
[8] Significant variation in point estimates with little overlap in confidence intervals
[9] Iriso, 2005
[10] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

WHO scoring system compared to culture-based reference standard in children and young people with suspected pulmonary tuberculosis

Test details	Number of evaluations	Quality assessment					Other considerations	Number of patients/ specimens	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
Sensitivity										
Scoring system	1[2]	cross-sectional	serious[3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	No information available on industry involvement Conducted in a high incidence country[7]	110	86%	LOW
Specificity										
Scoring system	1[2]	cross-sectional	serious[3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	No information available on industry involvement Conducted in a high incidence country[7]	110	22%	LOW

[1] WHO scoring system:
• duration of illness
• weight for age
• nutrition
• family history of tuberculosis
• tuberculin skin test
• unexplained fever and night sweats
• presence of lymph nodes, joint or bone swelling, abdominal mass or ascites, central nervous system signs, or kyphosis of the spine

[2] Iriso, 2005

[3] Both index test and reference standard performed in the every patient, with an appropriate period of time between the two

[4] Unclear if study avoided inappropriate exclusions

[5] Unclear if test interpretation was blinded

[6] Insufficient data to assess imprecision

[7] Countries/territories with an estimated incidence rate of 40 per 100,000 or greater are considered to have a high incidence of tuberculosis, as defined by Public Health England; current estimates of incidence for in the UK are 13.9 per 100,000

Time-to-detection

Test	Time	Reference
Time from obtaining specimen to reporting to clinician (median (interquartile range))		
Xpert MTB/RIF	0 (0–3) days	Zar, 2012
	1 (1–1) days	Zar, 2013
Culture	15 (12–20) days	Zar, 2012
	16 (13–19) days	Zar, 2013

A.4 RQ G

A.4.1 Diagnosis of active bone and joint tuberculosis

Use of interferon gamma release assays in the diagnosis of people with suspected bone and joint tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
1[1]	cross-sectional	serious[2]	serious[3]	no serious indirectness[3]	serious[4]	36	86.7% (95% CI 69.5 to 100%)	VERY LOW
Specificity								
1[1]	cross-sectional	serious[2]	serious[3]	no serious indirectness[3]	serious[4]	36	61.9% (95% CI 41.1 to 82.7%)	VERY LOW

[1] Lai, 2011
[2] Blinding of test interpretation unclear
[3] Patients received different reference standards
[4] Wide confidence interval

A.4.2 Diagnosis of active central nervous system tuberculosis

Use of microscopy in the diagnosis of people with suspected central nervous system tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
6[2,3,4,5,6,14]	cross-sectional	serious[8,9,10,11]	no serious inconsistency	no serious indirectness	serious[12,13]	706	Pooled sensitivity[1] (95% CI) = 68.8% (32.7 to 90.9%)	LOW
Specificity								
6[2,3,4,5,6,14]	cross-sectional	serious[8,9,10,11]	no serious inconsistency	no serious indirectness	no serious imprecision	706	See forest plot below[1,7]	MODERATE

[1] Forest plots for sensitivity and specificity:

[2] Al-Ateah, 2012
[3] Bonington, 2000
[4] Chedore and Jamieson, 2003
[5] Malbruny, 2011
[6] Teo, 2011
[7] Meta-analysis of relevant data not possible in STATA or R
[8] Unclear if a random or consecutive sample was used: Bonington, 2000; Chedore and Jamieson, 2003; Malbruny, 2011
[9] Unclear if test interpretation was blinded: Bonington, 2000; Chedore and Jamieson, 2003; Malbruny, 2011
[10] Interpretation of reference standard not blinded: Teo, 2011; Feng, 2014
[11] Unclear if a threshold for test interpretation was used and predefined: Bonington, 2000; Chedore and Jamieson, 2003; Malbruny, 2011
[12] Significant variation in point estimates with little overlap in confidence intervals
[13] Wide confidence interval
[14] Feng, 2014

Use of commercial nucleic acid amplification tests in the diagnosis of people with suspected central nervous system tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
29[1,7]	cross-sectional	serious[2,3,4]	serious[5]	no serious indirectness	serious[6]	2810	Pooled sensitivity (95% CI) = 70.6% (53.3 to 83.5%)	LOW
Specificity								
29[1,7,8]	cross-	serious[2,3,4]	serious[5]	no serious	serious[6]	2810	See forest plot below[1,8]	LOW

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
	sectional			indirectness				

[1] Forest plots for sensitivity and specificity:

[2] Unclear if a random or consecutive sample was used in a number of studies
[3] Unclear if test interpretation was blinded in a number of studies
[4] Unclear if a threshold for test interpretation was used and predefined in a number of studies
[5] A number of different reference standards were used, both across and within studies
[6] Significant variation in point estimates with little overlap in confidence intervals
[7] Systematic reviews: Denkinger (2014), Pai (2003); additional studies: Al-Ateah (2012), Bemer-Melchior (1998), Chedore and Jamieson (2003), Malbruny (2011), Teo (2011)
[8] Meta-analysis of relevant data not possible in STATA or R

Use of interferon gamma release assays in the diagnosis of people with suspected central nervous system tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
3[1,2,3,4]	cross-sectional	serious[6,7,8,9,10]	no serious inconsistency	no serious indirectness	serious[11]	141	Pooled sensitivity (95% CI) = 84.2% (71.9 to 91.7%)	LOW
Specificity								
3[1,2,3,4,5]	cross-sectional	serious[6,7,8,9,10]	no serious inconsistency	no serious indirectness	serious[11]	141	See forest plot below[1,5]	LOW

[1] Forest plots for sensitivity and specificity:

[2] Kim, 2008
[3] Liao, 2009
[4] Patel, 2010
[5] Meta-analysis of relevant data not possible in STATA or R
[6] Unclear if a random or consecutive sample was used: Liao, 2009
[7] Unclear if inappropriate exclusions were avoided
[8] Test interpretation unblinded: Kim, 2008
[9] Unclear if test interpretation was blinded: Liao, 2009
[10] PCR (one of the reference standards in Kim (2008)) is not a validated reference standard
[11] Some variation in point estimates with little overlap in confidence intervals

Use of tuberculin skin test in the diagnosis of people with suspected central nervous system tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
1[1]	cross-sectional	serious[2,3,4]	serious[5]	no serious indirectness	serious[6]	35	45.5% (95% CI 16.0 to 74.9%)	VERY LOW
Specificity								
1[1]	cross-sectional	serious[2,3,4]	serious[5]	no serious indirectness	serious[6]	35	66.7% (95% CI 47.8 to 85.5%)	VERY LOW

[1] Kim, 2008
[2] Unclear if inappropriate exclusions were avoided
[3] Test interpretation unblinded

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

[4] PCR is not a validated reference standard
[5] Patients did not all receive the same reference standard (PCR or culture)
[6] Wide confidence interval

Use of adenosine deanimase assays in the diagnosis of people with suspected central nervous system tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
Threshold for positivity: 4 U/l								
13[1,4]	cross-sectional	very serious[5,6,7]	serious[8]	no serious indirectness	no serious imprecision	1092	Pooled sensitivity (95% CI) = 92.7% (89.1 to 95.4%)	VERY LOW
Threshold for positivity: 8 U/l								
13[2,4]	cross-sectional	very serious[5,6,7]	serious[8]	no serious indirectness	serious[9]	1092	Pooled sensitivity (95% CI) = 63.0% (57.1 to 68.6%)	VERY LOW
Threshold for positivity: 10 U/l								
13[3,4]	cross-sectional	very serious[5,6,7]	serious[8]	no serious indirectness	serious[9]	1092	Pooled sensitivity (95% CI) = 49.5% (43.6 to 55.4%)	VERY LOW
Specificity								
Threshold for positivity: 4 U/l								
13[1,4]	cross-sectional	very serious[5,6,7]	serious[8]	no serious indirectness	serious[9]	1092	Pooled specificity (95% CI) = 72.3% (69.0 to 75.4%)	VERY LOW
Threshold for positivity: 8 U/l								
13[2,4]	cross-sectional	very serious[5,6,7]	serious[8]	no serious indirectness	serious[9]	1092	Pooled specificity (95% CI) = 84.8% (82.1 to 87.3%)	VERY LOW
Threshold for positivity: 10 U/l								
13[3,4]	cross-sectional	very serious[5,6,7]	serious[8]	no serious indirectness	serious[9]	1092	Pooled specificity (95% CI) = 90.7% (88.5 to 92.7%)	VERY LOW

[1] Forest plots for sensitivity and specificity at a threshold for positivity of 4 U/l:

[2] Forest plots for sensitivity and specificity at a threshold for positivity of 8 U/l:

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
[3] Forest plots for sensitivity and specificity at a threshold for positivity of 10 U/l: [4] Systematic review: Tuon (2010) [5] Included studies that used a case-control design [6] Unclear if inappropriate exclusions were avoided [7] 10 of the 13 studies included were not blinded [8] Different reference standards used in each study [9] Significant variation in point estimates with little overlap in confidence intervals								

A.4.3 Diagnosis of active genitourinary tuberculosis

Use of microscopy in the diagnosis of people with suspected genitourinary tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
2[1,2,3]	cross-sectional	serious[5,6,7,8]	no serious inconsistency	no serious indirectness	serious[9]	72	Pooled sensitivity (95% CI) = 36.3% (19.2 to 57.8%)	LOW
Specificity								
2[1,2,3,4]	cross-sectional	serious[5,6,7,8]	no serious inconsistency	no serious indirectness	no serious imprecision	72	See forest plot below[1,4]	MODERATE

[1] Forest plots for sensitivity and specificity:

[2] Hemal, 2000
[3] Lai, 2010
[4] Meta-analysis of relevant data not possible in STATA or R
[5] Unclear if a consecutive or random sample was used
[6] Unclear if inappropriate exclusions were avoided
[7] Unclear if test interpretation was blinded
[8] Unclear if a threshold for index test interpretation was used
[9] Wide confidence interval

Use of radiology[1] in the diagnosis of people with suspected genitourinary tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
1[2]	cross-sectional	serious[3,4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	42	91.4% (95% CI 82.2 to 100%)	MODERATE
Specificity								
1[2]	cross-sectional	serious[3,4,5,6]	no serious inconsistency	no serious indirectness	serious7	42	28.6% (95% CI 0.0 to 62.0%)	LOW

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

[1] Includes renal calcification, caliceal destruction, infundibular stenosis, cavitation, ureteral stricture, vesicoureteral reflux and small capacity bladder
[2] Hemal, 2000
[3] Unclear if a consecutive or random sample was used
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded
[6] Unclear if a threshold for index test interpretation was used
[7] Wide confidence interval

Use of commercial nucleic acid amplification tests in the diagnosis of people with suspected genitourinary tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
4[1,2,3,4,6]	cross-sectional	serious[7,8,9,10]	no serious inconsistency	no serious indirectness	serious[11,12]	208	Pooled sensitivity (95% CI) = 56.9% (34.9 to 76.4%)	LOW
Specificity								
4[1,2,3,4,5,6]	cross-sectional	serious[7,8,9,10]	no serious inconsistency	no serious indirectness	serious[11,12]	208	See forest plot below[1,5]	LOW

[1] Forest plots for sensitivity and specificity:

[2] Gamboa, 1997
[3] Gamboa, 1998 (2 evaluations)
[4] Zambardi, 1995
[5] Meta-analysis of relevant data not possible in available statistical software
[6] Systematic review: Dinnes (2007)
[7] Unclear if a consecutive or random sample was used
[8] Unclear if inappropriate exclusions were avoided
[9] Unclear if test interpretation was blinded in 3 of the 4 evaluations
[10] Unclear if a threshold for index test interpretation was used
[11] Significant variation in point estimates with little overlap in confidence intervals
[12] Wide confidence interval

Use of interferon gamma release assays in the diagnosis of people with suspected genitourinary tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	30	91.7% (95% CI 76.0 to 100%)	LOW
Specificity								
1[1]	cross-sectional	serious[2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	30	88.9% (95% CI 74.4 to 100%)	LOW

[1] Lai, 2010
[2] Unclear if a consecutive or random sample was used
[3] Unclear if inappropriate exclusions were avoided
[4] Unclear if test interpretation was blinded
[5] Wide confidence interval

A.4.4 Diagnosis of active gastrointestinal tuberculosis

Use of microscopy in the diagnosis of people with suspected gastrointestinal tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
3[1,2,3]	cross-sectional	serious[5,6,7,8,9]	serious12	no serious indirectness	serious[10,11]	124	Pooled sensitivity (95% CI) = 42.4% (12.2 to 79.6%)	VERY LOW
HIV-negative								
1[3]	cross-sectional	serious[5,6,7,8,9]	no serious inconsistency	no serious indirectness	serious[11]	41	85.7% (95% CI 35.6 to 98.5%)	LOW
Specificity								
3[1,2,3,4]	cross-sectional	serious[5,6,7,8,9]	serious[12]	no serious indirectness	serious[10]	124	See forest plot below[1,4]	VERY LOW
HIV-negative								
1[3]	cross-sectional	serious[5,6,7,8,9]	no serious inconsistency	no serious indirectness	serious[11]	41	71.1% (95% CI 55.2 to 83.0%)	LOW

[1] Forest plots for sensitivity and specificity:

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

[2] Cho, 2011
[3] Saleh, 2012
[4] Meta-analysis of relevant data not possible in available statistical software
[5] Unclear if a consecutive or random sample used: Saleh (2012)
[6] Unclear if inappropriate exclusions were avoided
[7] Unclear if index test interpretation was blinded
[8] Unclear if reference standard interpretation was blinded: Saleh (2012)
[9] Unclear if a threshold for test interpretation was used
[10] Significant variation in point estimates with little overlap in confidence intervals
[11] Wide confidence interval
[12] Patients received different reference standards: Cho (2011)

Use of interferon gamma release assays in the diagnosis of people with suspected gastrointestinal tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
14[1,2]	cross-sectional	serious[3,4,5]	serious[6]	no serious indirectness	serious[7]	965	Pooled sensitivity (95% CI) = 89.7% (82.6 to 94.1%)	VERY LOW
Specificity								
14[1,2]	cross-sectional	serious[3,4,5]	serious[6]	no serious indirectness	serious[7,8]	965	Pooled specificity (95% CI) = 93.3% (82.9 to 97.6%)	VERY LOW

[1] Forest plots for sensitivity and specificity:

[2] Systematic review: Su (2013); additional studies: Cho (2011), Liao (2009)
[3] Unclear if a consecutive or random sample used: Liao (2009)
[4] Unclear if inappropriate exclusions were avoided
[5] Unclear if test interpretation was blinded in a number of studies
[6] Patients received different reference standards, both within and between studies
[7] Significant variation in point estimates with little overlap in confidence intervals
[8] Wide confidence interval

Use of adenosine deanimase assays in the diagnosis of people with suspected gastrointestinal tuberculosis

Number of	Quality assessment	Number of	Summary of findings	Quality

evaluations	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	patients/ specimens		
Sensitivity								
17[1,5,6]	cross-sectional	serious[7,8,9,10]	serious[11]	serious[12]	serious[13]	1617	Pooled sensitivity (95% CI) = 94.9% (89.7 to 97.5%)	VERY LOW
Threshold for positivity: <10 U/l								
1[2,5]	cross-sectional	serious[7,8,9,10]	no serious inconsistency	serious[12]	serious[14]	368	58.8% (95% CI 35.4 to 82.2%)	VERY LOW
Threshold for positivity: 20 to 29 U/l								
1[3,5]	cross-sectional	serious[7,8,9,10]	no serious inconsistency	serious[12]	no serious imprecision	52	92.6% (95% CI 82.7 to 100%)	LOW
Threshold for positivity: >30 U/l								
15[4,5,6]	cross-sectional	serious[7,8,9,10]	serious[11]	serious[12]	serious[13]	1197	Pooled sensitivity (95% CI) = 94.7% (91.5 to 96.7%)	VERY LOW
Specificity								
17[1,5,6]	cross-sectional	serious[7,8,9,10]	serious[11]	serious[12]	serious[13]	1617	Pooled specificity (95% CI) = 96.2% (93.9 to 97.7%)	VERY LOW
Threshold for positivity: <10 U/l								
1[2,5]	cross-sectional	serious*	serious*	serious*	no serious imprecision	368	95.4% (95% CI 93.3 to 97.6%)	VERY LOW
Threshold for positivity: 20 to 29 U/l								
1[3,5]	cross-sectional	serious*	serious*	serious*	serious[14]	52	84.0% (95% CI 69.6 to 98.4%)	VERY LOW
Threshold for positivity: >30 U/l								
15[4,5,6]	cross-sectional	serious[7,8,9,10]	serious[11]	serious[12]	serious[13]	1197	Pooled specificity (95% CI) = 96.7% (94.3 to 98.1%)	VERY LOW

[1] Forest plots for sensitivity and specificity:

[2] Hillebrand, 1996
[3] Kang, 2012
[4] Forest plots for sensitivity and specificity at a threshold for positivity of >30 U/l

[5] Systematic review: Shen (2013)
[6] Additional study: Brant (1995)
[7] Unclear if a consecutive or random sample used
[8] Unclear if inappropriate exclusions were avoided

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

[9] Unclear if test interpretation was blinded
[10] Unclear if a threshold for test interpretation was used
[11] Patients received different reference standards, both within and between studies
[12] Review included inappropriate reference standards
[13] Significant variation in point estimates with little overlap in confidence intervals
[14] Wide confidence interval

A.4.5 Diagnosis of active lymph node tuberculosis

Use of microscopy in the diagnosis of people with suspected lymph node tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
7[1,2]	cross-sectional	serious[5,6,7]	no serious inconsistency	no serious indirectness	serious[8,9]	799	Pooled sensitivity (95% CI) = 36.4% (27.5 to 46.5%)	LOW
Children								
1[3]	cross-sectional	serious[5,6,7]	no serious inconsistency	no serious indirectness	serious[9]	129	44.3% (95% CI 33.9 to 54.7%)	LOW
HIV-positive								
1[4]	cross-sectional	serious[6,7]	no serious inconsistency	no serious indirectness	serious[9]	344	51.0% (95% CI 43.0 to 59.0%)	LOW
Specificity								
7[1,2]	cross-sectional	serious[5,6,7]	no serious inconsistency	no serious indirectness	serious[8,9]	799	Pooled specificity (95% CI) = 94.4% (78.4 to 98.8%)	LOW
Children								
1[3]	cross-sectional	serious[5,6,7]	no serious inconsistency	no serious indirectness	serious[9]	129	58.5% (95% CI 43.5 to 73.6%)	LOW
HIV-positive								
1[4]	cross-sectional	serious[6,7]	no serious inconsistency	no serious indirectness	serious[9]	344	96.0% (95% CI 93.1 to 98.7%)	LOW

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

[1] Forest plots for sensitivity and specificity:

[2] Fanny, 2012; Gamboa, 1997a; Gamboa, 1997b; Kerleguer, 2004; Malbruny, 2011; Rimek, 2002; Van Rie, 2013
[3] Fanny, 2012
[4] Van Rie, 2013
[5] Unclear if a random or consecutive sample used in a number of studies
[6] Unclear if a inappropriate exclusions were avoided in a number of studies
[7] Unclear if test interpretation was blinded in any of the studies
[8] Significant variation in point estimates with little overlap in confidence intervals
[9] Wide confidence interval

Use of cytology[1] in the diagnosis of people with suspected lymph node tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	250	99.2% (95% CI 97.7 to 100%)	HIGH
Specificity								
1[2]	cross-sectional	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	250	49.2% (95% CI 40.2 to 58.1%)	MODERATE

[1] Including the presence or absence of granulomas, Langerhan's giant cells, plasma cells, lymphocytes, macrophages, neutrophils and necrosis; the cytological criteria for diagnosis of tuberculous lymphadenitis were defined as epithelioid cell granulomas with or without multinucleate giant cells and caseation necrosis
[2] Nataraj, 2002
[3] Wide confidence interval

Use of commercial nucleic acid amplification tests in the diagnosis of people with suspected lymph node tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
26[1,2]	cross-sectional	serious[4,5,6,7]	serious[8]	no serious indirectness	serious[9,10]	1824	Pooled sensitivity (95% CI) = 86.5% (78.5 to 91.8%)	VERY LOW

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
HIV-positive								
1[3]	cross-sectional	serious[5,6,7]	no serious inconsistency	no serious indirectness	no serious imprecision	344	85.8% (95% CI 80.4 to 91.2%)	MODERATE
Specificity								
26[1,2]	cross-sectional	serious[4,5,6,7]	serious[8]	no serious indirectness	serious[9,10]	1824	Pooled specificity (95% CI) = 92.4% (88.7 to 95.0%)	VERY LOW
HIV-positive								
1[3]	cross-sectional	serious[5,6,7]	no serious inconsistency	no serious indirectness	no serious imprecision	344	94.5% (95% CI 91.2 to 97.8%)	MODERATE

[1] Forest plots for sensitivity and specificity:

[2] Systematic reviews: Denkinger (2014), Dinnes (2007); additional studies: Gamboa (1997b), Kerleguer (2004), Lithelm (2011), Malbruny (2011), Osores (2006), Pfyffer (1996), Van Rie (2013)
[3] Van Rie, 2013
[4] Unclear if a random or consecutive sample used in a number of studies
[5] Unclear if inappropriate exclusions were avoided in a number of studies
[6] Unclear if test interpretation was blinded in any of the studies
[7] Unclear if the threshold for test positivity was predefined in a number of studies
[8] Patients received different reference standards, both within and across studies
[9] Significant variation in point estimates with little overlap in confidence intervals
[10] Wide confidence interval

A.4.6 Diagnosis of active pericardial tuberculosis

Use of commercial nucleic acid amplification tests in the diagnosis of people with suspected pericardial tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
2[1,3,4]	cross-sectional	serious[5,6,7]	serious[8]	no serious indirectness	serious[9,10]	115	Pooled sensitivity (95% CI) = 51.5% (13.8 to 87.6%)	VERY LOW
Specificity								
2[1,2,3,4]	cross-sectional	serious[5,6,7]	serious[8]	no serious indirectness	no serious imprecision	115	See forest plot below[1,4]	LOW

[1] Forest plots for sensitivity and specificity:

[2] Meta-analysis of relevant data not possible in available statistical software
[3] Lee, 2002
[4] Reuter, 2006
[5] Unclear if index test interpretation was blinded: Reuter (2006)
[6] Unclear if reference standard was blinded
[7] Unclear if a threshold for test positivity was predefined: Reuter (2006)
[8] Different reference standards used, both within and across studies
[9] Significant variation in point estimates with little overlap in confidence intervals
[10] Wide confidence interval

Use of adenosine deanimase assays in the diagnosis of people with suspected pericardial tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
5[1,2]	cross-sectional	serious[3,4,5]	serious[6]	no serious indirectness	serious[7]	421	Pooled sensitivity (95% CI) = 88% (82 to 91%)	VERY LOW

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Specificity								
5[1,2]	cross-sectional	serious[3,4,5]	serious[6]	no serious indirectness	serious[7]	421	Pooled specificity (95% CI) = 83% (78 to 88%)	VERY LOW

[1] Forest plots for sensitivity and specificity:

[2] Tuon, 2006
[3] Unclear if inappropriate exclusions were avoided
[4] Unclear if test interpretation was blinded
[5] Thresholds for test positivity not always predefined
[6] Different reference standards used, both within and across studies
[7] Some variation in point estimates with little overlap in confidence intervals

Use of tuberculin skin tests in the diagnosis of people with suspected pericardial tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
Threshold for positivity: 10 mm								
1[1]	cross-sectional	serious[2,3,4]	serious[5]	no serious indirectness	serious[6]	52	88.9% (95% CI 78.6 to 99.2%)	VERY LOW
Threshold for positivity: 15 mm								
1[1]	cross-sectional	serious[2,3,4]	serious[5]	no serious indirectness	serious[6]	52	44.4% (95% CI 28.2 to 60.7%)	VERY LOW
Specificity								
Threshold for positivity: 10 mm								
1[1]	cross-sectional	serious[2,3,4]	serious[5]	no serious indirectness	serious[6]	52	56.3% (95% CI 31.9 to 80.6%)	VERY LOW
Threshold for positivity: 15 mm								
1[1]	cross-sectional	serious[2,3,4]	serious[5]	no serious indirectness	no serious imprecision	52	93.8% (95% CI 81.9 to 100%)	LOW

[1] Reuter, 2006
[2] Unclear if index test interpretation was blinded
[3] Unclear if reference standard was blinded

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

[4] Unclear if a threshold for test positivity was predefined
[5] Different reference standards used
[6] Wide confidence interval

A.4.7 Diagnosis of active pleural tuberculosis

Use of microscopy in the diagnosis of people with suspected pleural tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity[1]								
6[2,3,4,5,6]	cross-sectional	serious[8,9,10,11]	no serious inconsistency	no serious indirectness	serious[12,13]	294	Pooled sensitivity (95% CI) = 10.5% (3.7 to 26.4%)	LOW
Specificity[1,7]								
6[2,3,4,5,6]	cross-sectional	serious[8,9,10,11]	no serious inconsistency	no serious indirectness	no serious imprecision	294	See forest plot below[1,7]	MODERATE

[1] Forest plots for sensitivity and specificity:

[2] Gamboa, 1997a
[3] Gamboa, 1997b
[4] Hasaneen, 2003 (2 evaluations)
[5] Malbruny, 2011
[6] Maurya, 2011
[7] Meta-analysis of relevant data not possible in STATA or R
[8] Unclear if a random or consecutive sample was used
[9] Unclear if inappropriate exclusions were avoided in all studies
[10] Unclear if test interpretation was blinded in most studies
[11] Unclear if the test positivity threshold was predefined in most studies
[12] Some variation in point estimates with little overlap in confidence intervals
[13] Wide confidence interval

Use of commercial nucleic acid amplification tests in the diagnosis of people with suspected pleural tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity[1]								
26[2 to 14]	cross-sectional	serious[16,17,18,19]	serious[20]	no serious indirectness	serious[21,22]	1686	Pooled sensitivity (95% CI) = 53.0% (33.2 to 71.9%)	VERY LOW
Specificity[1]								
26[2 to 14]	cross-sectional	serious[16,17,18,19]	serious[20]	no serious indirectness	no serious imprecision	1686	Pooled specificity (95% CI) = 99.4% (98.1 to 99.8%)	LOW

[1] Forest plots for sensitivity and specificity:

[2] Artiles, 2001
[3] Bemer-Melchior, 1998
[4] D'Amato, 1996
[5] Dheda, 2009
[6] Ehlers, 1996
[7] Gamboa, 1997a
[8] Gamboa, 1997b
[9] Malbruny, 2011
[10] Mitarai, 2000
[11] Pfyffer, 1996
[12] Reischl, 1998
[13] Shah, 1998
[14] Vlaspolder, 1995
[15] Systematic reviews: Denkinger (2014), Pai (2004)
[16] Random or consecutive sample not used in all studies
[17] Unclear if inappropriate exclusions avoided
[18] Blinding of test interpretation not performed in all studies
[19] Unclear if threshold for test positivity predefined in all studies
[20] Different reference standards used, both within and across studies
[21] Significant variation in point estimates with little overlap in confidence intervals
[22] Wide confidence interval

Use of cytology[1] in the diagnosis of people with suspected pleural tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
1[2]	cross-	serious[3,4]	no serious	no serious	serious[5]	45	Sensitivity (95% CI) = 53.9%	LOW

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
	sectional		inconsistency	indirectness			(34.7 to 73.0%)	
Specificity								
1[2]	cross-sectional	serious[3,4]	no serious inconsistency	no serious indirectness	no serious imprecision	45	Specificity (95% CI) = 97.4% (90.4 to 100%)	MODERATE

[1] Histopathologic examination of pleural biopsy specimen fixed in formalin for caseating granuloma
[2] Hasaneen, 2003
[3] Unclear if a random or consecutive sample was used
[4] Unclear if the test positivity threshold was predefined
[5] Wide confidence interval

Use of interferon gamma release assays in the diagnosis of people with suspected pleural tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity[1]								
5[2,3,4]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[13,14]	150	Pooled sensitivity (95% CI) = 75.5% (60.9 to 85.8%)	VERY LOW
Specificity[1,5]								
5[2,3,4]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[13,14]	16[2] 23[2] 40[3] 39[3] 32[4]	See forest plot below[1,5]	VERY LOW

[1] Forest plots for sensitivity and specificity:

[2] Baba, 2008 (2 evaluations)
[3] Lee, 2009 (2 evaluations)
[4] Liao, 2009
[5] Meta-analysis of relevant data not possible in STATA or R
[8] Unclear if a random or consecutive sample was used
[9] Unclear if inappropriate exclusions were avoided in all studies
[10] Unclear if test interpretation was blinded
[11] Unclear if the test positivity threshold was predefined in Lee (2009)
[12] Different reference standards used, both within and across studies
[13] Significant variation in point estimates with little overlap in confidence intervals

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
[14] Wide confidence interval								

Use of lipoarabinomannan assays in the diagnosis of people with suspected pleural tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
Threshold for positivity: 30 g/l								
1[1]	cross-sectional	serious[2,3]	serious[4]	no serious indirectness	serious[5]	50	Sensitivity (95% CI) = 93.8% (86.9 to 100%)	VERY LOW
Threshold for positivity: 60 g/l								
1[1]	cross-sectional	serious[2,3]	serious[4]	no serious indirectness	serious[5]	50	Sensitivity (95% CI) = 91.7% (80.6 to 100%)	VERY LOW
Specificity								
Threshold for positivity: 30 g/l								
1[1]	cross-sectional	serious[2,3]	serious[4]	no serious indirectness	serious[5]	50	Specificity (95% CI) = 11.5% (0.0 to 23.8%)	VERY LOW
Threshold for positivity: 60 g/l								
1[1]	cross-sectional	serious[2,3]	serious[4]	no serious indirectness	serious[5]	50	Specificity (95% CI) = 92.3% (82.1 to 100%)	VERY LOW

[1] Dheda, 2009
[2] Unclear if test interpretation was blinded
[3] Unclear if the test positivity threshold was predefined
[4] Different reference standards used
[5] Wide confidence interval

Use of adenosine deanimase assays in the diagnosis of people with suspected pleural tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
65[1,11]	cross-	serious[8,9,10,11]	serious[12]	no serious	serious[13]	8222	Pooled sensitivity (95% CI) =	VERY LOW

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
	sectional			indirectness			94.2% (91.5 to 96.0%)	
Threshold for positivity: 10 to <15 U/l2								
1[3,11]	cross-sectional	serious[10,11]	serious[12]	no serious indirectness	no serious imprecision	74	Sensitivity (95% CI) = 99.0% (90.9 to 99.9%)	LOW
Threshold for positivity: 15 to <20 U/l4								
1[5,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	no serious imprecision	69	Sensitivity (95% CI) = 95.7% (85.8 to 98.8%)	LOW
Threshold for positivity: 30 to 35 U/l								
19[6,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[13]	1461	Pooled sensitivity (95% CI) = 94.2% (88.2 to 97.2%)	VERY LOW
Threshold for positivity: >35 to 40 U/l								
15[7,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[13]	1951	Pooled sensitivity (95% CI) = 94.3% (89.1 to 97.1%)	VERY LOW
Threshold for positivity: >40 to 45 U/l								
9[8,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[14]	1203	Pooled sensitivity (95% CI) = 89.5% (79.7 to 94.9%)	VERY LOW
Threshold for positivity: >45 to 50 U/l								
14[9,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[13]	2072	Pooled sensitivity (95% CI) = 92.6% (84.1 to 96.8%)	VERY LOW
Threshold for positivity: >50 U/l								
7[10,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	no serious imprecision	1448	Pooled sensitivity (95% CI) = 98.1% (88.3 to 99.7%)	LOW
Specificity[1]								
65[1,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[13]	8222	Pooled specificity (95% CI) = 91.3% (89.1 to 93.1%)	VERY LOW
Threshold for positivity: 10 to <15 U/l2								
1[3,11]	cross-sectional	serious[10,11]	serious[12]	no serious indirectness	serious[14]	74	Specificity (95% CI) = 38.5% (22.4 to 57.5%)	VERY LOW
Threshold for positivity: 15 to <20 U/l4								
1[5,11]	cross-	serious[8,9,10,11]	serious[12]	no serious	serious[14]	69	Specificity (95% CI) = 90.9%	VERY LOW

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
	sectional			indirectness			(72.2 to 97.5%)	
Threshold for positivity: 30 to 35 U/l								
19[6,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[13]	1461	Pooled specificity (95% CI) = 94.0% (89.3 to 96.7%)	VERY LOW
Threshold for positivity: >35 to 40 U/l								
15[7,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[13]	1951	Pooled specificity (95% CI) = 90.4% (83.3 to 94.7%)	VERY LOW
Threshold for positivity: >40 to 45 U/l								
9[8,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	no serious imprecision	1203	Pooled specificity (95% CI) = 93.0% (89.4 to 95.4%)	LOW
Threshold for positivity: >45 to 50 U/l								
14[9,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	serious[13]	2072	Pooled specificity (95% CI) = 87.7% (82.1 to 91.7%)	VERY LOW
Threshold for positivity: >50 U/l								
7[10,11]	cross-sectional	serious[8,9,10,11]	serious[12]	no serious indirectness	no serious imprecision	1448	Pooled specificity (95% CI) = 91.7% (87.8 to 94.4%)	LOW

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

[1] Forest plots for sensitivity and specificity:

[2] 13 U/l
[3] Dheda, 2009
[4] 20 U/l
[5] Andreasyan
[6] Forest plots for sensitivity and specificity at a threshold for positivity of 30 to 35 U/l:

[7] Forest plots for sensitivity and specificity at a threshold for positivity of >35 to 40 U/l:

[8] Forest plots for sensitivity and specificity at a threshold for positivity of >40 to 45 U/l:

[9] Forest plots for sensitivity and specificity at a threshold for positivity of >45 to 50 U/l:

[10] Forest plots for sensitivity and specificity at a threshold for positivity of >50 U/l:

[11] Systematic review: Liang, 2008
[8] Unclear if a random or consecutive sample was used
[9] Unclear if inappropriate exclusions were avoided in all studies
[10] Unclear if test interpretation was blinded
[11] Unclear if the test positivity threshold was predefined
[12] Different reference standards used, both within and across studies
[13] Significant variation in point estimates with little overlap in confidence intervals
[14] Wide confidence interval

Use of adenosine deanimase assays in conjunction with the lymphocyte-neutrophil ratio in the diagnosis of people with suspected pleural tuberculosis

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Sensitivity								
1[1]	cross-sectional	serious[2,3]	serious[4]	no serious indirectness	no serious imprecision	303	88.1% (95% CI 82.8 to 93.4%)	LOW
Specificity								
1[1]	cross-sectional	serious[2,3]	serious[4]	no serious indirectness	no serious imprecision	303	95.0% (95% CI 91.6 to 98.4%)	LOW

[1] Burgess, 1996

Number of evaluations	Quality assessment					Number of patients/ specimens	Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

[2] Unclear if test interpretation was blinded
[3] Test positivity threshold was not predefined
[4] Different reference standards used

Appendix E: GRADE profiles

A.5 RQ I

A.5.1 Dosing frequencies in children

Intervention: **daily (unsupervised) dosing**

Comparator: **intermittent (DOT) dosing**

Site of tuberculosis: **pulmonary/intrathoracic**

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Daily (unsupervised) dosing	Intermittent (DOT) dosing	Relative (95% CI)	Absolute (95% CI)	Quality
Relapse (number to experience clinical or radiological recurrence; follow-up 24 to 60 months[1])										
2[2,3]	randomised trials	very serious[4,5,6,7,8,9]	very serious[10,11,12]	serious[13]	very serious[14,15]	1/184 (0.54%)	1/155 (0.65%)	OR 0.87 (0.08 to 9.85)[16,17]	0 fewer per 100 (from 1 fewer to 5 more)	VERY LOW

[1] *After treatment initiation*
[2] *Te Water Naude et al, 2000*
[3] *Swaminathan et al, 2005*
[4] *Te Water Naude et al, 2000: randomisation not appropriate; conducted by household unit, analysis is at the level of the individual (i.e. unit-of-analysis error); insufficient data to correct*
[5] *Swaminathan et al, 2005: method of randomisation unclear*
[6] *Allocation concealment unclear*
[7] *Te Water Naude et al, 2000: blinding absent or unclear*
[8] *Swaminathan et al, 2005: aside from the blinding of the radiologist and paediatrician assessed the children's chest x-rays, blinding is unclear*
[9] *Analysis did not follow the intent-to-treat principle*
[10] *Te Water Naude et al, 2000: 'weight for age' and the 'number who were culture positive' was significantly lower in the intermittent group*
[11] *In addition to the use of different treatments, the two groups received different care: the intermittent regimens were supervised in the clinic, whereas the daily regimens were not supervised except on the day of medication collection*
[12] *Point estimate varies widely across studies*
[13] *Interventions and comparators vary by more than dosing frequency; that is, the intervention studied does not precisely match the intervention of interest*
[14] *GRADE rule of thumb event number <300*
[15] *Wide confidence intervals*
[16] *Odds ratio and 95% confidence intervals not provided by authors; calculated by reviewer*
[17] *Forest plot:*

Abbreviations: CI, confidence interval; DOT, directly observed therapy; OR, odds ratio

Intervention: **daily (unsupervised) dosing**

Comparator: **twice-weekly (DOT) followed by thrice-weekly (DOT) dosing**

Site of tuberculosis: **pulmonary**

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Daily (unsupervised) dosing	Twice-weekly (DOT) followed by thrice-weekly (DOT) dosing	Relative (95% CI)	Absolute (95% CI)	Quality
Mortality (number of tuberculosis-related deaths during the study; follow-up 24 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,8]	serious[9]	very serious[10,11]	1/68 (1.5%)	2/69 (2.9%)	OR 0.50 (0.04 to 5.65)[12]	1 fewer per 100 (from 3 fewer to 12 more)	VERY LOW
Response to treatment - disease resolution (% of participants with a normal chest radiograph at treatment completion; follow-up 6-9 months[13])										
1[14]	randomised trials	very serious[3,4,5,15]	serious[7,8]	very serious[9,16]	serious[10]	61%	48%	OR 1.69 (0.97 to 2.97)[12]	13 more per 100 (from 1 fewer to 25 more)	VERY LOW
Response to treatment - disease resolution (% of participants with a normal chest radiograph at 60 months)										
1[14]	randomised trials	very serious[3,4,5,15]	serious[7,8]	very serious[9,16]	serious[10]	82%	89.5%	OR 0.54 (0.20 to 1.48)[12]	7 fewer per 100 (from 27 fewer to 3 more)	VERY LOW
Response to treatment - disease resolution (% of participants with residual lesions at treatment completion; follow-up 6-9 months[13])										
1[14]	randomised trials	very serious[3,4,5,15]	serious[7,8]	very serious[9,16]	serious[10]	39%	49%	OR 0.67 (0.38 to 1.17)[12]	10 fewer per 100 (from 22 fewer to 4 more)	VERY LOW
Response to treatment - disease resolution (% of participants with residual lesions at 60 months)										
1[14]	randomised trials	very serious[3,4,5,15]	serious[7,8]	very serious[9,16]	very serious[10,11]	15%	1.5%	OR 11.40 (1.42 to 91.85)[12]	13 more per 100 (from 1 more to 57 more)	VERY LOW
Response to treatment - disease resolution (number of participants to require treatment extension due to incomplete resolution)										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,8]	very serious[9,16]	very serious[10,11]	5/68 (7.4%)	4/69 (5.8%)	OR 1.29 (0.33 to 5.02)[12]	2 more per 100 (from 4 fewer to 18 more)	VERY LOW
Relapse (number to experience clinical or radiological recurrence; follow-up 60 months)										
1[14]	randomised trials	very serious[3,4,5,15]	serious[7,8]	serious[9]	very serious[10,11]	1/67 (1.5%)	0/66 (0%)	OR 3.00 (0.12 to 74.98)[12]	-	VERY LOW
Adverse events - hepatotoxicity (number of patients to experience hepatotoxicity; follow-up 24 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,8]	serious[9]	very serious[10,11]	2/68 (2.9%)	1/69 (1.4%)	OR 2.06 (0.18 to 23.27)[12]	1 more per 100 (from 1 fewer to 24 more)	VERY LOW

[1] *After treatment completion*
[2] *Ramachrandan et al, 1998*
[3] *Method of randomisation and the use of allocation concealment was unclear*
[4] *The groups were not comparable at baseline – more patients in the intermittent group had cavitatory disease at baseline, a sign that the disease in this group may have been more severe at treatment initiation*
[5] *Aside from the blinding of the radiologist and paediatrician assessed the children's chest x-rays, blinding is unclear*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Daily (unsupervised) dosing	Twice-weekly (DOT) followed by thrice-weekly (DOT) dosing	Relative (95% CI)	Absolute (95% CI)	Quality

[6] *Unclear if analysis follows the intent-to-treat principle*
[7] *In addition to the use of different treatments, the two groups received different care: the thrice-weekly followed by twice-weekly regimen was supervised in the clinic, whereas the daily regimen was not supervised except on the day of medication collection*
[8] *The loss to follow-up in each group is unclear*
[9] *Intervention and comparator vary by more than dosing frequency; that is, the intervention studied does not precisely match the intervention of interest*
[10] *GRADE rule of thumb event number <300*
[11] *Wide confidence intervals*
[12] *Odds ratio and 95% confidence intervals not provided by authors; calculated by reviewer*
[13] *Treatment period*
[14] *Swaminathan et al, 2005*
[15] *Analysis did not follow the intent-to-treat principle*
[16] *Outcome is a substitute for the outcome of interest (cure, treatment success and treatment failure)*
Abbreviations: CI, confidence interval; DOT, directly observed therapy; OR, odds ratio

Intervention: **twice-weekly (DOT) dosing**

Comparator: **daily (Monday-Friday) (unsupervised) dosing**

Site of tuberculosis: **intrathoracic**

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Daily (Monday-Friday) (unsupervised) dosing	Twice-weekly (DOT) dosing	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment 3 months after treatment initiation (measured with: composite score obtained from parent assessment, clinical symptoms, weight gain and chest radiograph; range of scores: -4-8; better indicated by higher values)										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	very serious[8,9]	no serious imprecision[10]	89	70	-	median difference 0[11]	VERY LOW
Response to treatment at treatment completion (measured with: composite score obtained from parent assessment, clinical symptoms, weight gain and chest radiograph; range of scores: -4-8; better indicated by higher values; follow-up 6 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	very serious[8,9]	no serious imprecision[10]	93	70	-	median difference 0[11]	VERY LOW
Response to treatment 6 months after treatment completion (measured with: composite score obtained from parent assessment, clinical symptoms, weight gain and chest radiograph; range of scores: -4-8; better indicated by higher values; follow-up 12 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	very serious[8,9]	no serious imprecision[10]	74	65	-	median 1 higher[11]	VERY LOW
Response to treatment 12-24 months after treatment completion (measured with: composite score obtained from parent assessment, clinical symptoms, weight gain and chest radiograph; range of scores: -4-8; better indicated by higher values; follow-up 18-30 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	very serious[8,9]	no serious imprecision[10]	74	71	-	median difference 0[11]	VERY LOW
Symptom improvement - weight gain (weight gain from treatment initiation until treatment completion; better indicated by higher values; follow-up 6 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	serious[8]	no serious imprecision[10]	-[12]	-[12]	-	median 0.25 kg higher[11]	VERY LOW
Relapse (number to experience clinical or radiological recurrence; follow-up 30 months[1])										

Appendix E: GRADE profiles

Quality assessment						Number of patients		Effect		
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Daily (Monday-Friday) (unsupervised) dosing**	**Twice-weekly (DOT) dosing**	**Relative (95% CI)**	**Absolute (95% CI)**	**Quality**
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,13]	serious[8]	very serious[14,15]	0/117 (0%)	1/89 (1.1%)	OR 0.25 (0.01 to 6.24)[16]	1 fewer per 100 (from 1 fewer to 5 more)	VERY LOW
Adherence - treatment completion (number to complete treatment on schedule; follow-up 6 months[1,17])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,13]	serious[8]	serious[14]	114/117 (97.4%)	85/89 (95.5%)	OR 1.79 (0.39 to 8.20)[16]	2 more per 100 (from 6 fewer to 4 more)	VERY LOW
Adherence - number adherent (number of children taking ≥75% of the prescribed doses; follow-up 6 months[1,17])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,13]	serious[8]	serious[14]	90/117 (76.9%)	70/89 (78.7%)	OR 0.90 (0.47 to 1.76)[16]	2 fewer per 100 (from 15 fewer to 8 more)	VERY LOW
Adherence - number partially adherent (number of children taking ≥75% of the prescribed doses but <75% during any single 4-week period; follow-up 6 months[1,17])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,13]	serious[8]	serious[14]	30/117 (25.6%)	21/89 (23.6%)	OR 1.12 (0.59 to 2.12)[16]	2 more per 100 (from 8 fewer to 16 more)	VERY LOW
Adherence - time to default by non-adherers (days to default by non-adherers, defined as those taking <75% of the prescribed doses; better indicated by higher values; follow-up 6 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,13]	serious[8]	no serious imprecision[10]	117	89	-	median 30 days lower[11]	VERY LOW
Adherence - time to default by partial adherers (days to default by partial adherers, defined those taking ≥75% of the prescribed doses but <75% during any single 4-week period; better indicated by higher values; follow-up 6 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,13]	serious[8]	no serious imprecision[10]	117	89	-	median 23 days lower[11]	VERY LOW
Adherence - proportion of prescribed doses taken (% of prescribed doses taken; better indicated by higher values; follow-up 6 months[1,17])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7,13]	serious[8]	no serious imprecision[10]	117	89	-	median 2% lower[11]	VERY LOW

[1] *After treatment initiation*
[2] *Te Water Naude et al, 2000*
[3] *Randomisation not appropriate: randomisation is by household unit, analysis is at the level of the individual (i.e. unit-of-analysis error); insufficient data to correct*
[4] *Allocation concealment unclear*
[5] *Blinding absent or unclear*
[6] *Analysis did not follow the intent-to-treat principle*
[7] *'Weight for age' and the 'number who were culture positive' was significantly lower in the intermittent group – may indicate that the intermittent group were less likely to have tuberculosis, or that their tuberculosis was less severe than the daily group*
[8] *Intervention and comparator vary by more than dosing frequency; that is, the intervention studied does not precisely match the intervention of interest*
[9] *Outcome is a substitute for the outcome of interest (cure, treatment success and treatment failure)*
[10] *Data is given as median and interquartile range; imprecision cannot be judged*
[11] *Difference in the medians not provided by authors; calculated by reviewer as ($median_{twice\text{-}weekly}$ – $median_{daily}$)*
[12] *Total number of participants not stated*
[13] *In addition to the use of different treatments, the two groups received different care: the twice-weekly regimen was supervised in the clinic, whereas the daily regimen was not supervised except on the day of medication collection*
[14] *GRADE rule of thumb event number <300*
[15] *Wide confidence intervals*
[16] *Odds ratio and 95% confidence intervals not provided by authors; calculated by reviewer*
[17] *Treatment period*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Daily (Monday-Friday) (unsupervised) dosing	Twice-weekly (DOT) dosing	Relative (95% CI)	Absolute (95% CI)	Quality
Abbreviations: CI, confidence interval; DOT, directly observed therapy; OR, odds ratio										

Intervention: **daily (unsupervised) followed by twice-weekly (unsupervised) dosing**

Comparator: **daily (unsupervised) dosing**

Site of tuberculosis: **pulmonary**

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Daily (unsupervised) dosing	Daily (unsupervised) followed by twice-weekly (unsupervised) dosing	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - disease resolution (number of participants to completely resolve; follow-up 12 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	very serious[8,9]	serious[10]	9/15 (60%)	8/18 (44.4%)	OR 1.88 (0.47 to 7.53)[11]	16 more per 100 (from 17 fewer to 41 more)	VERY LOW
Response to treatment - radiologic improvement (number of participants to show radiologic improvement; follow-up 12 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	very serious[8,9]	very serious[10,12]	15/15 (100%)	18/18 (100%)	OR 0.84 (0.12 to 44.73)[11]	-	VERY LOW
Response to treatment - time to clinical response (therapy period for an early clinical response; better indicated by lower values; follow-up 12 months[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	very serious[8,9]	serious[12]	15	18	-	MD 1.6 months lower in the daily group (from 6.56 lower to 3.36 higher)[13]	VERY LOW
Symptom improvement - weight gain (number to experience weight gain; follow-up 12 months[1]; better indicated by higher values)										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	serious[8]	no serious imprecision	15	18	-	MD 0.09 kg higher in the daily group (from 1.15 lower to 1.33 higher)[13]	VERY LOW
Relapse (number to experience clinical or radiological recurrence in the 12 months after treatment completion[1])										
1[2]	randomised trial	very serious[3,4,5,6]	serious[7]	serious[8]	very serious[10,12]	0/15 (0%)	0/18 (0%)	OR 1.19 (0.02 to 63.73)[11]	-	VERY LOW
Adverse events - hepatotoxicity (number to experience elevated levels of serum aspartate aminotransferase and alanine aminotransferase; follow-up 12 months[1])										
1[2]	randomised trial	very serious[3,4,5,6,14]	serious[7]	serious[8]	very serious[10,12]	1/15 (6.7%)	0/18 (0%)	OR 3.83 (0.14 to 101.08)[11]	-	VERY LOW
Adherence (number excluded due to "poor compliance"; follow-up 12 months[1])										
1[2]	randomised trial	very serious[3,4,5,15]	no serious inconsistency	serious[8]	serious[10]	3/18 (16.7%)	0/18 (0%)	OR 8.35 (0.40 to 174.51)[11]	-	VERY LOW

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Daily (unsupervised) dosing	Daily (unsupervised) followed by twice-weekly (unsupervised) dosing	Relative (95% CI)	Absolute (95% CI)	Quality

[1] *After treatment completion*
[2] *Kansoy et al, 1998*
[3] *Method of randomisation unclear*
[4] *Allocation concealment unclear*
[5] *Blinding unclear*
[6] *Analysis did not follow the intent-to-treat principle*
[7] *Loss to follow-up varied between the two arms: 3 of 18 patients were excluded from the analysis in the daily followed by twice-weekly group for "poor compliance", none were excluded from the daily group*
[8] *Intervention and comparator vary by more than dosing frequency; that is, the intervention studied does not precisely match the intervention of interest*
[9] *Outcome is a substitute for the outcome of interest (cure, treatment success and treatment failure)*
[10] *GRADE rule of thumb event number <300*
[11] *Odds ratio and 95% confidence intervals not provided by authors; calculated by reviewer*
[12] *Wide confidence intervals*
[13] *Mean difference and 95% confidence intervals not provided by authors; calculated by reviewer; mean difference = ($mean_{daily+twice-weekly}$ – $mean_{daily}$)*
[14] *Outcome not clearly defined - thresholds for 'elevated' aspartate aminotransferase and alanine aminotransferase not given*
[15] *Outcome definition not provided*
Abbreviations: CI, confidence interval; DOT, directly observed therapy; MD, mean difference; OR, odds ratio

Intervention: **twice-weekly (DOT) dosing**

Comparator: **daily (unsupervised) followed by twice-weekly (DOT) dosing**

Site of tuberculosis: **cross-site**

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Twice-weekly (DOT) dosing	Daily (unsupervised) followed by twice-weekly (DOT) dosing	Relative (95% CI)	Absolute (95% CI)	Quality
Mortality (number of deaths during the study; follow-up <12-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	1/37 (2.7%)[8]	1/39 (2.6%)[8]	OR 1.06 (0.06 to 17.52)[9]	0 more per 100 (from 2 fewer to 29 more)	VERY LOW
Response to treatment - marked response (number of patients with marked response to treatment[12]; follow-up <12-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,10]	serious[11]	very serious[6,7]	25/37 (67.6%)[8]	28/39 (71.8%)[8]	OR 0.82 (0.31 to 2.18)[9]	4 fewer per 100 (from 28 fewer to 13 more)	VERY LOW
Response to treatment - moderate response (number of patients with moderate response to treatment[12]; follow-up <12-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,10]	serious[11]	very serious[6,7]	11/37 (29.7%)[8]	3/39 (7.7%)[8]	OR 5.08 (1.29 to 20.03)[9]	22 more per 100 (from 2 more to 55 more)	VERY LOW

Appendix E: GRADE profiles

Quality assessment						Number of patients		Effect		
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Twice-weekly (DOT) dosing**	**Daily (unsupervised) followed by twice-weekly (DOT) dosing**	**Relative (95% CI)**	**Absolute (95% CI)**	**Quality**
Response to treatment - poor response (number of patients with poor response to treatment[12]; follow-up <12-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,10]	serious[11]	very serious[6,7]	1/37 (2.7%)[8]	1/39 (2.6%)[8]	OR 1.06 (0.06 to 17.52)[9]	0 more per 100 (from 2 fewer to 29 more)	VERY LOW
Relapse (number to experience clinical or radiological recurrence; follow-up <12-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	0/35 (0%)[8]	0/35 (0%)[8]	OR 1.00 (0.02 to 51.81)[9]	-	VERY LOW
Adverse events - side effects requiring modification of treatment (number of participants that experienced side effects that required modification of treatment; follow-up <12-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	0/37 (0%)	0/39 (0%)	OR 1.05 (0.02 to 54.45)[9]	-	VERY LOW
Adverse events - hypersensitivity reactions (number of participants that experienced a hypersensitivity reaction; follow-up <12-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	0/37 (0%)	0/39 (0%)	OR 1.05 (0.02 to 54.45)[9]	-	VERY LOW
Adverse events - haematologic effects (number of participants that experienced haematologic effects; follow-up <12-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	0/37 (0%)	0/39 (0%)	OR 1.05 (0.02 to 54.45)[9]	-	VERY LOW

[1] *Kumar et al, 1990*
[2] *Allocation concealment unclear*
[3] *Blinding unclear*
[4] *In addition to the use of different treatments, the two groups received different care: twice-weekly dosing was supervised in the clinic, whereas the daily part of the daily followed by twice-weekly regimen was not supervised except on the day of medication collection*
[5] *Follow-up varied considerably between participants*
[6] *GRADE rule of thumb event number <300*
[7] *Wide confidence intervals*
[8] *Data for pulmonary tuberculosis, lymph node tuberculosis and disseminated tuberculosis was pooled by reviewer*
[9] *Odds ratio and 95% confidence intervals not provided by authors; calculated by reviewer*
[10] *Unclear length of follow-up*
[11] *Outcome is a substitute for the outcome of interest (cure, treatment success and treatment failure)*
[12] *See evidence table for criteria*
Abbreviations: CI, confidence interval; DOT, directly observed therapy; OR, odds ratio

Appendix E: GRADE profiles

Intervention: **twice-weekly (DOT) dosing**

Comparator: **daily (unsupervised) followed by twice-weekly (DOT) dosing**

Site of tuberculosis: **pulmonary**

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Daily (unsupervised) followed by twice-weekly (unsupervised) dosing	Daily (unsupervised) dosing	Relative (95% CI)	Absolute (95% CI)	
Mortality (number of deaths during the study; follow-up <12-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	1/20 (5%)	1/23 (4.3%)	OR 1.16 (0.07 to 19.80)[8]	1 more per 100 (from 4 fewer to 43 more)	VERY LOW
Response to treatment - marked response (number of patients with marked response to treatment[11]; follow-up <12-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,9]	serious[10]	serious[6]	13/20 (65%)	16/23 (69.6%)	OR 0.81 (0.23 to 2.92)[8]	5 fewer per 100 (from 35 fewer to 17 more)	VERY LOW
Response to treatment - moderate response (number of patients with moderate response to treatment[11]; follow-up <12-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,9]	serious[10]	very serious[6,7]	1/20 (5%)	0/23 (0%)	OR 3.62 (0.14 to 93.85)[8]	-	VERY LOW
Response to treatment - poor response (number of patients with poor response to treatment[11]; follow-up <12-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,9]	serious[10]	very serious[6,7]	1/20 (5%)	0/23 (0%)	OR 3.62 (0.14 to 93.85)[8]	-	VERY LOW
Relapse (number to experience clinical or radiological recurrence; follow-up <12-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	0/20 (0%)	0/23 (0%)	OR 1.15 (0.02 to 60.41)[8]	-	VERY LOW

[1] *Kumar et al, 1990*
[2] *Allocation concealment unclear*
[3] *Blinding unclear*
[4] *In addition to the use of different treatments, the two groups received different care: twice-weekly dosing was supervised in the clinic, whereas the daily part of the daily followed by twice-weekly regimen was not supervised except on the day of medication collection*
[5] *Follow-up varied considerably between participants*
[6] *GRADE rule of thumb event number <300*
[7] *Wide confidence intervals*
[8] *Odds ratio and 95% confidence intervals not provided by authors; calculated by reviewer*
[9] *Unclear length of follow-up*
[10] *Outcome is a substitute for the outcome of interest (cure, treatment success and treatment failure)*
[11] *See evidence table for criteria*
Abbreviations: CI, confidence interval; DOT, directly observed therapy; OR, odds ratio

Appendix E: GRADE profiles

Intervention: **twice-weekly (DOT) dosing**

Comparator: **daily (unsupervised) followed by twice-weekly (DOT) dosing**

Site of tuberculosis: **lymph node**

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Twice-weekly (DOT) dosing	Daily (unsupervised) followed by twice-weekly (DOT) dosing	Relative (95% CI)	Absolute (95% CI)	Quality
Mortality (number of deaths during the study; follow-up 15-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	0/15 (0%)	0/12 (0%)	OR 0.81 (0.01 to 43.60)[8]	-	VERY LOW
Response to treatment - marked response (number of patients with marked response to treatment[11]; follow-up 15-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,9]	serious[10]	serious[6]	10/15 (66.7%)	8/12 (66.7%)	OR 1.00 (0.20 to 5.00)[8]	0 fewer per 100 (from 38 fewer to 24 more)	VERY LOW
Response to treatment - moderate response (number of patients with moderate response to treatment[11]; follow-up 15-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,9]	serious[10]	very serious[6,7]	5/15 (33.3%)	3/12 (25%)	OR 1.50 (0.28 to 8.14)[8]	8 more per 100 (from 16 fewer to 48 more)	VERY LOW
Response to treatment - poor response (number of patients with poor response to treatment[11]; follow-up 15-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,9]	serious[10]	very serious[6,7]	0/15 (0%)	1/12 (8.3%)	OR 0.25 (0.01 to 6.64)[8]	6 fewer per 100 (from 8 fewer to 29 more)	VERY LOW
Relapse (number to experience clinical or radiological recurrence; follow-up 15-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	0/15 (0%)	0/12 (0%)	OR 0.81 (0.01 to 43.60)[8]	-	VERY LOW

[1] *Kumar et al, 1990*
[2] *Allocation concealment unclear*
[3] *Blinding unclear*
[4] *In addition to the use of different treatments, the two groups received different care: twice-weekly dosing was supervised in the clinic, whereas the daily part of the daily followed by twice-weekly regimen was not supervised except on the day of medication collection*
[5] *Follow-up varied considerably between participants*
[6] *GRADE rule of thumb event number <300*
[7] *Wide confidence intervals*
[8] *Odds ratio and 95% confidence intervals not provided by authors; calculated by reviewer*
[9] *Unclear length of follow-up*
[10] *Outcome is a substitute for the outcome of interest (cure, treatment success and treatment failure)*
[11] *See evidence table for criteria*
Abbreviations: CI, confidence interval; DOT, directly observed therapy; OR, odds ratio

Intervention: **twice-weekly (DOT) dosing**

Comparator: **daily (unsupervised) followed by twice-weekly (DOT) dosing**

Site of tuberculosis: **disseminated**

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Twice-weekly (DOT) dosing	Daily (unsupervised) followed by twice-weekly (DOT) dosing	Relative (95% CI)	Absolute (95% CI)	Quality
Mortality (number of deaths during the study; follow-up <12-24 months)										
1[1]	randomised trials	serious[2,3]	very serious[4,5]	no serious indirectness	very serious[6,7]	0/2 (0%)	0/4 (0%)	OR 1.80 (0.03 to 121.71)[8]	-	VERY LOW
Response to treatment - marked response (number of patients with marked response to treatment[11]; follow-up <12-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,9]	serious[10]	serious[6]	2/2 (100%)	4/4 (100%)	OR 0.56 (0.01 to 37.57)[8]	-	VERY LOW
Response to treatment - moderate response (number of patients with moderate response to treatment[11]; follow-up <12-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,9]	serious[10]	very serious[6,7]	0/2 (0%)	0/4 (0%)	OR 1.80 (0.03 to 121.71)[8]	-	VERY LOW
Response to treatment - poor response (number of patients with poor response to treatment[11]; follow-up <12-24 months)										
1[1]	randomised trials	very serious[2,3]	very serious[4,5,9]	serious[10]	very serious[6,7]	0/2 (0%)	0/4 (0%)	OR 1.80 (0.03 to 121.71)[8]	-	VERY LOW

[1] *Kumar et al, 1990*
[2] *Allocation concealment unclear*
[3] *Blinding unclear*
[4] *In addition to the use of different treatments, the two groups received different care: twice-weekly dosing was supervised in the clinic, whereas the daily part of the daily followed by twice-weekly regimen was not supervised except on the day of medication collection*
[5] *Follow-up varied considerably between participants*
[6] *GRADE rule of thumb event number <300*
[7] *Wide confidence intervals*
[8] *Odds ratio and 95% confidence intervals not provided by authors; calculated by reviewer*
[9] *Unclear length of follow-up*
[10] *Outcome is a substitute for the outcome of interest (cure, treatment success and treatment failure)*
[11] *See evidence table for criteria*
Abbreviations: CI, confidence interval; DOT, directly observed therapy; OR, odds ratio

A.6 RQ K

A.6.1 People coinfected with tuberculosis and HIV

Rifabutin-containing regimens compared with the standard recommended regimen

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rifabutin-containing regimen (2HRbZE/4HRb)	Standard recommended regimen (2HRZE/4HR)	Relative (95% CI)	Absolute		
Mortality (follow-up 6 months; assessed with: number of deaths during the study period)												
1[1]	randomised trials	no serious risk of bias[4]	no serious inconsistency	no serious indirectness	very serious[5,6]	none	4/25 (16%)	2/25 (8%)	OR 2.19 (0.36 to 13.22)[3]	8 more per 100 (from 5 fewer to 45 more)	LOW	
Changes in signs and symptoms – radiographic change (follow-up 6 months; assessed with: number of patients in whom radiographic improvement was observed)												
1[1]	randomised trials	no serious risk of bias[4]	no serious inconsistency	no serious indirectness	very serious[5,6]	none	24/25 (96%)	25/25 (100%)	OR 0.32 (0.01 to 8.25)[3]	-	LOW	
Response to treatment – sputum conversion (follow-up 6 months; assessed with: number of patients to undergo sputum conversion, defined as 3 consecutive negative sputum smears and cultures from the initiation of therapy or a negative smear followed by a consistent absence of sputum production)												
1[1]	randomised trials	no serious risk of bias[4]	no serious inconsistency	serious[2]	serious[6]	none	22/25 (88%)	22/25 (88%)	OR 1.00 (0.18 to 5.51)[3]	0 fewer per 100 (from 31 fewer to 10 more)	LOW	

[1] *Schwander et al, 1995*
[2] *Substitute outcome*
[3] *Odds ratio and 95% confidence interval not provided by authors; calculated by reviewer*
[4] *Patients were able to see the different shapes of tablets, but they were not informed about their content; study nurses and physicians were advised not to request information about medication from patients and remained blind to treatment throughout the study; the only individuals administering care not to be blinded were the drug dispensers*
[5] *Wide confidence interval*
[6] *GRADE rule of thumb: <300 events*

Ciprofloxacin-containing regimens compared with the standard recommended regimen

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Ciprofloxacin-containing regimen (4HRC/2HR)	Standard recommended regimen (2HRZE/2HRZ/2HR)	Relative (95% CI)	Absolute	Quality	Importance
Relapse (follow-up 12 months (6 months after treatment completion); assessed with: number of patients to experience culture-confirmed relapse)												
1[1]	randomised trials	very serious[2,3]	very serious[4,5,6]	no serious indirectness	very serious[7,8]	none	4/26 (15.4%)	0/32 (0%)	OR 13.00 (0.67 to 253.61)[9]	-	VERY LOW	
Response to treatment – culture conversion (follow-up 12 months (6 months after treatment completion); measured with: time to first negative test results; better indicated by lower values)												
1[1]	randomised trials	very serious[2,3]	very serious[4,5,6]	serious[10]	no serious imprecision[11]	none	26	32	-	MD 0.9 higher[12,13]	VERY LOW	

[1] *Kennedy et al 1996*
[2] *Unblinded, except for the radiographers*
[3] *Precise definition of outcome not provided, and it is unclear if the method used to determine the outcome was valid and reliable*
[4] *Unclear if the groups were comparable at baseline*
[5] *Unclear the comparison groups received the same care apart from the interventions studied*
[6] *Unclear if the groups were comparable for treatment completion and availability of outcome data*

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Ciprofloxacin-containing regimen (4HRC/2HR)	Standard recommended regimen (2HRZE/2HRZ/2HR)	Relative (95% CI)	Absolute	Quality	Importance
[7] Wide confidence interval *[8] GRADE rule of thumb: <300 events* *[9] Odds ratio and 95% confidence interval not provided by authors; calculated by reviewer* *[10] Substitute outcome* *[11] Unable to calculate confidence interval; insufficient data* *[12] Mean difference not provided by authors; calculated by reviewer* *[13] p = 0.0003*												

Non-rifampicin-containing regimens compared with rifampicin-containing regimens

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Non-rifampicin-containing regimens	Rifampicin-containing regimens	Relative (95% CI)	Absolute	Quality	Importance
Mortality (univariate analysis) (follow-up 1 years; assessed with: number of deaths during study period)												
1[1]	observational studies[2]	serious[3,4]	very serious[5,6,7]	serious[8]	no serious imprecision[9]	none	-	-	OR 1.82 (1.17 to 2.84)[10]	-	VERY LOW	
Mortality (multivariate analysis) (follow-up 1 years; assessed with: number of deaths during study period)												
1[1]	observational studies[2]	serious[3,4]	very serious[5,6,7]	serious[8]	no serious imprecision[9]	none	-	-	OR 1.21 (0.74 to 1.97)[11,12]	-	VERY LOW	
[1] HIV/TB Study Writing Group, 2009 *[2] Prospective* *[3] Unclear if method of allocation to treatment groups is related to potential confounding factors* *[4] Unclear if blinded, though unlikely* *[5] Unclear if the groups were comparable at baseline* *[6] Unclear the comparison groups received the same care apart from the interventions studied* *[7] Unclear if the groups were comparable for treatment completion and availability of outcome data* *[8] Unclear if the intervention exactly matches the intervention of interest; details provided are limited* *[9] Unclear if GRADE rule of thumb (300 events) met* *[10] p = 0.0079* *[11] Model was adjusted for the following a priori chosen variables: region of residence, age, sex, country of birth, risk factors for HIV and TB acquisition, HIV diagnosis preceding the date of TB diagnosis, CD4 cell count, prior AIDS, initiation of cART, date of TB diagnosis, previous TB, symptoms duration, resistance to anti-TB drugs, and TB location* *[12] p = 0.447*												

Ethambutol-containing continuation phase compared with the standard recommended regimen

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Ethambutol-containing continuation phase	Standard recommended regimen	Relative (95% CI)	Absolute	Quality	Importance
Mortality ($2HRZE_7/6HE_7$ or $2HRZE_3/6HE_7$ compared to $2HRZE_7/4HR_7$) (follow-up 12 months after treatment completion; assessed with: number of deaths)												

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Ethambutol-containing continuation phase	Standard recommended regimen	Relative (95% CI)	Absolute	Quality	Importance
1[1]	randomised trials[2]	serious[4]	serious[17,18]	very serious[19,20]	serious[12]	none	13/90 (14.4%)	4/37 (10.8%)	OR 1.39 (0.52 to 4.59)[13]	4 more per 100 (from 6 fewer to 25 more)	VERY LOW	
Mortality ($2HRZE_7/6HE_7$ compared to $2HRZE_7/4HR_7$) (follow-up 12 months after treatment completion; assessed with: number of deaths)												
1[1]	randomised trials[2]	serious[4]	serious[17,18]	very serious[19,20]	very serious[12,21]	none	10/45 (22.2%)	4/37 (10.8%)	OR 2.36 (0.67 to 8.25)[13]	11 more per 100 (from 3 fewer to 39 more)	VERY LOW	
Mortality (2HRZE/6HE compared to 2HRZE/6HR or 2HRZE/4HR) (follow-up 2 years; assessed with: number of deaths)												
1[16]	observational studies[2]	very serious[3,4,5]	very serious[6,7,8,9]	serious[10,11]	serious[12]	none	27/136 (19.9%)	113/413 (27.4%)	OR 0.66 (0.41 to 1.06)[13]	7 fewer per 100 (from 14 fewer to 1 more)	VERY LOW	
Mortality (2HRZE/6HE compared to 2HRZE/6HR) (follow-up 2 years; assessed with: number of deaths)												
1[16]	observational studies[2]	very serious[3,4,5]	very serious[6,7,8,9]	serious[10,11]	serious[12]	none	27/136 (19.9%)	62/266 (23.3%)	OR 0.82 (0.49 to 1.35)[13]	3 fewer per 100 (from 10 fewer to 6 more)	VERY LOW	
Treatment failure (2HRZE/6HE compared to 2HRZE/6HR or 2HRZE/4HR) (follow-up 2 years; assessed with: number of patients to experience treatment failure)												
1[16]	observational studies[2]	very serious[3,4,5,14]	very serious[6,7,8,9]	serious[10,11]	serious[12]	none	8/136 (5.9%)	12/413 (2.9%)	OR 2.09 (0.84 to 5.22)[13]	3 more per 100 (from 0 fewer to 11 more)	VERY LOW	
Treatment failure (2HRZE/6HE compared to 2HRZE/6HR) (follow-up 2 years; assessed with: number of patients to experience treatment failure)												
1[16]	observational studies[2]	very serious[3,4,5,14]	very serious[6,7,8,9]	serious[10,11]	serious[12]	none	8/136 (5.9%)	7/266 (2.6%)	OR 2.31 (0.82 to 6.52)[13]	3 more per 100 (from 0 fewer to 12 more)	VERY LOW	
Relapse (2HRZE/6HE compared to 2HRZE/6HR or 2HRZE/4HR) (follow-up 2 years; assessed with: number of patients to experience relapse, defined as the development of active tuberculosis after successful completion of an initial course of treatment during 24 months of follow-up after cure)												
1[16]	observational studies[2]	very serious[3,4,5]	very serious[6,7,8,9]	serious[10,11]	serious[12]	none	23/136 (16.9%)	30/413 (7.3%)	OR 2.60 (1.45 to 4.65)[13]	10 more per 100 (from 3 more to 19 more)	VERY LOW	
Relapse (2HRZE/6HE compared to 2HRZE/6HR) (follow-up 2 years; assessed with: number of patients to experience relapse, defined as the development of active tuberculosis after successful completion of an initial course of treatment during 24 months of follow-up after cure)												
1[16]	observational studies[2]	very serious[3,4,5]	very serious[6,7,8,9]	serious[10,11]	serious[12]	none	23/136 (16.9%)	14/266 (5.3%)	OR 3.66 (1.82 to	12 more per 100	VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Ethambutol-containing continuation phase	Standard recommended regimen	Relative (95% CI)	Absolute	Quality	Importance
									7.38)[13]	(from 4 more to 24 more)		
Response to treatment – unfavourable outcome ($2HRZE_7/6HE_7$ or $2HRZE_3/6HE_7$ compared to $2HRZE_7/4HR_7$) (follow-up 12 months after treatment completion; assessed with: number of patients to have an unfavourable outcome, defined as failure[22] or relapse[23], at the end of follow-up)												
1[1]	randomised trials[2]	serious[4]	serious[17,18]	very serious[19,20,15]	very serious[12,21]	none	13/90 (14.4%)	1/37 (2.7%)	OR 6.08 (0.77 to 48.27)[13]	12 more per 100 (from 1 fewer to 55 more)	VERY LOW	
Response to treatment - culture conversion (2HRZE/6HE compared to 2HRZE/6HR or 2HRZE/4HR) (follow-up 2 years; assessed with: number of patients to be culture-negative after 2 months of treatment)												
1[16]	observational studies[2]	very serious[3,4,5,14]	very serious[6,7,8,9]	very serious[10,11,15]	serious[12]	none	101/136 (74.3%)	364/413 (88.1%)	OR 0.39 (0.24 to 0.63)[13]	14 fewer per 100 (from 6 fewer to 24 fewer)	VERY LOW	
Adherence - treatment completion (2HRZE/6HE compared to 2HRZE/6HR or 2HRZE/4HR) (follow-up 2 years; assessed with: number of patients to complete therapy)												
1[16]	observational studies[2]	very serious[3,4,5,14]	very serious[6,7,8,9]	serious[10,11]	serious[12]	none	89/136 (65.4%)	317/413 (76.8%)	OR 0.57 (0.39 to 0.87)[13]	11 fewer per 100 (from 3 fewer to 20 fewer)	VERY LOW	
Adherence - treatment completion (2HRZE/6HE compared to 2HRZE/6HR) (follow-up 2 years; assessed with: number of patients to complete therapy)												
1[16]	observational studies[2]	very serious[3,4,5,14]	very serious[6,7,8,9]	serious[10,11]	serious[12]	none	89/136 (65.4%)	195/266 (73.3%)	OR 0.69 (0.44 to 1.08)[13]	8 fewer per 100 (from 19 fewer to 1 more)	VERY LOW	

[1] *Jindani et al 2004*
[2] *Prospective*
[3] *Unclear if method of allocation to treatment groups is related to potential confounding factors - allocation was based on the time of treatment*
[4] *Unclear if blinded, though unlikely*
[5] *Attempts were not made within the design or analysis to balance the groups for potential confounders*
[6] *Groups were not comparable at baseline - 2HRZE/4HR group were significantly older, 2HRZE/6HR group had significantly higher levels of haemoglobin, and 2HRZE/6HE group had significantly higher total white blood cell counts*
[7] *Groups did not receive the same care apart from the intervention(s) studied - rifampicin regimens (2HRZE/4HR and 2HRZE/6HR) were self-administered, non-rifampicin regimen (2HRZE/6HE) was directly observed*
[8] *Groups not followed up for an equal and appropriate length of time - median follow-up in the 2HRZE/4HR group was 512 days, 533 days in the 2HRZE/6HR group, and 661 days in the 2HRZE/6HE group*
[9] *Groups not comparable for treatment completion and availability of outcome data - 83% completed treatment in the 2HRZE/4HR group, 73% completed treatment in the 2HRZE/6HR group, and 65% completed treatment in the 2HRZE/6HE group*
[10] *Population appears to match the population of interest, although unclear if there was any drug resistance at baseline*
[11] *Interventions varied by more than the combination of drugs used (also varied by dosing frequency and the use of DOT, as well as the duration of treatment with regards to the 2HRZE/4HR group)*
[12] *GRADE rule of thumb: <300 events*

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Ethambutol-containing continuation phase	Standard recommended regimen	Relative (95% CI)	Absolute	Quality	Importance

[13] *Odds ratio and 95% confidence interval not provided by authors; calculated by reviewer*
[14] *Study did not provide a precise definition of the outcome*
[15] *Substitute outcome*
[16] *Okwera et al, 2006*
[17] *Unclear if groups were comparable at baseline as baseline characteristics not reported by HIV status*
[18] *Groups had comparable rates of attrition, though rates were high in both groups*
[19] *Population may not exactly match the population of interest: some drug resistance at baseline, although unclear if any within the HIV subgroup as baseline characteristics not reported by HIV status*
[20] *Intervention may not exactly match the intervention of interest: intervention varies by more than the combination of antituberculosis drugs – regimens with an E-continuation phase were 2 months longer than those with an R-continuation phase, and some patients receiving an E-continuation phase had an initial dosing schedule of 3-times weekly and some had a daily dosing schedule, whereas all*
[21] *Wide confidence interval*
[22] *Failure was defined as a culture of 20 or more colonies at month 6 or 8, or a change of treatment by the local investigator owing to treatment failure*
[23] *Relapse was defined as a culture of 20 or more colonies at any point after the end of treatment or, in the absence of culture confirmation, initiation by the local investigator of treatment for relapse*

A.6.2 People with tuberculosis and liver disease

Fluoroquinolone-containing regimen compared with rifampicin-containing regimen

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fluoroquinolone-containing regimen	Rifampicin-containing regimen	Relative (95% CI)	Absolute	Quality	Importance
Mortality - all-cause (2HZEO/10HEO compared with 2HRE/7HR) (follow-up 3 months after treatment was stopped; assessed with: number of patients to die from any cause)												
1[1]	randomised trials	very serious[2,3]	very serious[4,5]	serious[6]	very serious[7,8]	none	1/16 (6.3%)[9]	0/15 (0%)	OR 3.00 (0.11 to 79.50)[10]	-	VERY LOW	
Mortality - tuberculosis-related (2HZEO/10HEO compared with 2HRE/7HR) (follow-up 3 months after treatment was stopped; assessed with: number of tuberculosis-related deaths)												
1[1]	randomised trials	very serious[2,3]	very serious[4,5]	serious[6]	very serious[7,8]	none	0/16 (0%)	0/15 (0%)	OR 0.94 (0.02 to 50.32)[10]	-	VERY LOW	
Mortality - hepatotoxicity-related (2HZEO/10HEO compared with 2HRE/7HR) (follow-up 3 months after treatment was stopped; assessed with: number of hepatotoxicity-related deaths)												
1[1]	randomised trials	very serious[2,3]	very serious[4,5]	serious[6]	very serious[7,8]	none	0/16 (0%)	0/15 (0%)	OR 0.94 (0.02 to 50.32)[10]	-	VERY LOW	
Adverse events - hepatotoxicity (2HZEO/10HEO compared with 2HRE/7HR) (follow-up 3 months after treatment was stopped; assessed with: number of patients to experience hepatotoxicity, defined as ALT/AST levels >5-fold the baseline level or >400 IU/L, or if bilirubin increased by 2.5 mg/dl after exclusion of superimposed acute hepatitis)												
1[1]	randomised trials	very serious[2,3]	very serious[4,5]	serious[6]	serious[8]	none	0/16 (0%)	4/15 (26.7%)	OR 0.08 (0.00 to 1.58)[10]	24 fewer per 100 (from 27 fewer to 10 more)	VERY LOW	
Adverse events - hepatotoxicity (HRbAOL compared with HRZS/E) (follow-up unclear; assessed with: number of patients to experience liver dysfunction, defined as ALT >1336 IU/L 2-3 months												

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Fluoroquinolone-containing regimen	Rifampicin-containing regimen	Relative (95% CI)	Absolute	Quality	Importance
after initiation of antituberculosis chemotherapy)												
1[11]	observational studies[12]	very serious[13,14,15,16]	serious[5,17,18]	serious[19]	serious[20]	none	7/23 (30.4%)	19/24 (79.2%)	OR 0.12 (0.03 to 0.43)[10]	48 fewer per 100 (from 17 fewer to 69 fewer)	VERY LOW	

[1] *Saigal et al, 2001*
[2] *Unclear if there was adequate concealment of allocation*
[3] *Unblinded*
[4] *Groups not comparable at baseline - ofloxacin group had a significantly lower level of albumin and a greater prolongation of prothrombin time, which indicates that the underlying liver disease may have been more severe in this group; additionally, the aetiologies of the liver disease were not comparable in the 2 groups*
[5] *Unclear if groups received the same care apart from the intervention(s) studied; limited details provided*
[6] *Interventions varied by more than the combination of antituberculosis drugs used (regimens also varied by total duration of treatment); additionally, it is unclear if the doses used and the dosing frequencies were comparable in the 2 regimens*
[7] *Wide confidence interval*
[8] *GRADE rule of thumb: <300 events*
[9] *Death resulted from intracranial bleeding unrelated to the antituberculosis chemotherapy during the follow-up*
[10] *Odds ratio and 95% confidence interval not provided by authors; calculated by reviewer*
[11] *Pan et al, 2005*
[12] *Prospective*
[13] *Unclear if method of allocation to treatment groups is related to potential confounding factors*
[14] *Blinding unclear*
[15] *Attempts were not made within the design or analysis to balance the groups for potential confounders*
[16] *Unclear if follow-up was for an appropriate period of time*
[17] *Groups appear to be comparable at baseline - authors state that the 'general conditions of the 2 groups were not distinguishable (p > 0.05)', although no further details are provided*
[18] *Unclear if groups were followed up for an equal length of time*
[19] *Regimens used vary by more than the combinations of drugs used (the 2 regimens used different dosing schedules; additionally, it is unclear if the total duration of treatment was comparable in the 2 groups)*
[20] *GRADE rule of thumb: <300 events*

A.7 RQ L

A.7.1 Duration of treatment in adults with respiratory tuberculosis

SMEAR-POSITIVE, CULTURE-POSITIVE

4 vs **6** months

Age: mix

HIV status: not specified – negative?

Disease status: smear- and culture-positive

Site of disease: pulmonary

Drug sensitivity: susceptible only

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	4 months	6 months	Relative (95% CI)	Absolute (95% CI)	
Response to treatment - favourable status (assessed with: number of smear-positive culture-positive patients to achieve a favourable status at the end of treatment)										
1[1]	randomised trials	very serious[2,3]	serious[4]	serious[5,6]	serious[7]	161/161 (100%)	169/169 (100%)	OR 0.95 (0.02 to 48.31)[8]	-	VERY LOW
Relapse (follow-up 5 to 8 years after treatment initiation; assessed with: number of smear-positive culture-positive patients to experience relapse[9])										
1[1]	randomised trials	very serious[2,3]	serious[4,10]	serious[5,6]	very serious[7,11]	20/131 (15.3%)	3/138 (2.2%)	OR 8.11 (2.35 to 28)[8]	13 more per 100 (from 3 more to 36 more)	VERY LOW

[1] *Singapore TB Service / British Medical Research Council, 1979/86*
[2] *Blinding unclear*
[3] *Analysis is not intent-to-treat*
[4] *Unclear if the loss to follow-up was similar in the 2 groups*
[5] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[6] *Population does not exactly match the population of interest: unclear if children are included*
[7] *Wide confidence intervals*
[8] *Odds ratio and 95% confidence intervals calculated by reviewer*
[9] *See evidence table for the full definition*
[10] *Unclear if length of follow-up period was the same in the 2 groups*
[11] *GRADE rule of thumb: <300 events*
Abbreviations: CI, confidence interval; OR, odds ratio

3 vs **4.5** months

Age: mix

HIV status: not specified – negative?

Disease status: smear-positive, culture-positive

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	3 months	4.5 months	Relative (95% CI)	Absolute (95% CI)	
Response to treatment - culture status (intent-to-treat) (assessed with: number of smear-positive, culture-positive patients to be culture-negative at the end of treatment)										

Appendix E: GRADE profiles

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	3 months	4.5 months	Relative (95% CI)	Absolute (95% CI)	Quality
1[1]	randomised trials	serious[2,3,4]	serious[5]	very serious[6,7,12]	very serious[8,9]	58/91 (63.7%)	68/89 (76.4%)	OR 0.54 (0.28 to 1.04)[10]	13 fewer per 100 (from 29 fewer to 1 more)	VERY LOW
Response to treatment - culture status (among those that completed treatment) (assessed with: number of smear-positive, culture-positive patients to be culture-negative at the end of treatment)										
1[1]	randomised trials	very serious[2,3,4,11]	serious[5]	very serious[6,7,12]	very serious[8,9]	58/58 (100%)	68/68 (100%)	OR 0.85 (0.02 to 43.72)[10]	-	VERY LOW
Changes in signs and symptoms – deterioration in radiographic status (assessed with: number of smear-positive, culture-positive patients to experience deterioration in radiographic appearance 6 months after treatment initiation)										
1[1]	randomised trials	serious[2,3,4]	serious[5]	serious[6,12]	very serious[8,9]	0/91 (0%)	0/89 (0%)	OR 0.98 (0.02 to 49.83)[10]	-	VERY LOW
Changes in signs and symptoms – no change in radiographic status (assessed with: number of smear-positive, culture-positive patients to experience no change in radiographic appearance 6 months after treatment initiation)										
1[1]	randomised trials	serious[2,3,4]	serious[5]	serious[6,12]	very serious[8,9]	0/91 (0%)	1/89 (1.1%)	OR 0.32 (0.01 to 8.02)[10]	1 fewer per 100 (from 1 fewer to 7 more)	VERY LOW
Changes in signs and symptoms – moderate improvement in radiographic status (assessed with: number of smear-positive, culture-positive patients to experience moderate improvement in radiographic appearance 6 months after treatment initiation)										
1[1]	randomised trials	serious[2,3,4]	serious[5]	serious[6,12]	serious[9]	31/91 (34.1%)	39/89 (43.8%)	OR 0.66 (0.36 to 1.21)[10]	10 fewer per 100 (from 22 fewer to 5 more)	VERY LOW
Changes in signs and symptoms – marked improvement in radiographic status (assessed with: number of smear-positive, culture-positive patients to experience marked improvement in radiographic appearance 6 months after treatment initiation)										
1[1]	randomised trials	serious[2,3,4]	serious[5]	serious[6,12]	serious[9]	24/91 (26.4%)	24/89 (27%)	OR 0.97 (0.5 to 1.88)[10]	1 fewer per 100 (from 11 fewer to 14 more)	VERY LOW
Changes in signs and symptoms – marked improvement in radiographic status (assessed with: number of smear-positive, culture-positive patients to experience marked improvement in radiographic appearance 12 months after treatment initiation)										
1[1]	randomised trials	serious[2,3,4]	serious[5]	serious[6,12]	serious[9]	19/91 (20.9%)	15/89 (16.9%)	OR 1.30 (0.61 to 2.76)[10]	4 more per 100 (from 6 fewer to 19 more)	VERY LOW
Changes in signs and symptoms – marked improvement in radiographic status (assessed with: number of smear-positive, culture-positive patients to experience marked improvement in radiographic appearance 18 months after treatment initiation)										
1[1]	randomised trials	serious[2,3,4]	serious[5]	serious[6,12]	serious[9]	20/91 (22%)	16/89 (18%)	OR 1.29 (0.62 to 2.68)[10]	4 more per 100 (from 6 fewer to 19 more)	VERY LOW
Adverse events leading to treatment interruption (assessed with: number of smear-positive, culture-positive patients to experience adverse events leading to treatment interruption)										
1[1]	randomised trials	serious[2,3,4]	serious[5]	serious[6,12]	very serious[8,9]	5/91 (5.5%)	2/89 (2.2%)	OR 2.53 (0.48 to 13.39)[10]	3 more per 100 (from 1 fewer to 21 more)	VERY LOW

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	3 months	4.5 months	Relative (95% CI)	Absolute (95% CI)	
Adherence – treatment default (assessed with: number of smear-positive, culture-positive patients to default[13])										
1[1]	randomised trials	serious[2,3,4]	serious[5]	serious[6,12]	serious[9]	8/91 (8.8%)	7/89 (7.9%)	OR 1.13 (0.39 to 3.26)[10]	1 more per 100 (from 5 fewer to 14 more)	VERY LOW
Relapse (follow-up 1 year after treatment completion; assessed with: number of smear-positive, culture-positive patients to experience relapse[13])										
1[1]	randomised trials	serious[2,3,4]	serious[5]	serious[6,12]	very serious[8,9]	1/91 (1.1%)	1/89 (1.1%)	OR 0.98 (0.06 to 15.88)[10]	0 fewer per 100 (from 1 fewer to 14 more)	VERY LOW

[1] *Mehotra et al, 1982*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Although not statistically significant, there was a higher number who did not complete treatment and for whom data was not available amongst the 3-month group (36%) than the 4.5-month group (24%)*
[6] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[7] *Outcome is a substitute for an outcome of interest*
[8] *Wide confidence intervals*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Analysis not intent-to-treat*
[12] *Population does not exactly match the population of interest: may include some children (inclusion criteria: aged 12 years or more)*
[13] *See evidence table for the full definition*
Abbreviations: CI, confidence interval; OR, odds ratio

6 vs **8** months

Age: unclear

HIV status: not specified – negative?

Disease status: smear- and culture-positive (?), symptomatic

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	8 months	Relative (95% CI)	Absolute (95% CI)	
Relapse (follow-up 12 months after treatment completion; assessed with: number of smear-positive, culture-positive patients to experience relapse)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[10]	serious[6,7]	serious[8]	1/97 (1%)	3/96 (3.1%)	OR 0.32 (0.03 to 3.16)[9]	2 fewer per 100 (from 3	VERY LOW

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	8 months	Relative (95% CI)	Absolute (95% CI)	Quality
									fewer to 6 more)	

[1] *Nayar et al, 1988*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Analysis does not follow intent-to-treat principle*
[6] *Unclear if population includes children*
[7] *Intervention does not exactly match the intervention of interest: did not contain all of the 4 standard recommended drugs*
[8] *GRADE rule of thumb: <300 events*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*
[10] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
Abbreviations: CI, confidence interval; OR, odds ratio

SMEAR-POSITIVE, MIXED/UNSPECIFIED CULTURE

3 vs **6** months

Age: mix

HIV status: not specified – negative?

Disease status: smear-positive → negative, culture not specified

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	3 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - culture status (assessed with: number of initially smear-positive patients who were smear-negative after 8 weeks of treatment to be culture-negative at the end of treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7]	very serious[8,9,14,15]	very serious[10,11]	56/56 (100%)	70/70 (100%)	OR 0.80 (0.02 to 41.03)[12]	-	VERY LOW
Changes in signs and symptoms – marked improvement in radiographic status (assessed with: number of initially smear-positive patients who were smear-negative after 8 weeks of treatment to experience marked improvement in radiographic appearance by the end of treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7]	very serious[8,14,15]	serious[10]	36/56 (64.3%)	57/70 (81.4%)	OR 0.52 (0.23 to 1.2)[12]	12 fewer per 100 (from 31 fewer to 3 more)	VERY LOW
Changes in signs and symptoms – slight improvement in radiographic status (assessed with: number of initially smear-positive patients who were smear-negative after 8 weeks of treatment to experience slight improvement in radiographic appearance by the end of treatment)										

Appendix E: GRADE profiles

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	3 months	6 months	Relative (95% CI)	Absolute (95% CI)	
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7]	very serious[8,14,15]	serious[10]	9/56 (16.1%)	9/70 (12.9%)	OR 1.30 (0.48 to 3.53)[12]	3 more per 100 (from 6 fewer to 21 more)	VERY LOW
Changes in signs and symptoms – no change in radiographic status (assessed with: number of initially smear-positive patients who were smear-negative after 8 weeks of treatment to experience no change in radiographic appearance by the end of treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7]	very serious[8,14,15]	serious[10]	2/56 (3.6%)	4/70 (5.7%)	OR 0.61 (0.11 to 3.46)[12]	2 fewer per 100 (from 5 fewer to 12 more)	VERY LOW
Changes in signs and symptoms – deterioration in radiographic status (assessed with: number of initially smear-positive patients who were smear-negative after 8 weeks of treatment to experience deterioration in radiographic appearance by the end of treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7]	very serious[8,14,15]	very serious[10,11]	6/56 (10.7%)	0/70 (0%)	OR 18.15 (0.9995 to 329.54)[12]	-	VERY LOW
Relapse (follow-up 104 weeks after treatment initiation; assessed with: number of initially smear-positive patients who were smear-negative after 8 weeks of treatment to experience relapse)										
1[1]	randomised trials	very serious[2,3,4,5]	very serious[6,7,13]	very serious[8,14,15]	very serious[10,11]	12/56 (21.4%)	1/70 (1.4%)	OR 18.82 (2.36 to 149.85)[12]	20 more per 100 (from 2 more to 67 more)	VERY LOW

[1] *Research Committee of the Tuberculosis Association of India, 1984*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Analysis is not intent-to-treat*
[6] *Comparability of patients at baseline unclear*
[7] *Number of patients lost to follow-up in each group is unclear*
[8] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[9] *Outcome is a substitute for an outcome of interest*
[10] *GRADE rule of thumb: <300 events*
[11] *Wide confidence intervals*
[12] *Odds ratio and 95% confidence intervals calculated by reviewer*
[13] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[14] *Population does not exactly match the population of interest: may include some children (inclusion criteria = 15-45 years)*
[15] *Doses used are inconsistent with those recommended in the British National Formulary*
Abbreviations: CI, confidence interval; OR, odds ratio

6 vs **9** months

Age: mix

HIV status: unspecified – negative?

Disease status: smear-positive

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Cure (assessed with: number of sputum-smear-positive patients to be smear-negative in the last month of treatment and on at least one previous occasion)										
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	25/93 (26.9%)	19/107 (17.8%)	OR 1.73 (0.88 to 3.4)[7]	9 more per 100 (from 2 fewer to 25 more)	VERY LOW
Treatment failure (assessed with: number of sputum-smear-positive patients to be smear-positive at 5 months or later during treatment)										
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	very serious[6,8]	0/93 (0%)	1/107 (0.93%)	OR 0.38 (0.02 to 9.43)[7]	1 fewer per 100 (from 1 fewer to 7 more)	VERY LOW
Bacteriological relapse (assessed with: number of sputum-smear-positive patients to experience bacteriological relapse)										
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	very serious[6,8]	5/93 (5.4%)	0/107 (0%)	OR 13.36 (0.73 to 244.96)[7]	-	VERY LOW

[1] *Ziaullah et al, 2004*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Population does not exactly match the population of interest: includes children (33% aged 5 to 14 years, 33% aged 15 to 29 years)*
[6] *GRADE rule of thumb: <300 events*
[7] *Odds ratio and 95% confidence intervals calculated by reviewer*
[8] *Wide confidence intervals*
[9] *Analysis did not follow intent-to-treat principle*
Abbreviations: CI, confidence interval; OR, odds ratio

CULTURE-POSITIVE, MIXED/UNSPECIFIED SMEAR, CAVITATORY

9 vs **18** months

Age: mix

HIV status: not specified – negative?

Disease status: culture-positive, cavitatory

Site of disease: pulmonary

Drug sensitivity: some DR-TB

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	18 months	Relative (95% CI)	Absolute (95% CI)	
Treatment failure (assessed with: number of culture-positive patients with cavities >2 cm to experience treatment failure[1])										
1[2]	randomised trials	serious[3,4,5]	no serious inconsistency	very serious[6,7]	very serious[8,9]	0/187 (0%)	0/194 (0%)	1.04 (0.02 to 52.55)[10]	-	VERY LOW
'Alive and well' (assessed with: number of culture-positive patients with cavities >2 cm to be considered alive and well after 54 months of follow-up[1])										
1[2]	randomised trials	serious[3,4,5]	serious[11]	very serious[6,7,12]	serious[9]	116/187 (62%)	108/194 (55.7%)	OR 1.30 (0.86 to 1.96)[10]	6 more per 100 (from 4 fewer to 15 more)	VERY LOW
Relapse (follow-up 54 months; assessed with: number of culture-positive patients with cavities >2 cm to experience relapse[1])										
1[2]	randomised trials	serious[3,4,5]	serious[11]	very serious[6,7]	very serious[8,9]	0/187 (0%)	0/194 (0%)	1.04 (0.02 to 52.55)[10]	-	VERY LOW

[1] See evidence table for the full definition
[2] British Thoracic Society, 1975/80
[3] Method of randomisation unclear
[4] Allocation concealment possible - "random allocations of treatment were made centrally by coordinators"
[5] Radiographer blinded to treatment allocation, but unclear if to prognostic factors or if other investigators were blinded
[6] Intervention does not exactly match the intervention of interest: does not contain all of or just the 4 standard recommended drugs
[7] Population does not exactly match the population of interest: 3.4% drug resistance at baseline, and may include some children (inclusion criteria = 15 to 70 years)
[8] Wide confidence intervals
[9] GRADE rule of thumb: <300 events
[10] Odds ratio and 95% confidence intervals calculated by reviewer
[11] Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths
[12] Outcome is a substitute for an outcome of interest
Abbreviations: CI, confidence interval; OR, odds ratio

CULTURE-POSITIVE, MIXED/UNSPECIFIED SMEAR, NON-CAVITATORY

6 vs **12** months

Age: mix

HIV status: not specified – negative?

Disease status: culture-positive, non-cavitatory

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	12 months	Relative (95% CI)	Absolute (95% CI)	
Treatment failure (assessed with: number of culture-positive patients without HIV or cavities or no cavity >2 cm to experience treatment failure[1])										
1[2]	randomised trials	serious[3,4,5]	no serious inconsistency	very serious[6,7]	very serious[8,9]	1/214 (0.47%)	0/217 (0%)	OR 3.06 (0.12 to 75.45)[10]	-	VERY LOW
'Alive and well' (assessed with: number of culture-positive patients without HIV or cavities or no cavity >2 cm to be considered alive and well after 54 months of follow-up[1])										
1[2]	randomised trials	serious[3,4,5]	serious[11]	very serious[6,7,12]	serious[9]	129/214 (60.3%)	140/217 (64.5%)	OR 0.83 (0.57 to 1.23)[10]	4 fewer per 100 (from 14 fewer to 5 more)	VERY LOW
Relapse (follow-up 54 months; assessed with: number of culture-positive patients without HIV or cavities or no cavity >2 cm to experience relapse[1])										
1[2]	randomised trials	serious[3,4,5]	serious[11]	very serious[6,7]	very serious[8,9]	9/214 (4.2%)	2/217 (0.92%)	OR 4.72 (1.01 to 22.11)[10]	3 more per 100 (from 0 more to 16 more)	VERY LOW

[1] *See evidence table for the full definition*
[2] *British Thoracic Society, 1975/80*
[3] *Method of randomisation unclear*
[4] *Allocation concealment possible - "random allocations of treatment were made centrally by coordinators"*
[5] *Radiographer blinded to treatment allocation, but unclear if to prognostic factors or if other investigators were blinded*
[6] *Intervention does not exactly match the intervention of interest: does not contain all of or just the 4 standard recommended drugs*
[7] *Population does not exactly match the population of interest: 3.4% drug resistance at baseline, and may include some children (inclusion criteria = 15 to 70 years)*
[8] *Wide confidence intervals*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[12] *Outcome is a substitute for an outcome of interest*
Abbreviations: CI, confidence interval; OR, odds ratio

CULTURE-POSITIVE, MIXED/UNSPECIFIED SMEAR

6 vs **9** months

Age: adult-only

HIV status: not specified – negative?

Disease status: culture-positive

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Response to treatment - culture status (assessed with: number of culture-positive patients to be culture-negative after 6 months of treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7,8]	very serious[9,10]	serious[11]	287/287 (100%)	157/157 (100%)	OR 1.83 (0.04 to 92.44)[12]	-	VERY LOW
Relapse (follow-up a minimum of 3 years after treatment completion; assessed with: number of culture-positive patients to experience relapse)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7,8]	serious[9]	very serious[11,13]	6/287 (2.1%)	2/157 (1.3%)	OR 1.65 (0.33 to 8.3)[12]	1 more per 100 (from 1 fewer to 8 more)	VERY LOW
Adverse events requiring modification or withdrawal of treatment (assessed with: number of culture-positive patients to experience an adverse event requiring modification or withdrawal of treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7,8]	serious[9]	serious[13]	19/344 (5.5%)	7/177 (4%)	OR 1.42 (0.59 to 3.44)[12]	2 more per 100 (from 2 fewer to 8 more)	VERY LOW
Adverse events - hepatic (assessed with: number of culture-positive patients to experience a hepatic adverse event)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7,8]	serious[9]	serious[13]	14/287 (4.9%)	7/157 (4.5%)	OR 1.10 (0.43 to 2.78)[12]	0 more per 100 (from 2 fewer to 7 more)	VERY LOW
Adverse events - rash (assessed with: number of culture-positive patients to experience rash)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7,8]	serious[9]	very serious[11,13]	13/287 (4.5%)	1/157 (0.64%)	OR 7.40 (0.96 to 57.12)[12]	4 more per 100 (from 0 fewer to 26 more)	VERY LOW
Adverse events - arthralgia (assessed with: number of culture-positive patients to experience arthralgia)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7,8]	serious[9]	very serious[11,13]	2/287 (0.7%)	0/157 (0%)	OR 2.76 (0.13 to 57.82)[12]	-	VERY LOW
Adherence - treatment default (assessed with: number of culture-positive patients to default treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7,8]	serious[9]	serious[13]	11/344 (3.2%)	4/177 (2.3%)	OR 1.43 (0.45 to 4.55)[12]	1 more per 100 (from 1 fewer to 7 more)	VERY LOW

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Adherence - isoniazid metabolites (assessed with: number of urine samples from culture-positive patients that were positive for isoniazid metabolites[14])										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7,8]	very serious[9,10]	no serious imprecision	1334/1379 (96.7%)	1128/1166 (96.7%)	OR 1.00 (0.64 to 1.55)[12]	0 fewer per 100 (from 2 fewer to 1 more)	VERY LOW

[1] *British Thoracic Society, 1981/2/4*
[2] *Method of randomisation unclear*
[3] *Allocation concealment*
[4] *Blinding unclear*
[5] *Analysis did not follow the intent-to-treat principle*
[6] *Unclear if groups received the same care except for the intervention*
[7] *Unclear if the groups were comparable for treatment completion*
[8] *High attrition rate with regards to the number of participants for whom data is available*
[9] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs, and the 2 regimens vary by more than duration*
[10] *Outcome is a substitute for an outcome of interest*
[11] *Wide confidence intervals*
[12] *Odds ratio and 95% confidence intervals calculated by reviewer*
[13] *GRADE rule of thumb: <300 events*
[14] *See evidence table for the full definition*
Abbreviations: CI, confidence interval; OR, odds ratio

SMEAR-NEGATIVE, MIXED/UNSPECIFIED CULTURE

<6 vs **6** months

Age: mix

HIV status: not specified – negative?

Disease status: smear-megative

Site of disease: pulmonary

Drug sensitivity: susceptible / unclear (pooled)

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	<6 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality
Relapse (follow-up 5 years after treatment initiation; assessed with: number of smear-negative patients to experience relapse)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7]	very serious[8,9]	very serious[10,12]	72/1502 (4.8%)	10/190 (5.3%)	OR 0.91 (0.46 to 1.79)[11]	0 fewer per 100 (from 3 fewer to 4 more)	VERY LOW
Bacteriological relapse (follow-up 5 years after treatment initiation; assessed with: number of smear-negative patients to experience bacteriologically confirmed relapse)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6,7]	very serious[8,9]	very serious[10,12]	32/1502 (2.1%)	4/190 (2.1%)	OR 1.01 (0.35 to 2.89)[11]	0 more per 100 (from 1 fewer to 4 more)	VERY LOW
Adverse events (any) (assessed with: number of smear-negative patients to experience any adverse event)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6]	very serious[8,9]	very serious[10,12]	462/1502 (30.8%)	81/190 (42.6%)	OR 0.60 (0.44 to 0.81)[11]	12 fewer per 100 (from 5 fewer to 18 fewer)	VERY LOW
Adverse events requiring withdrawal of one or more drug (assessed with: number of smear-negative patients to experience an adverse event requiring withdrawal of one or more drug)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6]	very serious[8,9]	very serious[10,12]	71/1502 (4.7%)	6/190 (3.2%)	OR 1.52 (0.65 to 3.55)[11]	2 more per 100 (from 1 fewer to 7 more)	VERY LOW
Adverse events leading to a temporary interruption in treatment (assessed with: number of smear-negative patients to experience an adverse event leading to a temporary interruption in treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6]	very serious[8,9]	very serious[10,12]	153/1502 (10.2%)	25/190 (13.2%)	OR 0.75 (0.48 to 1.18)[11]	3 fewer per 100 (from 6 fewer to 2 more)	VERY LOW
Adverse events - cutaneous (assessed with: number of smear-negative patients to experience a cutaneous adverse reaction)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6]	very serious[8,9]	very serious[10,12]	110/1502 (7.3%)	16/190 (8.4%)	OR 0.86 (0.5 to 1.49)[11]	1 fewer per 100 (from 4 fewer to 4 more)	VERY LOW
Adverse events - gastrointestinal (assessed with: number of smear-negative patients to experience a gastrointestinal adverse reaction)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6]	very serious[8,9]	very serious[10,12]	87/1502 (5.8%)	20/190 (10.5%)	OR 0.52 (0.31 to 0.87)[11]	5 fewer per 100 (from 1 fewer to 7 fewer)	VERY LOW
Adverse events - vestibular (assessed with: number of smear-negative patients to experience a vestibular adverse reaction)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6]	very serious[8,9]	very serious[10,12]	69/1502 (4.6%)	7/190 (3.7%)	OR 1.26 (0.57 to 2.78)[11]	1 more per 100 (from 2 fewer to 6 more)	VERY LOW
Adverse events - hepatic (assessed with: number of smear-negative patients to experience a hepatic adverse reaction)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6]	very serious[8,9]	very serious[10,12]	18/1502 (1.2%)	0/190 (0%)	OR 4.75 (0.29 to 79.11)[11]	-	VERY LOW

[1] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	<6 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality

[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Analysis did not follow intent-to-treat principle*
[6] *Unclear if loss to follow-up was the same in the 2 groups*
[7] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[8] *Intervention does not exactly match the intervention of interest: did not contain all of the 4 standard recommended drugs*
[9] *Population does not exactly match the population of interest: includes some children (inclusion criteria = 15 to 75 years), and some cases were possibly 'inactive'*
[10] *Wide confidence intervals*
[11] *Odds ratio and 95% confidence intervals calculated by reviewer*
[12] *GRADE rule of thumb: <300 events*
Abbreviations: CI, confidence interval; OR, odds ratio

4 vs **6** months

Age: mix

HIV status: not specified – negative?

Disease status: smear-negative, radiographically active

Site of disease: pulmonary

Drug sensitivity: susceptible / unclear (pooled)

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	4 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality
Relapse (follow-up 5 years after treatment initiation; assessed with: number of smear-negative patients to experience relapse[1])										
2[2,3]	randomised trials	very serious[4,5,6]	very serious[7,8,9]	very serious[10,11,12,15]	serious[13]	7/384 (1.8%)	8/231 (3.5%)	OR 0.47 (0.17 to 1.3)[14,16]	2 fewer per 100 (from 3 fewer to 1 more)	VERY LOW

[1] *See evidence table for the full definitions*
[2] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989*
[3] *Teo et al, 2002*
[4] *Blinding unclear*
[5] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: method of randomisation unclear*
[6] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: allocation concealment unclear*
[7] *Teo et al, 2002: comparability of patients at baseline was unclear*
[8] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: number of patients lost to follow-up in each group is unclear*
[9] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[10] *Teo et al, 2002: intervention does not exactly match the intervention of interest: did not contain all of the 4 standard recommended drugs, and the two arms vary by more than duration alone*
[11] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: population does not exactly match the population of interest: includes some*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	4 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality

children, and some cases were possibly 'inactive'
[12] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[13] *GRADE rule of thumb: <300 events*
[14] *Pooled odds ratio and 95% confidence intervals calculated by reviewer*
[15] *Teo et al, 2002: population does not exactly match the population of interest: unclear if children are included*
[16] *Forest plot (relapse):*

Abbreviations: CI, confidence interval; OR, odds ratio

Age: mix

HIV status: not specified – negative?

Disease status: smear-negative, culture unspecified, radiographically active

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	4 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality
Treatment failure (assessed with: number of smear-negative patients to experience treatment failure[1])										
1[2]	randomised trials	serious[3]	serious[4]	serious[5,10]	very serious[6,7]	0/59 (0%)	1/54 (1.9%)	OR 0.30 (0.01 to 7.52)[8]	1 fewer per 100 (from 2 fewer to 11 more)	VERY LOW
Changes in signs and symptoms – no change in radiographic status (assessed with: number of smear-negative patients to experience no change in radiographic appearance at the end of treatment)										
1[2]	randomised trials	serious[3]	serious[4]	serious[5,10]	serious[6]	0/59 (0%)	0/54 (0%)	OR 0.92 (0.02 to 49.97)[8]	-	VERY LOW
Changes in signs and symptoms – <50% radiographic clearing (assessed with: number of smear-negative patients to experience less than 50% radiographic clearing at the end of treatment)										
1[2]	randomised trials	serious[3]	serious[4]	serious[5,10]	serious[6]	0/59 (0%)	0/54 (0%)	OR 0.92 (0.02 to 49.97)[8]	-	VERY LOW
Changes in signs and symptoms – >50% radiographic clearing (assessed with: number of smear-negative patients to experience more than 50% radiographic clearing at the end of treatment)										
1[2]	randomised trials	serious[3]	serious[4]	serious[5,10]	serious[6]	52/59 (88.1%)	52/54 (96.3%)	OR 0.29 (0.06 to 1.44)[8]	8 fewer per 100 (from 35 fewer to 1 more)	VERY LOW
Changes in signs and symptoms – complete radiographic clearing (assessed with: number of smear-negative patients to demonstrate radiographic clearing at the end of treatment)										
1[2]	randomised trials	serious[3]	serious[4]	serious[5,10]	very serious[6,7]	52/59 (88.1%)	52/54 (96.3%)	OR 0.29 (0.06 to 1.44)[8]	8 fewer per 100 (from 35 fewer to 1 more)	VERY LOW
Relapse (follow-up 60 months after treatment initiation; assessed with: number of smear-negative patients to experience relapse[1])										
1[2]	randomised trials	serious[3]	very serious[4,9]	serious[5,10]	serious[6]	0/59 (0%)	0/54 (0%)	OR 0.92 (0.02 to 49.97)[8]	-	VERY LOW

[1] *See evidence table for the full definition*
[2] *Teo et al, 2002*
[3] *Blinding unclear*
[4] *Comparability of patients at baseline was unclear*
[5] *Intervention does not exactly match the intervention of interest: did not contain all of the 4 standard recommended drugs, and the two arms vary by more than duration alone*
[6] *GRADE rule of thumb: <300 events*
[7] *Wide confidence intervals*
[8] *Odds ratio and 95% confidence intervals calculated by reviewer*
[9] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[10] *Population does not exactly match the population of interest: unclear if children are included*
Abbreviations: CI, confidence interval; OR, odds ratio

2 vs **3** months

Age: mix

HIV status: not specified – negative?

Disease status: smear-negative, culture-positive or negative (i.e. all patients in trial)

Site of disease: pulmonary

Drug sensitivity: some DR-TB

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	2 months	3 months	Relative (95% CI)	Absolute (95% CI)	
Response to treatment - culture status (assessed with: number of smear-negative patients to be culture-negative at the end of treatment)										
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	very serious[5,6,7]	serious[9]	303/303 (100%)	307/307 (100%)	OR 0.98 (0.02 to 49.9)[10]	-	VERY LOW
Relapse (follow-up 60 months after treatment initiation; assessed with: number of smear-negative patients to experience bacteriological, radiographic or clinical relapse[11])										
1[1]	randomised trials	serious[2,3,4]	serious[12]	very serious[5,6]	serious[8]	45/303 (14.9%)	21/307 (6.8%)	OR 2.38 (1.38 to 4.1)[10]	8 more per 100 (from 2 more to 16 more)	VERY LOW
Bacteriological relapse (follow-up 60 months after treatment initiation; assessed with: number of smear-negative patients to experience bacteriologically confirmed relapse[11])										
1[1]	randomised trials	serious[2,3,4]	serious[12]	very serious[5,6]	serious[8]	30/303 (9.9%)	13/307 (4.2%)	OR 2.49 (1.27 to 4.86)[10]	6 more per 100 (from 1 more to 13 more)	VERY LOW
Adverse events (any) (assessed with: number of smear-negative patients to experience any adverse reaction during chemotherapy)										
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	very serious[5,6]	serious[8]	76/303 (25.1%)	98/307 (31.9%)	OR 0.71 (0.5 to 1.02)[10,13]	7 fewer per 100 (from 13 fewer to 0 more)	VERY LOW
Adverse events requiring withdrawal of chemotherapy (assessed with: number of smear-negative patients to experience any adverse reaction requiring withdrawal of one or more drug)										
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	very serious[5,6]	serious[8]	6/303 (2%)	9/307 (2.9%)	OR 0.67 (0.24 to 1.9)[10]	1 fewer per 100 (from 2 fewer to 2 more)	VERY LOW

[1] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1979/84*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Population does not exactly match the population of interest: some cases were drug resistant, and the population may include some children (inclusion criteria = 15-75 years)*
[6] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[7] *Outcome is a substitute for an outcome of interest*
[8] *GRADE rule of thumb: <300 events*
[9] *Wide confidence intervals*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *See evidence table for the full definition*
[12] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	2 months	3 months	Relative (95% CI)	Absolute (95% CI)	
[13] *note: most adverse reactions were reported by the authors to be "trivial or mild cutaneous, vestibular or gastrointestinal episodes"*										
Abbreviations: CI, confidence interval; OR, odds ratio										

SMEAR-NEGATIVE, CULTURE-POSITIVE

4 vs **6** months

Age: mix

HIV status: not specified – negative?

Disease status: smear-negative, 1 or more positive culture, radiographically active

Site of disease: pulmonary

Drug sensitivity: susceptible only

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	4 months	6 months	Relative (95% CI)	Absolute (95% CI)	
Response to treatment - culture status (assessed with: number of smear-negative patients with 1 or more initial culture positive to be culture-negative at the end of treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[6]	very serious[7,8,9]	serious[10]	325/325 (100%)	177/177 (100%)	OR 1.83 (0.04 to 92.82)[11]	-	VERY LOW
Relapse (follow-up 5 years after treatment initiation; assessed with: number of smear-negative patients with 1 or more initial culture positive to experience relapse)										
1[1]	randomised trials	very serious[2,3,4,5]	very serious[6,12]	very serious[7,8]	serious[10]	7/325 (2.2%)	8/177 (4.5%)	OR 0.47 (0.17 to 1.3)[11]	2 fewer per 100 (from 4 fewer to 1 more)	VERY LOW
Bacteriological relapse (follow-up 5 years after treatment initiation; assessed with: number of smear-negative patients with 1 or more initial culture positive to experience bacteriologically confirmed relapse)										
1[1]	randomised trials	very serious[2,3,4,5]	very serious[6,12]	very serious[7,8]	serious[10]	5/325 (1.5%)	3/177 (1.7%)	OR 0.90 (0.21 to 3.84)[11]	0 fewer per 100 (from 1 fewer to 5 more)	VERY LOW

[1] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Analysis is not intent-to-treat*
[6] *Number of patients lost to follow-up in each group is unclear*
[7] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[8] *Population does not exactly match the population of interest: includes some children (inclusion criteria = 15-75 years), and some cases were possibly 'inactive'*
[9] *Outcome is a substitute for an outcome of interest*
[10] *GRADE rule of thumb: <300 events*
[11] *Odds ratio and 95% confidence intervals calculated by reviewer*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	4 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality
[12] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths* *Abbreviations: CI, confidence interval; OR, odds ratio*										

2 vs **3** months

Age: mix

HIV status: not specified – negative?

Disease status: smear-negative, culture-positive

Site of disease: pulmonary

Drug sensitivity: some DR-TB

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	2 months	3 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - culture status (assessed with: number of smear-negative patients with 1 or more initial culture positive to be culture-negative at the end of treatment)										
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	very serious[5,6,7]	very serious[8,9]	71/71 (100%)	68/68 (100%)	OR 1.04 (0.02 to 53.35)[10]	-	VERY LOW
Relapse (follow-up 60 months after treatment initiation; assessed with: number of smear-negative patients with 1 or more initial culture positive to experience bacteriological, radiographic or clinical relapse[11])										
1[1]	randomised trials	serious[2,3,4]	serious[12]	very serious[5,6]	serious[8]	23/71 (32.4%)	9/68 (13.2%)	OR 3.14 (1.33 to 7.42)[10]	19 more per 100 (from 4 more to 40 more)	VERY LOW
Bacteriological relapse (follow-up 60 months after treatment initiation; assessed with: number of smear-negative patients with 1 or more initial culture positive to experience bacteriologically confirmed relapse[11])										
1[1]	randomised trials	serious[2,3,4]	serious[12]	very serious[5,6]	serious[8]	16/71 (22.5%)	7/68 (10.3%)	OR 2.54 (0.97 to 6.62)[10]	12 more per 100 (from 0 fewer to 33 more)	VERY LOW

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	2 months	3 months	Relative (95% CI)	Absolute (95% CI)	Quality

[1] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1979/84*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Population does not exactly match the population of interest: may include some children (inclusion criteria = 15-75 years)*
[6] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[7] *Outcome is a substitute for an outcome of interest*
[8] *GRADE rule of thumb: <300 events*
[9] *Wide confidence intervals*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *See evidence table for the full definition*
[12] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
Abbreviations: CI, confidence interval; OR, odds ratio

SMEAR-NEGATIVE, CULTURE-NEGATIVE

3 vs **4** months

Age: mix

HIV status: not specified – negative?

Disease status: smear-negative, culture-negative

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	3 months	4 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - culture status (assessed with: number of smear-negative patients with all initial cultures negative to be culture-negative at the end of treatment)										
1[1]	randomised trials	serious[2,3,4]	serious[5]	very serious[6,7,8]	serious[9]	759/759 (100%)	359/359 (100%)	OR 2.11 (0.04 to 106.69)[10]	-	VERY LOW
Relapse (follow-up 5 years after treatment initiation; assessed with: number of smear-negative patients with all initial cultures negative to experience relapse)										
1[1]	randomised trials	serious[2,3,4]	very serious[5,11]	very serious[6,7]	serious[12]	48/759 (6.3%)	12/359 (3.3%)	OR 1.95 (1.02 to 3.72)[10]	3 more per 100 (from 0 more to 8 more)	VERY LOW
Bacteriological relapse (assessed with: number of smear-negative patients with all initial cultures negative to experience bacteriologically confirmed relapse)										
1[1]	randomised trials	serious[2,3,4]	very serious[5,11]	very serious[6,7]	serious[12]	20/759 (2.6%)	4/359 (1.1%)	OR 2.40 (0.81 to 7.08)[10]	2 more per 100 (from 0 fewer to 6 more)	VERY LOW

[1] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	3 months	4 months	Relative (95% CI)	Absolute (95% CI)	Quality

[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Unclear if loss to follow-up was similar in the 2 groups*
[6] *Population does not exactly match the population of interest: may include some children (inclusion criteria = 15-75 years), and some cases were possibly 'inactive'*
[7] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[8] *Outcome is a substitute for an outcome of interest*
[9] *Wide confidence intervals*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[12] *GRADE rule of thumb: <300 events*
Abbreviations: CI, confidence interval; OR, odds ratio

2 vs **3** months

Age: mix

HIV status: not specified – negative?

Disease status: smear-negative, culture-negative

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	2 months	3 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - culture status (assessed with: number of smear-negative patients with all initial cultures negative to be culture-negative at the end of treatment)										
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	very serious[5,6,7]	very serious[8,9]	161/161 (100%)	161/161 (100%)	OR 1.00 (0.02 to 50.71)[10]	-	VERY LOW
Relapse (follow-up 60 months after treatment initiation; assessed with: number of smear-negative patients with all initial cultures negative to experience bacteriological, radiographic or clinical relapse[11])										
1[1]	randomised trials	serious[2,3,4]	serious[12]	very serious[5,6]	serious[8]	17/161 (10.6%)	11/161 (6.8%)	OR 1.61 (0.73 to 3.55)[10]	4 more per 100 (from 2 fewer to 14 more)	VERY LOW
Bacteriological relapse (follow-up 60 months after treatment initiation; assessed with: number of smear-negative patients with all initial cultures negative to experience bacteriologically confirmed relapse[11])										
1[1]	randomised trials	serious[2,3,4]	serious[12]	very serious[5,6]	serious[8]	10/161 (6.2%)	5/161 (3.1%)	OR 2.07 (0.69 to 6.19)[10]	3 more per 100 (from 1 fewer to 13 more)	VERY LOW

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	2 months	3 months	Relative (95% CI)	Absolute (95% CI)	Quality

[1] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1979/84*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Population does not exactly match the population of interest: may include some children (inclusion criteria = 15-75 years)*
[6] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[7] *Outcome is a substitute for an outcome of interest*
[8] *GRADE rule of thumb: <300 events*
[9] *Wide confidence intervals*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *See evidence table for the full definition*
[12] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
Abbreviations: CI, confidence interval; OR, odds ratio

MIXED POPULATIONS

<6 vs **6** months

Age: mix

HIV status: not specified – negative?

Disease status: various

Site of disease: pulmonary

Drug sensitivity: susceptible / unclear (pooled)

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	<6 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - culture status (assessed with: number of patients to be culture-negative the end of treatment)										
2[1,2]	randomised trials	very serious[3,4,5,6]	very serious[7,8]	very serious[9,10,11,12]	serious[13]	1558/1558 (100%)	260/260 (100%)	OR 5.98 (0.12 to 302.19)[14]	-	VERY LOW

[1] *Research Committee of the Tuberculosis Association of India, 1984*
[2] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989*
[3] *Method of randomisation unclear*
[4] *Allocation concealment unclear*
[5] *Blinding unclear*
[6] *Analysis did not follow intent-to-treat principle*
[7] *Research Committee of the Tuberculosis Association of India, 1984: comparability of patients at baseline was unclear*
[8] *Unclear if loss to follow-up was the same in the 2 groups*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	<6 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality

[9] *Research Committee of the Tuberculosis Association of India, 1984: intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs, and doses used are inconsistent with those recommended in the British National Formulary*
[10] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: intervention does not exactly match the intervention of interest: did not contain all of the 4 standard recommended drugs*
[11] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: population does not exactly match the population of interest: includes some children (inclusion criteria = 15 to 75 years), and some cases were possibly 'inactive'*
[12] *Outcome is a substitute for an outcome of interest*
[13] *Wide confidence intervals*
[14] *Pooled odds ratio and 95% confidence intervals calculated by reviewer*
[15] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[16] *GRADE rule of thumb: <300 events*
[17] *Research Committee of the Tuberculosis Association of India, 1984: population does not exactly match the population of interest: includes some children (inclusion criteria = 15 to 45 years)*
Abbreviations: CI, confidence interval; OR, odds ratio

4 vs **6** months

Age: mix

HIV status: not specified – negative?

Disease status: various

Site of disease: pulmonary

Drug sensitivity: susceptible/unclear

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	4 months	6 months	Relative (95% CI)	Absolute (95% CI)	Quality
Relapse (follow-up 5 to 8 years after treatment initiation; assessed with: number of patients to experience relapse)										
3[1,2,3]	randomised trials	very serious[4,5,6,7]	very serious[8,9,10,17]	very serious[11,12,13]	very serious[14,15]	27/515 (5.2%)	11/369 (3%)	OR 1.90 (0.11 to 32.97)[16,18]	3 more per 100 (from 3 fewer to 47 more)	VERY LOW

[1] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989*
[2] *Singapore TB Service / British Medical Research Council, 1979/86*
[3] *Teo et al, 2002*
[4] *Method of randomisation unclear*
[5] *Allocation concealment unclear*
[6] *Blinding unclear*
[7] *Analysis is not intent-to-treat*
[8] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[9] *Teo et al, 2002: comparability of patients at baseline unclear*
[10] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council (1989) and Singapore TB Service / British Medical Research Council (1979/86): number of patients lost to follow-up in each group is unclear*
[11] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[12] *Unclear if populations contain children (inclusion criteria = 15 years or more)*
[13] *Teo et al, 2002: intervention does not exactly match the intervention of interest: varied by more than duration – the continuation phase of the 6-month regimen was intermittent (3 times weekly), whereas the 4-month regimen was daily throughout*
[14] *Wide confidence intervals*
[15] *GRADE rule of thumb: <300 events*
[16] *Pooled odds ratio and 95% confidence intervals calculated by reviewer*
[17] *Point estimates vary widely, with no overlap of the confidence intervals*
[18] *Forest plot (relapse):*

Abbreviations: CI, confidence interval; OR, odds ratio

3 vs **4 or 4.5** months

Age: mix

HIV status: not specified – negative?

Disease status: various

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	3 months	4 or 4.5 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - culture status (assessed with: number of patients to be culture-negative at the end of treatment)										
2[1,2]	randomised trials	serious[3,4,5]	serious[6,7]	very serious[8,9,10,15]	serious[11]	850/850 (100%)	448/448 (100%)	OR 1.90 (0.04 to 95.74)[12]	-	VERY LOW
Relapse (follow-up 1 year after treatment completion to 5 years after treatment initiation; assessed with: number of patients to experience relapse)										
2[1,2]	randomised trials	serious[3,4,5]	very serious[6,7,13]	very serious[8,9,15]	serious[14]	49/850 (5.8%)	13/448 (2.9%)	OR 1.88 (1 to 3.53)[12,16]	2 more per 100 (from 0 more to 7 more)	VERY LOW

[1] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989*
[2] *Mehotra et al, 1982*
[3] *Method of randomisation unclear*
[4] *Allocation concealment unclear*
[5] *Blinding unclear*
[6] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: unclear if loss to follow-up was similar in the 2 groups*
[7] *Mehotra et al, 1982: although not statistically significant, there was a higher number who did not complete treatment and for whom data was not available amongst the 3-month group (36%) than the 4.5-month group (24%)*
[8] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: population does not exactly match the population of interest: may include some children (inclusion criteria = 15-75 years), and some cases were possibly 'inactive'*
[9] *Intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs*
[10] *Outcome is a substitute for an outcome of interest*
[11] *Wide confidence intervals*
[12] *Pooled odds ratio and 95% confidence intervals calculated by reviewer*
[13] *Hong Kong Chest Service / Tuberculosis Research Centre, Madras / British Medical Research Council, 1989: follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[14] *GRADE rule of thumb: <300 events*
[15] *Mehotra et al, 1982: population does not exactly match the population of interest: may include some children (inclusion criteria: aged 12 years or more)*
[16] *Forest plot (relapse):*

Abbreviations: CI, confidence interval; OR, odds ratio

6 vs **>6** months

Appendix E: GRADE profiles

Age: mix

HIV status: not specified – negative?

Disease status: various

Site of disease: pulmonary

Drug sensitivity: some DR-TB

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	>6 months	Relative (95% CI)	Absolute (95% CI)	
Treatment failure (assessed with: number of patients to experience treatment failure)										
2[1,2]	randomised trials	serious[3,4,5,6,7]	no serious inconsistency	very serious[8,9,10,18]	very serious[11,12]	1/307 (0.33%)	1/324 (0.31%)	OR 1.08 (0.11 to 10.44)[13,18]	0 more per 100 (from 0 fewer to 3 more)	VERY LOW
Relapse (follow-up 12 months after treatment completion to 54 months after treatment initiation; assessed with: number of patients to experience relapse)										
4[1,2,20,21]	randomised trials	very serious[3,4,5,6,7]	very serious[14,15,16,17]	very serious[8,9,10,18]	very serious[11,12]	21/691 (3%)	7/577 (1.2%)	OR 2.26 (0.61 to 8.39)[13,19]	1 more per 100 (from 0 fewer to 8 more)	VERY LOW

[1] *British Thoracic Society, 1975/80*
[2] *Ziaullah et al, 2004*
[3] *Method of randomisation unclear*
[4] *British Thoracic Society, 1975/80: allocation concealment possible - "random allocations of treatment were made centrally by coordinators"*
[5] *British Thoracic Society, 1975/80: radiographer blinded to treatment allocation, but unclear if to prognostic factors or if other investigators were blinded*
[6] *Ziaullah et al (2004), British Thoracic Society (1981/2/4) and Nayar et al (1988): allocation concealment unclear*
[7] *Ziaullah et al (2004), British Thoracic Society (1981/2/4) and Nayar et al (1988): blinding unclear*
[8] *British Thoracic Society (1975/80 and 1981/2/4) and Nayar et al (1988): intervention does not exactly match the intervention of interest: does not contain all of or just the 4 standard recommended drugs*
[9] *British Thoracic Society, 1975/80: population does not exactly match the population of interest: 3.4% drug resistance at baseline*
[10] *Ziaullah et al, 2004: population does not exactly match the population of interest: includes children (33% aged 5 to 14 years, 33% aged 15 to 29 years)*
[11] *GRADE rule of thumb: <300 events*
[12] *Wide confidence intervals*
[13] *Pooled odds ratio and 95% confidence intervals calculated by reviewer*
[14] *Follow-up varies considerably between studies and between groups*
[15] *British Thoracic Society, 1981/2/4: unclear if groups received the same care except for the intervention*
[16] *British Thoracic Society, 1981/2/4: unclear if the groups were comparable for treatment completion*
[17] *British Thoracic Society, 1981/2/4: high attrition rate with regards to the number of participants for whom data is available*
[18] *Forest plot (treatment failure):*

[19] *Forest plot (relapse):*

[20] *Nayar et al, 1988*
[21] *British Thoracic Society, 1981/2/4*
Abbreviations: CI, confidence interval; OR, odds ratio

6 vs **9** months

Age: mix

HIV status: unspecified – negative?

Disease status: various

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Relapse (follow-up a minimum of 18 months after treatment completion; assessed with: number of patients to experience relapse)										
2[1,2]	randomised trials	very serious[3,4,5,6]	serious[7,8,9]	very serious[10,11]	very serious[12,13]	11/380 (2.9%)	2/264 (0.76%)	OR 3.34 (0.45 to 24.5)[14,15]	2 more per 100 (from 0 fewer to 15 more)	VERY LOW

[1] *British Thoracic Society, 1981/2/4*
[2] *Ziaullah et al, 2004*
[3] *Method of randomisation unclear*
[4] *Allocation concealment*
[5] *Blinding unclear*
[6] *Analysis did not follow the intent-to-treat principle*
[7] *British Thoracic Society, 1981/2/4: unclear if groups received the same care except for the intervention*
[8] *British Thoracic Society, 1981/2/4: unclear if the groups were comparable for treatment completion*
[9] *British Thoracic Society, 1981/2/4: high attrition rate with regards to the number of participants for whom data is available*
[10] *British Thoracic Society, 1981/2/4: intervention does not exactly match the intervention of interest: did not contain all of or just the 4 standard recommended drugs, and the 2 regimens vary by more than duration*
[11] *Ziaullah et al, 2004: population does not exactly match the population of interest: includes children (33% aged 5 to 14 years, 33% aged 15 to 29 years)*
[12] *Wide confidence intervals*
[13] *GRADE rule of thumb: <300 events*
[14] *Pooled odds ratio and 95% confidence intervals calculated by reviewer*
[15] *Forest plot (relapse):*

Abbreviations: CI, confidence interval; OR, odds ratio

HIV-POSITIVE

6 vs **>6** months

Age: mix

HIV status: positive

Disease status: various

Site of disease: respiratory

Drug sensitivity: some DR-TB

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	>6 months	Relative (95% CI)	Absolute (95% CI)	
Relapse (follow-up 24 to 36 months after treatment initiation; assessed with: number of patients with HIV to experience relapse)										
2[1,11]	randomised trials	very serious[2,3,4,5]	very serious[6,7,12]	serious[8,13]	very serious[9]	22/290 (7.6%)	17/284 (6.0%)	OR 0.61 (0.02 to 16.51)[10,14]	2 fewer per 100 (from 6 fewer to 45 more)	VERY LOW

[1] *Perriens et al, 1995*
[2] *Perriens et al, 1995: method of randomisation unclear*
[3] *Perriens et al, 1995: allocation concealment unclear*
[4] *Perriens et al, 1995: patients blinded, but investigators and those administering care were not*
[5] *Swaminathan et al, 2010: unblinded*
[6] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[7] *Perriens et al, 1995: groups were statistically comparable at baseline, but the 12-month arm has a higher CD4 count at baseline*
[8] *Swaminathan et al, 2010: population does not exactly match the population of interest: some DR-TB, and there may be some children (inclusion = 15 years and above)*
[9] *Wide confidence intervals; GRADE rule of thumb: <300 events*
[10] *Pooled odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Swaminathan et al, 2010*
[12] *Point estimates vary widely, with no overlap of the confidence intervals*
[13] *Swaminathan et al, 2010: doses used are inconsistent with those listed in the British National Formulary*
[14] *Forest plot (relapse):*

Abbreviations: CI, confidence interval; OR, odds ratio

6 vs **9** months

Age: mix

HIV status: positive

Disease status: smear-positive or radiographically active

Site of disease: respiratory (91% pulmonary, 9% pleural or lymph node)

Drug sensitivity: some DR-TB

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Mortality (all cause) (follow-up 36 months after treatment initiation; assessed with: number of patients with HIV to die (all cause))										
1[1]	randomised trials	serious[2]	serious[3]	serious[4,10]	serious[5]	33/167 (19.8%)	37/160 (23.1%)	OR 0.82 (0.48 to 1.38)	3 fewer per 100 (from 11 fewer to 6 more)	VERY LOW

Appendix E: GRADE profiles

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Cure (assessed with: number of patients with HIV to achieve a 'favourable response' by the end of treatment[6])										
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[4,10]	serious[5]	138/167 (82.6%)	122/160 (76.3%)	OR 1.48 (0.86 to 2.55)[7]	6 more per 100 (from 3 fewer to 13 more)	VERY LOW
Treatment failure (assessed with: number of patients with HIV to experience treatment failure[6])										
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[4,10]	serious[5]	8/167 (4.8%)	11/160 (6.9%)	OR 0.68 (0.27 to 1.74)[7]	2 fewer per 100 (from 5 fewer to 5 more)	VERY LOW
Bacteriological relapse (follow-up 36 months after treatment initiation; assessed with: number of patients with HIV to experience bacteriologically confirmed relapse[6])										
1[1]	randomised trials	serious[2]	serious[3]	serious[4,10]	serious[5]	21/167 (12.6%)	8/160 (5.0%)	OR 2.73 (1.17 to 6.36)[7]	8 more per 100 (from 1 more to 20 more)	VERY LOW
Adverse events (any) (assessed with: number of patients with HIV to experience any adverse event resulting from drug toxicity)										
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[4,10]	very serious[5,9]	1/167 (0.6%)	1/160 (0.63%)	OR 0.96 (0.06 to 15.45)[7]	0 fewer per 100 (from 1 fewer to 8 more)	VERY LOW
Adherence - treatment default (assessed with: number of patients with HIV to default treatment)										
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[4,10]	serious[5]	11/167 (6.6%)	16/160 (10%)	OR 0.63 (0.29 to 1.41)[7]	3 fewer per 100 (from 7 fewer to 4 more)	VERY LOW

[1] Swaminathan et al, 2010
[2] Unblinded
[3] Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths
[4] Population does not exactly match the population of interest: some DR-TB (12% at baseline), and there may be some children (inclusion = 15 years and above)
[5] GRADE rule of thumb: <300 events
[6] See evidence table for the full definition
[7] Odds ratio and 95% confidence intervals calculated by reviewer
[8] Analysis did not follow intent-to-treat principle
[9] Wide confidence intervals
[10] Doses used are inconsistent with those listed in the British National Formulary
Abbreviations: CI, confidence interval; OR, odds ratio

Age: mix

HIV status: positive

Disease status: smear- or culture-positive or radiographically active

Site of disease: respiratory (91% pulmonary, 9% pleural or lymph node)

Drug sensitivity: susceptible only

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Mortality (follow-up 36 months after treatment initiation; assessed with: number of drug susceptible patients with HIV to die (all cause))										
1[1]	randomised trials	very serious[2,3]	serious[4]	serious[5,11]	very serious[6,7]	3/100 (3%)	10/97 (10.3%)	OR 0.27 (0.07 to 1.01)[9]	7 fewer per 100 (from 10 fewer to 0 more)	VERY LOW
Cure (assessed with: number of drug susceptible patients with HIV to achieve a 'favourable response' by the end of treatment[8])										
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	serious[5,11]	serious[6]	3/100 (3%)	10/97 (10.3%)	OR 0.27 (0.07 to 1.01)[9]	7 fewer per 100 (from 10 fewer to 0 more)	VERY LOW
Treatment failure (assessed with: number of drug susceptible patients with HIV to experience treatment failure[8])										
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	serious[5,11]	serious[6]	3/100 (3%)	7/97 (7.2%)	OR 0.40 (0.1 to 1.58)[9]	4 fewer per 100 (from 6 fewer to 4 more)	VERY LOW
Adverse events (any) (assessed with: number of drug susceptible patients with HIV to experience any adverse event resulting from drug toxicity)										
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	serious[5,11]	very serious[6,7]	1/100 (1%)	0/97 (0%)	OR 2.93 (0.12 to 73.05)[9]	-	VERY LOW
Adherence - treatment default (assessed with: number of drug susceptible patients with HIV to default treatment)										
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	serious[5,11]	serious[6]	5/100 (5%)	4/97 (4.1%)	OR 1.22 (0.32 to 4.7)[9]	1 more per 100 (from 3 fewer to 13 more)	VERY LOW

[1] *Swaminathan et al, 2010*
[2] *Unblinded*
[3] *Analysis did not follow intent-to-treat principle*
[4] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[5] *Population does not exactly match the population of interest: may be some children (inclusion = 15 years and above)*
[6] *GRADE rule of thumb: <300 events*
[7] *Wide confidence intervals*
[8] *See evidence table for the full definition*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*
[10] *Population does not exactly match the population of interest: some extrapulmonary TB*
[11] *Doses used are inconsistent with those listed in the British National Formulary*
Abbreviations: CI, confidence interval; OR, odds ratio

Age: mix

HIV status: positive

Disease status: culture-positive

Site of disease: pulmonary

Drug sensitivity: some DR-TB

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Cure (assessed with: number of culture-positive patients with HIV to achieve a 'favourable response' by the end of treatment[1])										
1[2]	randomised trials	very serious[3,4]	no serious inconsistency	serious[5,9]	serious[6]	96/117 (82.1%)	81/110 (73.6%)	OR 1.64 (0.87 to 3.09)[7]	8 more per 100 (from 3 fewer to 16 more)	VERY LOW
Treatment failure (assessed with: number of culture-positive patients with HIV to experience treatment failure[1])										
1[2]	randomised trials	very serious[3,4]	no serious inconsistency	serious[5,9]	serious[6]	8/117 (6.8%)	11/110 (10%)	OR 0.66 (0.26 to 1.71)[7]	3 fewer per 100 (from 7 fewer to 6 more)	VERY LOW
Adverse events (any) (assessed with: number of culture-positive patients with HIV to experience any adverse event resulting from drug toxicity)										
1[2]	randomised trials	very serious[3,4]	no serious inconsistency	serious[5,9]	very serious[6,8]	1/117 (0.85%)	0/110 (0%)	OR 2.85 (0.11 to 70.6)[7]	-	VERY LOW

[1] *See evidence table for the full definition*
[2] *Swaminathan et al, 2010*
[3] *Unblinded*
[4] *Analysis did not follow intent-to-treat principle*
[5] *Population does not exactly match the population of interest: some DR-TB, and there may be some children (inclusion = 15 years and above)*
[6] *GRADE rule of thumb: <300 events*
[7] *Odds ratio and 95% confidence intervals calculated by reviewer*
[8] *Wide confidence intervals*
[9] *Doses used are inconsistent with those listed in the British National Formulary*
Abbreviations: CI, confidence interval; OR, odds ratio

6 vs **12** months

Age: adults only (?)

HIV status: positive

Disease status: smear- and culture-positive

Site of disease: pulmonary

Drug sensitivity: unclear

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	12 months	Relative (95% CI)	Absolute (95% CI)	
Mortality (follow-up 24 months after treatment initiation; assessed with: number of smear- and culture-positive patients with HIV to die from tuberculosis)										
1[1]	randomised trials	very serious[2,3,4]	serious[5,6]	no serious indirectness	very serious[7,8]	1/123 (0.81%)	0/124 (0%)	OR 3.05 (0.12 to 75.58)[9]	-	VERY LOW
Relapse (follow-up 24 months after treatment initiation; assessed with: number of smear- and culture-positive patients with HIV to experience relapse)										
1[1]	randomised	very	serious[5,6]	no serious	serious[7]	1/123	9/124	OR 0.10 (0.01	6 fewer per	VERY LOW

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	12 months	Relative (95% CI)	Absolute (95% CI)	Quality
	trials	serious[2,3,4]		indirectness		(0.81%)	(7.3%)	to 0.84)[9]	100 (from 1 fewer to 7 fewer)	

[1] *Perriens et al, 1995*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Patients blinded, but investigators and those administering care were not*
[5] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
[6] *Groups were statistically comparable at baseline, but the 12-month arm has a higher CD4 count at baseline*
[7] *GRADE rule of thumb: <300 events*
[8] *Wide confidence intervals*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*
Abbreviations: CI, confidence interval; OR, odds ratio

HIV-POSITIVE, CULTURE-NEGATIVE

6 vs **9** months

Age: mix

HIV status: positive

Disease status: culture-negative

Site of disease: pulmonary

Drug sensitivity: some DR-TB

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Cure (assessed with: number of culture-negative patients with HIV to achieve a 'favourable response' by the end of treatment[1])										
1[2]	randomised trials	very serious[3,4]	no serious inconsistency	serious[5,9]	serious[6]	28/34 (82.4%)	31/38 (81.6%)	OR 1.05 (0.32 to 3.51)[7]	1 more per 100 (from 23 fewer to 12 more)	VERY LOW
Treatment failure (assessed with: number of culture-negative patients with HIV to experience treatment failure[1])										
1[2]	randomised trials	very serious[3,4]	no serious inconsistency	serious[5,9]	very serious[6,8]	0/34 (0%)	0/38 (0%)	OR 1.12 (0.02 to 57.77)[7]	-	VERY LOW
Adverse events (any) (assessed with: number of culture-negative patients with HIV to experience any adverse event resulting from drug toxicity)										
1[2]	randomised trials	very serious[3,4]	no serious inconsistency	serious[5,9]	very serious[6,8]	0/34 (0%)	1/38 (2.6%)	OR 0.36 (0.01 to 9.2)[7]	2 fewer per 100 (from 3 fewer to 17 more)	VERY LOW

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality

[1] *See evidence table for the full definition*
[2] *Swaminathan et al, 2010*
[3] *Unblinded*
[4] *Analysis did not follow intent-to-treat principle*
[5] *Population does not exactly match the population of interest: some DR-TB, and there may be some children (inclusion = 15 years and above)*
[6] *GRADE rule of thumb: <300 events*
[7] *Odds ratio and 95% confidence intervals calculated by reviewer*
[8] *Wide confidence intervals*
[9] *Doses used are inconsistent with those listed in the British National Formulary*
Abbreviations: CI, confidence interval; OR, odds ratio

A.8 RQ M

A.8.1 Duration of treatment in children with respiratory tuberculosis

Intervention: **9 months**

Comparator: **12 months**

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	12 months	Relative (95% CI)	Absolute (95% CI)	Quality
Recurrence (number to experience clinical or radiological recurrence in the 12 months after treatment completion; follow-up 12 months after treatment completion)										
1[1]	randomised trials	serious[2,3,4]	serious[6]	very serious[7,12,13]	very serious[8,9]	0/18 (0%)	0/18 (0%)	OR 1.00 (0.02 to 53.12)[10]	-	VERY LOW
Adverse events - hepatotoxicity (number to experience elevated levels of serum aspartate aminotransferase and alanine aminotransferase)										
1[1]	randomised trials	very serious[2,3,4,11]	serious[6]	very serious[7,13]	very serious[8,9]	0/18 (0%)	1/18 (5.6%)	OR 0.32 (0.01 to 8.27)[10]	4 fewer per 100 (from 5 fewer to 27 more)	VERY LOW
Adherence (number excluded due to "poor compliance")										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	very serious[7,13]	serious[8]	0/18 (0%)	3/18 (16.7%)	OR 0.12 (0.01 to 2.5)[10]	14 fewer per 100 (from 16 fewer to 17 more)	VERY LOW

[1] *Kansoy et al, 1998*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Outcome definition not provided*
[6] *Loss to follow-up varied between the two arms: 3 of 18 patients were excluded from the 9-month group for "poor compliance", none were excluded from the 12-month group*
[7] *Intervention is not the same as the intervention of interest: combination was not the 4 drugs in the standard recommended regimen, and intervention and comparator varied by more than duration alone*
[8] *GRADE rule of thumb event number <300*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	12 months	Relative (95% CI)	Absolute (95% CI)	Quality

[9] *Wide confidence intervals*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Outcome not clearly defined - thresholds for 'elevated' aspartate aminotransferase and alanine aminotransferase not given*
[12] *Substitute for outcome of interest (relapse)*
[13] *Prescribed doses of isoniazid and streptomycin are above that recommended by the British National Formulary*
Abbreviations: CI, confidence interval; OR, odds ratio

A.9 RQs N and Q

A.9.1 Use of adjunctive corticosteroids in people with active tuberculosis

PULMONARY TUBERCULOSIS

Prednisolone *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (follow-up 1 to 3 years; assessed with: number of deaths)												
2[1,2]	randomised trials	serious[3,4]	very serious[5,6]	serious[7]	serious[8]	none	17/184 (9.2%)	14/181 (7.7%)	OR 1.28 (0.59 to 2.77)[9,20]	2 more per 100 (from 3 fewer to 11 more)	⊙○○○ VERY LOW	
Response to treatment – sputum conversion at 1 month (assessed with: number of patients to have a sputum culture negative for M. tuberculosis after 1 month of treatment)												
2[2,10]	randomised trials	serious[11,12]	very serious[6,13]	very serious[14,15]	serious[8]	none	139/354 (39.3%)	115/363 (31.7%)	OR 1.67 (0.65 to 4.31)[9,21]	12 more per 100 (from 9 fewer to 35 more)	⊙○○○ VERY LOW	
Response to treatment – sputum conversion at 2 months (assessed with: number of patients to have a sputum culture negative for M. tuberculosis after 2 months of treatment)												
2[2,10]	randomised trials	serious[11,12]	very serious[6,13]	very serious[14,15]	serious[8]	none	247/354 (69.8%)	247/363 (68%)	OR 1.08 (0.78 to 1.5)[9,22]	2 more per 100 (from 6 fewer to 8 more)	⊙○○○ VERY LOW	
Response to treatment – sputum conversion at 3 months (assessed with: number of patients to have a sputum culture negative for M. tuberculosis after 3 months of treatment)												
1[10]	randomised trials	serious[11,12]	serious[13]	very serious[14,15]	serious[8]	none	187/261 (71.6%)	183/269 (68%)	OR 1.19 (0.82 to 1.72)[9]	4 more per 100 (from 4 fewer to	⊙○○○ VERY LOW	

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
										11 more)		
Change in signs and symptoms - disappearance of cavitation (assessed with: number of patients in whom cavitation was present on admission but disappeared by the end of treatment)												
1[10]	randomised trials	serious[11,12]	serious[13]	very serious[14,15]	serious[8]	none	103/245 (42%)	88/250 (35.2%)	OR 1.34 (0.93 to 1.92)[9]	7 more per 100 (from 2 fewer to 16 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - radiographic improvement (assessed with: number of patients to achieve moderate or greater radiographic improvement after 2 months of treatment)												
1[10]	randomised trials	serious[11,12]	serious[13]	very serious[14,15]	serious[8]	none	130/261 (49.8%)	107/269 (39.8%)	OR 1.50 (1.06 to 2.12)[9]	10 more per 100 (from 1 more to 19 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - lessening of cavitation (assessed with: number of patients in whom the cavitation that was present on admission had lessened by the end of treatment)												
1[10]	randomised trials	serious[11,12]	serious[13]	very serious[14,15]	serious[8]	none	97/245 (39.6%)	111/250 (44.4%)	OR 0.82 (0.57 to 1.17)[9]	5 fewer per 100 (from 13 fewer to 4 more)	⊙○○○ VERY LOW	
Change in signs and symptoms – endobronchial lesions (assessed with: number of endobronchial lesions identified using bronchoscopy before treatment to have improved after 2 months of treatment[16])												
1[17]	randomised trials	serious[11,12]	serious[18]	no serious indirectness	serious[8]	none	24/35 (68.6%)	22/30 (73.3%)	OR 0.79 (0.27 to 2.33)[9]	5 fewer per 100 (from 31 fewer to 13 more)	⊙○○○ VERY LOW	
Change in signs and symptoms – pulmonary lesions (assessed with: number of lesions identified using chest-x-ray before treatment to have improved after 2 months of treatment[16])												
1[17]	randomised trials	serious[11,12]	serious[18]	no serious indirectness	serious[8]	none	22/35 (62.9%)	23/30 (76.7%)	OR 0.68 (0.19 to 2.48)[9]	8 fewer per 100 (from 38 fewer to 12 more)	⊙○○○ VERY LOW	
Relapse (HIV-negative) (follow-up 1 to 3 years; assessed with: number of patients to experience relapse during follow-up)												
2[1,10]	randomised trials	very serious[3,4,11,12]	serious[5,13]	serious[7]	very serious[8,19]	none	5/352 (1.4%)	6/356 (1.7%)	OR 0.86 (0.26 to 2.84)[9,23]	0 fewer per 100 (from 1 fewer to 3 more)	⊙○○○ VERY LOW	

[1] *Bilaçeroglu et al, 1999*
[2] *Mayanja-Kizza et al, 2005*
[3] *Bilaçeroglu et al, 1999: method of randomisation and use of allocation concealment is unclear*
[4] *Bilaçeroglu et al, 1999: only laboratory staff and those reading chest scans were blinded*
[5] *Bilaçeroglu et al, 1999: follow-up period was appropriate (1 to 3 years), although it is unclear if it was the same in each group*

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		

[6] *Mayanja-Kizza et al, 2005: fever and night sweats were present at baseline in significantly more patients who went on to receive prednisolone than amongst those that went on to receive placebo*
[7] *Bilaçeroglu et al, 1999: antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[8] *GRADE rule of thumb: <300 events*
[9] *Odds ratio and 95% confidence interval calculated by reviewer*
[10] *Tuberculosis Research Centre (Madras), 1983*
[11] *Tuberculosis Research Centre (Madras) (1983) and Park et al (1997): method of randomisation and use of allocation concealment and blinding is unclear*
[12] *Unclear if the analysis follows the intent-to-treat principle*
[13] *Tuberculosis Research Centre (Madras), 1983: unclear if the groups were comparable for treatment completion and availability of outcome data*
[14] *Outcome is a substitute for an outcome of interest*
[15] *Tuberculosis Research Centre (Madras), 1983: antituberculosis regimens do not use all of or just the 4 standard recommended drugs; in particular, rifampicin is not used throughout*
[16] *See evidence table for full definition*
[17] *Park et al, 1997*
[18] *Follow-up not for the full treatment period*
[19] *Wide confidence interval*
[20] *Forest plot (mortality):*

[21] *Forest plot (response to treatment – sputum conversion at 1 month):*

[22] *Forest plot (response to treatment – sputum conversion at 2 month):*

[23] *Forest plot (relapse):*

Prednisolone *vs* **antituberculosis chemotherapy alone or plus placebo** in **people with HIV**

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (HIV-positive) (follow-up 1 to 3 years; assessed with: number of deaths)												
1[1]	randomised trials	no serious risk of bias	serious[2]	no serious indirectness	serious[3]	none	17/93 (18.3%)	14/94 (14.9%)	OR 1.28 (0.59 to 2.77)[4]	3 more per 100 (from 6 fewer to 18 more)	⊙⊙○○ LOW	
Event- free survival (HIV-positive) (follow-up 1 to 3 years; assessed with: number of patients to survive to 36 months without significant adverse event)												
1[1]	randomised trials	no serious risk of bias	serious[2]	serious[5]	serious[3]	none	36/93 (38.7%)	40/94 (42.6%)	OR 0.85 (0.48 to 1.53)[4]	4 fewer per 100 (from 16 fewer to 11 more)	⊙○○○ VERY LOW	
Treatment failure (HIV-positive) (assessed with: number of patients to experience treatment failure[6])												

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
1[1]	randomised trials	no serious risk of bias	serious[2]	no serious indirectness	very serious[3,7]	none	1/93 (1.1%)	1/94 (1.1%)	OR 1.01 (0.06 to 16.41)[4]	0 more per 100 (from 1 fewer to 14 more)	⊙○○○ VERY LOW	
Response to treatment – sputum conversion at 1 month (HIV-positive) (assessed with: number of patients to have a sputum culture negative for M. tuberculosis after 1 month of treatment)												
1[1]	randomised trials	no serious risk of bias	serious[2]	serious[5]	serious[3]	none	58/93 (62.4%)	35/94 (37.2%)	OR 2.79 (1.54 to 5.05)[4]	25 more per 100 (from 11 more to 38 more)	⊙○○○ VERY LOW	
Recurrence (HIV-positive) (assessed with: number of patients to experience recurrence within 2 years of initiating treatment[6])												
1[1]	randomised trials	no serious risk of bias	serious[2]	serious[5]	serious[3]	none	8/93 (8.6%)	11/94 (11.7%)	OR 0.71 (0.27 to 1.85)[4]	3 fewer per 100 (from 8 fewer to 8 more)	⊙○○○ VERY LOW	
Adverse events (HIV-positive) (assessed with: number of patients to experience any adverse event)												
1[1]	randomised trials	no serious risk of bias	serious[2]	no serious indirectness	very serious[3,7]	none	87/93 (93.5%)	82/94 (87.2%)	OR 2.55 (0.86 to 7.54)[4]	7 more per 100 (from 2 fewer to 11 more)	⊙○○○ VERY LOW	
Response to treatment – sputum conversion at 2 months (HIV-positive) (assessed with: number of patients to have a sputum culture negative for M. tuberculosis after 2 months of treatment)												
1[1]	randomised trials	no serious risk of bias	serious[2]	serious[5]	serious[3]	none	80/93 (86%)	80/94 (85.1%)	OR 1.08 (0.48 to 2.44)[4]	1 more per 100 (from 12 fewer to 8 more)	⊙○○○ VERY LOW	
Adverse events (severe or life threatening) (HIV-positive) (assessed with: of patients to experience a severe or life-threatening adverse event)												
1[1]	randomised trials	no serious risk of bias	serious[2]	no serious indirectness	serious[3]	none	22/93 (23.7%)	18/94 (19.1%)	OR 1.31 (0.65 to 2.64)[4]	5 more per 100 (from 6 fewer to 19 more)	⊙⊙○○ LOW	

[1] *Mayanja-Kizza et al, 2005*
[2] *Fever and night sweats were present at baseline in significantly more patients who went on to receive prednisolone than amongst those that went on to receive placebo*
[3] *GRADE rule of thumb: <300 events*
[4] *Odds ratio and 95% confidence interval calculated by reviewer*
[5] *Outcome is a substitute for an outcome of interest*
[6] *See evidence table for full definition*
[7] *Wide confidence interval*

Prednisolone *vs* **antituberculosis chemotherapy alone or plus placebo** in **people without HIV**

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (HIV-negative) (follow-up 1 to 3 years; assessed with: number of deaths)												
1[1]	randomised trials	serious[2,3]	serious[4]	serious[5]	very serious[6,7]	none	0/91 (0%)	0/87 (0%)	OR 0.96 (0.02 to 48.73)[8]	-	⊙○○○ VERY LOW	
Response to treatment - decrease in bacillary count (HIV-negative) (follow-up 50 days; assessed with: number of to experience a drop in bacillary count 50 days after prednisolone was initiated)												
1[1]	randomised trials	serious[2,3]	serious[4]	very serious[5,9]	very serious[6,7]	none	91/91 (100%)	81/87 (93.1%)	OR 14.60 (0.81 to 263.12)[8]	6 more per 100 (from 1 fewer to 7 more)	⊙○○○ VERY LOW	
Response to treatment - marked decrease in bacillary count (HIV-negative) (follow-up 50 days; assessed with: number of to experience a marked drop in bacillary count 50 days after prednisolone was initiated)												
1[1]	randomised trials	serious[2,3]	serious[4]	very serious[5,9]	serious[7]	none	78/91 (85.7%)	54/87 (62.1%)	OR 3.67 (1.77 to 7.61)[8]	24 more per 100 (from 12 more to 30 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - fever (HIV-negative) (measured with: change in temperature within 72 hours; better indicated by lower values)												
1[1]	randomised trials	serious[2,3]	serious[4]	serious[5]	serious[10]	none	91	87	-	MD 1.4°C higher[11]	⊙○○○ VERY LOW	
Change in signs and symptoms - weight (HIV-negative) (measured with: weight change during treatment; better indicated by lower values)												
1[1]	randomised trials	serious[2,3]	serious[4]	serious[5]	serious[10]	none	91	87	-	MD 1.4kg higher[11]		
Change in signs and symptoms - marked radiographic improvement (HIV-negative) (assessed with: number of patients to experience marked radiographic improvement 50 days after prednisolone initiation[12])												
1[1]	randomised trials	serious[2,3]	serious[4]	serious[5]	serious[7]	none	15/91 (16.5%)	8/87 (9.2%)	OR 1.95 (0.78 to 4.86)[11]	7 more per 100 (from 2 fewer to 24 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - radiographic improvement (HIV-negative) (assessed with: number of patients to experience radiographic improvement (marked, moderate or slight) 50 days after prednisolone initiation[12])												
1[1]	randomised trials	serious[2,3]	serious[4]	serious[5]	very serious[6,7]	none	91/91 (100%)	83/87 (95.4%)	OR 9.86 (0.52 to 185.96)[1][1]	4 more per 100 (from 4 fewer to 5 more)	⊙○○○ VERY LOW	

[1] *Bilaçeroglu et al, 1999*
[2] *Bilaçeroglu et al, 1999: method of randomisation and use of allocation concealment is unclear*
[3] *Bilaçeroglu et al, 1999: only laboratory staff and those reading chest scans were blinded*
[4] *Bilaçeroglu et al, 1999: follow-up period was appropriate (1 to 3 years), although it is unclear if it was the same in each group*
[5] *Bilaçeroglu et al (1999) and Tuberculosis Research Centre (Madras) (1983): antituberculosis regimens do not use all of or just the 4 standard recommended drugs*

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
[6] *Wide confidence interval* [7] *GRADE rule of thumb: <300 events* [8] *Odds ratio and 95% confidence interval calculated by reviewer* [9] *Outcome is a substitute for an outcome of interest* [10] *Authors did not provide sufficient data to calculate a confidence interval* [11] *Mean difference calculated by reviewer* [12] *See evidence table for full definition*												

PLEURAL TUBERCULOSIS

Dexamethasone *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus dexamethasone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Changes in signs and symptoms – weight (follow-up unclear; measured with: weight at the end of follow-up; better indicated by higher values)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	30	20	-	MD 1.6kg higher[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – weight change (follow-up unclear; measured with: change in mean weight from baseline to the end of follow-up; better indicated by higher values)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	30	20	-	MD 0.5kg higher[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – cough (follow-up unclear; measured with: time to relief of cough; better indicated by lower values)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	30	20	-	MD 12.1days lower[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – pleural effusion (follow-up unclear; measured with: time taken for complete absorption of pleural effusion; better indicated by lower values)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	30	20	-	MD 47.7 days lower[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – large pleural effusion (follow-up unclear; measured with: time taken for complete absorption of a large pleural effusion; better indicated by lower values)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	9	4	-	MD 63.8 days lower[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – medium pleural effusion (follow-up unclear; measured with: time taken for complete absorption of a medium pleural effusion; better indicated by lower values)												
1[1]	non-randomised	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	16	12	-	MD 50.0 days	⊙○○○ VERY	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus dexamethasone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
	trials									lower[9]	LOW	
Changes in signs and symptoms – small pleural effusion (follow-up unclear; measured with: time taken for complete absorption of a small pleural effusion; better indicated by lower values)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	5	4	-	MD 30.0 days lower[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – chest pain (follow-up unclear; measured with: time to relief of chest pain; better indicated by lower values)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	30	20	-	MD 13.8 days lower[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – shortness of breath (follow-up unclear; measured with: time to relief of shortness of breath; better indicated by lower values)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	30	20	-	MD 12.6 days lower[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – temperature (follow-up unclear; measured with: time to normalisation of temperature; better indicated by lower values)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[8]	none	30	20	-	MD 19.8 days lower[9]	⊙○○○ VERY LOW	
Recurrence (follow-up unclear; assessed with: number of patients to experience recurrence)												
1[1]	non-randomised trials	very serious[2,3,4]	serious[5,6]	serious[7]	serious[10]	none	0/39 (0%)	4/20 (20%)	OR 0.06 (0 to 1.19)[11]	19 fewer per 100 (from 20 fewer to 3 more)	⊙○○○ VERY LOW	

[1] *Singh and Yesikar, 1965*
[2] *No randomisation*
[3] *No allocation concealment*
[4] *No blinding*
[5] *Unclear if the groups were comparable at baseline*
[6] *Unclear if the groups received the same care apart from the intervention(s) studied; details provided are limited*
[7] *Antituberculosis regimens do not use all of or just the 4 standard recommended drugs; of particular note is that rifampicin is not used, and that only a 2-drug regimen was used*
[8] *Authors did not provide sufficient data to calculate a confidence interval*
[9] *Mean difference calculated by reviewer*
[10] *GRADE rule of thumb: <300 events*
[11] *Odds ratio and 95% confidence intervals*

Prednisolone *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Changes in signs and symptoms – disappearance of clinical signs and symptoms (follow-up unclear; measured with: time to disappearance of clinical signs and symptoms (including fever, chest pain and dyspnea); better indicated by lower values)												
1[1]	randomised trials	very serious[2,3,4]	no serious inconsistency[5]	serious[6]	no serious imprecision	none	21	19	-	MD 6.8 days lower (14.3 lower to 0.07 higher)[7]	⊙○○○ VERY LOW	
Changes in signs and symptoms – pleural effusion (follow-up unclear; measured with: time to clearance of pleural effusion (as defined by roentgenologic evidence of clearing of the lung field, with visualisation of the diaphragm and costophrenic angle); better indicated by lower values)												
1[1]	randomised trials	very serious[2,3,4]	no serious inconsistency[5]	serious[6]	serious[8]	none	21	19	-	MD 68.7 days lower[7]	⊙○○○ VERY LOW	
Changes in signs and symptoms – fever (follow-up unclear; measured with: duration of fever at 46 months; better indicated by lower values)												
1[9]	randomised trials	serious[2]	no serious inconsistency[5]	serious[6]	serious[8]	none	57	60	-	MD 0.83 days lower[7]	⊙○○○ VERY LOW	
Changes in signs and symptoms – pleural adhesions (follow-up unclear; assessed with: number of patients to experience pleural adhesions)												
1[1]	randomised trials	very serious[2,3,4]	no serious inconsistency[5]	serious[6]	serious[10]	none	1/21 (4.8%)	3/19 (15.8%)	OR 0.27 (0.03 to 2.82)[14]	11 fewer per 100 (from 15 fewer to 19 more)	⊙○○○ VERY LOW	

[1] *Lee et al, 1988*
[2] *Method of randomisation and use of allocation concealment is unclear*
[3] *Analysis did not follow the intent-to-treat principle*
[4] *Blinding is unclear*
[5] *Unclear if the groups received the same care apart from the intervention(s) studied; details provided are limited*
[6] *Antituberculosis regimens do not use all of or just the 4 standard recommended drugs; of particular note is that Galarza et al (1995) used only a 2-drug antituberculosis regimen*
[7] *Mean difference calculated by reviewer*
[8] *Authors did not provide sufficient data to calculate a confidence interval*
[9] *Galarza et al, 1995*
[10] *GRADE rule of thumb: <300 events*
[11] *Wide confidence intervals*
[12] *Wyser et al, 1996*
[13] *Wyser et al, 1996: although not statistically significant (p = 0.06), more patients receiving placebo (44.4%) had pleuritis and pulmonary tuberculosis than amongst those receiving prednisolone (21.2%)*
[14] *Odds ratio and 95% confidence interval calculated by reviewer*

Prednisolone *vs* **antituberculosis chemotherapy alone or plus placebo** in **people with HIV**

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (HIV-positive) (measured with: mortality rate; better indicated by lower values)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	99	98	rate ratio 0.84 deaths/100 person years higher (0.53 to 1.32 higher)	-	⊙⊙⊙⊙ HIGH	
Changes in signs and symptoms – anorexia (HIV-positive) (assessed with: number of patients to be anorexic after 24 weeks of treatment)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	none	12/99 (12.1%)	3/98 (3.1%)	OR 4.37 (1.19 to 16)[4]	9 more per 100 (from 1 more to 31 more)	⊙⊙○○ LOW	
Changes in signs and symptoms – weight (HIV-positive) (measured with: weight after 24 weeks of treatment; better indicated by higher values)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	99	98	-	MD 3kg higher[6]	⊙⊙⊙○ MODERATE	
Changes in signs and symptoms – cough (HIV-positive) (assessed with: number of patients with a cough after 24 weeks of treatment)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	26/99 (26.3%)	14/98 (14.3%)	OR 2.14 (1.04 to 4.4)[6]	12 more per 100 (from 0 more to 28 more)	⊙⊙⊙○ MODERATE	
Changes in signs and symptoms – pleural effusion (HIV-positive) (assessed with: number of patients with pleural effusion after 24 weeks of treatment)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	7/99 (7.1%)	17/98 (17.3%)	OR 0.36 (0.14 to 0.92)[4]	10 fewer per 100 (from 1 fewer to 14 fewer)	⊙⊙⊙○ MODERATE	
Recurrence (HIV-positive) (measured with: recurrence rate; better indicated by lower values)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	serious[7]	no serious imprecision	none	99	98	-	recurrence rate 2.3 higher (0.6 to 9 higher)	⊙⊙⊙○ MODERATE	

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Adverse events requiring treatment discontinuation (HIV-positive) (assessed with: number of patients to experience an adverse event that required discontinuation of placebo/prednisolone)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	none	9/99 (9.1%)	2/98 (2%)	OR 4.80 (1.01 to 22.82)[4]	7 more per 100 (from 0 more to 30 more)	⊙⊙⊙○ MODERATE	
Adverse events - incidence of HIV-related disease: Kaposi sarcoma (HIV-positive) (assessed with: number of patients to experience Kaposi sarcoma)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2,3]	none	9/99 (9.1%)	2/98 (2%)	OR 13.70 (0.76 to 246.52)[4]	20 more per 100 (from 0 fewer to 82 more)	⊙⊙○○ LOW	
Adverse events - incidence of HIV-related disease: cryptococcal meningitis (HIV-positive) (assessed with: number of patients to experience cryptococcal meningitis)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	3/99 (3%)	5/98 (5.1%)	OR 0.58 (0.14 to 2.5)[4]	2 fewer per 100 (from 4 fewer to 7 more)	⊙⊙⊙○ MODERATE	
Adverse events - incidence of HIV-related disease: oesophageal candidiasis (HIV-positive) (assessed with: number of patients to experience oesophageal candidiasis)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	35/99 (35.4%)	23/98 (23.5%)	OR 1.78 (0.96 to 3.32)[4]	12 more per 100 (from 1 fewer to 27 more)	⊙⊙⊙○ MODERATE	
Adverse events - incidence of HIV-related disease: herpes zoster (HIV-positive) (assessed with: number of patients to experience herpes zoster)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	22/99 (22.2%)	19/98 (19.4%)	OR 1.19 (0.6 to 2.37)[4]	3 more per 100 (from 7 fewer to 17 more)	⊙⊙⊙○ MODERATE	
Adverse events - incidence of HIV-related disease: oral or genital herpes simplex (HIV-positive) (assessed with: number of patients to experience oral or genital herpes simplex)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	22/99 (22.2%)	20/98 (20.4%)	OR 1.11 (0.56 to 2.21)[4]	2 more per 100 (from 8 fewer to 16 more)	⊙⊙⊙○ MODERATE	
Adverse events - incidence of HIV-related disease: oral thrush (HIV-positive) (assessed with: number of patients to experience oral thrush)												
1[1]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	31/99 (31.3%)	31/98 (31.6%)	OR 1.43 (0.79 to 2.56)[4]	8 more per 100 (from 5 fewer to 23 more)	⊙⊙⊙○ MODERATE	
Adverse events - incidence of HIV-related disease: gastroenteritis (HIV-positive) (assessed with: number of patients to experience gastroenteritis)												
1[1]	randomised trials	no serious risk of	no serious inconsistency	no serious indirectness	serious[2]	none	34/99 (34.3%)	28/98 (28.6%)	OR 1.32 (0.72 to	6 more per 100	⊙⊙⊙○ MODERATE	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
		bias							2.39)[4]	(from 6 fewer to 20 more)		

[1] *Elliott et al, 2004*
[2] *GRADE rule of thumb: <300 events*
[3] *Wide confidence intervals*
[4] *Odds ratio and 95% confidence interval calculated by reviewer*
[5] *Authors did not provide sufficient data to calculate a confidence interval*
[6] *Mean difference calculated by reviewer*
[7] *Outcome is a substitute for an outcome of interest*

Prednisolone *vs* **antituberculosis chemotherapy alone or plus placebo** in **people without HIV**

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Changes in signs and symptoms – pleural hemithorax (HIV-negative) (follow-up unclear; measured with: Index of reabsorption of pleural hemithorax at 12 months; better indicated by lower values)												
1[10]	randomised trials	serious[2]	no serious inconsistency[4]	serious[6]	very serious[10,11]	none	57	60	-	MD 4% higher (18 lower to 26 higher)[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – pleural thickening on x-ray (HIV-negative) (assessed with: number of patients with residual pleural thickening, as assessed using a chest x-ray)												
2[1,10]	randomised trials	very serious[2,3]	serious[4,5]	serious[6]	serious[10]	none	18/91 (19.8%)	23/96 (24%)	OR 0.60 (0.13 to 2.67)[8,11]	8 fewer per 100 (from 20 fewer to 22 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – pleural thickening on CT scan (HIV-negative) (assessed with: number of patients with residual pleural thickening, as assessed using a CT scan)												
1[1]	randomised trials	very serious[2,3]	serious[4,5]	serious[6]	serious[7]	none	17/34 (50%)	21/36 (58.3%)	OR 0.71 (0.28 to 1.84)[8]	8 fewer per 100 (from 30 fewer to 14 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – pleural thickening on x-ray (HIV-negative) (measured with: pleural thickening at 24 weeks, as assessed using a chest x-ray; better indicated by lower values)												
1[1]	randomised trials	very serious[2,3]	serious[4,5]	serious[6]	no serious imprecision	none	34	36	-	MD 0.4mm lower (1.9 lower to 1.1	⊙○○○ VERY LOW	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
										higher)[9]		
Changes in signs and symptoms – pleural thickening on x-ray (HIV-negative) (measured with: change in pleural thickening from baseline to 24 weeks, as assessed using a chest x-ray; better indicated by lower values)												
1[1]	randomised trials	very serious[2,3]	serious[4,5]	serious[6]	no serious imprecision	none	34	36	-	difference in change in means 0.6mm lower[9]	⊙○○○ VERY LOW	
Changes in signs and symptoms – pleural thickening on CT scan (HIV-negative) (measured with: pleural thickening at 24 weeks, as assessed using a CT scan; better indicated by lower values)												
1[1]	randomised trials	very serious[2,3]	serious[4,5]	serious[6]	no serious imprecision	none	34	36	-	MD 1.3mm lower (3.4 lower to 0.8 higher)[9]	⊙○○○ VERY LOW	
Adverse events (HIV-negative) (assessed with: number of patients to experience an adverse event)												
1[1]	randomised trials	very serious[2,3]	serious[4,5]	serious[6]	serious[7]	none	17/34 (50%)	21/36 (58.3%)	OR 1.47 (0.3 to 7.1)[8]	9 more per 100 (from 29 fewer to 33 more)	⊙○○○ VERY LOW	

[1] *Wyser et al, 1996*
[2] *Method of randomisation and use of allocation concealment is unclear*
[3] *Analysis did not follow the intent-to-treat principle*
[4] *Unclear if the groups received the same care apart from the intervention(s) studied; details provided are limited*
[5] *Although not statistically significant (p = 0.06), more patients receiving placebo (44.4%) had pleuritis and pulmonary tuberculosis than amongst those receiving prednisolone (21.2)*
[6] *Antituberculosis regimens do not use all of or just the 4 standard recommended drugs; of particular note is that Galarza et al (1995) used only a 2-drug antituberculosis regimen*
[7] *GRADE rule of thumb: <300 events*
[8] *Odds ratio and 95% confidence interval calculated by reviewer*
[9] *Mean difference and confidence interval calculated by reviewer*
[10] *Galarza et al, 1995*
[11] *Forest plot (changes in signs and symptoms – pleural thickening):*

TUBERCULOSIS WITH SEVERE BRONCHIAL OBSTRUCTION

Prednisolone *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Changes in signs and symptoms – normalisation of radiological status (prednisolone; children) (assessed with: number of patients whose radiological score normalised during												

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
treatment)												
1[1]	randomised trials	serious[2,3]	serious[4]	no serious indirectness	very serious[6,7]	none	13/15 (86.7%)	9/14 (64.3%)	OR 6.61 (0.57 to 22.9)[8]	28 more per 100 (from 14 fewer to 33 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – improvement in radiological status (prednisolone; children) (assessed with: number of patients whose radiological score improved within 1 month)												
1[1]	randomised trials	serious[2,3]	serious[4]	no serious indirectness	very serious[6,7]	none	7/15 (46.7%)	0/14 (0%)	OR 22.59 (1.29 to 506.48)[8]	-	⊙○○○ VERY LOW	
Changes in signs and symptoms – deterioration in radiological status (prednisolone; children) (assessed with: number of patients whose radiological score deteriorated during treatment)												
1[1]	randomised trials	serious[2,3]	serious[4]	no serious indirectness	serious[6]	none	2/15 (13.3%)	5/14 (35.7%)	OR 0.58 (0.04 to 1.76)[8]	11 fewer per 100 (from 34 fewer to 14 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – bronchoscopy score (prednisolone; children) (measured with: change in bronchoscopy score from baseline to 1 month post-treatment[9]; better indicated by higher values)												
1[1]	randomised trials	serious[2,3]	serious[4]	no serious indirectness	very serious[6,7]	none	15	14	-	MD 6.20 higher (1.83 to 10.57 higher)[10]	⊙○○○ VERY LOW	
Response to treatment – need for multiple bronchoscopies (prednisolone; children) (assessed with: number of patients to require >2 bronchoscopies)												
1[1]	randomised trials	serious[2,3]	serious[4]	serious[5]	serious[6]	none	1/15 (6.7%)	6/14 (42.9%)	OR 0.10 (0.01 to 0.94)[8]	36 fewer per 100 (from 2 fewer to 42 fewer)	⊙○○○ VERY LOW	

[1] *Toppet et al, 1990*
[2] *Unclear if allocation concealment was used*
[3] *'Open' trial, although examination of bronchoscopy and radiographs blinded*
[4] *Unclear if the groups received the same care apart from the intervention(s) studied: those receiving steroids were recommended a sodium-restricted diet, potassium glucoconate supplements and gastric protection by aluminium phosphate, but it is unclear if those on antituberculosis chemotherapy alone received these*
[5] *Outcome is a surrogate for an outcome of interest*
[6] *GRADE rule of thumb: <300 events*
[7] *Wide confidence intervals*
[8] *Odds ratio and 95% confidence interval calculated by reviewer*
[9] *See evidence table for full definition*
[10] *Mean difference and 95% confidence interval calculated by reviewer*

CENTRAL NERVOUS SYSTEM TUBERCULOSIS

Dexamethasone *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus dexamethasone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (dexamethasone) (follow-up 3 months to 5 years; assessed with: number of deaths)												
5[1,2,3,4,5]	randomised trials	very serious[6,7,8]	very serious[9,10,11,12]	serious[13]	no serious imprecision	none	216/486 (44.4%)	232/457 (50.8%)	OR 0.79 (0.61 to 1.02)[14,33]	6 fewer per 100 (from 12 fewer to 0 more)	⊙○○○ VERY LOW	
Mortality (dexamethasone; rifampicin-containing antituberculosis regimens only) (follow-up 3 months to 5 years; assessed with: number of deaths)												
3[2,4,5]	randomised trials	very serious[6,7]	very serious[10,11,12]	serious[13]	no serious imprecision	none	138/330 (41.8%)	144/310 (46.5%)	OR 0.85 (0.62 to 1.16)[14,33]	4 fewer per 100 (from 11 fewer to 4 more	⊙○○○ VERY LOW	
Mortality (dexamethasone; non-randomised) (follow-up unclear; assessed with: number of deaths)												
1[15]	non-randomised trials	very serious[16,17,18]	serious[19,20]	serious[13]	serious[21]	none	39/66 (59.1%)	42/70 (60%)	OR 0.96 (0.49 to 1.91)[14]	1 fewer per 100 (from 18 fewer to 14 more)	⊙○○○ VERY LOW	
Response to treatment - full or partial recovery (dexamethasone) (assessed with: number of patients to achieve a full or partial recovery)												
1[4]	randomised trials	very serious[22,23]	no serious inconsistency	very serious[24,25]	serious[21]	none	15/24 (62.5%)	13/23 (56.5%)	OR 1.28 (0.4 to 4.12)[14]	6 more per 100 (from 22 fewer to 28 more)	⊙○○○ VERY LOW	
Response to treatment - poor outcome (dexamethasone) (assessed with: number of patients to experience a poor outcome (death or survival with major sequelae (persistent vegetative state, blindness, symptomatic hydrocephalus, moderate-to-severe intellectual impairment, severe functional disability (totally dependent), or uncontrolled seizures)))												
1[4]	randomised trials	serious[22]	no serious inconsistency	very serious[24,25]	serious[21]	none	5/24 (20.8%)	8/23 (34.8%)	OR 0.49 (0.13 to 1.82)[14]	14 fewer per 100 (from 28 fewer to 14 more)	⊙○○○ VERY LOW	
Response to treatment - good outcome (dexamethasone) (assessed with: number of patients to experience a good outcome (survival with minor (mild intellectual impairment, mild-to-moderate functional disability (able to enact the activities of daily living with minimal or no assistance)) or no sequelae))												
1[4]	randomised trials	serious[22]	no serious inconsistency	very serious[24,25]	serious[21]	none	15/24 (62.5%)	13/23 (56.5%)	OR 1.28 (0.4 to 4.12)[14]	6 more per 100 (from 22 fewer to 28 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - fever (dexamethasone) (measured with: time to recovery of fever amongst surviving patients; better indicated by lower values)												
1[4]	randomised trials	very serious[22,26]	no serious inconsistency	serious[24]	serious[27]	none	15	14	-	MD 2.7 days higher	⊙○○○ VERY LOW	

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus dexamethasone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Changes in signs and symptoms - fever (dexamethasone) (measured with: time to fever clearance (days from randomisation to observation of a maximal daily temperature of less than 37.5°C for more than five consecutive days); better indicated by lower values)												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[27]	none	274	271	-	difference between the medians 2 days lower	⊙⊙⊙○ MODERATE	
Changes in signs and symptoms - coma (dexamethasone) (measured with: time to coma clearance (median, days from randomization until observation of a Glasgow coma score of 15 for more than two consecutive days); better indicated by lower values)												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[27]	none	274	271	-	difference between the medians 2 days lower	⊙⊙⊙○ MODERATE	
Changes in signs and symptoms - hemiparesis (dexamethasone) (assessed with: number of patients with hemiparesis at baseline to resolve after 9 months of treatment)												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[21]	none	36/48 (75%)	30/37 (81.1%)	OR 0.70 (0.24 to 2)[14]	6 fewer per 100 (from 30 fewer to 8 more)	⊙⊙⊙○ MODERATE	
Changes in signs and symptoms - hemiparesis (dexamethasone) (assessed with: number of patients without hemiparesis at baseline to be experiencing hemiparesis after 9 months of treatment)												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[21]	none	14/226 (6.2%)	11/234 (4.7%)	OR 1.34 (0.59 to 3.01)[14]	1 more per 100 (from 2 fewer to 8 more)	⊙⊙⊙○ MODERATE	
Changes in signs and symptoms - paraparesis (dexamethasone) (assessed with: number of patients with paraparesis at baseline to resolve after 9 months of treatment)												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[21]	none	19/28 (67.9%)	9/11 (81.8%)	OR 0.47 (0.08 to 2.63)[14]	14 fewer per 100 (from 55 fewer to 10 more)	⊙⊙⊙○ MODERATE	
Changes in signs and symptoms - paraparesis (dexamethasone) (assessed with: number of patients without paraparesis at baseline to be experiencing paraparesis after 9 months of treatment)												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[21]	none	11/246 (4.5%)	11/260 (4.2%)	OR 1.06 (0.45 to 2.49)[14]	0 more per 100 (from 2 fewer to 6 more)	⊙⊙⊙○ MODERATE	
Changes in signs and symptoms - tuberculoma (dexamethasone) (assessed with: number of patients to experience a tuberculoma during 9 months of treatment)												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[21]	none	9/246 (3.7%)	5/260 (1.9%)	OR 1.81 (0.6 to 5.46)[14]	2 more per 100 (from 1 fewer to 8 more)	⊙⊙⊙○ MODERATE	

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus dexamethasone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Changes in signs and symptoms - hydrocephalus (dexamethasone) (assessed with: number of patients to experience a hydrocephalus during 9 months of treatment)												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[21]	none	10/246 (4.1%)	7/260 (2.7%)	OR 1.43 (0.54 to 3.81)[14]	1 more per 100 (from 1 fewer to 7 more)	⊙⊙⊙○ MODERATE	
Changes in signs and symptoms - good disability status (dexamethasone) (assessed with: number of patients in a good disability status 5 years after randomisation)												
2[2,5]	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[21]	none	84/306 (27.5%)	65/287 (22.6%)	OR 1.36 (0.72 to 2.58)[14,34]	6 more per 100 (from 5 fewer to 20 more)	⊙⊙○○ LOW	
Changes in signs and symptoms – intermediate or severe disability status (dexamethasone) (assessed with: number of patients in an intermediate or severe disability status 5 years after randomisation)												
2[2,5]	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	serious[21]	none	68/306 (22.2%)	58/287 (20.2%)	OR 1.12 (0.75 to 1.66)[14,36]	2 more per 100 (from 4 fewer to 9 more)	⊙⊙○○ LOW	
Changes in signs and symptoms - cognitive status (dexamethasone) (measured with: time to improvement in mini-mental score amongst surviving patients[28]; better indicated by lower values)												
1[4]	randomised trials	serious[22]	no serious inconsistency	serious[24]	serious[27]	none	15	14	-	MD 3.4 days higher	⊙○○○ VERY LOW	
Changes in signs and symptoms - neurological abnormalities during treatment (dexamethasone) (assessed with: number of patients to develop neurologic abnormalities (fundus, hemiparesis or hydrocephalus) during treatment)												
1[3]	randomised trials	serious[8]	no serious inconsistency	serious[13]	serious[21]	none	8/145 (5.5%)	15/135 (11.1%)	OR 0.47 (0.19 to 1.14)[14]	6 fewer per 100 (from 9 fewer to 1 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - residual neurological abnormalities (dexamethasone) (assessed with: number of patients to with permanent residual neurologic abnormalities (fundus, hemiparesis or hydrocephalus))												
1[3]	randomised trials	serious[8]	no serious inconsistency	serious[13]	serious[21]	none	14/145 (9.7%)	27/135 (20%)	OR 0.43 (0.21 to 0.86)[14]	10 fewer per 100 (from 2 fewer to 15 fewer)	⊙○○○ VERY LOW	
Changes in signs and symptoms - headache (dexamethasone) (measured with: time to recovery of headache amongst surviving patients; better indicated by lower values)												
1[4]	randomised trials	serious[22]	no serious inconsistency	serious[24]	serious[27]	none	15	14	-	MD 7.4 days higher	⊙○○○ VERY LOW	
Changes in signs and symptoms - activity of daily living (dexamethasone; children) (measured with: time to improvement in Barthel score amongst surviving patients[28]; better indicated by lower values)												
1[4]	randomised trials	serious[22]	no serious inconsistency	serious[24]	serious[27]	none	15	14	-	MD 5.3 days	⊙○○○ VERY	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus dexamethasone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
										higher	LOW	
Relapse (dexamethasone) (assessed with: number of patients to experience relapse[28])												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[21,27]	none	41/274 (15%)	48/271 (17.7%)	OR 0.82 (0.52 to 1.29)[14]	3 fewer per 100 (from 8 fewer to 4 more)	⊙⊙⊙○ MODERATE	
Adverse events - ocular (dexamethasone; non-randomised) (follow-up unclear; assessed with: number of patients with ocular complications)												
1[15]	non-randomised trials	very serious[16,17,18]	serious[19,20]	serious[13]	serious[21]	none	2/66 (3%)	7/70 (10%)	OR 0.28 (0.06 to 1.41)[14]	7 fewer per 100 (from 9 fewer to 4 more)	⊙○○○ VERY LOW	
Adverse events - severe (dexamethasone) (assessed with: number of patients to experience a severe event (any event causing or threatening to cause prolonged hospital stay, disability, or death))												
1[2]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[21,27]	none	26/274 (9.5%)	45/271 (16.6%)	OR 0.53 (0.31 to 0.88)[14]	7 fewer per 100 (from 2 fewer to 11 fewer)	⊙⊙⊙○ MODERATE	
Adverse events - hepatitis (dexamethasone; HIV-negative) (assessed with: number of patients to experience clinical or subclinical hepatitis[29])												
1[5]	randomised trials	very serious[30,31]	no serious inconsistency	no serious indirectness	serious[21]	none	5/32 (15.6%)	4/16 (25%)	OR 0.56 (0.13 to 2.44)[14]	9 fewer per 100 (from 21 fewer to 20 more)	⊙○○○ VERY LOW	
Adverse events - gastrointestinal bleeding (dexamethasone; HIV-negative) (assessed with: of patients to experience gastrointestinal bleeding[29])												
1[5]	randomised trials	very serious[30,31]	no serious inconsistency	no serious indirectness	very serious[21,32]	none	4/32 (12.5%)	0/16 (0%)	OR 5.21 (0.26 to 103)[14]	-	⊙○○○ VERY LOW	
Adverse events - paradoxical tuberculoma (dexamethasone; HIV-negative) (assessed with: number of patients to experience paradoxical tuberculoma[29])												
1[5]	randomised trials	very serious[30,31]	no serious inconsistency	no serious indirectness	serious[21]	none	2/32 (6.3%)	2/16 (12.5%)	OR 0.47 (0.06 to 3.66)[14]	6 fewer per 100 (from 12 fewer to 22 more)	⊙○○○ VERY LOW	

[1] *O'Toole et al, 1969*
[2] *Thwaites et al, 2004/7 / Török et al, 2011*
[3] *Girgis et al, 1991*
[4] *Kumarvelu et al, 1994*
[5] *Malhotra et al, 2009*
[6] *Unclear if analysis followed the intent-to-treat principle*
[7] *Malhotra et al, 2009: unclear if alloctation concealment used, and blinding not used*
[8] *Girgis et al, 1991: use of allocation concealment and blinding unclear*
[9] *O'Toole et al, 1969: unclear if groups were comparable at baseline, or if they were comparable for treatment completion and availability of outcome data*
[10] *Kumarvelu et al, 1994: follow-up only 3 months after treatment initiation*

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus dexamethasone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance

[11] *Follow-up varied widely between groups*
[12] *Estimates of effect very widely across the studies*
[13] *O'Toole et al (1969), Girgis et al (1983 and 1991), Kumarvelu et al (1994): antituberculosis regimens do not use all of or just the 4 standard recommended drugs; of particular not is the lack of rifampicin in O'Toole et al (1969) and Girgis et al (1983 and 1991)*
[14] *Odds ratio and 95% confidence interval calculated by reviewer*
[15] *Girgis et al, 1983*
[16] *Non-randomised; patients were alternately assigned to receive antituberculosis chemotherapy plus dexamethasone or antituberculosis chemotherapy alone*
[17] *No allocation concealment*
[18] *Use of blinding unclear*
[19] *Authors state that groups were comparable with respect to age, sex and disease severity on admission to hospital; however, although not statistically significant, more patients in the dexamethasone group (32/70) were comatose on admission than in the antituberculosis chemotherapy alone group (41/66) - that is, the condition of those in the dexamethasone group could be considered to be more severe*
[20] *Unclear if groups received the same care except for the intervention(s) studied; limited information available*
[21] *GRADE rule of thumb: <300 events*
[22] *Kumarvelu et al, 1994: use of allocation concealment and blinding is unclear*
[23] *Authors do not provide a definition*
[24] *Kumarvelu et al, 1994: antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[25] *Outcome is a surrogate for an outcome of interest*
[26] *Some data was only available for patients with either 'severe' or 'mild-to-moderate' disease on admission who survived; since the authors do not provide the number of patients with either 'severe' or 'mild-to-moderate' disease on admission who were randomised to each intervention, this data could not be analysed in accordance with the intent-to-treat principle*
[27] *Insufficient data to calculate confidence intervals*
[28] *For full definition, see evidence table*
[29] *For full definition, see evidence tables*
[30] *Malhotra et al, 2009: use of allocation concealment unclear*
[31] *Malhotra et al, 2009: unblinded*
[32] *Wide confidence intervals*
[33] *Forest plot (mortality):*

[34] *Forest plot (changes in signs and symptoms - good disability status):*

[35] *Forest plot (changes in signs and symptoms - intermediate or severe disability status):*

Prednisolone *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (prednisolone) (follow-up 3 to 18 months; assessed with: number of deaths)												
2[1,2]	randomised trials	very serious[3,4,5]	very serious[6,7]	serious[8]	very serious[9,10]	none	9/99 (9.1%)	15/101 (14.9%)	OR 0.81 (0.08 to 8.31)[11,19]	2 fewer per 100 (from 13 fewer to 44 more)	⊙○○○ VERY LOW	
Response to treatment - need for additional intervention (prednisolone; HIV-negative) (follow-up 18 months; assessed with: number of patients to require ventricular shunting, as indicated by persistent high cerebrospinal fluid pressure after 4 weeks of repeated lumbar puncture)												
1[2]	randomised trials	very serious[12]	very serious[6]	very serious[8,13]	very serious[9,10]	none	5/29 (17.2%)	4/30 (13.3%)	OR 1.35 (0.33 to 5.64)[11]	4 more per 100 (from 9 fewer to 33 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - disability (prednisolone; children) (assessed with: number of patients to be disabled (severely or mildly) at 6 months)												
1[1]	randomised trials	very serious[3,5,14]	serious[15]	serious[16]	serious[9]	none	54/70 (77.1%)	49/71 (69%)	OR 1.52 (0.71 to 3.21)[11]	8 more per 100 (from 8 fewer to 19 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - neurological abnormalities during treatment (prednisolone; HIV-negative) (assessed with: number of patients to develop neurological abnormalities during treatment)												
1[2]	randomised trials	very serious[12]	serious[6]	serious[8]	very serious[9,10]	none	2/29 (6.9%)	4/30 (13.3%)	OR 0.48 (0.08 to 2.86)[11]	6 fewer per 100 (from 12 fewer to 17 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - hearing (prednisolone; children) (assessed with: number of patients with deterioration in their hearing (decreased hearing, though not deaf) at 6 months)												
1[1]	randomised trials	very serious[3,5,14]	serious[15]	serious[16]	serious[9]	none	3/70 (4.3%)	6/71 (8.5%)	OR 0.49 (0.12 to 2.02)[11]	4 fewer per 100 (from 7 fewer to 7 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - severe disability (prednisolone; children) (assessed with: number of patients to be severely disabled at 6 months)												
1[1]	randomised trials	very serious[3,5,14]	serious[15]	serious[16]	serious[9]	none	14/70 (20%)	19/71 (26.8%)	OR 0.68 (0.31 to 1.5)[11]	7 fewer per 100 (from 17 fewer to 9 more)	⊙○○○ VERY LOW	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Changes in signs and symptoms - tuberculoma (prednisolone; children) (assessed with: number of patients to develop tuberculomas in the first month of treatment)												
1[1]	randomised trials	very serious[3,5,14]	serious[15]	serious[16]	serious[9]	none	2/70 (2.9%)	9/71 (12.7%)	OR 0.20 (0.04 to 0.97)[11]	10 fewer per 100 (from 0 fewer to 12 fewer)	⊙○○○ VERY LOW	
Changes in signs and symptoms - IQ (prednisolone; children) (assessed with: number of patients to have an IQ of less than 75 at 6 months)												
1[1]	randomised trials	very serious[3,5,14]	serious[15]	serious[16]	serious[9]	none	31/70 (44.3%)	36/71 (50.7%)	OR 0.77 (0.4 to 1.5)[11]	7 fewer per 100 (from 22 fewer to 10 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - motor function (prednisolone; children) (assessed with: number of patients to be experience hemiplegia or quadriplegia at 6 months)												
1[1]	randomised trials	very serious[3,5,14]	serious[15]	serious[16]	serious[9]	none	24/70 (34.3%)	24/71 (33.8%)	OR 1.02 (0.51 to 2.05)[11]	0 more per 100 (from 13 fewer to 17 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - vision (prednisolone; children) (assessed with: number of patients with visual deterioration (decreased vision or blindness) at 6 months)												
1[1]	randomised trials	very serious[3,5,14]	serious[15]	serious[16]	serious[9]	none	9/70 (12.9%)	7/71 (9.9%)	OR 1.35 (0.47 to 3.85)[11]	3 more per 100 (from 5 fewer to 20 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - vision (prednisolone; children) (assessed with: number of patients to be blind at 6 months)												
1[1]	randomised trials	very serious[3,5,14]	serious[15]	serious[16]	serious[9]	none	3/70 (4.3%)	3/71 (4.2%)	OR 1.01 (0.2 to 5.21)[11]	0 more per 100 (from 3 fewer to 14 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - hearing (prednisolone; children) (assessed with: number of patients to be deaf at 6 months)												
1[1]	randomised trials	very serious[3,5,14]	serious[15]	serious[16]	very serious[9,17]	none	0/70 (0%)	0/71 (0%)	1.01 (0.02 to 51.82)[11]	-	⊙○○○ VERY LOW	
Change in signs and symptoms - neurological abnormalities after treatment (prednisolone; HIV-negative) (follow-up 18 months; assessed with: number of patients to develop neurological abnormalities after treatment)												
1[2]	randomised trials	very serious[12]	serious[6]	serious[8]	very serious[9,10]	none	4/29 (13.8%)	2/30 (6.7%)	OR 2.24 (0.38 to 13.3)[11]	7 more per 100 (from 4 fewer to 42 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - headache (prednisolone; HIV-negative) (measured with: time until disappearance of headache; better indicated by lower values)												
1[2]	randomised trials	very serious[12]	serious[6]	serious[8]	serious[18]	none	29	30	-	MD 2.6 days higher[11]	⊙○○○ VERY LOW	
Change in signs and symptoms - fever (prednisolone; HIV-negative) (measured with: time until normalisation of body temperature; better indicated by lower values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
1[2]	randomised trials	very serious[12]	serious[6]	serious[8]	serious[18]	none	29	30	-	MD 3.7 days lower[11]	⊙○○○ VERY LOW	
Recurrence (prednisolone; HIV-negative) (follow-up 18 months; assessed with: number of patients to experience recurrence of meningitis during follow-up)												
1[2]	randomised trials	very serious[12]	serious[6]	very serious[8,13]	very serious[9,10]	none	0/29 (0%)	0/30 (0%)	OR 1.03 (0.02 to 53.83)[11]	-	⊙○○○ VERY LOW	
Adverse events - hyperglycaemia (prednisolone; HIV-negative) (follow-up 18 months; assessed with: number of patients to experience hyperglycaemia)												
1[2]	randomised trials	very serious[12]	serious[6]	serious[8]	very serious[9,10]	none	0/29 (0%)	0/30 (0%)	OR 1.03 (0.02 to 53.83)[11]	-	⊙○○○ VERY LOW	
Adverse events - gastrointestinal bleeding (prednisolone; HIV-negative) (follow-up 18 months; assessed with: number of patients to experience gastrointestinal bleeding)												
1[2]	randomised trials	very serious[12]	serious[6]	serious[8]	very serious[9,10]	none	0/29 (0%)	0/30 (0%)	OR 1.03 (0.02 to 53.83)[11]	-	⊙○○○ VERY LOW	

[1] *Schoeman et al, 1997*
[2] *Chotmongkol et al, 1996*
[3] *Method of randomisation and use of allocation concealment is unclear*
[4] *Schoeman et al, 1997: blinded = clinical psychologist assessing intelligence, clinician testing hearing, ophthalmologist testing vision, and physical therapist testing motor function; unclear if patients or other health professionals were blinded*
[5] *Unclear if analysis followed the intent-to-treat principle*
[6] *Chotmongkol et al, 1996: groups not comparable at baseline - clinical presentations and staging were similar in the intervention and comparator groups at randomisation; however, although not statistically significant, more patients in the prednisolone group (17%) had motor weakness than in the placebo group (3%), and more patients in the prednisolone group (17%) had motor weakness than in the placebo group (10%); additionally, there were more patients with severe (stage 3) disease and fewer patients with less severe (stage 1) disease in the prednisolone group than in the placebo group, although again this was not statistically significant*
[7] *Follow-up varied widely between groups*
[8] *Chotmongkol et al, 1996: antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[9] *GRADE rule of thumb: <300 events*
[10] *Wide confidence intervals*
[11] *Odds ratio and 95% confidence interval calculated by reviewer*
[12] *Chotmongkol et al, 1996: method of randomisation and use of allocation concealment is unclear*
[13] *Outcome is a surrogate for an outcome of interest*
[14] *Blinded = clinical psychologist assessing intelligence, clinician testing hearing, ophthalmologist testing vision, and physical therapist testing motor function; unclear if patients or other health professionals were blinded*
[15] *Follow-up only 3 months after treatment initiation*
[16] *Antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[17] *Wide confidence interval*
[18] *Authors did not provide sufficient data to calculate confidence interval*
[19] *Forest plot (mortality):*

Methylprednisolone *vs* **antituberculosis chemotherapy alone or plus placebo**

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus methylprednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (methylprednisolone; HIV-negative) (assessed with: number of deaths after 6 months of treatment)												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	no serious indirectness	serious[4]	none	9/33 (27.3%)	7/16 (43.8%)	OR 0.48 (0.14 to 1.68)[5]	17 fewer per 100 (from 34 fewer to 13 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – severe disability (methylprednisolone; HIV-negative) (assessed with: number of patients to experience severe disability after 6 months of treatment[6])												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	no serious indirectness	serious[4]	none	6/33 (18.2%)	3/16 (18.8%)	OR 0.96 (0.21 to 4.47)[5]	1 fewer per 100 (from 14 fewer to 32 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – intermediate disability (methylprednisolone; HIV-negative) (assessed with: number of patients to experience intermediate disability after 6 months of treatment[6])												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	no serious indirectness	serious[4]	none	0/33 (0%)	2/16 (12.5%)	OR 0.09 (0 to 1.92)[5]	11 fewer per 100 (from 12 fewer to 9 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – good disability status (methylprednisolone; HIV-negative) (assessed with: number of patients to achieve a good disability status after 6 months of treatment[6])												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	no serious indirectness	serious[4]	none	15/33 (45.5%)	4/16 (25%)	OR 2.50 (0.67 to 9.39)[5]	20 more per 100 (from 7 fewer to 51 more)	⊙○○○ VERY LOW	
Adverse events - hepatitis (methylprednisolone; HIV-negative) (assessed with: number of patients to experience clinical or subclinical hepatitis[6])												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	no serious indirectness	serious[4]	none	7/33 (21.2%)	4/16 (25%)	OR 0.81 (0.2 to 3.3)[5]	4 fewer per 100 (from 19 fewer to 27 more)	⊙○○○ VERY LOW	
Adverse events - gastrointestinal bleeding (methylprednisolone; HIV-negative) (assessed with: number of patients to experience gastrointestinal bleeding[6])												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	no serious indirectness	very serious[4,7]	none	2/33 (6.1%)	1/16 (6.3%)	OR 0.97 (0.08 to 11.54)[5]	0 fewer per 100 (from 6 fewer to 37 more)	⊙○○○ VERY LOW	
Adverse events - paradoxical tuberculoma (methylprednisolone; HIV-negative) (assessed with: number of patients to experience paradoxical tuberculoma[6])												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	no serious indirectness	serious[4]	none	2/33 (6.1%)	3/16 (18.8%)	OR 0.14 (0.01 to 1.42)[5]	16 fewer per 100 (from 19 fewer to 6 more)	⊙○○○ VERY LOW	

[1] *Malhotra et al, 2009*

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus methylprednisolone	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance

[2] *Use of allocation concealment unclear*
[3] *Unblinded*
[4] *GRADE rule of thumb: <300 events*
[5] *Odds ratio and 95% confidence interval calculated by reviewer*
[6] *For full definition, see evidence tables*
[7] *Wide confidence interval*

Any corticosteroid *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (follow-up 3 months to 5 years; assessed with: number of deaths)												
7[1,2,3,4,5,6,7]	randomised trials	very serious[8,9,10,11]	very serious[12,13,14,15,16]	serious[17]	no serious imprecision	none	234/618 (37.9%)	253/574 (44.1%)	OR 0.75 (0.56 to 0.99)[18,20]	7 fewer per 100 (from 13 fewer to 0 fewer)	⊙○○○ VERY LOW	
Mortality (follow-up <1 year) (follow-up 3 to 10 months; assessed with: number of deaths)												
3[1,6,7]	randomised trials	very serious[8,9,10,11]	very serious[14,15,16]	serious[17]	serious[19]	none	22/127 (17.3%)	35/126 (27.8%)	OR 0.52 (0.26 to 1.02)[18,21]	11 fewer per 100 (from 19 fewer to 0 more)	⊙○○○ VERY LOW	
Mortality (follow-up >1 year) (follow-up 18 months to 5 years; assessed with: number of deaths)												
3[2,4,5]	randomised trials	very serious[8,10]	very serious[12,15,16]	serious[17]	no serious imprecision	none	198/448 (44.2%)	209/436 (47.9%)	OR 0.85 (0.6 to 1.21)[18,21]	4 fewer per 100 (from 12 fewer to 5 more)	⊙○○○ VERY LOW	
Mortality (rifampicin-containing antituberculosis regimens only) (follow-up 3 months to 5 years; assessed with: number of deaths)												
5[1,2,4,6,7]	randomised trials	very serious[8,9,10,11]	very serious[12,14,15,16]	serious[17]	no serious imprecision	none	148/430 (34.4%)	165/427 (38.6%)	OR 0.76 (0.45 to 1.28)[18,22]	6 fewer per 100 (from 17 fewer to 6 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - neurological abnormalities (follow-up 18 to 24 months; assessed with: number of patients to develop neurological abnormalities during treatment)												
2[2,5]	randomised trials	serious[8]	serious[12]	serious[17]	serious[19]	none	10/174 (5.7%)	19/165 (11.5%)	OR 0.47 (0.21 to 1.04)[18,23]	6 fewer per 100 (from 9 fewer to 0 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - neurological abnormalities (rifampicin-containing antituberculosis regimens only) (follow-up 18 months; assessed with: number of patients to develop neurological abnormalities during treatment)												
1[2]	randomised trials	serious[8]	serious[12]	serious[17]	serious[19]	none	8/145 (5.5%)	15/135 (11.1%)	OR 0.47 (0.19 to 1.14)[14]	6 fewer per 100 (from 9 fewer to 1 more)	⊙○○○ VERY LOW	

[1] *Schoeman et al, 1997*
[2] *Chotmongkol et al, 1996*
[3] *O'Toole et al, 1969*
[4] *Thwaites et al, 2004/7 / Török et al, 2011*
[5] *Girgis et al, 1991*
[6] *Kumarvelu et al, 1994*
[7] *Malhotra et al, 2009*

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance

[8] *Schoeman et al (1997) and Chotmongkol et al (1996): method of randomisation and use of allocation concealment is unclear*
[9] *Schoeman et al, 1997: blinded = clinical psychologist assessing intelligence, clinician testing hearing, ophthalmologist testing vision, and physical therapist testing motor function; unclear if patients or other health professionals were blinded*
[10] *Unclear if analysis followed the intent-to-treat principle*
[11] *Malhotra et al, 2009: blinding not used*
[12] *Chotmongkol et al, 1996: groups not comparable at baseline - clinical presentations and staging were similar in the intervention and comparator groups at randomisation; however, although not statistically significant, more patients in the prednisolone group (17%) had motor weakness than in the placebo group (3%), and more patients in the prednisolone group (17%) had motor weakness than in the placebo group (10%); additionally, there were more patients with severe (stage 3) disease and fewer patients with less severe (stage 1) disease in the prednisolone group than in the placebo group, although again this was not statistically significant*
[13] *O'Toole et al, 1969: unclear if groups were comparable at baseline, or if they were comparable for treatment completion and availability of outcome data*
[14] *Kumarvelu et al, 1994: follow-up only 3 months after treatment initiation*
[15] *Follow-up varied widely between groups*
[16] *Estimates of effect very widely across the studies*
[17] *Chotmongkol et al (1996), O'Toole et al (1969), Girgis et al (1991), Kumarvelu et al (1994): antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[18] *Odds ratio and 95% confidence interval calculated by reviewer*
[19] *GRADE rule of thumb: <300 events*
[20] *Forest plot (mortality):*

[21] *Forest plot (mortality; follow-up subgroups):*

[22] *Forest plot (mortality; rifampicin-containing antituberculosis regimens only):*

[23] *Forest plot (change in signs and symptoms - neurological abnormalities):*

Any corticosteroid *vs* **antituberculosis chemotherapy alone or plus placebo** in **people without HIV**

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (HIV-negative) (follow-up 18 months; assessed with: number of deaths)												
2[1,2]	randomised trials	very serious[3,4,5]	serious[6]	serious[7]	serious[8]	none	22/94 (23.4%)	15/62 (24.2%)	OR 1.04 (0.2 to 5.53)[9,14]	1 more per 100 (from 18 fewer to 40 more)	⊙○○○ VERY LOW	
Response to treatment - need for additional intervention (HIV-negative) (follow-up 18 months; assessed with: number of deaths)												
1[1]	randomised trials	very serious[3]	very serious[6]	very serious[7,10]	very serious[8,11]	none	5/29 (17.2%)	4/30 (13.3%)	OR 1.35 (0.33 to 5.64)[9]	4 more per 100 (from 9 fewer to 33 more)	⊙○○○ VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Change in signs and symptoms - neurological abnormalities during treatment (HIV-negative) (assessed with: number of patients to develop neurological abnormalities during treatment)												
1[1]	randomised trials	very serious[3]	serious[6]	serious[7]	very serious[8,11]	none	2/29 (6.9%)	4/30 (13.3%)	OR 0.48 (0.08 to 2.86)[9]	6 fewer per 100 (from 12 fewer to 17 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - neurological abnormalities after treatment (HIV-negative) (follow-up 18 months; assessed with: number of patients to develop neurological abnormalities after treatment)												
1[1]	randomised trials	very serious[3]	serious[6]	serious[7]	very serious[8,11]	none	4/29 (13.8%)	2/30 (6.7%)	OR 2.24 (0.38 to 13.3)[9]	7 more per 100 (from 4 fewer to 42 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - headache (HIV-negative) (measured with: time until disappearance of headache; better indicated by lower values)												
1[1]	randomised trials	very serious[3]	serious[6]	serious[7]	serious[12]	none	29	30	-	MD 2.6 days higher[9]	⊙○○○ VERY LOW	
Change in signs and symptoms - severe disability (HIV-negative) (assessed with: number of patients to experience severe disability after 6 months of treatment[13])												
1[2]	randomised trials	very serious[4,5]	no serious inconsistency	no serious indirectness	serious[8]	none	11/65 (16.9%)	5/32 (15.6%)	OR 1.10 (0.35 to 3.49)[9]	1 more per 100 (from 10 fewer to 24 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - intermediate disability (HIV-negative) (assessed with: number of patients to experience intermediate disability after 6 months of treatment[13])												
1[2]	randomised trials	very serious[4,5]	no serious inconsistency	no serious indirectness	serious[8]	none	3/65 (4.6%)	4/32 (12.5%)	OR 0.34 (0.07 to 1.62)[9]	8 fewer per 100 (from 12 fewer to 6 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - no disability (HIV-negative) (assessed with: number of patients with a good outcome after 6 months of treatment[13])												
1[2]	randomised trials	very serious[4,5]	no serious inconsistency	no serious indirectness	very serious[8,11]	none	30/65 (46.2%)	8/32 (25%)	OR 2.57 (1.01 to 6.56)[9]	21 more per 100 (from 0 more to 44 more)	⊙○○○ VERY LOW	
Change in signs and symptoms - fever (HIV-negative) (measured with: time until normalisation of body temperature; better indicated by lower values)												
1[1]	randomised trials	very serious[3]	serious[6]	serious[7]	serious[12]	none	29	30	-	MD 3.7 days lower[9]	⊙○○○ VERY LOW	
Recurrence (HIV-negative) (follow-up 18 months; assessed with: number of patients to experience recurrence of meningitis during follow-up)												
1[1]	randomised trials	very serious[3]	serious[6]	very serious[7,10]	very serious[8,11]	none	0/29 (0%)	0/30 (0%)	OR 1.03 (0.02 to 53.83)[9]	-	⊙○○○ VERY LOW	
Adverse events - hyperglycaemia (HIV-negative) (follow-up 18 months; assessed with: number of patients to experience gastrointestinal bleeding)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
1[1]	randomised trials	very serious[3]	serious[6]	serious[7]	very serious[8,11]	none	0/29 (0%)	0/30 (0%)	OR 1.03 (0.02 to 53.83)[9]	-	⊙○○○ VERY LOW	
Adverse events - hepatitis (HIV-negative) (assessed with: number of patients to experience clinical or subclinical hepatitis[13])												
1[2]	randomised trials	very serious[4,5]	no serious inconsistency	no serious indirectness	serious[8]	none	12/65 (18.5%)	8/32 (25%)	OR 0.68 (0.25 to 1.88)[9]	7 fewer per 100 (from 17 fewer to 14 more)	⊙○○○ VERY LOW	
Adverse events - gastrointestinal bleeding (HIV-negative) (assessed with: number of patients to experience gastrointestinal bleeding)												
2[1,2]	randomised trials	very serious[3,4,5]	serious[6]	serious[7]	very serious[8,11]	none	6/94 (6.4%)	1/62 (1.6%)	OR 3.15 (0.36 to 27.37)[9,15]	3 more per 100 (from 1 fewer to 29 more)	⊙○○○ VERY LOW	
Adverse events - paradoxical tuberculoma (HIV-negative) (assessed with: number of patients to experience paradoxical tuberculoma[13])												
1[2]	randomised trials	very serious[4,5]	no serious inconsistency	no serious indirectness	serious[8]	none	3/65 (4.6%)	5/32 (15.6%)	OR 0.26 (0.06 to 1.17)[9]	11 fewer per 100 (from 15 fewer to 2 more)	⊙○○○ VERY LOW	

[1] *Chotmongkol et al, 1996*
[2] *Malhotra et al, 2009*
[3] *Chotmongkol et al, 1996: method of randomisation and use of allocation concealment is unclear*
[4] *Malhotra et al, 2009: use of allocation concealment unclear*
[5] *Malhotra et al, 2009: unblinded*
[6] *Chotmongkol et al, 1996: groups not comparable at baseline - clinical presentations and staging were similar in the intervention and comparator groups at randomisation; however, although not statistically significant, more patients in the prednisolone group (17%) had motor weakness than in the placebo group (3%), and more patients in the prednisolone group (17%) had motor weakness than in the placebo group (10%); additionally, there were more patients with severe (stage 3) disease and fewer patients with less severe (stage 1) disease in the prednisolone group than in the placebo group, although again this was not statistically significant*
[7] *Chotmongkol et al, 1996: antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[8] *GRADE rule of thumb: <300 events*
[9] *Odds ratio and 95% confidence interval calculated by reviewer*
[10] *Outcome is a surrogate for an outcome of interest*
[11] *Wide confidence intervals*
[12] *Authors did not provide sufficient data to calculate confidence interval*
[13] *For full definition, see evidence tables*
[14] *Forest plot (mortality):*

[15] *Forest plot (adverse events - gastrointestinal bleeding):*

Corticosteroids *vs* **antituberculosis chemotherapy alone or plus placebo** in **children**

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or with placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (prednisolone; children) (follow-up 3 months to 6 months; assessed with: number of deaths)												
1[1]	randomised trials	very serious[4,5,6]	no serious inconsistency	serious[8]	serious[9]	none	4/70 (5.7%)	13/71 (18.3%)	OR 0.27 (0.08 to 0.88)[10]	-	⊙○○○ VERY LOW	
Changes in signs and symptoms - disability (prednisolone; children) (assessed with: number of patients to be disabled (severely or mildly) at 6 months)												
1[1]	randomised trials	very serious[4,5,6]	serious[11]	serious[12]	serious[9]	none	54/70 (77.1%)	49/71 (69%)	OR 1.52 (0.71 to 3.21)[10]	8 more per 100 (from 8 fewer to 19 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - severe disability (prednisolone; children) (assessed with: number of patients to be severely disabled at 6 months)												
1[1]	randomised trials	very serious[4,5,6]	serious[11]	serious[12]	serious[9]	none	14/70 (20%)	19/71 (26.8%)	OR 0.68 (0.31 to 1.5)[10]	7 fewer per 100 (from 17 fewer to 9 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - tuberculoma (prednisolone; children) (assessed with: number of patients to develop tuberculomas in the first month of treatment)												
1[1]	randomised trials	very serious[4,5,6]	serious[11]	serious[12]	serious[9]	none	2/70 (2.9%)	9/71 (12.7%)	OR 0.20 (0.04 to 0.97)[10]	10 fewer per 100 (from 0 fewer to 12 fewer)	⊙○○○ VERY LOW	
Changes in signs and symptoms - IQ (prednisolone; children) (assessed with: number of patients to have an IQ of less than 75 at 6 months)												
1[1]	randomised trials	very serious[4,5,6]	serious[11]	serious[12]	serious[9]	none	31/70 (44.3%)	36/71 (50.7%)	OR 0.77 (0.4 to 1.5)[10]	7 fewer per 100 (from 22 fewer to 10 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - motor function (prednisolone; children) (assessed with: number of patients to be experience hemiplegia or quadriplegia at 6 months)												
1[1]	randomised trials	very serious[4,5,6]	serious[11]	serious[12]	serious[9]	none	24/70 (34.3%)	24/71 (33.8%)	OR 1.02 (0.51 to 2.05)[10]	0 more per 100 (from 13 fewer to 17 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - vision (prednisolone; children) (assessed with: number of patients with visual deterioration (decreased vision or blindness) at 6 months)												
1[1]	randomised trials	very serious[4,5,6]	serious[11]	serious[12]	serious[9]	none	9/70 (12.9%)	7/71 (9.9%)	OR 1.35 (0.47 to 3.85)[10]	3 more per 100 (from 5 fewer to 20 more)	⊙○○○ VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or with placebo	Relative (95% CI)	Absolute (95% CI)		
Changes in signs and symptoms - vision (prednisolone; children) (assessed with: number of patients to be blind at 6 months)												
1[1]	randomised trials	very serious[4,5,6]	serious[11]	serious[12]	serious[9]	none	3/70 (4.3%)	3/71 (4.2%)	OR 1.01 (0.2 to 5.21)[10]	0 more per 100 (from 3 fewer to 14 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - hearing (prednisolone; children) (assessed with: number of patients with deterioration in their hearing (decreased hearing, though not deaf) at 6 months)												
1[1]	randomised trials	very serious[4,5,6]	serious[11]	serious[12]	serious[9]	none	3/70 (4.3%)	6/71 (8.5%)	OR 0.49 (0.12 to 2.02)[10]	4 fewer per 100 (from 7 fewer to 7 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - hearing (prednisolone; children) (assessed with: number of patients to be deaf at 6 months)												
1[1]	randomised trials	very serious[4,5,6]	serious[11]	serious[12]	very serious[9,13]	none	0/70 (0%)	0/71 (0%)	OR 1.01 (0.02 to 51.82)[10]	-	⊙○○○ VERY LOW	

[1] *Schoeman et al, 1997*
[4] *Unclear if analysis followed the intent-to-treat principle*
[5] *Schoeman et al, 1997: method of randomisation and use of allocation concealment is unclear*
[6] *Schoeman et al, 1997: blinded = clinical psychologist assessing intelligence, clinician testing hearing, ophthalmologist testing vision, and physical therapist testing motor function; unclear if patients or other health professionals were blinded*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence interval calculated by reviewer*
[11] *Follow-up only 3 months after treatment initiation*
[12] *Antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[13] *Wide confidence interval*
[14] *Insufficient data to calculate confidence intervals*
[15] *Authors do not provide a definition*
[16] *Outcome is a surrogate for an outcome of interest*
[17] *For full definition, see evidence table*

Corticosteroids *vs* **antituberculosis chemotherapy alone or plus placebo** in **stage 1 CNS tuberculosis**

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (dexamethasone; stage 1) (measured with: survival rate at 5 years amongst those classified as stage 1 on admission; better indicated by higher values)												
1[1]	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	90	86	-	difference in survival rates 0.14 higher (0.01 lower to 0.29 higher)	⊙⊙⊙○ MODERATE	
Mortality (stage 1) (follow-up 10 to 18 months; assessed with: number of deaths amongst those classified as stage 1 on admission)												
2[3,4]	randomised trials	very serious[5,6,7]	serious[8]	serious[9]	very serious[10,11]	none	0/17 (0%)	1/13 (7.7%)	OR 0.15 (0.01 to 4.18)[12,15]	6 fewer per 100 (from 8 fewer to 18 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - severe disability status (stage 1) (follow-up 10 months; assessed with: number of deaths amongst those with stage 1 CNS TB on admission)												
1[4]	randomised trials	serious[13]	no serious inconsistency	serious[14]	very serious[10,11]	none	2/14 (14.3%)	1/7 (14.3%)	OR 1.00 (0.07 to 13.37)[12]	0 fewer per 100 (from 13 fewer to 55 more)	⊙○○○ VERY LOW	

[1] *Thwaites et al, 2004/7 / Török et al, 2011*
[2] *Analysis does not follow intent-to-treat principle*
[3] *Chotmongkol et al, 1996*
[4] *Malhotra et al, 2009*
[5] *Chotmongkol et al, 1996: method of randomisation and use of allocation concealment is unclear*
[6] *Malhotra et al, 2009: use of allocation concealment unclear*
[7] *Malhotra et al, 2009: unblinded*
[8] *Chotmongkol et al, 1996: groups not comparable at baseline - clinical presentations and staging were similar in the intervention and comparator groups at randomisation; however, although not statistically significant, more patients in the prednisolone group (17%) had motor weakness than in the placebo group (3%), and more patients in the prednisolone group (17%) had motor weakness than in the placebo group (10%); additionally, there were more patients with severe (stage 3) disease and fewer patients with less severe (stage 1) disease in the prednisolone group than in the placebo group, although again this was not statistically significant*
[9] *Chotmongkol et al, 1996: antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[10] *GRADE rule of thumb: <300 events*
[11] *Wide confidence intervals*
[12] *Odds ratio and 95% confidence interval calculated by reviewer*
[13] *Malhotra et al, 2009: use of allocation concealment is unclear; blinding not used*
[14] *Chotmongkol et al (1996), Malhotra et al (2009): antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[15] *Forest plot (mortality):*

Corticosteroids *vs* **antituberculosis chemotherapy alone or plus placebo** in **stage 2 CNS tuberculosis**

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (stage 2) (follow-up 6 to 18 months; assessed with: number of deaths amongst those with stage 2 CNS TB on admission)												
4[1,2,3,4]	randomised trials	very serious[5,6,7,8]	very serious[9,10,11,12,13]	serious[14]	serious[15]	none	16/98 (16.3%)	14/82 (17.1%)	OR 0.70 (0.28 to 1.77)[16,22]	4 fewer per 100 (from 12 fewer to 10 more)	⊙○○○ VERY LOW	
Mortality (stage 2; rifampicin-containing antituberculosis regimens only) (follow-up 6 to 18 months; assessed with: number of deaths amongst those with stage 2 CNS TB on admission)												
3[1,2,4]	randomised trials	very serious[5,6,7,8]	very serious[9,11,12,13]	serious[14]	serious[15]	none	7/75	7/73	OR 0.91 (0.28 to 2.99)[16]	1 fewer per 100 (from 7 fewer to 14 more)	⊙○○○ VERY LOW	
Mortality (dexamethasone; stage 2) (measured with: survival rate at 5 years amongst those classified as stage 2 on admission; better indicated by higher values)												
1[17]	randomised trials	serious[18]	no serious inconsistency	no serious indirectness	no serious imprecision	none	122	125	-	difference in survival rates 0.02 lower (0.15 lower to 0.11 higher)	⊙⊙⊙○ MODERATE	
Mortality (prednisolone; stage 2) (follow-up 6 to 18 months; assessed with: number of deaths amongst those classified as stage 2 on admission)												
2[1,2]	randomised trials	very serious[6,7,19]	very serious[9,12]	serious[20]	very serious[15,21]	none	2/57 (3.5%)	1/56 (1.8%)	OR 1.61 (0.19 to 13.49)[16,23]	1 more per 100 (from 1 fewer to 18 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - severe disability status (stage 2) (follow-up 10 months; assessed with: number of deaths amongst those with stage 2 CNS TB on admission)												
1[4]	randomised trials	serious[8]	no serious inconsistency	serious[20]	serious[15]	none	6/35 (17.1%)	3/18 (16.7%)	OR 1.03 (0.23 to 4.73)[16]	0 more per 100 (from 12 fewer to 32 more)	⊙○○○ VERY LOW	

[1] *Schoeman et al, 1997*
[2] *Chotmongkol et al, 1996*
[3] *O'Toole et al, 1969*
[4] *Malhotra et al, 2009*
[5] *Schoeman et al (1997) and Chotmongkol et al (1996): method of randomisation and use of allocation concealment is unclear*
[6] *Schoeman et al, 1997: blinded = clinical psychologist assessing intelligence, clinician testing hearing, ophthalmologist testing vision, and physical therapist testing motor function; unclear if patients or other health professionals were blinded*
[7] *Unclear if analysis followed the intent-to-treat principle*
[8] *Malhotra et al, 2009: use of allocation concealment is unclear; blinding not used*
[9] *Chotmongkol et al, 1996: groups not comparable at baseline - clinical presentations and staging were similar in the intervention and comparator groups at randomisation; however, although not statistically significant, more patients in the prednisolone group (17%) had motor weakness than in the placebo group (3%), and more patients in the prednisolone group (17%) had motor weakness*

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
than in the placebo group (10%); additionally, there were more patients with severe (stage 3) disease and fewer patients with less severe (stage 1) disease in the prednisolone group than in the placebo group, although again this was not statistically significant												

[10] O'Toole et al, 1969: unclear if groups were comparable at baseline, or if they were comparable for treatment completion and availability of outcome data
[11] Kumarvelu et al, 1994: follow-up only 3 months after treatment initiation
[12] Follow-up varied widely between groups
[13] Estimates of effect very widely across the studies
[14] Chotmongkol et al (1996), O'Toole et al (1969), Girgis et al (1991), Kumarvelu et al (1994): antituberculosis regimens do not use all of or just the 4 standard recommended drugs
[15] GRADE rule of thumb: <300 events
[16] Odds ratio and 95% confidence interval calculated by reviewer
[17] Thwaites et al, 2004/7 / Török et al, 2011
[18] Analysis does not follow intent-to-treat principle
[19] Method of randomisation and use of allocation concealment is unclear
[20] Chotmongkol et al (1996), Malhotra et al (2009): antituberculosis regimens do not use all of or just the 4 standard recommended drugs
[21] Wide confidence intervals
[22] Forest plot (mortality):

[23] Forest plot (mortality; prednisolone):

Corticosteroids *vs* **antituberculosis chemotherapy alone or plus placebo** in **stage 3 CNS tuberculosis**

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (stage 3) (follow-up 6 to 18 months; assessed with: number of deaths amongst those with stage 3 CNS TB on admission)												
4[1,2,3,4]	randomised trials	very serious[5,6,7,8]	very serious[9,10,11,12,13]	serious[14]	serious[15]	none	13/49 (26.5%)	21/49 (42.9%)	OR 0.42 (0.14 to 1.27)[16,24]	19 fewer per 100 (from 33 fewer to 6 more)	⊙○○○ VERY LOW	
Mortality (stage 3; rifampicin-containing antituberculosis regimens only) (follow-up 6 to 18 months; assessed with: number of deaths amongst those with stage 3 CNS TB on admission)												
3[1,2,4]	randomised trials	very serious[5,6,7,8]	very serious[9,11,12,13]	serious[14]	serious[15]	none	10/45 (22.2%)	17/45 (37.8%)	OR 0.53 (0.12 to 2.27)[16]	13 fewer per 100 (from 31 fewer to 20 more)	⊙○○○ VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (dexamethasone; stage 3) (measured with: survival rate at 5 years amongst those classified as stage 3 on admission; better indicated by higher values)												
1[17]	randomised trials	serious[18]	no serious inconsistency	no serious indirectness	no serious imprecision	none	62	60	-	difference in survival rates 0.02 lower (0.2 lower to 0.15 higher)	⊙⊙⊙○ MODERATE	
Mortality (prednisolone; stage 3) (follow-up 6 to 18 months; assessed with: number of deaths amongst those classified as stage 3 on admission)												
2[1,2]	randomised trials	very serious[6,7,19]	very serious[9,12]	serious[20]	serious[15]	none	7/39 (17.9%)	14/39 (35.9%)	OR 0.47 (0.05 to 4.44)[16,25]	15 fewer per 100 (from 33 fewer to 35 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - severe disability status (stage 3) (follow-up 10 months; assessed with: number of deaths amongst those with stage 3 CNS TB on admission)												
1[4]	randomised trials	serious[21]	no serious inconsistency	serious[22]	very serious[15,23]	none	3/12 (25%)	1/5 (20%)	OR 1.22 (0.1 to 17.1)[16]	3 more per 100 (from 18 fewer to 61 more)	⊙○○○ VERY LOW	

[1] *Schoeman et al, 1997*
[2] *Chotmongkol et al, 1996*
[3] *O'Toole et al, 1969*
[4] *Malhotra et al, 2009*
[5] *Schoeman et al (1997) and Chotmongkol et al (1996): method of randomisation and use of allocation concealment is unclear*
[6] *Schoeman et al, 1997: blinded = clinical psychologist assessing intelligence, clinician testing hearing, ophthalmologist testing vision, and physical therapist testing motor function; unclear if patients or other health professionals were blinded*
[7] *Unclear if analysis followed the intent-to-treat principle*
[8] *Malhotra et al, 2009: blinding not used*
[9] *Chotmongkol et al, 1996: groups not comparable at baseline - clinical presentations and staging were similar in the intervention and comparator groups at randomisation; however, although not statistically significant, more patients in the prednisolone group (17%) had motor weakness than in the placebo group (3%), and more patients in the prednisolone group (17%) had motor weakness than in the placebo group (10%); additionally, there were more patients with severe (stage 3) disease and fewer patients with less severe (stage 1) disease in the prednisolone group than in the placebo group, although again this was not statistically significant*
[10] *O'Toole et al, 1969: unclear if groups were comparable at baseline, or if they were comparable for treatment completion and availability of outcome data*
[11] *Kumarvelu et al, 1994: follow-up only 3 months after treatment initiation*
[12] *Follow-up varied widely between groups*
[13] *Estimates of effect very widely across the studies*
[14] *Chotmongkol et al (1996), O'Toole et al (1969), Girgis et al (1991), Kumarvelu et al (1994): antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[15] *GRADE rule of thumb: <300 events*
[16] *Odds ratio and 95% confidence interval calculated by reviewer*
[17] *Thwaites et al, 2004/7 / Török et al, 2011*
[18] *Analysis does not follow intent-to-treat principle*
[19] *Method of randomisation and use of allocation concealment is unclear*
[20] *Chotmongkol et al, 1996: antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[21] *Malhotra et al, 2009: use of allocation concealment is unclear; blinding not used*
[22] *Chotmongkol et al (1996), Malhotra et al (2009): antituberculosis regimens do not use all of or just the 4 standard recommended drugs*

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance

[23] *Wide confidence intervals*
[24] *Forest plot (mortality):*

[25] *Forest plot (mortality; prednisolone):*

Corticosteroids *vs* **antituberculosis chemotherapy alone or plus placebo** in **culture-positive CNS tuberculosis**

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (dexamethasone; culture-positive) (follow-up 2 years; assessed with: number of deaths amongst those classified as culture-positive on admission)												
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[3]	serious[4]	none	32/75 (42.7%)	50/85 (58.8%)	OR 0.52 (0.28 to 0.98)[5]	16 fewer per 100 (from 0 fewer to 30 fewer)	⊙○○○ VERY LOW	
Changes in signs and symptoms - neurological abnormalities during treatment (dexamethasone; culture-positive) (follow-up ; assessed with: number of patients to develop neurologic abnormalities (fundus, hemiparesis or hydrocephalus) during treatment amongst those classified as culture-positive on admission)												
1[1]	randomised trials	serious[2]	no serious inconsistency	serious	serious[4]	none	4/75 (5.3%)	10/85 (11.8%)	OR 0.42 (0.13 to 1.41)[5]	6 fewer per 100 (from 10 fewer to 4 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - residual neurological abnormalities (dexamethasone; culture-positive) (assessed with: number of patients to with permanent residual neurologic abnormalities (fundus, hemiparesis or hydrocephalus) amongst those classified as culture-positive on admission)												
1[1]	randomised trials	serious[2]	no serious inconsistency	serious	serious[4]	none	6/75 (8%)	13/85 (15.3%)	OR 0.48 (0.17 to 1.34)[5]	7 fewer per 100 (from 12 fewer to 4 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - fever (dexamethasone; culture-positive) (measured with: time to become afebrile (defined as a temperature of <37.5'C) amongst those classified as culture-positive on admission; better indicated by lower values)												
1[1]	randomised trials	serious[2]	no serious inconsistency	serious	no serious imprecision	none	75	85	-	MD 3.0 days lower (6.9 lower to 0.9 higher)[6]	⊙⊙○○ LOW	
Changes in signs and symptoms - responsiveness (dexamethasone; culture-positive) (measured with: time to become fully alert amongst those classified as culture-positive on admission[7]; better indicated by lower values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[3]	serious[8]	none	75	85	-	MD 4 days higher (4.9 lower to 12.9 higher)[6]	⊙○○○ VERY LOW	
Adverse events - ocular (dexamethasone; culture-positive; non-randomised) (follow-up unclear; assessed with: number of patients with ocular complications amongst those classified as culture-positive on admission)												
1[9]	non-randomised trials	very serious[10,11,12]	serious[13,14]	serious[3]	very serious[4,8]	none	2/30 (6.7%)	4/34 (11.8%)	OR 2.46 (0.42 to 14.52)[5]	13 more per 100 (from 6 fewer to 54 more)	⊙○○○ VERY LOW	

[1] *Girgis et al, 1991*
[2] *Girgis et al, 1991: use of allocation concealment and blinding unclear*
[3] *Girgis et al (1983 and 1991): antituberculosis regimens do not use all of or just the 4 standard recommended drugs; of particular note is the lack of rifampicin*
[4] *GRADE rule of thumb: <300 events*
[5] *Odds ratio and 95% confidence interval calculated by reviewer*
[6] *Mean difference and 95% confidence interval calculated by reviewer*
[7] *For full definition, see evidence table*
[8] *Wide confidence interval*
[9] *Girgis et al, 1983*
[10] *Non-randomised; patients were alternately assigned to receive antituberculosis chemotherapy plus dexamethasone or antituberculosis chemotherapy alone*
[11] *No allocation concealment*
[12] *Use of blinding unclear*
[13] *Authors state that groups were comparable with respect to age, sex and disease severity on admission to hospital; however, although not statistically significant, more patients in the dexamethasone group (32/70) were comatose on admission than in the antituberculosis chemotherapy alone group (41/66) - that is, the condition of those in the dexamethasone group could be considered to be more severe*
[14] *Unclear if groups received the same care except for the intervention(s) studied; limited information available*

Corticosteroids *vs* **antituberculosis chemotherapy alone or plus placebo** in **culture-negative CNS tuberculosis**

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (dexamethasone; culture-negative) (follow-up 2 years; assessed with: number of deaths amongst those classified as culture-negative on admission)												
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[3]	serious[4]	none	40/70 (57.1%)	29/50 (58%)	OR 0.97 (0.46 to 2.01)[5]	1 fewer per 100 (from 19 fewer to 16 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - neurological abnormalities during treatment (dexamethasone; culture-negative) (follow-up ; assessed with: number of patients to develop												

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy alone or plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
neurologic abnormalities (fundus, hemiparesis or hydrocephalus) during treatment amongst those classified as culture-negative on admission)												
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[3]	serious[4]	none	4/70 (5.7%)	5/50 (10%)	OR 0.67 (0.17 to 2.6)[5]	3 fewer per 100 (from 8 fewer to 12 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - residual neurological abnormalities (dexamethasone; culture-negative) (assessed with: number of patients to with permanent residual neurologic abnormalities (fundus, hemiparesis or hydrocephalus) amongst those classified as culture-negative on admission)												
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[3]	serious[4]	none	8/70 (11.4%)	14/50 (28%)	OR 0.33 (0.13 to 0.87)[5]	17 fewer per 100 (from 3 fewer to 23 fewer)	⊙○○○ VERY LOW	

[1] *Girgis et al, 1991*
[2] *Use of allocation concealment and blinding unclear*
[3] *Antituberculosis regimens do not use all of or just the 4 standard recommended drugs; of particular note is the lack of rifampicin*
[4] *GRADE rule of thumb: <300 events*
[5] *Odds ratio and 95% confidence interval calculated by reviewer*

BONE & JOINT, INCLUDING SPINAL, TUBERCULOSIS

Prednisolone *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute (95% CI)		
Response to treatment – need for additional surgical intervention (assessed with: number of patients requiring surgery due to insufficient shrinkage of the swollen joint)												
1[1]	randomised trials	serious[2]	serious[3,4]	very serious[5,6,7,8]	very serious[9,10]	none	9/10 (90%)	5/6 (83.3%)	OR 1.80 (0.09 to 35.43)[11]	67 more per 1000 (from 523 fewer to 161 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – weight (assessed with: number of patients that failed to gain weight)												
1[1]	randomised trials	serious[2]	serious[3,4]	very serious[5,6,7]	very serious[9,10]	none	1/10 (10%)	1/6 (16.7%)	OR 0.56 (0.03 to 10.93)[11]	66 fewer per 1000 (from 161 fewer to 519 more)	⊙○○○ VERY LOW	

[1] *Cathro, 1958*
[2] *Method of randomisation, and use of allocation concealment and blinding, is unclear*
[3] *Details provided are limited, but site of disease varies between the 2 groups: prednisolone group = 7 spinal, 2 knee, 1 hip; antituberculosis chemotherapy alone = 4 hip, 2 knee*
[4] *It is unclear if the groups received the same care apart from the intervention(s) studied as authors provided only limited information*
[5] *Only limited details of the study population available; therefore the directness of the study population cannot be confirmed*
[6] *Antituberculosis regimens do not use all of or just the 4 standard recommended drugs; of particular note was the lack of rifampicin*
[7] *Only 2 antituberculosis drugs used*
[8] *Outcome is a surrogate for the outcomes of interest*
[9] *GRADE rule of thumb: <300 events*
[10] *Wide confidence interval*
[11] *Odds ratio and 95% confidence interval calculated by reviewer*

PERICARDIAL TUBERCULOSIS

Prednisolone *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow-up 1 to 10 years; assessed with: number of deaths)												
4[1,2,3,4]	randomised trials	very serious[5,6,7,8]	very serious[9,10,11]	serious[12]	serious[13]	none	47/224 (21%)	64/249 (25.7%)	OR 0.70 (0.45 to 1.08)[14,17]	6 fewer per 100 (from 12 fewer to 1 more)	⊙○○○ VERY LOW	
Response to treatment - favourable (assessed with: number of patients to be considered in a favourable status after 24 months of follow-up)												
2[3,4]	randomised	very	serious[10]	very	serious[13]	none	141/187	140/196	OR 1.23	4 more	⊙○○○	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
	trials	serious[7,8]		serious[12,15]			(75.4%)	(71.4%)	(0.78 to 1.93)[14,18]	per 100 (from 5 fewer to 11 more)	VERY LOW	
Response to treatment - need for surgical intervention (assessed with: number of patients to require surgical intervention (pericardectomy))												
3[1,3,4]	randomised trials	very serious[5,7,8]	serious[10]	very serious[12,15]	[13]	none	31/220 (14.1%)	29/220 (13.2%)	OR 1.12 (0.6 to 2.09)[14,19]	1 more per 100 (from 5 fewer to 11 more)		
Changes in signs and symptoms - unrestricted physical activity (assessed with: number of patients with unrestricted physical activity after 10 years of follow-up)												
2[3,4]	randomised trials	very serious[7,8]	serious[10]	serious[12]	serious[13]	none	30/187 (16%)	60/196 (30.6%)	OR 0.43 (0.26 to 0.71)[14,20]	15 fewer per 100 (from 7 fewer to 20 fewer)	⊙○○○ VERY LOW	
Changes in signs and symptoms - 'out and about' but restricted physical activity (assessed with: number of patients to be 'out and about' but with restricted physical activity after 10 years of follow-up)												
2[3,4]	randomised trials	very serious[7,8]	serious[10]	serious[12]	serious[13]	none	94/187 (50.3%)	78/196 (39.8%)	OR 1.53 (1.02 to 2.3)[14,21]	10 more per 100 (from 0 more to 21 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - confined, restricted physical activity (assessed with: number of patients to confined to home or hospital after 10 years of follow-up)												
2[3,4]	randomised trials	very serious[7,8]	serious[10]	serious[12]	very serious[13,16]	none	96/187 (51.3%)	140/196 (71.4%)	OR 0.21 (0 to 9.34)[14,22]	37 fewer per 100 (from 71 fewer to 24 more)	⊙○○○ VERY LOW	

[1] *Reuter et al, 2006*
[2] *Hakim et al, 2000*
[3] *Strang et al, 1987/2004*
[4] *Strang et al, 1988/2004*
[5] *Reuter et al (2006): randomisation code remained concealed and was not revealed to the investigators or the study subjects until completion of the study; however, physician administering the intrapericardial steroids/placebo was unblinded*
[6] *Hakim et al, 2000: use of allocation concealment unclear*
[7] *Analysis does not follow the intent-to-treat principle*
[8] *Strang et al (1987/2004 and 1988/2004): quasi-randomised*
[9] *Hakim et al, 2000: unclear if the groups were comparable in terms of treatment completion and availability of outcome data*
[10] *Strang et al, 1988/2004: unclear if the groups were comparable at the baseline*
[11] *Follow-up periods varied widely*
[12] *Strang et al (1987/2004 and 1988/2004): antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[13] *GRADE rule of thumb: <300 events*
[14] *Odds ratio and 95% confidence intervals calculated by reviewer*
[15] *Outcome is a surrogate for an outcome of interest*
[16] *Wide confidence intervals*

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
[17] *Forest plot (mortality):*												
[18] *Forest plot (response to treatment – favourable):*												
[19] *Forest plot (response to treatment – need for surgical intervention):*												
[20] *Forest plot (changes in signs and symptoms - unrestricted physical activity):*												
[21] *Forest plot (changes in signs and symptoms - 'out and about' but restricted physical activity):*												
[22] *Forest plot (changes in signs and symptoms - confined, restricted physical activity):*												

Prednisolone *vs* **antituberculosis chemotherapy alone or plus placebo** for **effusive pericardial tuberculosis**

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (effusive TB) (follow-up 1 to 10 years; assessed with: number of deaths)												
3[1,2,3]	randomised trials	very serious[4,5,6,7]	very serious[8,9,10]	serious[11]	serious[12]	none	31/154 (20.1%)	43/176 (24.4%)	OR 0.69 (0.4 to 1.17)[13,15]	6 fewer per 100 (from 13 fewer to 3 more)	⊙○○○ VERY LOW	
Response to treatment - favourable (effusive TB) (follow-up 10 years; assessed with: number of patients to be considered in a favourable status after 24 months of follow-up)												
1[3]	randomised trials	very serious[6,7]	serious[9]	serious[11]	serious[12]	none	91/117 (77.8%)	88/123 (71.5%)	OR 1.39 (0.77 to 2.5)[13]	6 more per 100 (from 6 fewer to 15 more)	⊙○○○ VERY LOW	
Response to treatment - need for surgical intervention (effusive TB) (follow-up 1 to 10 years; assessed with: number of patients to require surgical intervention (pericardectomy))												
2[1,3]	randomised trials	very serious[4,6,7]	very serious[8,9,10]	serious[11]	serious[12]	none	12/125 (9.6%)	7/147 (4.8%)	OR 1.98 (0.77 to 5.09)[13,16]	4 more per 100 (from 1 fewer to 16 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - unrestricted physical activity (effusive TB) (follow-up 10 years; assessed with: number of patients to with unrestricted physical activity after 10 years of follow-up)												
1[3]	randomised trials	very serious[6,7]	serious[9]	serious[11]	serious[12]	none	21/116 (18.1%)	20/123 (16.3%)	OR 0.68 (0.36 to 1.27)[13]	5 fewer per 100 (from 10 fewer to 4	⊙○○○ VERY LOW	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
										more)		
Changes in signs and symptoms - 'out and about' but restricted physical activity (effusive TB) (follow-up 10 years; assessed with: number of patients to be 'out and about' but with restricted physical activity after 10 years of follow-up)												
1[3]	randomised trials	very serious[6,7]	serious[9]	serious[11]	serious[12]	none	57/117 (48.7%)	46/123 (37.4%)	OR 1.59 (0.95 to 2.66)[13]	11 more per 100 (from 1 fewer to 24 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - confined, restricted physical activity (effusive TB) (follow-up 10 years; assessed with: number of patients to confined to home or hospital after 10 years of follow-up)												
1[3]	randomised trials	very serious[6,7]	serious[9]	serious[11]	serious[12]	none	8/117 (6.8%)	7/123 (5.7%)	OR 1.22 (0.43 to 3.47)[13]	1 more per 100 (from 3 fewer to 12 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - reduced activity levels (follow-up 1 years; assessed with: number of patients to experience reduced levels of activity at 1-year of follow-up)												
1[1]	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[12,14]	none	1/8 (12.5%)	3/24 (12.5%)	OR 1.00 (0.09 to 11.24)[13]	0 fewer per 100 (from 11 fewer to 49 more)	⊙○○○ VERY LOW	

[1] *Reuter et al, 2006*
[2] *Hakim et al, 2000*
[3] *Strang et al, 1988/2004*
[4] *Reuter et al (2006): randomisation code remained concealed and was not revealed to the investigators or the study subjects until completion of the study; however, physician administering the intrapericardial steroids/placebo was unblinded*
[5] *Hakim et al, 2000: use of allocation concealment unclear*
[6] *Analysis does not follow the intent-to-treat principle*
[7] *Strang et al, 1988/2004: quasi-randomised*
[8] *Hakim et al, 2000: unclear if the groups were comparable in terms of treatment completion and availability of outcome data*
[9] *Strang et al, 1988/2004: unclear if the groups were comparable at the baseline*
[10] *Follow-up periods varied widely*
[11] *Strang et al, 1988/2004: antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[12] *GRADE rule of thumb: <300 events*
[13] *Odds ratio and 95% confidence intervals calculated by reviewer*
[14] *Wide confidence intervals*
[15] *Forest plot (mortality):*

[16] *Forest plot (response to treatment - need for surgical intervention):*

Prednisolone *vs* **antituberculosis chemotherapy alone or plus placebo** for **constrictive tuberculous pericarditis**

Quality assessment							No of patients		Effect		Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (constrictive tuberculous pericarditis) (follow-up 10 years; assessed with: number of deaths)												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	serious[4]	serious[5]	none	16/70 (22.9%)	21/73 (28.8%)	OR 0.73 (0.35 to 1.56)[6]	6 fewer per 100 (from 16 fewer to 10 more)	⊙○○○ VERY LOW	
Response to treatment - favourable (constrictive tuberculous pericarditis) (assessed with: number of patients to be considered in a favourable status after 24 months of follow-up)												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	very serious[4,7]	serious[5]	none	50/70 (71.4%)	52/73 (71.2%)	OR 1.01 (0.49 to 2.08)[6]	0 more per 100 (from 16 fewer to 13 more)	⊙○○○ VERY LOW	
Response to treatment - need for surgical intervention (constrictive tuberculous pericarditis) (Copy) (assessed with: number of patients to require surgical intervention (pericardectomy))												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	very serious[4,7]	serious[5]	none	18/70 (25.7%)	22/73 (30.1%)	OR 0.80 (0.39 to 1.67)[6]	4 fewer per 100 (from 16 fewer to 12 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - unrestricted physical activity (constrictive tuberculous pericarditis) (follow-up 10 years; assessed with: number of patients to with unrestricted physical activity after 10 years of follow-up)												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	serious[4]	serious[5]	none	9/70 (12.9%)	14/73 (19.2%)	OR 0.62 (0.22 to 1.55)[6]	6 fewer per 100 (from 14 fewer to 8 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - 'out and about' but restricted physical activity (constrictive tuberculous pericarditis) (follow-up 10 years; assessed with: number of patients to be 'out and about' but with restricted physical activity after 10 years of follow-up)												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	serious[4]	serious[5]	none	37/70 (52.9%)	32/73 (43.8%)	OR 1.44 (0.74 to 2.78)[6]	9 more per 100 (from 7 fewer to 25 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - confined, restricted physical activity (constrictive tuberculous pericarditis) (follow-up 10 years; assessed with: number of patients to confined to home or hospital after 10 years of follow-up)												
1[1]	randomised trials	very serious[2,3]	no serious inconsistency	serious[4]	very serious[5,8]	none	5/70 (7.1%)	2/73 (2.7%)	OR 2.73 (0.51 to 14.56)[6]	4 more per 100 (from 1 fewer to 26 more)	⊙○○○ VERY LOW	

[1] *Strang et al, 1987/2004*
[2] *Analysis does not follow the intent-to-treat principle*
[3] *Quasi-randomised*
[4] *Antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[5] *GRADE rule of thumb: <300 events*
[6] *Odds ratio and 95% confidence intervals calculated by reviewer*
[7] *Outcome is a surrogate for an outcome of interest*

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance

[8] *Wide confidence interval*

Any corticosteroid *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow-up 1 to 10 years; assessed with: number of deaths)												
4[1,2,3,4]	randomised trials	very serious[5,6,7,8]	very serious[9,10,11]	serious[12]	serious[13]	none	47/249 (18.9%)	64/249 (25.7%)	OR 0.67 (0.44 to 1.03)[14,16]	6 fewer per 100 (from 12 fewer to 1 more)	⊙○○○ VERY LOW	
Response to treatment - need for surgical intervention (follow-up 1 to 10 years; assessed with: number of patients to require surgical intervention (pericardectomy))												
3[1,3,4]	randomised trials	very serious[5,7,8]	very serious[10,11]	serious[12]	serious[13]	none	31/220 (14.1%)	29/220 (13.2%)	OR 1.12 (0.6 to 2.09)[14,17]	1 more per 100 (from 5 fewer to 11 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - reduced activity levels (follow-up 1 years; assessed with: number of patients to experience reduced levels of activity at 1-year of follow-up)												
1[1]	randomised trials	serious[5]	no serious inconsistency	no serious indirectness	very serious[13,15]	none	4/33 (12.1%)	3/24 (12.5%)	OR 0.97 (0.20 to 4.78)[14]	12 fewer per 100 (from 10 fewer to 28 more)	⊙○○○ VERY LOW	

[1] *Reuter et al, 2006*
[2] *Hakim et al, 2000*
[3] *Strang et al, 1987/2004*
[4] *Strang et al, 1988/2004*
[5] *Reuter et al (2006): randomisation code remained concealed and was not revealed to the investigators or the study subjects until completion of the study; however, physician administering the intrapericardial steroids/placebo was unblinded*
[6] *Hakim et al, 2000: use of allocation concealment unclear*
[7] *Analysis does not follow the intent-to-treat principle*
[8] *Strang et al (1987/2004 and 1988/2004): quasi-randomised*
[9] *Hakim et al, 2000: unclear if the groups were comparable in terms of treatment completion and availability of outcome data*
[10] *Strang et al, 1988/2004: unclear if the groups were comparable at the baseline*
[11] *Follow-up periods varied widely*
[12] *Strang et al (1987/2004 and 1988/2004): antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[13] *GRADE rule of thumb: <300 events*
[14] *Odds ratio and 95% confidence intervals calculated by reviewer*
[15] *Wide confidence intervals*
[16] *Forest plot (mortality):*

[17] *Forest plot (response to treatment - need for surgical intervention):*

Any corticosteroid *vs* antituberculosis chemotherapy alone or plus placebo for effusive pericardial tuberculosis

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus any corticosteroid	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (effusive TB) (follow-up 1 to 10 years; assessed with: number of deaths)												
3[1,2,3]	randomised trials	very serious[4,5,6,7]	very serious[8,9,10]	serious[11]	serious[12]	none	31/179 (17.3%)	43/176 (24.4%)	OR 0.69 (0.4 to 1.17)[13,14]	6 fewer per 100 (from 13 fewer to 3 more)	⊙○○○ VERY LOW	
Response to treatment - need for surgical intervention (effusive TB) (follow-up 1 to 10 years; assessed with: number of patients to require surgical intervention (pericardectomy))												
2[1,3]	randomised trials	very serious[4,6,7]	very serious[8,9,10]	serious[11]	serious[12]	none	13/150 (8.7%)	7/147 (4.8%)	OR 1.85 (0.73 to 4.73)[13,15]	4 more per 100 (from 1 fewer to 14 more)	⊙○○○ VERY LOW	

[1] *Reuter et al, 2006*
[2] *Hakim et al, 2000*
[3] *Strang et al, 1988/2004*
[4] *Reuter et al (2006): randomisation code remained concealed and was not revealed to the investigators or the study subjects until completion of the study; however, physician administering the intrapericardial steroids/placebo was unblinded*
[5] *Hakim et al, 2000: use of allocation concealment unclear*
[6] *Analysis does not follow the intent-to-treat principle*
[7] *Strang et al, 1988/2004: quasi-randomised*
[8] *Hakim et al, 2000: unclear if the groups were comparable in terms of treatment completion and availability of outcome data*
[9] *Strang et al, 1988/2004: unclear if the groups were comparable at the baseline*
[10] *Follow-up periods varied widely*
[11] *Strang et al, 1988/2004: antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[12] *GRADE rule of thumb: <300 events*
[13] *Odds ratio and 95% confidence intervals calculated by reviewer*
[14] *Forest plot (mortality):*

[15] *Forest plot (response to treatment - need for surgical intervention):*

Prednisolone *vs* **triamcinalone**

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus triamcinalone	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Mortality (effusive TB) (follow-up 1 years; assessed with: number of deaths)												
1[1]	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	very serious[3,4]	none	0/8 (0%)	0/17 (0%)	2.06 (0.04 to 112.94)[5]	-	⊙○○○ VERY LOW	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus triamcinalone	Relative (95% CI)	Absolute (95% CI)		
Response to treatment – need for additional intervention (effusive TB) (follow-up 1 years; assessed with: number of patients to require surgery)												
1[1]	randomised trials	serious[2]	no serious inconsistency	serious[6]	very serious[3,4]	none	1/8 (12.5%)	0/17 (0%)	OR 6.18 (0.23 to 168.11)[5]	-	⊙○○○ VERY LOW	
Changes in signs and symptoms – activity levels (effusive TB) (follow-up 1 years; assessed with: number of patients to experience reduced levels of activity at 1-year of follow-up)												
1[1]	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	very serious[3,4]	none	1/8 (12.5%)	2/17 (11.8%)	OR 1.07 (0.08 to 13.9)[5]	1 more per 100 (from 11 fewer to 53 more)	⊙○○○ VERY LOW	

[1] *Reuter et al, 2006*
[2] *Randomisation code remained concealed and was not revealed to the investigators or the study subjects until completion of the study; however, physician administering the intrapericardial steroids/placebo was unblinded*
[3] *GRADE rule of thumb: <300 events*
[4] *Wide confidence intervals*
[5] *Odds ratio and 95% confidence intervals calculated by reviewer*
[6] *Outcome is a surrogate for an outcome of interest*

Prednisolone *vs* **antituberculosis chemotherapy alone or plus placebo** for **effusive pericardial tuberculosis** in **people with HIV**

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (HIV-positive; effusive TB) (follow-up 18 months; assessed with: number of deaths)												
1[1]	randomised trials	serious[2,3]	serious[4]	no serious indirectness	serious[5]	none	5/29 (17.2%)	10/29 (34.5%)	OR 0.40 (0.12 to 1.36)[6]	17 fewer per 100 (from 29 fewer to 7 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms - constrictive pericarditis (HIV-positive; effusive TB) (follow-up 18 months; assessed with: number of patients to experience constrictive pericarditis)												
1[1]	randomised trials	serious[2,3]	serious[4]	no serious indirectness	very serious[5,7]	none	2/29 (6.9%)	2/29 (6.9%)	OR 1.00 (0.13 to 7.62)[6]	0 fewer per 100 (from 6 fewer to 29 more)	⊙○○○ VERY LOW	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
Adherence (HIV-positive; effusive TB) (follow-up 18 months; assessed with: number of pill counts showing that >90% of tablets had been consumed)												
1[1]	randomised trials	serious[2,3]	serious[4]	serious[8]	no serious imprecision	none	169/230 (73.5%)	119/182 (65.4%)	OR 1.47 (0.96 to 2.24)[6]	8 more per 100 (from 1 fewer to 15 more)	⊙○○○ VERY LOW	

[1] *Hakim et al, 2000*
[2] *Use of allocation concealment unclear*
[3] *Unclear if analysis follows the intent-to-treat principle*
[4] *Unclear if the groups were comparable in terms of treatment completion and availability of outcome data*
[5] *GRADE rule of thumb: <300 events*
[6] *Odds ratio and 95% confidence intervals calculated by reviewer*
[7] *Wide confidence intervals*
[8] *Outcome is a surrogate for an outcome of interest*

TB- ASSOCIATED IMMUNE RECONSTITUTION INFLAMMATORY SYNDROME

Prednisolone *vs* antituberculosis chemotherapy alone or plus placebo

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)		
Mortality (follow-up 12 weeks; assessed with: number of deaths)												
1[1]	randomised trials	serious[2]	serious[3]	no serious indirectness	very serious[5,6]	none	3/55 (5.5%)	2/55 (3.6%)	OR 1.53 (0.25 to 9.52)[7]	2 more per 100 (from 3 fewer to 23 more)	⊙○○○ VERY LOW	
Change in signs and symptoms – improvement (assessed with: number of patients in whom symptoms improved or were resolved after 4 weeks[4])												
1[1]	randomised trials	serious[2]	serious[3]	no serious indirectness	serious[5]	none	44/55 (80%)	31/55 (56.4%)	OR 1.81 (0.72 to 4.5)[7]	14 more per 100 (from 8 fewer to 29 more)	⊙○○○ VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus prednisolone	Antituberculosis chemotherapy plus placebo	Relative (95% CI)	Absolute (95% CI)	Quality	Importance
Change in signs and symptoms – deterioration (assessed with: number of patients in whom symptoms deteriorated after 4 weeks[4])												
1[1]	randomised trials	serious[2]	serious[3]	no serious indirectness	serious[5]	none	7/55 (12.7%)	9/55 (16.4%)	OR 0.75 (0.26 to 2.17)[7]	4 fewer per 100 (from 12 fewer to 13 more)	⊙○○○ VERY LOW	
Change in signs and symptoms – improvement of chest radiograph (assessed with: number of patients whose chest radiographs improved or were resolved after 4 weeks[4])												
1[1]	randomised trials	serious[2]	serious[3]	no serious indirectness	very serious[5,6]	none	40/55 (72.7%)	25/55 (45.5%)	OR 3.20 (1.44 to 7.09)[7]	27 more per 100 (from 9 more to 40 more)	⊙○○○ VERY LOW	
Change in signs and symptoms – deterioration of chest radiograph (assessed with: number of patients whose chest radiographs deteriorated after 4 weeks[4])												
1[1]	randomised trials	serious[2]	serious[3]	no serious indirectness	serious[5]	none	4/55 (7.3%)	18/55 (32.7%)	OR 0.16 (0.05 to 0.52)[7]	26 fewer per 100 (from 13 fewer to 30 fewer)	⊙○○○ VERY LOW	
Adverse events - drug reactions (assessed with: number of patients to experience adverse drug reactions)												
1[1]	randomised trials	serious[2]	serious[3]	no serious indirectness	very serious[5,6]	none	8/55 (14.5%)	3/55 (5.5%)	OR 2.95 (0.74 to 11.78)[7]	9 more per 100 (from 1 fewer to 35 more)	⊙○○○ VERY LOW	
Adverse events - infections (assessed with: number of patients to experience infections)												
1[1]	randomised trials	serious[2]	serious[3]	no serious indirectness	serious[5]	none	27/55 (49.1%)	17/55 (30.9%)	OR 2.16 (0.99 to 4.7)[7]	18 more per 100 (from 0 fewer to 37 more)	⊙○○○ VERY LOW	

[1] *Meintjes et al, 2010*
[2] *Unclear if allocation concealment was used*
[3] *Groups were not comparable at baseline: there was a longer period (p = 0.02) between taking antituberculosis chemotherapy and initiating ART amongst patients in the prednisolone arm (66 days) than the placebo arm (43.5 days)*
[4] *For full definition, see evidence table*
[5] *GRADE rule of thumb: <300 events*
[6] *Wide confidence intervals*
[7] *Odds ratio and 95% confidence intervals calculated by reviewer*

A.10 RQ P

A.10.1 Duration of treatment in people with non-respiratory tuberculosis

CENTRAL NERVOUS SYSTEM TB

6 *MONTHS* vs **9** MONTHS

Mortality

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Mortality (children only; antituberculosis chemotherapy + corticosteroids) (number of deaths during treatment)										
1[1]	non-randomised trials/observational studies	very serious[2,3,4,6]	serious[5,6]	very serious[7,8,11]	serious[9]	7/45 (15.6%)	2/4 (50%)	OR 0.18 (0.02 to 1.53)[10]	35 fewer per 100 (from 48 fewer to 10 more)	VERY LOW

[1] *Jacobs et al, 1992*
[2] *No randomisation or blinding*
[3] *Allocation concealment unclear*
[4] *No blinding*
[5] *Unclear if the groups were comparable at baseline*
[6] *Unclear if the groups were followed up for the same length of time*
[7] *Regimens do not contain all of/just the 4 standard recommended drugs*
[8] *All patients received corticosteroids*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Doses used are inconsistent with those recommended in the British National Formulary*
Abbreviations: CI, confidence interval; OR, odds ratio

Change in signs and symptoms

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Change in signs and symptoms – neurological sequelae (children only; antituberculosis chemotherapy + corticosteroids) (number of patients to experience neurological sequelae (hydrocephalus, cerebral palsy with mental retardation, hemiparesis, long-term seizures, or behavioural changes))										
1[1]	non-randomised trials/observational studies	very serious[2,3,4,6]	serious[5,6]	very serious[7,8,11]	serious[9]	11/45 (24.4%)	2/4 (50%)	OR 0.32 (0.04 to 2.58)[10]	26 fewer per 100 (from 46 fewer to 22 more)	VERY LOW

[1] *Jacobs et al, 1992*
[2] *No randomisation or blinding*
[3] *Allocation concealment unclear*
[4] *No blinding*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality

[5] *Unclear if the groups were comparable at baseline*
[6] *Unclear if the groups were followed up for the same length of time, or if follow-up was for an appropriate length of time*
[7] *Regimens do not contain all of/just the 4 standard recommended drugs*
[8] *All patients received corticosteroids*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Doses used are inconsistent with those recommended in the British National Formulary*
Abbreviations: CI, confidence interval; OR, odds ratio

8 *MONTHS* vs **12 to 16** MONTHS

Change in signs and symptoms

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	8 months	12 to 16 months	Relative (95% CI)	Absolute (95% CI)	Quality
Change in signs and symptoms – neurological sequelae (antituberculosis chemotherapy + corticosteroids) (follow-up 8 months (median months (IQR)) = 10 (6–24); 12–16 months (median months (IQR)) = 13 (4–36); assessed with: number of patients with residual neurological sequelae (hydrocephalus, cerebral atrophy, hemiparesis/monoparesis, visual impairment, imbalance, sense or hearing loss))										
1[1]	non-randomised trials/observational studies	serious[2,3,4]	very serious[2,5,6,7]	serious[8]	serious[9]	8/37 (21.6%)	10/35 (28.6%)	OR 0.69 (0.24 to 2.02)[10]	7 fewer per 100 (from 20 fewer to 16 more)	VERY LOW

[1] *Doğanay et al, 1995*
[2] *Allocation was based upon the centre attended by the patient - potential systematic differences between clinics (for example, differences in delivery of care)*
[3] *Unclear if attempts were made within the study design or analysis to balance potential confounders*
[4] *No randomisation, and blinding unclear*
[5] *Retreatment and default cases excluded from 8-month group but not the 12-to-16-month group*
[6] *Differences between groups in the corticosteroid regimens used*
[7] *Wide variations in duration of follow-up*
[8] *Intervention does not exactly match the intervention of interest: does not contain all of or just the 4 standard recommended drugs, and the 2 arms vary by more than duration alone*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
Abbreviations: CI, confidence interval; OR, odds ratio

Relapse

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	8 months	12 to 16 months	Relative (95% CI)	Absolute (95% CI)	Quality
Relapse (follow-up: 8-month group (median months (IQR)) = 10 (6–24); 12-to-16-month group (median months (IQR)) = 13 (4–36); assessed with: number of patients to experience relapse)										
1[1]	non-randomised	serious[2,3,4]	very	serious[8]	very serious[9,12]	0/100	0/100	OR 1.00 (0.02	-	VERY LOW

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	8 months	12 to 16 months	Relative (95% CI)	Absolute (95% CI)	Quality
	trials/observational studies		serious[2,5,6,7,11]			(0%)[13]	(0%)[13]	to 50.89)[10]		

[1] *Doğanay et al, 1995*
[2] *Allocation was based upon the centre attended by the patient - potential systematic differences between clinics (for example, differences in delivery of care)*
[3] *Unclear if attempts were made within the study design or analysis to balance potential confounders*
[4] *Blinding unclear*
[5] *Retreatment and default cases excluded from 8-month group but not the 12-to-16-month group*
[6] *Differences between groups in the corticosteroid regimens used*
[7] *Wide variations in duration of follow-up*
[8] *Intervention does not exactly match the intervention of interest: does not contain all of or just the 4 standard recommended drugs, and the 2 arms vary by more than duration alone*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Unclear if the 2 arms were comparable for the availability of outcome data*
[12] *Wide confidence intervals*
[13] *It is unclear how many patients in each group had relapse data available - authors report only that no patient in either arm experienced relapse; therefore reviewer treated this as a 0% relapse rate in each group*
Abbreviations: CI, confidence interval; OR, odds ratio

Adverse events

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	8 months	12 to 16 months	Relative (95% CI)	Absolute (95% CI)	Quality
Adverse events (any) (assessed with: number of patients to experience any adverse event)										
1[1]	non-randomised trials/observational studies	serious[2,3,4]	very serious[2,5,6]	serious[8]	serious[9]	6/37 (16.2%)	8/35 (22.9%)	OR 0.65 (0.20 to 2.12)[10]	7 fewer per 100 (from 17 fewer to 16 more)	VERY LOW

[1] *Doğanay et al, 1995*
[2] *Allocation was based upon the centre attended by the patient - potential systematic differences between clinics (for example, differences in delivery of care)*
[3] *Unclear if attempts were made within the study design or analysis to balance potential confounders*
[4] *Blinding unclear*
[5] *Retreatment and default cases excluded from 8-month group but not the 12-to-16-month group*
[6] *Differences between groups in the corticosteroid regimens used*
[7] *Wide variations in duration of follow-up*
[8] *Intervention does not exactly match the intervention of interest: does not contain all of or just the 4 standard recommended drugs, and the 2 arms vary by more than duration alone*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
Abbreviations: CI, confidence interval; OR, odds ratio

SPINAL TB

6 MONTHS vs 9 MONTHS

Mortality

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Mortality (antituberculosis chemotherapy + surgery) (follow-up 60 months; number of deaths associated with spinal tuberculosis)										
1[1]	randomised trials	very serious[2,3,4]	serious[10]	very serious[5,6]	very serious[7,8]	0/24 (0%)	0/26 (0%)	OR 1.08 (0.02 to 56.64)[9]	-	VERY LOW

[1] *Darbyshire, 1999*
[2] *Allocation concealment unclear*
[3] *Blinding unclear*
[4] *Analysis does not follow the intent-to-treat principle*
[5] *Intervention does not exactly match the intervention of interest: regimens do not contain all of/just the 4 standard recommended drugs, and all patients underwent surgery in addition to receiving antituberculosis chemotherapy*
[6] *Population does not exactly match the population of interest: 6 of the 43 patients tested had single or combined drug resistance*
[7] *GRADE rule of thumb: <300 events*
[8] *Wide confidence intervals*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*
[10] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
Abbreviations: CI, confidence interval; OR, odds ratio

Change in signs and symptoms

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Change in signs and symptoms – complete bony fusion (antituberculosis chemotherapy + surgery) (follow-up 36 months; number of patients with complete bony fusion[1])										
1[11]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	very serious[9,13]	25/25 (100%)	26/26 (100%)	OR 0.96 (0.02 to 50.35)[10]	-	VERY LOW
Change in signs and symptoms – kyphosis (antituberculosis chemotherapy + surgery) (follow-up minimum 10 years; mean increase in the angle of kyphosis from baseline to end of follow-up; better indicated by lower values)										
1[21]	randomised trials	serious[4,5,6]	no serious inconsistency	very serious[8,12]	no serious imprecision[16]	25	26	-	MD 0.7 lower (5.31 lower to 3.91 higher)[22]	VERY LOW
Change in signs and symptoms – kyphosis (antituberculosis chemotherapy + surgery) (follow-up 60 months; increase in the mean angle of kyphosis from baseline to end of follow-up[19]; better indicated by lower values)										
1[23]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	no serious imprecision[16]	14	14	-	MD 14.1 higher[17,18,20]	VERY LOW
Change in signs and symptoms – kyphosis (antituberculosis chemotherapy + surgery) (number of patients with improvement in their angle of kyphosis (reduction of 11° or more) from baseline to 60 months)										
1[23]	randomised	very serious[5,6,7]	serious[24]	very serious[8,12]	very serious[9,13]	0/14	1/14	OR 0.31 (0.01	5 fewer per	VERY LOW

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
	trials					(0%)	(7.1%)	to 8.29)[10]	100 (7 fewer to 32 more)	
Change in signs and symptoms – kyphosis (antituberculosis chemotherapy + surgery) (number of patients with no change in their angle of kyphosis (within ±10°) from baseline to 60 months)										
1[23]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	very serious[9,13]	5/14 (35.7%)	11/14 (78.6%)	OR 0.15 (0.03 to 0.81)[10]	43 fewer per 100 (4 fewer to 69 fewer)	VERY LOW
Change in signs and symptoms – kyphosis (antituberculosis chemotherapy + surgery) (number of patients with deterioration in their angle of kyphosis (increase of 11° or more) from baseline to 60 months)										
1[23]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	very serious[9,13]	9/14 (64.3%)	2/14 (14.3%)	OR 10.80 (1.69 to 68.94)[10]	50 more per 100 (from 8 more to 78 more)	VERY LOW
Change in signs and symptoms – vertebral loss (antituberculosis chemotherapy + surgery) (follow-up 60 months; number of patients with no change in their vertebral loss (an increase or decrease of within 0.24 vertebrae))										
1[23]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	serious[9]	13/24 (54.2%)	14/25 (56%)	OR 0.93 (0.30 to 2.86)[10,20]	2 fewer per 100 (from 28 fewer to 22 more)	VERY LOW
Change in signs and symptoms – vertebral loss (antituberculosis chemotherapy + surgery) (follow-up 60 months; number of patients with improvement in their vertebral loss (reduction in loss of more than 0.25 vertebrae))										
1[23]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	serious[9]	2/24 (8.3%)	5/25 (20%)	OR 0.36 (0.06 to 2.09)[10]	12 fewer per 100 (from 19 fewer to 14 more)	VERY LOW
Change in signs and symptoms – vertebral loss (antituberculosis chemotherapy + surgery) (follow-up 60 months; number of patients with deterioration in their vertebral loss (increase in loss of more than 0.25 vertebrae))										
1[23]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	serious[9]	6/24 (25%)	9/25 (37.5%)	OR 0.59 (0.17 to 2.03)[10]	11 fewer per 100 (from 27 fewer to 17 more)	VERY LOW
Change in signs and symptoms – vertebral loss (antituberculosis chemotherapy + surgery) (mean vertebral loss from treatment initiation to 60 months)										
1[23]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	no serious imprecision[16]	24	25	-	MD 0.06 higher[17,18,20]	VERY LOW
Change in signs and symptoms – sinuses (antituberculosis chemotherapy + surgery) (follow-up 36 months; number of patients with sinus and/or clinically evident abscesses on admission which had resolved during follow-up)										
1[11]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	very serious[9,13]	4/5 (80%)	2/2 (100%)	OR 0.60 (0.02 to 20.98)[10]	-	VERY LOW
Change in signs and symptoms – sinuses (antituberculosis chemotherapy + surgery) (follow-up 36 months; number of patients with new sinus and/or clinically evident abscesses that resolved during follow-up)										
1[11]	randomised trials	very serious[5,6,7]	serious[24]	very serious[8,12]	very serious[9,13]	1/1 (100%)	2/3 (66.7%)	OR 1.80 (0.04 to 79.43)[10]	12 more per 100 (from 59 fewer to 33 more)	VERY LOW
Change in signs and symptoms – nervous system involvement (antituberculosis chemotherapy + surgery) (follow-up 36 months; number of patients with nervous system involvement on admission which had resolved during follow-up)										
1[11]	randomised	very serious[5,6,7]	serious[24]	very serious[8,12]	very serious[9,13]	1/1	2/2	OR 0.60 (0.01	-	VERY LOW

Appendix E: GRADE profiles

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
	trials					(100%)	(100%)	to 49.45)[10]		

[1] *For full definition, see evidence tables in the appendices*
[4] *Method of randomisation unclear*
[5] *Allocation concealment unclear*
[6] *Blinding unclear*
[7] *Analysis does not follow the intent-to-treat principle*
[8] *Intervention does not exactly match the intervention of interest: regimens do not contain all of/just the 4 standard recommended drugs, and all patients underwent surgery in addition to receiving antituberculosis chemotherapy*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Medical Research Council Working Party on Tuberculosis of the Spine (Griffiths et al), 1986*
[12] *Population does not exactly match the population of interest: 6 of the 43 patients tested had single or combined drug resistance (Medical Research Council Working Party on Tuberculosis of the Spine (Griffiths et al) (1986) and Darbyshire (1999)), or some patients also had respiratory TB (Upadhyay et al (1986))*
[13] *Wide confidence intervals*
[14] *Patients in Medical Research Council Working Party on Tuberculosis of the Spine (Griffiths et al) (1986) underwent surgery in addition to receiving antituberculosis chemotherapy*
[15] *Individual point estimates vary widely*
[16] *Authors did not give standard deviations or standard errors of the means; reviewer could not assess imprecision*
[17] *Authors did not give standard deviations or standard errors of the means; reviewer could not calculate 95% confidence intervals*
[18] *Mean difference calculated by reviewer*
[19] *Calculated by the reviewer*
[20] *The authors state that the differences in the change from baseline to 36 months are unlikely to be due to the different durations of treatment because they occurred mainly in the first 6 months - that is, when there was no difference between the regimens of the two groups*
[21] *Upadhyay et al, 1986*
[22] *Mean difference and 95% confidence intervals calculated by reviewer*
[23] *Darbyshire, 1999*
[24] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
Abbreviations: CI, confidence interval; MD, mean difference; OR, odds ratio

Response to treatment

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment – favourable response (antituberculosis chemotherapy + surgery) (follow-up 60 months; number of patients who had a 'favourable' response to treatment[12])										
1[1]	randomised trials	very serious[4,5,6]	serious[12]	very serious[7,8,13]	very serious[9,10]	23/24 (95.8%)	25/26 (96.2%)	OR 0.92 (0.05 to 15.58)[11]	0 fewer per 100 (from 41 fewer to 4 more)	VERY LOW
Response to treatment – unfavourable response (antituberculosis chemotherapy + surgery) (number of patients who had an unfavourable response to treatment that required additional chemotherapy and/or surgery during the 60-month follow-up)										
1[1]	randomised trials	very serious[4,5,6]	serious[12]	very serious[7,8,13]	very serious[9,10]	1/24 (4.2%)	1/26 (3.8%)	OR 1.09 (0.06 to 18.40)[11]	0 more per 100 (from 4 fewer to 39 more)	VERY LOW

[1] Darbyshire, 1999
[4] Allocation concealment unclear

Appendix E: GRADE profiles

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality

[5] Blinding unclear
[6] Analysis does not follow the intent-to-treat principle
[7] Intervention does not exactly match the intervention of interest: regimens do not contain all of/just the 4 standard recommended drugs, and all patients underwent surgery in addition to receiving antituberculosis chemotherapy
[8] Substitute for outcome of interest
[9] GRADE rule of thumb: <300 events
[10] Wide confidence intervals
[11] Odds ratio and 95% confidence intervals calculated by reviewer
[12] Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths
[13] Population does not exactly match the population of interest: 6 of the 43 patients tested had single or combined drug resistance
Abbreviations: CI, confidence interval; OR, odds ratio

Relapse

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Recurrence (antituberculosis chemotherapy + surgery) (follow-up minimum 10 years; number of patients to experience recurrence or reactivation of tuberculosis during follow-up)										
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency[12]	very serious[6,10,13]	very serious[8,11]	0/25 (0%)	0/26 (0%)	OR 1.04 (0.02 to 54.38)[9]	-	VERY LOW

[1] *Upadhyay et al, 1986*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Analysis does not follow the intent-to-treat principle*
[6] *Intervention does not exactly match the intervention of interest: regimens do not contain all of/just the 4 standard recommended drugs, and all patients underwent surgery in addition to receiving antituberculosis chemotherapy*
[7] *Substitute for outcome of interest*
[8] *GRADE rule of thumb: <300 events*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*
[10] *Population does not exactly match the population of interest: some patients also had respiratory TB*
[11] *Wide confidence intervals*
[12] *Unclear if follow-up was the same in each group*
[13] *Substitute for outcome of interest (relapse)*
Abbreviations: CI, confidence interval; OR, odds ratio

Adverse events

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	
Adverse events leading to treatment modification (antituberculosis chemotherapy + surgery) (follow-up for the full treatment period; number of patients to experience adverse events that led to modification of the allocated regimen)										
1[1]	randomised trials	serious[3,4]	no serious inconsistency	very serious[6,11]	very serious[7,8]	2/31 (6.5%)	0/29 (0%)	OR 5.00 (0.23 to 108.68)[9]	-	VERY LOW
Adverse events - any (antituberculosis chemotherapy + surgery) (follow-up minimum 10 years; number of patients to experience an adverse event)										
1[13]	randomised trials	serious[2,3,4]	no serious inconsistency	very serious[6,11]	very serious[7,8]	6/25 (24%)	5/26 (19.2%)	OR 1.33 (0.35 to 5.06)[9,14]	5 more per 100 (from 12 fewer to 35 more)	VERY LOW

[1] *Medical Research Council Working Party on Tuberculosis of the Spine (Griffiths et al), 1986*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[6] *Intervention does not exactly match the intervention of interest: regimens do not contain all of/just the 4 standard recommended drugs, and all patients underwent surgery in addition to receiving antituberculosis chemotherapy*
[7] *GRADE rule of thumb: <300 events*
[8] *Wide confidence intervals*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*
[11] *Population does not exactly match the population of interest: 6 of the 43 patients tested had single or combined drug resistance (Medical Research Council Working Party on Tuberculosis of the Spine (Griffiths et al) (1986) and Darbyshire (1999)), or some patients also had respiratory TB (Upadhyay et al (1986))*
[13] *Upadhyay et al, 1986*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality

[14] The authors note that the incidence of drug reactions is not related to the duration of chemotherapy because most of the adverse events were observed in the earlier period of drug therapy
Abbreviations: CI, confidence interval; OR, odds ratio

LYMPH NODE TB

6 MONTHS vs 9 MONTHS

Treatment success or failure

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Treatment success (number of patients to be defined as a treatment success after 5 years of follow-up (5-year actuarial remission rate)[1])										
1[2]	randomised trials	very serious[3,4,5,6]	no serious inconsistency	serious[7,11]	serious[8]	39/43 (90.7%)	47/48 (97.9%)	OR 0.21 (0.02 to 1.93)[9]	7 fewer per 100 (from 49 fewer to 1 more)	VERY LOW
Treatment failure (number of patients to be defined as a treatment failure at the end of treatment[1])										
1[2]	randomised trials	very serious[3,4,5,6]	no serious inconsistency	serious[7]	very serious[8,10]	2/43 (4.7%)	1/48 (2.1%)	OR 2.29 (0.20 to 26.22)[9]	3 more per 100 (from 2 fewer to 34 more)	VERY LOW

[1] For full definitions, see evidence tables in the appendices
[2] Yuen et al, 1997
[3] Method of randomisation unclear
[4] Allocation concealment unclear
[5] Blinding unclear
[6] Analyses did not follow the intent-to-treat principle
[7] Regimens does not contain all of or just the 4 standard recommended drugs
[8] GRADE rule of thumb: <300 events
[9] Odds ratio and 95% confidence intervals calculated by reviewer
[10] Wide confidence intervals
[11] Doses not consistent with those listed in the British National Formulary
Abbreviations: CI, confidence interval; OR, odds ratio

Change in signs and symptoms

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Change in signs and symptoms – residual nodes (follow-up 30 months; number of patients with residual nodes)										
1[1]	randomised trials	very serious[2,3,4]	serious[10]	serious[5]	serious[6]	10/58 (17.2%)	16/107 (15%)[7]	OR 1.18 (0.50 to 2.81)[8]	2 more per 100 (from 7 fewer to 18 more)	VERY LOW
Change in signs and symptoms – node enlargement (follow-up 30 months; number of patients with nodes that had enlarged in size)										
1[1]	randomised trials	very serious[2,3,4]	serious[10]	serious[5]	serious[6]	4/58 (6.9%)	8/107 (7.5%)[7]	OR 0.81 (0.24 to 2.77)[8]	1 fewer per 100 (from 6 fewer to 11 more)	VERY LOW
Change in signs and symptoms – sinuses (follow-up 30 months; number of patients with new sinuses)										
1[1]	randomised trials	very serious[2,3,4]	serious[10]	serious[5]	very serious[6,9]	2/58 (3.4%)	3/107 (2.8%)[7]	OR 1.24 (0.20 to 7.63)[8]	1 more per 100 (from 2 fewer to 15 more)	VERY LOW
Change in signs and symptoms – glands (follow-up 30 months; number of patients with new glands)										
1[1]	randomised trials	very serious[2,3,4]	serious[10]	serious[5]	serious[6]	2/58 (3.4%)	7/107 (6.5%)[7]	OR 0.51 (0.10 to 2.54)[8]	3 fewer per 100 (from 6 fewer to 9 more)	VERY LOW

[1] *Campbell et al, 1993*
[2] *Method of randomisation unclear*
[3] *Blinding unclear*
[4] *Analyses did not follow the intent-to-treat principle*
[5] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs*
[6] *GRADE rule of thumb: <300 events*
[7] *Data for multiple groups pooled by reviewer*
[8] *Odds ratio and 95% confidence intervals calculated by reviewer*
[9] *Wide confidence intervals*
[10] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
Abbreviations: CI, confidence interval; OR, odds ratio

Relapse

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Relapse (number of patients to experience relapse during follow-up[1])										
2[2,3]	randomised trials	very serious[4,5,6,7]	no serious inconsistency	very serious[8,9,12,13]	serious[10]	14/158 (8.9%)	16/207 (7.7%)	OR 1.05 (0.49 to 2.26)[11,14]	0 more per 100 (from 4 fewer to 8 more)	VERY LOW

[1] *For full definitions, see evidence tables in the appendices*
[2] *Campbell et al, 1993*
[3] *Yuen et al, 1997*
[4] *Method of randomisation unclear*
[5] *Allocation concealment unclear in Campbell et al (1993)*

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
[6] *Blinding unclear* [7] *Analyses did not follow the intent-to-treat principle* [8] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs* [9] *Different combinations of drugs in each arm in Campbell et al (1993)* [10] *GRADE rule of thumb: <300 events* [11] *Odds ratio and 95% confidence intervals calculated by reviewer* [12] *Yuen et al, 1997: doses not consistent with those listed in the British National Formulary* [13] *Forest plot (relapse):* *Abbreviations: CI, confidence interval; OR, odds ratio*										

Adverse events

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Adverse events leading to treatment modification (number of patients to experience adverse events that led to modification of the allocated regimen)										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	serious[6,9]	serious[7]	4/49 (8.2%)	13/64 (20.3%)	OR 0.35 (0.11 to 1.15)[8]	12 fewer per 100 (from 18 fewer to 2 more)	VERY LOW
[1] *Yuen et al, 1997* [2] *Method of randomisation unclear* [3] *Allocation concealment unclear* [4] *Blinding unclear* [5] *Analyses did not follow the intent-to-treat principle* [6] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs* [7] *GRADE rule of thumb: <300 events* [8] *Odds ratio and 95% confidence intervals calculated by reviewer* [9] *Doses not consistent with those listed in the British National Formulary* *Abbreviations: CI, confidence interval; OR, odds ratio*										

Adherence and treatment default

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Treatment default (follow-up for the full treatment period; number of patients to default treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	serious[6,10]	very serious[7,8]	2/49 (4.1%)	3/64 (4.7%)	OR 0.87 (0.14 to 5.39)[9]	1 fewer per 100 (from 4 fewer to 16 more)	VERY LOW
[1] *Yuen et al, 1997* [2] *Method of randomisation unclear* [3] *Allocation concealment unclear* [4] *Blinding unclear* [5] *Analyses did not follow the intent-to-treat principle*										

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
[6] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs* [7] *GRADE rule of thumb: <300 events* [8] *Wide confidence intervals* [9] *Odds ratio and 95% confidence intervals calculated by reviewer* [10] *Doses not consistent with those listed in the British National Formulary* *Abbreviations: CI, confidence interval; OR, odds ratio*										

9 *months vs* ***>9*** *months*

Adverse events

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	>9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Adverse events - hepatotoxicity (number of patients to experience hepatotoxicity during treatment)										
2[1,2]	randomised trials	very serious[3,4,5,6]	no serious inconsistency	serious[7]	serious[8]	1/110 (0.91%)	3/109 (2.8%)	OR 0.33 (0.01 to 8.20)[9,10]	2 fewer per 100 (from 3 fewer to 16 more)	VERY LOW
[1] *Al-Aska et al, 1992* [2] *Campbell et al, 1985* [3] *Method of randomisation unclear* [4] *Allocation concealment unclear* [5] *Blinding unclear* [6] *Analysis in Campell et al (1985) did not follow the intent-to-treat principle* [7] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs* [8] *GRADE rule of thumb: <300 events* [9] *Odds ratio and 95% confidence intervals calculated by reviewer* [10] *Forest plot (hepatotoxicity):* *Abbreviations: CI, confidence interval; OR, odds ratio*											

9 *MONTHS vs* ***12*** *MONTHS*

Response to treatment

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	12 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - favourable response (number of patients to achieve a favourable outcome)										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	very serious[6,7,8]	serious[9]	30/34 (88.2%)	32/33 (97%)	OR 0.23 (0.02 to 2.22)[10]	9 fewer per 100 (from 58	VERY LOW

Appendix E: GRADE profiles

Quality assessment						Number of patients		Effect		
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**9 months**	**12 months**	**Relative (95% CI)**	**Absolute (95% CI)**	**Quality**
									fewer to 2 more)	

[1] *Al-Aska et al, 1992*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *No clear definition of the outcome*
[6] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs*
[7] *Different combinations of drugs in each arm*
[8] *Substitute for an outcome of interest*
[9] *GRADE rule of thumb: <300 events*
[10] *Odds ratio and 95% confidence intervals calculated by reviewer*
Abbreviations: CI, confidence interval; OR, odds ratio

Adverse events

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	12 months	Relative (95% CI)	Absolute (95% CI)	
Adverse events - hepatotoxicity (number of patients to experience hepatotoxicity)										
1[1]	randomised trials	very serious[2,3,4]	no serious inconsistency	serious[5,6]	very serious[7,8]	1/34 (2.9%)	2/33 (6.1%)	OR 0.47 (0.04 to 5.44)[9]	3 fewer per 100 (from 6 fewer to 20 more)	VERY LOW

[1] *Al-Aska et al, 1992*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs*
[6] *Different combinations of drugs in each arm*
[7] *GRADE rule of thumb: <300 events*
[8] *Wide confidence intervals*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*
Abbreviations: CI, confidence interval; OR, odds ratio

9 *MONTHS vs* ***18*** *MONTHS*

Response to treatment

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	18 months	Relative (95% CI)	Absolute (95% CI)	
Change in signs and symptoms – residual nodes (number of patients with residual nodes at the end of treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	serious[6]	very serious[7,8]	7/56 (12.5%)	3/57 (5.3%)	OR 2.57 (0.63 to 10.50)[9]	7 more per 100 (from 2 fewer to 32 more)	VERY LOW
Change in signs and symptoms – residual nodes (follow-up 36 months; number of patients with residual nodes during follow-up)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[10]	serious[6]	very serious[7]	2/56 (3.6%)	3/57 (5.3%)	OR 0.67 (0.11 to 4.15)[9]	2 fewer per 100 (from 5 fewer to 13 more)	VERY LOW
Change in signs and symptoms – fresh nodes (number of patients with fresh nodes during treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	serious[6]	serious[7]	5/56 (8.9%)	8/57 (14%)	OR 0.60 (0.18 to 1.96)[9]	5 fewer per 100 (from 11 fewer to 10 more)	VERY LOW
Change in signs and symptoms – fresh nodes (follow-up 36 months; number of patients with fresh nodes during follow-up)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[10]	serious[6]	very serious[7,8]	2/56 (3.6%)	0/57 (0%)	OR 5.28 (0.25 to 112.39)[9]	-	VERY LOW
Change in signs and symptoms – node enlargement (number of patients with nodes that had enlarged in size during treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	serious[6]	very serious[7,8]	8/56 (14.3%)	5/57 (8.8%)	OR 1.73 (0.53 to 5.66)[9]	5 more per 100 (from 4 fewer	VERY LOW

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	18 months	Relative (95% CI)	Absolute (95% CI)	Quality
									to 26 more)	
Change in signs and symptoms – node enlargement (follow-up 36 months; number of patients with nodes that had enlarged in size during follow-up)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[10]	serious[6]	very serious[7,8]	6/56 (10.7%)	4/57 (7%)	OR 1.59 (0.42 to 5.97)[9]	4 more per 100 (from 4 fewer to 24 more)	VERY LOW
Change in signs and symptoms – sinuses (number of patients with new sinuses during treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	serious[6]	serious[7]	0/56 (0%)	3/57 (5.3%)	OR 0.14 (0.01 to 2.73)[9]	4 fewer per 100 (from 5 fewer to 8 more)	VERY LOW
Change in signs and symptoms – sinuses (follow-up 36 months; number of patients with new sinuses during follow-up)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[10]	serious[6]	very serious[7,8]	0/56 (0%)	0/57 (0%)	OR 1.02 (0.02 to 52.18)[9]	-	VERY LOW

[1] *Campbell et al, 1985*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Analysis did not follow the intent-to-treat principle*
[6] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs*
[7] *GRADE rule of thumb: <300 events*
[8] *Wide confidence intervals*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*
[10] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths*
Abbreviations: CI, confidence interval; OR, odds ratio

Response to treatment

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	18 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - need for surgical intervention (number of patients needing surgical intervention (e.g. aspiration of pus) during treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	very serious[6,7]	serious[8]	4/56 (7.1%)	6/57 (10.5%)	OR 0.65 (0.17 to 2.45)[9]	3 fewer per 100 (from 9 fewer to 12 more)	VERY LOW
Response to treatment - need for surgical intervention (follow-up 36 months; number of patients needing surgical intervention (e.g. aspiration of pus) during follow-up)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[11]	very serious[6,7]	very serious[8,10]	4/56 (7.1%)	6/57 (10.5%)	OR 0.65 (0.17 to 2.45)[9]	3 fewer per 100 (from 9 fewer to 12 more)	VERY LOW

[1] *Campbell et al, 1985*
[2] *Method of randomisation unclear*
[3] *Allocation concealment unclear*
[4] *Blinding unclear*
[5] *Analysis did not follow the intent-to-treat principle*

Appendix E: GRADE profiles

Quality assessment						Number of patients		Effect		
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**9 months**	**18 months**	**Relative (95% CI)**	**Absolute (95% CI)**	**Quality**
[6] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs* [7] *Substitute for an outcome of interest* [8] *GRADE rule of thumb: <300 events* [9] *Odds ratio and 95% confidence intervals calculated by reviewer* [10] *Wide confidence intervals* [11] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths* *Abbreviations: CI, confidence interval; OR, odds ratio*										

Relapse

Quality assessment						Number of patients		Effect		
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**9 months**	**18 months**	**Relative (95% CI)**	**Absolute (95% CI)**	**Quality**
Relapse (follow-up 5 years; number of patients to experience clinical or microbiological relapse during follow-up)										
1[1]	randomised trials	very serious[2,3,4,5]	serious[10]	serious[6]	very serious[7,8]	0/34 (0%)	0/39 (0%)	OR 1.14 (0.02 to 59.26)[9]	-	VERY LOW
[1] *Campbell et al, 1988* [2] *Method of randomisation unclear* [3] *Allocation concealment unclear* [4] *Blinding unclear* [5] *Analysis did not follow the intent-to-treat principle* [6] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs* [7] *GRADE rule of thumb: <300 events* [8] *Wide confidence intervals* [9] *Odds ratio and 95% confidence intervals calculated by reviewer* [10] *Follow-up began from treatment initiation; therefore, as different durations of treatment were used, follow-up was for different lengths* *Abbreviations: CI, confidence interval; OR, odds ratio*										

Adverse events

Quality assessment						Number of patients		Effect		
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**9 months**	**18 months**	**Relative (95% CI)**	**Absolute (95% CI)**	**Quality**
Adverse events - hepatotoxicity (number of patients to experience hepatotoxicity during treatment)										
1[1]	randomised trials	very serious[2,3,4,5]	no serious inconsistency	serious[6]	very serious[7,8]	0/76 (0%)	1/76 (1.3%)	OR 0.33 (0.01 to 8.20)[9]	1 fewer per 100 (from 1 fewer to 9 more)	VERY LOW
[1] *Campbell et al, 1985* [2] *Method of randomisation unclear* [3] *Allocation concealment unclear* [4] *Blinding unclear* [5] *Analysis did not follow the intent-to-treat principle* [6] *Intervention does not contain all of/contains drugs other than the 4 standard recommended drugs* [7] *GRADE rule of thumb: <300 events*										

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	9 months	18 months	Relative (95% CI)	Absolute (95% CI)	Quality
[8] Wide confidence intervals [9] Odds ratio and 95% confidence intervals calculated by reviewer Abbreviations: CI, confidence interval; OR, odds ratio										

GASTROINTESTINAL TB

6 months vs 9 months

Response to treatment

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Response to treatment - complete response (follow-up 1 year after treatment completion; number of patients to achieve a complete response during follow-up[1])										
1[2]	randomised trials	serious[3,4]	no serious inconsistency	serious[5,9]	very serious[6,7]	42/45 (93.3%)	41/45 (91.1%)	OR 1.37 (0.29 to 6.48)[8]	2 more per 100 (from 16 fewer to 7 more)	VERY LOW
Response to treatment - need for additional treatment (follow-up for the full treatment period; number of patients to need additional chemotherapy due to incomplete response[1])										
1[2]	randomised trials	serious[3,4]	no serious inconsistency	serious[5,9]	very serious[6,7]	1/45 (2.2%)	0/45 (0%)	OR 3.07 (0.12 to 77.33)[8]	-	VERY LOW
Response to treatment - need for additional treatment (follow-up for the full treatment period; number of patients to need surgery due to incomplete response[1])										
1[2]	randomised trials	serious[3,4]	no serious inconsistency	serious[5,9]	very serious[6,7]	0/45 (0%)	0/45 (0%)	OR 1.00 (0.02 to 51.49)[8]	-	VERY LOW
[1] For full definitions, see evidence tables in the appendices [2] Park et al, 2009 [3] Allocation concealment unclear [4] Investigators not blinded, unclear if others were blinded [5] Substitute for an outcome of interest [6] GRADE rule of thumb: <300 events [7] Wide confidence intervals [8] Odds ratio and 95% confidence intervals calculated by reviewer [9] Doses not consistent with those listed in the British National Formulary Abbreviations: CI, confidence interval; OR, odds ratio										

Relapse

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Recurrence (follow-up 1 year after treatment completion; number of patients to experience recurrence during follow-up[1])										
1[2]	randomised trials	serious[3,4]	no serious inconsistency	very serious[8,9]	very serious[5,6]	1/45 (2.2%)	0/45 (0%)	OR 3.07 (0.12 to 77.33)[7]	-	VERY LOW
[1] For full definitions, see evidence tables in the appendices										

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality

[2] *Park et al, 2009*
[3] *Allocation concealment unclear*
[4] *Investigators not blinded, unclear if others were blinded*
[5] *GRADE rule of thumb: <300 events*
[6] *Wide confidence intervals*
[7] *Odds ratio and 95% confidence intervals calculated by reviewer*
[8] *Doses not consistent with those listed in the British National Formulary*
[9] *Substitute for an outcome of interest (relapse)*
Abbreviations: CI, confidence interval; OR, odds ratio

Adverse events

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	6 months	9 months	Relative (95% CI)	Absolute (95% CI)	Quality
Adverse events leading to treatment discontinuation (follow-up up to the full treatment period; number of patients to experience adverse events that led to treatment discontinuation)										
1[1]	randomised trials	serious[2,3]	no serious inconsistency	serious[6]	serious[4]	2/45 (4.4%)	4/45 (8.9%)	OR 0.48 (0.08 to 2.74)[5]	4 fewer per 100 (from 8 fewer to 12 more)	VERY LOW

[1] *Park et al, 2009*
[2] *Allocation concealment unclear*
[3] *Investigators not blinded, unclear if others were blinded*
[4] *GRADE rule of thumb: <300 events*
[5] *Odds ratio and 95% confidence intervals calculated by reviewer*
[6] *Doses not consistent with those listed in the British National Formulary*
Abbreviations: CI, confidence interval; OR, odds ratio

9 *months vs* ***15*** *months*

Response to treatment

Quality assessment						Number of patients		Effect		Quality
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**9 months**	**15 months**	**Relative (95% CI)**	**Absolute (95% CI)**	
Response to treatment - complete response (follow-up 23-34 months; number of patients to achieve a complete response during follow-up[1])										
1[2]	randomised trials	serious[3,4]	serious[5]	serious[6,11]	very serious[7,8]	22/22 (100%)	18/18 (100%)	OR 1.22 (0.02 to 64.31)[9]	-	VERY LOW
Response to treatment - complete response (follow-up 23-34 months; mean interval (months) to complete response[1]; better indicated by lower values)										
1[2]	randomised trials	serious[3,4]	serious[5]	serious[6,11]	very serious[8]	22	18	-	MD 0.9 lower (2.6 lower to 0.80 higher)[10]	VERY LOW

[1] *For full definitions, see evidence tables in the appendices*
[2] *Kim et al, 2003*
[3] *Allocation concealment unclear*
[4] *Blinding of participants and those administering care unclear*
[5] *Follow-up not equal*
[6] *Substitute for outcome of interest*
[7] *GRADE rule of thumb: <300 events*
[8] *Wide confidence intervals*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*
[10] *Mean difference and 95% confidence intervals calculated by reviewer*
[11] *Doses not consistent with those listed in the British National Formulary*
Abbreviations: CI, confidence interval; MD, mean difference; OR, odds ratio

Relapse

Quality assessment						Number of patients		Effect		Quality
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**9 months**	**15 months**	**Relative (95% CI)**	**Absolute (95% CI)**	
Recurrence (follow-up 23-34 months; number of patients to experience recurrence during follow-up[1])										
1[2]	randomised trials	serious[3,4]	serious[5]	very serious[9,10]	very serious[6,7]	0/22 (0%)	0/18 (0%)	OR 0.82 (0.02 to 43.48)[8]	-	VERY LOW

[1] *For full definitions, see evidence tables in the appendices*
[2] *Kim et al, 2003*
[3] *Allocation concealment unclear*
[4] *Blinding of participants and those administering care unclear*
[5] *Follow-up not equal*
[6] *GRADE rule of thumb: <300 events*
[7] *Wide confidence intervals*
[8] *Odds ratio and 95% confidence intervals calculated by reviewer*
[9] *Doses not consistent with those listed in the British National Formulary*
[10] *Substitute for an outcome of interest (relapse)*
Abbreviations: CI, confidence interval; OR, odds ratio

A.11 RQs O, R and X

A.11.1 Adjunctive surgery in the treatment of active PULMONARY tuberculosis

RANDOMISED CONTROLLED TRIALS

No randomised controlled trials identified

NON-RANDOMISED CONTROLLED TRIALS

No non-randomised controlled trials identified

OBSERVATIONAL STUDIES

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Mortality (follow-up unclear; assessed with: number of deaths)												
1[1]	observational studies[2]	very serious[3,4,5,6,7]	very serious[8,9,10,11]	very serious[12,13]	serious[14]	none	6/184 (3.3%)	3/48 (6.3%)	OR 0.51 (0.12 to 2.10)[15]	3 fewer per 100 (from 5 fewer to 6 more)	VERY LOW	
Cure (follow-up unclear; assessed with: number of patients to be classified as a cure)												
1[1]	observational studies[2]	very serious[3,4,5,6,7]	very serious[8,9,10,11]	very serious[12,13]	very serious[14,16]	none	175/184 (95.1%)	35/48 (72.9%)	OR 7.22 (2.87 to 18.20)[15]	22 more per 100 (from 16 more to 25 more)	VERY LOW	
Treatment failure (follow-up unclear; assessed with: number of patients who still had active tuberculosis)												
1[1]	observational studies[2]	very serious[3,4,5,6,7]	very serious[8,9,10,11]	very serious[12,13]	very serious[14,16]	none	3/184 (1.6%)	10/48 (20.8%)	OR 0.06 (0.02 to 0.24)[15]	19 fewer per 100 (from 15 fewer to 20 fewer)	VERY LOW	
Functionality – return to work (follow-up unclear; assessed with: number of patients who still had active tuberculosis)												
1[1]	observational studies[2]	very serious[3,4,5,6,7]	very serious[8,9,10,11]	very serious[12,13,17]	very serious[14,16]	none	3/184 (1.6%)	10/48 (20.8%)	OR 0.06 (0.02 to 0.24)[15]	19 fewer per 100 (from 15 fewer to 20 fewer)	VERY LOW	

[1] *Jaworski, 1972*
[2] *retrospective*
[3] *allocation based on qualification for surgery and subsequent agreement or refusal to undergo surgery by the patient*
[4] *blinding unclear, though unlikely*

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance

[5] *attempts do not appear to have been made to balance confounders*
[6] *unclear of length of follow-up appropriate*
[7] *unclear if precise and reliable definitions of outcome used (diagnostic criteria for â€˜cureâ€™ and the number of patients to still have active tuberculosis not provided)*
[8] *unclear if comparable at baseline*
[9] *unclear if groups received the same â€˜otherâ€™ care*
[10] *unclear if groups were followed for an equal period*
[11] *groups comparable for treatment completion and availability of outcome data*
[12] *some drug resistant cases were included*
[13] *unclear which antituberculosis drugs were used, or if same regimens were used in the 2 groups*
[14] *GRADE rule of thumb: <300 events*
[15] *Odds ratio and 95% confidence intervals calculated by reviewer*
[16] *Wide confidence intervals*
[17] *substitute for outcome of interest*

POST-OPERATIVE COMPLICATIONS

Jaworski, 1972

Pleural empyema with fistula = 6%

Exacerbations = 4.4%

Bleeding into the operated space requiring thoracotomy = 1.6%

Jaundice = 3.3%

Psychoses = 1.6%

Early death resulting from fibrinolytic shock = 1.1%

By type of surgery

Fewest complications were found after segmentectomies (20%), and the most after pneumonectomies (56.3%)

By duration of disease

The influence of duration of disease was not negligible, with complications found in 15.5% of patients ill for 1 to 5 years, and in 50% ill over 5 years

By susceptibility status

Complications were most frequent in in patients resistant to 3 or more drugs (81.1%), occurring in 22.7% of those resistant to 2 drugs and in 9% of those resistant to 1 drug

A.11.2 Adjunctive surgery in the treatment of active ENDOBRONCHIAL tuberculosis

RANDOMISED CONTROLLED TRIALS

No randomised controlled trials identified

NON-RANDOMISED CONTROLLED TRIALS

No non-randomised controlled trials identified

OBSERVATIONAL STUDIES

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Changes in signs and symptoms - improvement in endobronchial lesions (follow-up 9 months after initiation of treatment; assessed with: number of patients in whom lesions were improved (the number and/or volume of lesions reduced) or healed (lesions removed completely) after 16 weeks)												
1[1]	observational studies[2]	very serious[3,4,5]	no serious inconsistency	serious[6]	very serious[7,8]	none	41/41 (100%)	62/74 (83.8%)	OR 6.60 (0.97 to 288.09)[9]	13 more per 100 (from 0 fewer to 16 more)	VERY LOW	
Changes in signs and symptoms - deterioration in endobronchial lesions (follow-up 9 months after initiation of treatment; assessed with: number of patients in whom lesions had deteriorated (the number and/or volume of lesions had increased) at 16 weeks)												
1[1]	observational studies[2]	very serious[3,4,5]	no serious inconsistency	serious[6]	very serious[7,8]	none	0/41 (0%)	3/74 (4.1%)	OR 0.25 (0.01 to 4.88)[9]	3 fewer per 100 (from 4 fewer to 13 more)	VERY LOW	
Changes in signs and symptoms - recurrence of endobronchial lesions (follow-up 9 months after initiation of treatment; assessed with: number of patients in whom lesions and recurred after 9 months of follow-up)												
1[1]	observational studies[2]	very serious[3,4,5]	no serious inconsistency	serious[6]	very serious[7,8]	none	0/41 (0%)	0/74 (0%)	OR 1.80 (0.04 to 92.15)[9]	-	VERY LOW	

[1] *Jin et al, 2013*
[2] *'historical controlled trial'; retrospective observational*
[3] *allocation was based upon the time at which the patient was treated*
[4] *blinding unclear, though unlikely*
[5] *attempts do not appear to have been made to balance confounders*
[6] *may have been some drug resistant cases were included (only patients with disease resistant to a combination of rifampicin, isoniazid or ethambutol were excluded)*
[7] *GRADE rule of thumb: <300 events*
[8] *wide confidence intervals*
[9] *Odds ratio and 95% confidence intervals calculated by reviewer*

POST-OPERATIVE COMPLICATIONS

Jin et al, 2013

Laryngeal spasm = 1 (2.4%)
Cough = 35 (85.4%)
5–10 ml bleeding = 5 (12.2%)
Secondary pulmonary infection = 0
Esphagotrachea fistula = 0
Pneumothorax = 0
Trachea perforation = 0
Death = 0

A.11.3 Adjunctive surgery in the treatment of active CHEST WALL tuberculosis

RANDOMISED CONTROLLED TRIALS

No randomised controlled trials identified

NON-RANDOMISED CONTROLLED TRIALS

No non-randomised controlled trials identified

OBSERVATIONAL STUDIES

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Response to treatment - good outcome (follow-up unclear; assessed with: number of patients to have a good outcome)												
1[1]	observational studies	very serious[2,3,4,5,6]	very serious[7,8,9,10]	very serious[11,12]	very serious[13,14]	none	6/6 (100%)	1/1 (100%)	OR 4.33 (0.06 to 320.42)	-	VERY LOW	

[1] *Hsu et al, 1995*
[2] *allocation to treatment groups related to potential confounding factors*
[3] *unblinded*
[4] *attempts do not appear to have been made to balance confounders*
[5] *if unclear length of follow-up was appropriate*
[6] *definition of 'good outcome' not provided*
[7] *significant variation in age, size and location of the chest wall mass, radiography, extent of bone and cartilage involvement, and histological status*
[8] *groups received different combinations of antituberculosis drugs for treatment periods of different duration*
[9] *unclear if groups were followed for an equal period*
[10] *groups appear to be comparable for treatment completion and availability of outcome data*
[11] *antituberculosis regimens did not use all of or just the 4 standard recommended drugs, and the intervention and comparator arms varied by more than the presence or absence of surgery*
[12] *'good outcome' is a substitute for cure and/or treatment success, and perhaps the changes in signs and symptoms of disease*
[13] *GRADE rule of thumb: <300 events*
[14] *wide confidence intervals*

POST-OPERATIVE COMPLICATIONS

Hsu et al, 1995

No details provided

A.11.4 Adjunctive surgery in the treatment of active BONE AND JOINT tuberculosis

RANDOMISED CONTROLLED TRIALS

No randomised controlled trials identified

NON-RANDOMISED CONTROLLED TRIALS

No non-randomised controlled trials identified

OBSERVATIONAL STUDIES

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Changes in signs and symptoms – bony fusion (follow-up mean 15 years[1]; assessed with: number of patients to experience bony fusion/ankylosis)												
1[2]	observational studies[3]	very serious[4,5,6,7,8]	very serious[9,10,11]	very serious[12,13]	very serious[14,15]	none	4/15 (26.7%)	0/15 (0%)	OR 12.13 (0.59 to 248.50)[16]	-	⊙○○○ VERY LOW	
Changes in signs and symptoms – bony fusion (follow-up mean 29.3 months[17]; assessed with: number of patients to experience fusion of the sacroiliac joint, as assessed using plain radiographs and confirmed using CT or MRI scans)												
1[18]	observational studies[3]	very serious[5,6,7,8,19]	very serious[10,20]	very serious[12,21]	very serious[14,15]	none	6/12 (50%)	0/4 (0%)	OR 9.00 (0.40 to 203.31)[16]	-	⊙○○○ VERY LOW	
Changes in signs and symptoms – healing (follow-up mean 29.3 months[17]; assessed with: number of patients to heal[22])												
1[18]	observational studies[3]	very serious[5,6,7,8,19]	very serious[10,20]	very serious[12,21]	serious[14]	none	6/12 (50%)	4/4 (100%)	OR 0.11 (0.00 to 2.51)[16]	-	⊙○○○ VERY LOW	
Changes in signs and symptoms – healing (follow-up mean 29.3 months[17]; measured with: time to healing[22]; better indicated by lower values)												
1[18]	observational studies[3]	very serious[5,6,7,8,19]	very serious[10,20]	very serious[12,21]	serious[14]	none	12	4	-	MD 1.0 higher (0.9 lower to 2.9 higher)[23]	⊙○○○ VERY LOW	
Recurrence (follow-up mean 15 years[1]; assessed with: number of patients to experience recurrence)												
1[2]	observational studies[3]	very serious[4,5,6,7,8]	very serious[9,10,11]	very serious[12,13,24]	very serious[14,15]	none	4/15 (26.7%)	0/14 (0%)	OR 12.13 (0.59 to 248.50)[16]	-	⊙○○○ VERY LOW	

[1] *antituberculosis chemotherapy plus surgery = 13 years; conservative management = 17 years*
[2] *Chow & Yau, 1980*
[3] *retrospective - review of clinical records and collection of additional data via interview*
[4] *unclear if allocation to treatment groups related to potential confounding factors*
[5] *blinding unclear, though unlikely*
[6] *attempts do not appear to have been made to balance confounders*
[7] *length of follow-up was appropriate*

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		

[8] *outcome definitions were valid and precise*
[9] *50% of the surgical group were treated before the age of 20, whereas 80% of those treated conservatively were treated before the age of 20*
[10] *groups appeared to receive the same care apart from the intervention(s) studied, although details were limited*
[11] *mean follow-up in the surgical group was 13 years, mean follow-up amongst those treated conservatively was 17 years*
[12] *population appears to match the population of interest, although details were limited*
[13] *unclear if the intervention exactly matches the intervention of interest (details of the antituberculosis regimen(s) used not provided)*
[14] *GRADE rule of thumb: <300 events*
[15] *wide confidence intervals*
[16] *odds ratio and 95% confidence intervals calculated by reviewer*
[17] *antituberculosis chemotherapy plus surgery = 28.3; antituberculosis chemotherapy alone = 32.4*
[18] *Kim et al, 1999*
[19] *unclear if allocation to treatment groups related to potential confounding factors, although it appears not (all those that underwent surgery had more advanced disease)*
[20] *mean follow-up was longer in those that received antituberculosis chemotherapy alone*
[21] *antituberculosis regimens do not use all of or just the 4 standard recommended drugs (lacked pyrazinamide and contained streptomycin)*
[22] *criteria for healing: no pain or tenderness over the lesion site, no pain or discomfort during walking, a return to normal value of the erythrocyte sedimentation rate, disappearance of the abscess, clearance of sclerosis of the joint margin, and fusion of the sacroiliac joint*
[23] *mean difference and 95% confidence intervals calculated by reviewer*
[24] *outcome is a substitute for an outcome of interest*

POST-OPERATIVE COMPLICATIONS

Chow & Yau, 1980

No details provided

A.11.5 Adjunctive surgery in the treatment of active SPINAL tuberculosis

RANDOMISED CONTROLLED TRIALS

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
Mortality - TB-related (follow-up 10 years; assessed with: number of deaths associated with spinal tuberculosis)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	very serious[6,7]	none	4/100 (4%)	0/204 (0%)	OR 19.07 (1.02 to 357.83)[8]	-	VERY LOW	
Changes in signs and symptoms - complete bony fusion (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to experience complete bony fusion within 10-year follow-up)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	64/100 (64%)	127/204 (62.3%)	OR 1.08 (0.66 to 1.77)[8]	2 more per 100 (from 10 fewer to 12 more)	VERY LOW	
Changes in signs and symptoms - complete bony fusion (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to experience complete bony fusion within 10-year follow-up)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	64/100 (64%)	61/101 (60.4%)	OR 1.17 (0.66 to 2.06)[8]	4 more per 100 (from 10 fewer to 15 more)	VERY LOW	
Changes in signs and symptoms - partial bony fusion (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to experience partial bony fusion within 10-year follow-up)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	5/100 (5%)	21/204 (10.3%)	OR 0.46 (0.17 to 1.25)[8]	5 fewer per 100 (from 8 fewer to 2 more)	VERY LOW	
Changes in signs and symptoms - partial bony fusion (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to experience partial bony fusion within 10-year follow-up)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	5/100 (5%)	11/101 (10.9%)	OR 0.43 (0.17 to 1.25)[8]	6 fewer per 100 (from 9 fewer to 2 more)	VERY LOW	
Changes in signs and symptoms - no bony fusion (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to have no bony fusion within 10-year follow-up)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	2/100 (2%)	5/204 (2.5%)	OR 0.81 (0.15 to 4.26)[8]	0 fewer per 100 (from 2 fewer to 7 more)	VERY LOW	
Changes in signs and symptoms - no bony fusion (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to have no bony fusion within 10-year follow-up)												

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Antituberculosis chemotherapy plus surgery**	**Antituberculosis chemotherapy alone**	**Relative (95% CI)**	**Absolute**		
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	2/100 (2%)	3/101 (3%)	OR 0.67 (0.11 to 4.08)[8]	1 fewer per 100 (from 3 fewer to 8 more)	VERY LOW	
Changes in signs and symptoms - spontaneous bony fusion (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to experience spontaneous bony fusion within 10-year follow-up)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	1/100 (1%)	7/204 (3.4%)	OR 0.28 (0.03 to 2.34)[8]	2 fewer per 100 (from 3 fewer to 4 more)	VERY LOW	
Changes in signs and symptoms - kyphosis (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (measured with: mean angle of kyphosis at 10-year follow-up amongst patients with thoracic or thoracolumbar lesions; better indicated by lower values)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[9]	none	28	79	-	MD 3 lower (0 to 0 higher)[10]	VERY LOW	
Changes in signs and symptoms - kyphosis (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (measured with: mean angle of kyphosis at 10-year follow-up amongst patients with thoracic or thoracolumbar lesions; better indicated by lower values)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[9]	none	28	41	-	MD 6 lower (0 to 0 higher)[10]	VERY LOW	
Changes in signs and symptoms - improvement in kyphosis (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (follow-up 5 years; assessed with: number of patients to experience an improvement of 11° or more in their angle of kyphosis)												
1[11]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	very serious[6,7]	none	1/100 (1%)	2/204 (0.98%)	OR 1.02 (0.09 to 11.39)[8]	0 more per 100 (from 1 fewer to 9 more)	VERY LOW	
Changes in signs and symptoms - improvement in kyphosis (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (follow-up 5 years; assessed with: number of patients to experience an improvement of 11Â° or more in their angle of kyphosis)												
1[11]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	very serious[6,7]	none	1/100 (1%)	0/101 (0%)	OR 3.06 (0.12 to 76.03)[8]	-	VERY LOW	
Changes in signs and symptoms - deterioration in kyphosis (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (follow-up 5 years; assessed with: number of patients to experience an deterioration of 11° or more in their angle of kyphosis)												
1[11]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	13/100 (13%)	40/204 (19.6%)	OR 0.61 (0.31 to 1.21)[8]	7 fewer per 100 (from 13 fewer to 3 more)	VERY LOW	
Changes in signs and symptoms - deterioration in kyphosis (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (follow-up 5 years; assessed with: number of patients to experience an deterioration of 11° or more in their angle of kyphosis)												
1[11]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	13/100 (13%)	17/101 (16.8%)	OR 0.74 (0.34 to 1.61)[8]	4 fewer per 100 (from 10 fewer to 8	VERY LOW	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
										more)		
Changes in signs and symptoms - increase in kyphosis (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (measured with: mean increase in angle of kyphosis from baseline to 10-year follow-up amongst patients with thoracic or thoracolumbar lesions; better indicated by lower values)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[9]	none	28	79	-	MD 0 higher (0 to 0 higher)[10]	VERY LOW	
Changes in signs and symptoms - increase in kyphosis (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (measured with: mean increase in angle of kyphosis from baseline to 10-year follow-up amongst patients with thoracic or thoracolumbar lesions; better indicated by lower values)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[9]	none	28	41	-	MD 2 lower (0 to 0 higher)[10]	VERY LOW	
Changes in signs and symptoms - improvement in vertebral loss (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (follow-up 5 years; assessed with: number of patients to experience an improvement of 0.25 vertebrae or more in their vertebral loss)												
1[11]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	very serious[6,7]	none	5/100 (5%)	0/204 (0%)	OR 23.56 (1.29 to 430.36)[8]	-	VERY LOW	
Changes in signs and symptoms - deterioration in vertebral loss (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (follow-up 5 years; assessed with: number of patients to experience a deterioration of 0.25 vertebrae or more in their vertebral loss)												
1[11]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	24/100 (24%)	66/204 (32.4%)	OR 0.66 (0.38 to 1.14)[8]	8 fewer per 100 (from 17 fewer to 3 more)	VERY LOW	
Changes in signs and symptoms - deterioration in vertebral loss (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (follow-up 5 years; assessed with: number of patients to experience a deterioration of 0.25 vertebrae or more in their vertebral loss)												
1[11]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	24/100 (24%)	37/101 (36.6%)	OR 0.55 (0.3 to 1.01)[8]	13 fewer per 100 (from 22 fewer to 0 more)	VERY LOW	
Changes in signs and symptoms - increase in vertebral loss (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (measured with: mean increase in vertebral loss from baseline to 5 years; better indicated by lower values)												
1[12]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[9]	none	75	157	-	MD 0.11 lower (0 to 0 higher)[10]	VERY LOW	
Changes in signs and symptoms - increase in vertebral loss (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (measured with: mean increase in vertebral loss from baseline to 5 years; better indicated by lower values)												
1[12]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[9]	none	75	75	-	MD 0.16 lower (0 to 0 higher)[10]	VERY LOW	
Changes in signs and symptoms - myelopathy (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to experience residual myelopathy during 3-year follow-up)												
1[11]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	very serious[6,7]	none	2/100 (2%)	0/204 (0%)	OR 10.38 (0.49 to	-	VERY LOW	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
									218.3)[8]			
Changes in signs and symptoms - new sinuses and/or abscesses (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients in whom the sinuses and/or clinically evident abscesses developed during 5-year follow-up)												
1[13]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	21/100 (21%)	60/204 (29.4%)	OR 0.64 (0.36 to 1.13)[8]	8 fewer per 100 (from 16 fewer to 3 more)	VERY LOW	
Changes in signs and symptoms - reactivation of spinal lesions (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients in whom spinal lesions reactivated during 5-year follow-up)												
1[12]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	very serious[6,7]	none	0/100 (0%)	0/204 (0%)	OR 2.03 (0.04 to 103.30)[8]	-	VERY LOW	
Response to treatment - favourable status (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to achieve a favourable status during 10-year follow-up[14])												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	serious[5]	serious[6]	none	70/100 (70%)	151/204 (74%)	OR 0.82 (0.48 to 1.39)[8]	4 fewer per 100 (from 16 fewer to 6 more)	VERY LOW	
Response to treatment - favourable status (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to achieve a favourable status during 10-year follow-up[14])												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	very serious[5,15]	serious[6]	none	70/100 (70%)	73/101 (72.3%)	OR 0.90 (0.49 to 1.65)[8]	2 fewer per 100 (from 16 fewer to 9 more)	VERY LOW	
Response to treatment - need for additional intervention (antituberculosis chemotherapy alone = 6 months or 9 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to require additional chemotherapy or surgery during 10-year follow-up)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	very serious[5,15]	serious[6]	none	5/100 (5%)	6/204 (2.9%)	OR 1.74 (0.52 to 5.83)[8]	2 more per 100 (from 1 fewer to 12 more)	VERY LOW	
Response to treatment - need for additional intervention (antituberculosis chemotherapy alone = 6 months; antituberculosis chemotherapy in surgery group = 6 months) (assessed with: number of patients to require additional chemotherapy or surgery during 10-year follow-up)												
1[1]	randomised trials	serious[2,3,4]	no serious inconsistency	very serious[5,15]	serious[6]	none	5/100 (5%)	5/101 (5%)	OR 1.01 (0.28 to 3.6)[8]	0 more per 100 (from 4 fewer to 11 more)	VERY LOW	

[1] *Parthasarathy et al, 1999*
[2] *unclear if appropriate method of randomisation was used*
[3] *allocation concealment unclear*
[4] *blinding unclear*
[5] *antituberculosis regimens do not use all of the 4 standard recommended drugs, and the intervention and comparator differ by more than the presence of absence of surgery (some patients in the chemotherapy alone group received antituberculosis drugs for a longer period (duration of treatment = 6 or 9 months) than in the surgery group (duration of treatment = 6 months for all patients))*

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
[6] *GRADE rule of thumb: <300 events* [7] *wide confidence interval* [8] *odds ratio and 95% confidence intervals calculated by reviewer* [9] *insufficient data to assess imprecision* [10] *mean difference calculated by reviewer* [11] *Darbyshire, 1999* [12] *Reetha et al, 1994* [13] *Balasubramanian et al, 1994* [14] *for full definition, see evidence table* [15] *outcome is a substitute for an outcome of interest*												

NON-RANDOMISED CONTROLLED TRIALS

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
Changes in signs and symptoms - myelopathy (follow-up 27 were followed up for 5 years. 1 for 15 months and 1 for 12 months; assessed with: number of patients to experience myelopathy during long-term follow-up)												
1[1]	randomised trials	very serious[2,3,4]	serious[5]	very serious[6,7]	very serious[8,9]	none	0/21 (0%)	0/8 (0%)	OR 0.40 (0.01 to 21.58)[10]	-	VERY LOW	
Changes in signs and symptoms - sinuses (follow-up 27 were followed up for 5 years. 1 for 15 months and 1 for 12 months; assessed with: number of patients to develop a sinus during long-term follow-up)												
1[1]	randomised trials	very serious[2,3,4]	serious[5]	very serious[6,7]	very serious[8,9]	none	0/21 (0%)	0/8 (0%)	OR 0.40 (0.01 to 21.58)[10]	-	VERY LOW	
Changes in signs and symptoms - abscesses (follow-up 27 were followed up for 5 years. 1 for 15 months and 1 for 12 months; assessed with: number of patients to develop an abscess during long-term follow-up)												
1[1]	randomised trials	very serious[2,3,4]	serious[5]	very serious[6,7]	very serious[8,9]	none	0/21 (0%)	0/8 (0%)	OR 0.40 (0.01 to 21.58)[10]	-	VERY LOW	
Changes in signs and symptoms - limitation of physical activity (follow-up 27 were followed up for 5 years. 1 for 15 months and 1 for 12 months; assessed with: number of patients to experience limitation to their physical activity due to a spinal lesion during long-term follow-up)												
1[1]	randomised trials	very serious[2,3,4]	serious[5]	very serious[6,7]	very serious[8,9]	none	0/21 (0%)	0/8 (0%)	OR 0.40 (0.01 to 21.58)[10]	-	VERY LOW	
Changes in signs and symptoms - limitation of physical activity (surgery at any time) (follow-up 27 were followed up for 5 years. 1 for 15 months and 1 for 12 months; measured with: mean interval to becoming ambulant; better indicated by lower values)												
1[1]	randomised trials	very serious[2,3,4]	serious[5]	very serious[6,7]	serious[11]	none	21	8	-	MD 60 higher (0 to 0 higher)[12]	VERY LOW	
Changes in signs and symptoms - limitation of physical activity (surgery within 10 days of initiating antituberculosis chemotherapy) (follow-up 27 were followed up for 5 years. 1 for 15 months and 1 for 12 months; measured with: mean interval to becoming ambulant; better indicated by lower values)												
1[1]	randomised	very	serious[5]	very	serious[11]	none	18	8	-	MD 44	VERY	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
	trials	serious[2,3,4]		serious[6,7]						higher (0 to 0 higher)[12]	LOW	
Relapse (follow-up 27 were followed up for 5 years. 1 for 15 months and 1 for 12 months; assessed with: number of patients to experience relapse during long-term follow-up)												
1[1]	randomised trials	very serious[2,3,4]	serious[5]	very serious[6,7]	very serious[8,9]	none	0/21 (0%)	0/8 (0%)	OR 0.40 (0.01 to 21.58)[10]	-	VERY LOW	
Hospitalisation (surgery at any time) (follow-up 27 were followed up for 5 years. 1 for 15 months and 1 for 12 months; measured with: mean duration of hospital stay; better indicated by lower values)												
1[1]	randomised trials	very serious[2,3,4]	serious[5]	very serious[6,7,13]	serious[11]	none	21	8	-	MD 55 higher (0 to 0 higher)[12]	VERY LOW	
Hospitalisation (surgery within 10 days of initiating antituberculosis chemotherapy) (follow-up 27 were followed up for 5 years. 1 for 15 months and 1 for 12 months; measured with: mean duration of hospital stay; better indicated by lower values)												
1[1]	randomised trials	very serious[2,3,4]	serious[5]	very serious[6,7,13]	serious[11]	none	18	8	-	MD 3 higher (0 to 0 higher)[12]	VERY LOW	

[1] *Rajeswari et al, 1997*
[2] *only 23 of the 33 patients included underwent randomisation*
[3] *allocation concealment unclear*
[4] *blinding unclear*
[5] *unclear if groups were comparable at baseline, although all 3 patients who had sinuses at baseline were in the surgery group*
[6] *3 cases of drug resistance (1 to streptomycin, 1 to isoniazid and 1 to isoniazid and rifampicin)*
[7] *antituberculosis regimens do not use all of or just the 4 standard recommended drugs*
[8] *GRADE rule of thumb: <300 events*
[9] *wide confidence intervals*
[10] *odds ratio and 95% confidence interval calculated by reviewer*
[11] *insufficient data to assess imprecision*
[12] *mean difference calculate by reviewer*
[13] *outcome is a substitute for an outcome of interest*

OBSERVATIONAL STUDIES

Mortality

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Mortality (follow-up median 24 months; assessed with: number of deaths)												
1[1]	observational studies[2]	very serious[3]	serious[4]	serious[5]	very serious[6,7]	none	0/5 (0%)	0/7 (0%)	OR 1.36 (0.02 to 79.97)[8]	-	VERY LOW	
Mortality (follow-up at least 1 year amongst those who survived; assessed with: number of deaths)												

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Antituberculosis chemotherapy plus surgery**	**Antituberculosis chemotherapy alone**	**Relative (95% CI)**	**Absolute**		
1[9]	observational studies[2]	very serious[10]	very serious[11]	very serious[12]	very serious[6,7]	none	2/11 (18.2%)	0/9 (0%)	OR 5.00 (0.21 to 118.66)[8]	-	VERY LOW	
Mortality (follow-up unclear; assessed with: number of deaths)												
1[13]	observational studies[14]	very serious[15]	very serious[16]	serious[17]	serious[6]	none	1/22 (4.5%)	1/6 (16.7%)	OR 0.24 (0.01 to 4.5)[8]	12 fewer per 100 (from 16 fewer to 31 more)	VERY LOW	
Mortality - TB-related (follow-up unclear; assessed with: number of TB-related deaths)												
1[13]	observational studies[14]	very serious[15]	very serious[16]	serious[17]	serious[6]	none	0/22 (0%)	1/6 (16.7%)	OR 0.08 (0 to 2.28)[8]	15 fewer per 100 (from 17 fewer to 15 more)	VERY LOW	
Mortality - treatment-related (follow-up unclear; assessed with: number of treatment-related deaths)												
1[13]	observational studies[14]	very serious[15]	very serious[16]	serious[17]	serious[6]	none	1/22 (4.5%)	0/6 (0%)	OR 0.91 (0.03 to 25.06)[8]	-	VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance

[1] *Eisen et al, 2012*
[2] *retrospective*
[3] *allocation to treatment groups was related to potential confounding factors (decision to use surgery was based on the presence of cord compression with neurological manifestations or spinal instability); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; length of follow-up was appropriate; outcome definitions were valid and precise*
[4] *groups appeared to be comparable at baseline, although some baseline characteristics are not reported by group; unclear if groups received the same care apart from the intervention(s) studied; unclear if follow-up was equal between the groups; groups appear to be comparable for treatment completion and availability of outcome data*
[5] *population appears to match the population of interest; antituberculosis regimens do not use all of or just the 4 standard recommended drugs; a number of patients received second-line antituberculosis drugs; some patients in the surgery group received antituberculosis chemotherapy for more than 12 months, whereas all patients in the antituberculosis chemotherapy alone group received antituberculosis chemotherapy for 12 months; outcome is not a substitute or surrogate outcome*
[6] *GRADE rule of thumb: <300 events*
[7] *wide confidence interval*
[8] *odds ratio and 95% confidence interval calculated by reviewer*
[9] *Rezai et al, 1995*
[10] *allocation to treatment groups was related to potential confounding factors, since the majority of 'non-operative' patients did not meet the clinical criteria for surgical management, which were based on potentially confounding factors (clinical signs and symptoms, responsiveness to antituberculosis chemotherapy, non-adherence); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; length of follow-up was appropriate; study used precise definitions and reliable measures of outcom*
[11] *the majority of 'non-operative' patients did not meet the clinical criteria for surgical management, which were based on potentially confounding factors (clinical signs and symptoms, responsiveness to antituberculosis chemotherapy, non-adherence), whereas all patients in the 'operative' group met these criteria; the 'operative' group generally had disease of a higher grade of severity; the 'operative' group consisted of both males and females, whereas the 'non-operative' group was all-male; groups appeared to receive the same care apart from the intervention(s) studied, although bracing was undertaken for a longer period in those who did not undergo surgery; unclear if the groups were followed up for an equal time; 2 patients died in the surgery group and therefore did not complete treatment or follow-up, whereas no loss to follow-up occurred in the 'non-operative' group*
[12] *2 patients had drug-resistant strains of tuberculosis; duration of antituberculosis chemotherapy is not reported; 2 patients in the 'non-operative' group underwent aspiration - although this is an invasive technique, the authors do not consider it a surgical technique*
[13] *Richardson et al, 1976*
[14] *unclear if prospective or retrospective*
[15] *allocation to treatment groups appears to be related to potential confounding factors, since not all patients in the antituberculosis chemotherapy alone group met the criteria for surgery; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate; neurological 'improvement' was not defined*
[16] *unclear if the groups were comparable at baseline; unclear if the groups received the same care apart from the intervention(s) studied; unclear if the groups were followed up for an equal time; groups appear to be comparable for treatment completion and availability of outcome data*
[17] *population appears to match the population of interest, although details were limited; unclear if the intervention exactly matches the intervention of interest (details of the antituberculosis regimen(s) used not provided)*

Changes in signs and symptoms

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Changes in signs and symptoms - neurological improvement (follow-up unclear; assessed with: number of patients with neurological improvement)												
1[1]	observational studies[2]	very serious[3]	very serious[4]	serious[5]	very serious[6,7]	none	21/22 (95.5%)	3/6 (50%)	OR 21.00 (1.61 to 273.35)[8]	45 more per 100 (from 12 more to 50 more)	VERY LOW	
Changes in signs and symptoms - neurological status (follow-up at least 1 year amongst those who survived; assessed with: number of patients to improve or remain neurologically intact)												
1[9]	observational studies[10]	very serious[11]	very serious[12]	very serious[13]	very serious[6,7]	none	10/11 (90.9%)	9/9 (100%)	OR 0.37 (0.11 to 10.18)[8]	-	VERY LOW	
Changes in signs and symptoms - neurological status (follow-up unclear; assessed with: number of patients with improved neurological status)												
1[14]	observational studies[10]	very serious[15]	very serious[16]	serious[17]	serious[6]	none	1/5 (20%)	3/4 (75%)	OR 0.08 (0 to 1.95)[8]	56 fewer per 100 (from 75 fewer to 10 more)	VERY LOW	
Changes in signs and symptoms - neural recovery (follow-up (mean (range). years) = 2.6 (2–5); assessed with: number of patients to experience complete neural recovery)												
1[18]	observational studies	very serious[19]	very serious[20]	serious[21]	very serious[6,7]	none	13/20 (65%)	2/2 (100%)	OR 0.36 (0.02 to 8.53)[8]	-	VERY LOW	
Changes in signs and symptoms - residual deformity (follow-up median 24 months; assessed with: number of patients to experience residual deformity)												
1[22]	observational studies[10]	very serious[23]	serious[24]	serious[25]	very serious[6,7]	none	0/5 (0%)	1/7 (14.3%)	OR 0.39 (0.01 to 11.76)	8 fewer per 100 (from 14 fewer to 52 more)	VERY LOW	
Changes in signs and symptoms - kyphosis (follow-up mean 20.2 years; assessed with: number of patients to have kyphosis)												
1[26]	observational studies[10]	very serious[27]	very serious[4]	serious[5]	serious[6]	none	6/18 (33.3%)	8/8 (100%)	OR 0.03 (0 to 0.62)[8]	-	VERY LOW	
Changes in signs and symptoms - kyphosis (follow-up mean 20.2 years; measured with: mean angle of kyphosis; better indicated by lower values)												
1[26]	observational studies[10]	very serious[27]	very serious[4]	serious[5]	serious[28]	none	6	8	-	MD 31.1 lower (0 to 0 higher)[29]	VERY LOW	
Changes in signs and symptoms - kyphosis (all ages) (follow-up at least 24 months; measured with: mean angle of kyphosis at end of follow-up; better indicated by lower values)												
1[30]	observational studies[10]	very serious[31]	very serious[32]	very serious[33]	serious[28]	none	31	23	-	MD 10 lower (0 to 0 higher)[29]	VERY LOW	

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
Changes in signs and symptoms - kyphosis (adults) (follow-up at least 24 months; measured with: mean angle of kyphosis amongst the adults at end of follow-up; better indicated by lower values)												
1[30]	observational studies[10]	very serious[31]	very serious[32]	very serious[33]	serious[28]	none	26	13	-	MD 11 lower (0 to 0 higher)[29]	VERY LOW	
Changes in signs and symptoms - kyphosis (children) (follow-up at least 24 months; measured with: mean angle of kyphosis amongst the children at end of follow-up; better indicated by lower values)												
1[30]	observational studies[10]	very serious[31]	very serious[32]	very serious[33]	serious[28]	none	5	10	-	MD 1 higher (0 to 0 higher)[29]	VERY LOW	
Changes in signs and symptoms - change in kyphosis (follow-up at least 1 year amongst those who survived; measured with: change in mean angle of kyphosis from baseline to follow-up; better indicated by lower values)												
1[9]	observational studies[10]	very serious[11]	very serious[12]	very serious[13]	serious[28]	none	11	9	-	MD 11 lower (0 to 0 higher)[29]	VERY LOW	
Changes in signs and symptoms - change in kyphosis (all ages) (follow-up at least 24 months; measured with: change in mean angle of kyphosis at end of follow-up; better indicated by lower values)												
1[30]	observational studies[10]	very serious[31]	very serious[32]	very serious[33]	serious[28]	none	31	23	-	MD 13 lower (0 to 0 higher)[29]	VERY LOW	
Changes in signs and symptoms - change in kyphosis (adults) (follow-up at least 24 months; measured with: change in mean angle of kyphosis amongst adults at end of follow-up; better indicated by lower values)												
1[30]	observational studies[10]	very serious[31]	very serious[32]	very serious[33]	serious[28]	none	26	13	-	MD 15 lower (0 to 0 higher)[29]	VERY LOW	
Changes in signs and symptoms - change in kyphosis (children) (follow-up at least 24 months; measured with: change in mean angle of kyphosis amongst children at end of follow-up; better indicated by lower values)												
1[30]	observational studies[10]	very serious[31]	very serious[32]	very serious[33]	serious[28]	none	5	10	-	MD 1 lower (0 to 0 higher)[29]	VERY LOW	
Changes in signs and symptoms - improvement in kyphosis (follow-up at least 72 months; assessed with: number of patients to experience an improvement (decrease) in their angle of kyphosis)												
1[34]	observational studies[10]	very serious[35]	very serious[36]	serious[37]	serious[6]	none	4/30 (13.3%)	7/60 (11.7%)	OR 1.16 (0.31 to 4.34)[8]	2 more per 100 (from 8 fewer to 25 more)	VERY LOW	
Changes in signs and symptoms - deterioration in kyphosis (follow-up at least 72 months; assessed with: number of patients to experience moderate or severe deterioration (an increase of more than 11°) in their angle of kyphosis)												
1[34]	observational studies[10]	very serious[35]	very serious[36]	serious[37]	serious[6]	none	14/30 (46.7%)	34/60 (56.7%)	OR 0.67 (0.28 to	10 fewer per 100	VERY LOW	

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
									1.61)[8]	(from 30 fewer to 11 more)		
Changes in signs and symptoms - improvement in kyphosis (<16 years) (follow-up at least 72 months; assessed with: number of patients below the age of 16 to experience an improvement (decrease) in their angle of kyphosis)												
1[34]	observational studies[10]	very serious[35]	very serious[36]	serious[37]	very serious[6,7]	none	4/7 (57.1%)	6/30 (20%)	OR 5.33 (0.93 to 30.51)[8]	37 more per 100 (from 1 fewer to 68 more)	VERY LOW	
Changes in signs and symptoms - deterioration in kyphosis (<16 years) (follow-up at least 72 months; assessed with: number of patients below the age of 16 to experience moderate or severe deterioration (an increase of more than 11°) in their angle of kyphosis)												
1[34]	observational studies[10]	very serious[35]	very serious[36]	serious[37]	serious[6]	none	3/7 (42.9%)	17/30 (56.7%)	OR 0.57 (0.11 to 3.02)[8]	14 fewer per 100 (from 44 fewer to 23 more)	VERY LOW	
Changes in signs and symptoms - improvement in kyphosis (>16 years) (follow-up at least 72 months; assessed with: number of patients aged 16 and above to experience an improvement (decrease) in their angle of kyphosis)												
1[34]	observational studies[10]	very serious[35]	very serious[36]	serious[37]	very serious[6,7]	none	0/23 (0%)	1/30 (3.3%)	OR 0.42 (0.02 to 10.75)[8]	2 fewer per 100 (from 3 fewer to 24 more)	VERY LOW	
Changes in signs and symptoms - deterioration in kyphosis (>16 years) (follow-up at least 72 months; assessed with: number of patients aged 16 and above to experience moderate or severe deterioration (an increase of more than 11°) in their angle of kyphosis)												
1[34]	observational studies[10]	very serious[35]	very serious[36]	serious[37]	serious[6]	none	11/23 (47.8%)	17/30 (56.7%)	OR 0.70 (0.24 to 2.09)[8]	9 fewer per 100 (from 33 fewer to 17 more)	VERY LOW	
Changes in signs and symptoms - spinal fusion (follow-up unclear; assessed with: number of patients with spinal fusion)												
1[1]	observational studies[2]	very serious[3]	very serious[4]	serious[5]	very serious[6,7]	none	22/22 (100%)	3/6 (50%)	OR 45.00 (1.89 to 1071.38)[8]	48 more per 100 (from 15 more to 50 more)	VERY LOW	
Changes in signs and symptoms - fusion (follow-up mean 20.2 years; assessed with: number of patients to experience radiographic fusion)												
1[26]	observational studies[10]	very serious[27]	very serious[4]	serious[5]	very serious[6,7]	none	18/18 (100%)	8/8 (100%)	OR 2.18 (0.04 to 119.22)[8]	-	VERY LOW	
Changes in signs and symptoms - fusion (all ages) (follow-up at least 24 months; assessed with: number of patients to experience intracorporeal fusion)												
1[30]	observational studies[10]	very serious[31]	very serious[32]	very serious[33]	very serious[6,7]	none	26/31 (83.9%)	15/23 (65.2%)	OR 2.77 (0.77 to 10.03)[8]	19 more per 100 (from 6 fewer to 30 more)	VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Changes in signs and symptoms - fusion (adults) (follow-up at least 24 months; assessed with: number of adult patients to experience intracorporeal fusion)												
1[30]	observational studies[10]	very serious[31]	very serious[32]	very serious[33]	very serious[6,7]	none	26/26 (100%)	13/13 (100%)	OR 1.96 (0.04 to 104.47)[8]	-	VERY LOW	
Changes in signs and symptoms - fusion (children) (follow-up at least 24 months; assessed with: number of children to experience intracorporeal fusion)												
1[30]	observational studies[10]	very serious[31]	very serious[32]	very serious[33]	very serious[6,7]	none	0/5 (0%)	2/10 (20%)	OR 0.31 (0.01 to 7.74)[8]	13 fewer per 100 (from 20 fewer to 46 more)	VERY LOW	
Changes in signs and symptoms - pain (follow-up at least 1 year amongst those who survived; assessed with: number of patients with persistent pain)												
1[9]	observational studies[10]	very serious[11]	very serious[12]	very serious[13]	serious[6]	none	0/11 (0%)	2/9 (22.2%)	OR 0.13 (0.01 to 3.11)[8]	19 fewer per 100 (from 22 fewer to 25 more)	VERY LOW	
Changes in signs and symptoms - functional independence (follow-up unclear; measured with: mean change in measure of functional independence; better indicated by higher values)												
1[14]	observational studies[10]	very serious[15]	very serious[16]	serious[17]	serious[7]	none	5	4	-	MD 0.50 higher (16.06 lower to 11.66 higher)[29]	VERY LOW	
Changes in signs and symptoms - functional independence (self-care) (follow-up unclear; measured with: mean change in self-care score; better indicated by higher values)												
1[14]	observational studies[10]	very serious[15]	very serious[16]	serious[17]	serious[7]	none	5	4	-	MD 5.5 lower (17.46 lower to 6.46 higher)[29]	VERY LOW	
Changes in signs and symptoms - functional independence (mobility) (follow-up unclear; measured with: mean change in mobility and transfer score; better indicated by higher values)												
1[14]	observational studies[10]	very serious[15]	very serious[16]	serious[17]	serious[7]	none	5	4	-	MD 3.00 higher (0.64 lower to 6.64 higher)[29]	VERY LOW	
Changes in signs and symptoms - functional independence (locomotion) (follow-up unclear; measured with: mean change in locomotion score; better indicated by higher values)												
1[14]	observational studies[10]	very serious[15]	very serious[16]	serious[17]	serious[7]	none	5	4	-	MD 0.20 lower (2.16 lower to 1.76 higher)[29]	VERY LOW	
Changes in signs and symptoms - walking ability (follow-up unclear; assessed with: number of patients able to walk on discharge)												
1[14]	observational studies[10]	very serious[15]	very serious[16]	serious[17]	very serious[6,7]	none	3/5 (60%)	3/4 (75%)	OR 0.50 (0.03 to	15 fewer per 100	VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
									8.95)[8]	(from 67 fewer to 21 more)		

[1] *Richardson et al, 1976*
[2] *unclear if prospective or retrospective*
[3] *allocation to treatment groups appears to be related to potential confounding factors, since not all patients in the antituberculosis chemotherapy alone group met the criteria for surgery; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate; neurological 'improvement' was not defined*
[4] *unclear if the groups were comparable at baseline; unclear if the groups received the same care apart from the intervention(s) studied; unclear if the groups were followed up for an equal time; groups appear to be comparable for treatment completion and availability of outcome data*
[5] *population appears to match the population of interest, although details were limited; unclear if the intervention exactly matches the intervention of interest (details of the antituberculosis regimen(s) used not provided)*
[6] *GRADE rule of thumb: <300 events*
[7] *wide confidence interval*
[8] *odds ratio and 95% confidence interval calculated by reviewer*
[9] *Rezai et al, 1995*
[10] *retrospective*
[11] *allocation to treatment groups was related to potential confounding factors, since the majority of 'non-operative' patients did not meet the clinical criteria for surgical management, which were based on potentially confounding factors (clinical signs and symptoms, responsiveness to antituberculosis chemotherapy, non-adherence); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; length of follow-up was appropriate; study used precise definitions and reliable measures of outcom*
[12] *the majority of 'non-operative' patients did not meet the clinical criteria for surgical management, which were based on potentially confounding factors (clinical signs and symptoms, responsiveness to antituberculosis chemotherapy, non-adherence), whereas all patients in the 'operative' group met these criteria; the 'operative' group generally had disease of a higher grade of severity; the 'operative' group consisted of both males and females, whereas the 'non-operative' group was all-male; groups appeared to receive the same care apart from the intervention(s) studied, although bracing was undertaken for a longer period in those who did not undergo surgery; unclear if the groups were followed up for an equal time; 2 patients died in the surgery group and therefore did not complete treatment or follow-up, whereas no loss to follow-up occurred in the 'non-operative' group*
[13] *2 patients had drug-resistant strains of tuberculosis; duration of antituberculosis chemotherapy is not reported; 2 patients in the 'non-operative' group underwent aspiration - although this is an invasive technique, the authors do not consider it a surgical technique*
[14] *Zaoui et al, 2012*
[15] *allocation to treatment groups related to potential confounding factors (allocation to surgery was based upon the presence of compressive abscess with neurological complications); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate; outcome definitions were valid and precise*
[16] *more patients that underwent surgery had complete neurological impairment; groups appeared to receive the same care apart from the intervention(s) studied, although details were limited; unclear of groups were followed up for an equal period; groups appear to be comparable for treatment completion and availability of outcome data*
[17] *population appears to match the population of interest, although details were limited; unclear if the intervention exactly matches the intervention of interest (details of the antituberculosis regimen(s) used not provided)*
[18] *Kumar et al, 2007*
[19] *allocation to treatment groups was related to potential confounding factors (decision to operate was based upon presence of extradural granuloma (19 patients), although 1 of the 3 patients without extradural granuloma, all of whom had intramedullary lesions, also underwent surgery - the indication for surgery in this patient is not reported); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; length of follow-up was appropriate; outcome definition unclear*
[20] *unclear if groups were comparable at baseline; groups appeared to receive the same care apart from the intervention(s) studied, although details were limited; unclear if follow-up was equal between the groups; groups appear to be comparable for treatment completion and availability of outcome data*
[21] *population appears to match the population of interest, although details were limited; unclear if the intervention exactly matches the intervention of interest (details of the antituberculosis regimen(s) used not provided); outcome not a substitute or surrogate outcome*
[22] *Eisen et al, 2012*
[23] *allocation to treatment groups was related to potential confounding factors (decision to use surgery was based on the presence of cord compression with neurological manifestations or spinal instability); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; length of follow-up was appropriate; outcome definitions were valid and precise*
[24] *groups appeared to be comparable at baseline, although some baseline characteristics are not reported by group; unclear if groups received the same care apart from the intervention(s) studied; unclear if follow-up was equal between the groups; groups appear to be comparable for treatment completion and availability of outcome data*
[25] *population appears to match the population of interest; antituberculosis regimens do not use all of or just the 4 standard recommended drugs; a number of patients received second-line antituberculosis drugs; some patients in the surgery group received antituberculosis chemotherapy for more than 12 months, whereas all patients in the antituberculosis chemotherapy alone*

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
group received antituberculosis chemotherapy for 12 months; outcome is not a substitute or surrogate outcome [26] *Pun et al, 1990* [27] *unclear if allocation to treatment groups related to potential confounding factors; blinding unclear, though unlikely; unclear if attempts were made to balance confounders; length of follow-up was appropriate; outcome definitions were valid and precise* [28] *insufficient data to assess imprecision* [29] *mean difference (and 95% confidence interval, where possible) calculated by reviewer* [30] *Moon et al, 2007* [31] *allocation to treatment groups was related to potential confounding factors (decision to operate was based upon clinical signs and symptoms); blinding unclear, though unlikely; unclear if attempts were made to balance confounders; length of follow-up was appropriate; outcome definitions were valid and precise* [32] *groups not comparable at baseline (angle of kyphosis higher in the surgical group (13.2 vs 12.6; adults: 13 vs 9; children: 14 vs 12)); unclear if groups received the same care apart from the intervention(s) studied; unclear if groups were followed up for an equal period; groups appear to be comparable for treatment completion and availability of outcome data* [33] *population appears to match the population of interest, although details were limited; antituberculosis regimens do not use all of the 4 standard recommended drugs, and some surgeries were undertaken for diagnostic rather than therapeutic purposes* [34] *Rajasekaran et al, 1987* [35] *unclear if allocation to treatment groups related to potential confounding factors; blinding unclear, though unlikely; unclear if attempts were made to balance confounders; length of follow-up was appropriate; outcome definitions were valid and precise* [36] *antituberculosis chemotherapy alone has significantly more patients <16 years of age than the surgery group (no further details are available for the groups' characteristics at baseline); unclear if the groups received the same care apart from the intervention(s) studied; unclear if the groups were followed for an equal period; groups appear to be comparable for treatment completion and availability of outcome data* [37] *population appears to match the population of interest, although details were limited; the antituberculosis regimens do not use all of the 4 standard recommended drugs, and the intervention and comparator differ by more than the presence of absence of surgery (some patients in the chemotherapy alone group received antituberculosis drugs for a longer period (duration of treatment = 6 or 9 months) than in the surgery group (duration of treatment = 6 months for all patients)); no substitute or surrogate outcomes were used*											

Response to treatment

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
Response to treatment - favourable (follow-up ranged from 1.5 to 3 years in the antituberculosis chemotherapy alone group. and from 1 month to 3 years in the group that also underwent surgery)												
1[1]	observational studies[2]	serious[3]	very serious[4]	very serious[5]	very serious[6,7]	none	19/20 (95%)	5/5 (100%)	OR 1.18 (0.04 to 33.27)[8]	-	VERY LOW	
Response to treatment - disease resolution (follow-up median 24 months; assessed with: number of patients in whom the disease fully resolved)												
1[9]	observational studies[10]	very serious[11]	serious[12]	very serious[13]	very serious[7,14]	none	5/5 (100%)	5/7 (71.4%)	OR 5.00 (0.19 to 130.03)	21 more per 100 (from 39 fewer to 28 more)	VERY LOW	
Response to treatment - hospitalisation (follow-up unclear; measured with: mean duration of hospitalisation; better indicated by lower values)												
1[15]	observational studies[2]	very serious[16]	very serious[17]	very serious[18,19]	serious[20]	none	22	6	-	MD 24.0 lower (0 to 0 higher)[21,2]	VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute [2]		
Response to treatment - hospitalisation (follow-up unclear; measured with: mean duration of hospitalisation; better indicated by higher values)												
1[23]	observational studies[10]	very serious[24]	very serious[25]	serious[18]	serious[7]	none	5	4	-	MD 4.00 higher (13.19 lower to 21.19 higher)[26]	VERY LOW	

[1] *Arthornthurasook, 1983*
[2] *unclear if prospective or retrospective*
[3] *unclear if allocation to treatment groups related to potential confounding factors; blinding unclear, though unlikely; unclear if attempts were made to balance confounders; length of follow-up was appropriate; outcome definitions were valid and precise*
[4] *unclear if the groups were comparable at baseline; unclear if groups received the same care apart from the intervention(s) studied; follow-up ranged from 1.5 to 3 years in the antituberculosis chemotherapy alone group, and from 1 month to 3 years in the group that also underwent surgery; unclear if groups were comparable for treatment completion and availability of outcome data*
[5] *population appears to match the population of interest, although details were limited; unclear if the intervention exactly matches the intervention of interest (details of the antituberculosis regimen(s) used not provided); 'response to treatment' is a substitute for cure / treatment success and changes in the signs and symptoms of the disease*
[6] *GRADE rule of thumb: 300 events*
[7] *wide confidence interval*
[8] *odds ratio and 95% confidence intervals calculated by reviewer*
[9] *Eisen et al, 2012*
[10] *retrospective*
[11] *allocation to treatment groups was related to potential confounding factors (decision to use surgery was based on the presence of cord compression with neurological manifestations or spinal instability); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; length of follow-up was appropriate; outcome definitions were valid and precise*
[12] *groups appeared to be comparable at baseline, although some baseline characteristics are not reported by group; unclear if groups received the same care apart from the intervention(s) studied; unclear if follow-up was equal between the groups; groups appear to be comparable for treatment completion and availability of outcome data*
[13] *population appears to match the population of interest; antituberculosis regimens do not use all of or just the 4 standard recommended drugs; a number of patients received second-line antituberculosis drugs; some patients in the surgery group received antituberculosis chemotherapy for more than 12 months, whereas all patients in the antituberculosis chemotherapy alone group received antituberculosis chemotherapy for 12 months; outcome is a substitute for an outcome of interest*
[14] *GRADE rule of thumb: <300 events*
[15] *Richardson et al, 1976*
[16] *allocation to treatment groups appears to be related to potential confounding factors, since not all patients in the antituberculosis chemotherapy alone group met the criteria for surgery; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate; neurological 'improvement' was not defined*
[17] *unclear if the groups were comparable at baseline; unclear if the groups received the same care apart from the intervention(s) studied; unclear if the groups were followed up for an equal time; groups appear to be comparable for treatment completion and availability of outcome data*
[18] *population appears to match the population of interest, although details were limited; unclear if the intervention exactly matches the intervention of interest (details of the antituberculosis regimen(s) used not provided)*
[19] *outcome is a substitute for an outcome of interest*
[20] *insufficient data to assess imprecision*
[21] *odds ratio and 95% confidence interval calculated by reviewer*
[22] *mean difference calculated by reviewer*
[23] *Zaoui et al, 2012*
[24] *allocation to treatment groups related to potential confounding factors (allocation to surgery was based upon the presence of compressive abscess with neurological complications); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate; outcome definitions were valid and precise*
[25] *more patients that underwent surgery had complete neurological impairment; groups appeared to receive the same care apart from the intervention(s) studied, although details were limited; unclear of groups were followed up for an equal period; groups appear to be comparable for treatment completion and availability of outcome data*
[26] *mean difference (and 95% confidence interval, where possible) calculated by reviewer*

Relapse

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Relapse (follow-up median 24 months; assessed with: number of patients to experience relapse)												
1[1]	observational studies[2]	very serious[3]	serious[4]	serious[5]	very serious[6,7]	none	1/5 (20%)	1/7 (14.3%)	OR 1.50 (0.07 to 31.58)[8]	6 more per 100 (from 13 fewer to 70 more)	VERY LOW	

[1] *Eisen et al, 2012*
[2] *retrospective*
[3] *allocation to treatment groups was related to potential confounding factors (decision to use surgery was based on the presence of cord compression with neurological manifestations or spinal instability); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; length of follow-up was appropriate; outcome definitions were valid and precise*
[4] *groups appeared to be comparable at baseline, although some baseline characteristics are not reported by group; unclear if groups received the same care apart from the intervention(s) studied; unclear if follow-up was equal between the groups; groups appear to be comparable for treatment completion and availability of outcome data*
[5] *population appears to match the population of interest; antituberculosis regimens do not use all of or just the 4 standard recommended drugs; a number of patients received second-line antituberculosis drugs; some patients in the surgery group received antituberculosis chemotherapy for more than 12 months, whereas all patients in the antituberculosis chemotherapy alone group received antituberculosis chemotherapy for 12 months; outcome is not a substitute or surrogate outcome*
[6] *GRADE rule of thumb: <300 events*
[7] *wide confidence interval*
[8] *odds ratio and 95% confidence interval calculated by reviewer*

POST-OPERATIVE COMPLICATIONS

Randomised controlled trials

ICMR/MRC, 1994a/4b/9a/9b

No details provided

Non-randomised controlled trials

Rajeswari et al, 1997

No details provided

Observational studies

Arthornthurasook, 1983

No details provided

Eisen et al, 2012

None

Kumar et al, 2007

No details provided

Moon et al, 2007

No details provided

Pun et al, 1990

No significant postoperative complications

Rajasekaran et al, 1987

No details provided

Rezai et al, 1995

No details provided

Richardson et al, 1976

Blood loss:

- excessive bleeding = 1
- mean blood loss:
  - adults = 380 ml
  - children = 80 ml
- need for transfusion = 5

Operative mortality = 1

Intraoperative neurological complications = 0

Wound infection = 1

Draining sinus tracts after chest tube removal = 2

Zaoui et al, 2012

Appendix E: GRADE profiles

No details provided

A.11.6 Adjunctive surgery in the treatment of active CENTRAL NERVOUS SYSTEM tuberculosis

RANDOMISED CONTROLLED TRIALS

No randomised controlled trials identified

NON-RANDOMISED CONTROLLED TRIALS

No non-randomised controlled trials identified

OBSERVATIONAL STUDIES

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
Mortality (follow-up minimum 1 year; assessed with: number of deaths)												
1[1]	observational studies[2]	very serious[3,4,5,6,7]	very serious[8,9,10]	serious[11]	serious[12]	none	3/28 (10.7%)	11/28 (39.3%)	OR 0.19 (0.04 to 0.77)[13]	28 fewer per 100 (from 6 fewer to 37 fewer)	⊙○○○ VERY LOW	
Changes in signs and symptoms – neurological sequelae (follow-up 1 years; assessed with: number of patients to experience neurological sequelae, including neurological deficit, cognitive impairment2, optic atrophy and/or motor deficit)												
1[14]	observational studies[15]	very serious[16,17,18]	very serious[19,20]	no serious indirectness	very serious[12,21]	none	9/12 (75%)	17/53 (32.1%)	OR 6.35 (1.52 to 26.50)[13]	43 more per 100 (from 10 more to 61 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – disability (follow-up minimum 1 year; assessed with: number of patients to experience disability)												
1[1]	observational studies[2]	very serious[3,4,5,6,7]	very serious[8,9,10]	serious[11]	serious[12]	none	16/28 (57.1%)	18/28 (64.3%)	OR 0.74 (0.25 to 2.17)[13]	7 fewer per 100 (from 33 fewer to 15 more)	⊙○○○ VERY LOW	
Changes in signs and symptoms – 'well' or minor physical abnormality (follow-up minimum 1 year; assessed with: number of patients to be considered 'well', or had a minor physical abnormality which did not interfere with his or her lifestyle)												
1[1]	observational studies[2]	very serious[3,4,5,6,7]	very serious[8,9,10]	serious[11]	very serious[12,21]	none	9/28 (32.1%)	2/28 (7.1%)	OR 6.16 (1.19 to 31.82)[13]	25 more per 100 (from 1 more to 64 more)	⊙○○○ VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
Response to treatment - poor outcome (follow-up unclear; assessed with: number of patients (stage II or III) to have a 'poor outcome' (severe neurologic deficit or death))												
1[22]	observational studies[2]	very serious[17,18,23,24,25]	very serious[26,27,28]	very serious[29,30]	serious[12]	none	85/147 (57.8%)	108/240 (45%)	OR 1.68 (1.11 to 2.54)[13]	13 more per 100 (from 3 more to 23 more)	⊙○○○ VERY LOW	
Response to treatment - poor outcome (stage II) (follow-up unclear; assessed with: number of patients with stage II disease to have a 'poor outcome' (severe neurologic deficit or death))												
1[22]	observational studies[2]	very serious[17,18,23,24,25]	very serious[26,27,28]	very serious[29,30]	serious[12]	none	17/54 (31.5%)	23/102 (22.5%)	OR 1.58 (0.75 to 3.30)[13]	9 more per 100 (from 5 fewer to 26 more)	⊙○○○ VERY LOW	
Response to treatment - poor outcome (stage III) (follow-up unclear; assessed with: number of patients with stage III disease to have a 'poor outcome' (severe neurologic deficit or death))												
1[22]	observational studies[2]	very serious[17,18,23,24,25]	very serious[26,27,28]	very serious[29,30]	serious[12]	none	68/93 (73.1%)	85/138 (61.6%)	OR 1.70 (0.97 to 3.01)[13]	12 more per 100 (from 1 fewer to 21 more)	⊙○○○ VERY LOW	
Response to treatment - poor outcome (follow-up 3 months; assessed with: number of patients to have a â€˜poor outcomeâ€™, as defined by death or a Barthel Index score of <12)												
1[31]	observational studies[15]	very serious[17,32,33,34]	very serious[20,35]	very serious[30,36]	very serious[12,21]	none	9/14 (64.3%)	11/35 (31.4%)	OR 3.93 (1.06 to 14.49)[13]	33 more per 100 (from 1 more to 55 more)	⊙○○○ VERY LOW	

[1] *Peacock & Deeny, 1984*
[2] *retrospective*
[3] *study did not explicitly report the question it addressed, therefore it is unclear if this was an appropriate and focussed question*
[4] *it is unclear if the same exclusion criteria was applied to cases and controls; cases and controls adequately differentiated*
[5] *unclear if measures taken to prevent knowledge of primary exposure from influencing case ascertainment*
[6] *exposure status measured in a standard, valid and reliable way*
[7] *unclear if the main potential confounders were identified and taken into account in the design and analysis*
[8] *unclear if the cases and controls were taken from comparable populations, although they were matched for age and severity of disease*
[9] *it is unclear if the 2 groups were matched in terms of the participation rate*
[10] *unclear if follow-up was equal in the 2 groups*
[11] *unclear if the intervention exactly matches the interevention of interest (antituberculosis regimens were not reported)*
[12] *GRADE rule of thumb: <300 events*
[13] *odds ratio and 95% confidence intervals calculated by reviewer*
[14] *Kalita et al, 2007*
[15] *prospective*
[16] *allocation to receive shunt was based on clinical status*
[17] *blinding unclear, though unlikely*
[18] *attempts do not appear to have been made to balance confounders*
[19] *those that received shunt were selected due to the presence of hydrocephalus and raised intracranial pressure, so groups not balanced at baseline*
[20] *unclear if groups received the same care apart from the intervention(s) studied*
[21] *wide confidence intervals*
[22] *Lee, 2000*
[23] *unclear if allocation to treatment groups was related to potential confounding factors*

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
[24] unclear if length of follow-up was appropriate [25] â€˜poor outcomeâ€™ defined only as the incident of severe neurologic deficit or death (â€˜severe neurologic deficitâ€™ not defined) [26] unclear if groups were comparable at baseline [27] unclear if the 2 groups received antituberculosis drugs in the same doses for the same durations [28] unclear if groups were followed for an equal period [29] antituberculosis regimens did not use all of or just the 4 standard recommended drugs, and the dosing and duration of the antituberculosis regimens was not reported [30] outcome is a substitute for an outcome of interest [31] Misra et al, 1996 [32] allocation to treatment groups related to potential confounding factors (allocation to receive shunt was based on presence of obstructive hydocephalus) [33] attempts were made to balance confounders, although this only benefits the p-value and z-statistic (odds ratio was calculated by the reviewer) [34] follow-up was only 3 months [35] those that received shunt were selected due to the presence of obstructive hydrocephalus, and therefore the groups were not comparable at baseline [36] duration of antituberculosis chemotherapy unclear, and children received streptomycin instead of ethambutol												

POST-OPERATIVE COMPLICATIONS

Kalita et al, 2007

No details provided

Lee, 2000

No details provided

Misra et al, 1996

Shunt surgery complications = 6 of 14

- obstruction = 2
- infection = 2
- slit ventricles = 2
- subdural haematoma = 1
- intracerebral haematoma = 1

Peacock & Deeny, 1984

No details provided

A.11.7 Adjunctive surgery in the treatment of active GENITOURINARY tuberculosis

RANDOMISED CONTROLLED TRIALS

No randomised controlled trials identified

NON-RANDOMISED CONTROLLED TRIALS

No non-randomised controlled trials identified

OBSERVATIONAL STUDIES

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Response to treatment - need for additional intervention (follow-up median (maximum), months = 34 (62); assessed with: number of patients in whom reconstructive surgery or nephrectomy was required)												
1[1]	observational studies[2]	very serious[3]	serious[4]	very serious[5]	serious[6]	none	39/47 (83%)[7]	30/37 (81.1%)[7]	OR 1.14 (0.37 to 3.49)[8]	2 more per 100 (from 20 fewer to 13 more)	VERY LOW	
Response to treatment - need for reconstructive surgery (follow-up median (maximum), months = 34 (62); assessed with: number of patients in whom reconstructive surgery was required)												
1[1]	observational studies[2]	very serious[3]	serious[4]	very serious[5]	very serious[6,9]	none	23/47 (48.9%)[7]	3/37 (8.1%)[7]	OR 10.86 (2.93 to 40.32)[8]	41 more per 100 (from 12 more to 70 more)	VERY LOW	
Response to treatment - need for nephrectomy (follow-up median (maximum), months = 34 (62); assessed with: number of patients in whom nephrectomy was required)												
1[1]	observational studies[2]	very serious[3]	serious[4]	very serious[5]	very serious[6,9]	none	16/47 (34%)[7]	27/37 (73%)[7]	OR 0.19 (0.07 to 0.49)[8]	39 fewer per 100 (from 16 fewer to 57 fewer)	VERY LOW	
Treatment failure (any surgery compared with no surgery) (follow-up 9 to 60 months; assessed with: number of patients to experience bacteriological failure)												
1[10]	observational studies[11]	very serious[12]	serious[13]	serious[14]	very serious[6,9]	none	0/74 (0%)	0/18 (0%)	OR 0.40 (0.01 to 20.42)[8]	-	VERY LOW	
Treatment failure (ablative surgery compared with no surgery) (follow-up 9 to 60 months; assessed with: number of patients to experience bacteriological failure)												
1[10]	observational studies[11]	very serious[12]	serious[13]	serious[14]	very serious[6,9]	none	0/45 (0%)	0/18 (0%)	OR 0.21 (0.00 to 11.19)[8]	-	VERY LOW	
Treatment failure (reconstructive surgery compared with no surgery) (follow-up 16 to 60 months; assessed with: number of patients to experience bacteriological failure)												
1[10]	observational studies[11]	very serious[12]	serious[13]	serious[14]	very serious[6,9]	none	0/29 (0%)	0/18 (0%)	OR 0.32 (0.01 to 17.37)[8]	-	VERY LOW	
Adverse events – drug toxicity leading to drug withdrawal (any surgery compared with no surgery) (follow-up 9 to 60 months; assessed with: number of patients to experience drug toxicity												

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
leading to withdrawal of drug (without change to duration of treatment))												
1[10]	observational studies[11]	very serious[12]	serious[13]	serious[14]	serious[6]	none	9/74 (12.2%)	2/18 (11.1%)	OR 1.11 (0.22 to 5.64)[8]	1 more per 100 (from 8 fewer to 30 more)	VERY LOW	
Adverse events – drug toxicity leading to drug withdrawal (ablative surgery compared with no surgery) (follow-up 9 to 60 months; assessed with: number of patients to experience drug toxicity leading to withdrawal of drug (without change to duration of treatment))												
1[10]	observational studies[11]	very serious[12]	serious[13]	serious[14]	very serious[6,9]	none	5/45 (11.1%)	2/18 (11.1%)	OR 1.00 (0.10 to 9.75)[8]	0 fewer per 100 (from 10 fewer to 44 more)	VERY LOW	
Adverse events – drug toxicity leading to drug withdrawal (reconstructive surgery compared with no surgery) (follow-up 16 to 60 months; assessed with: number of patients to experience drug toxicity leading to withdrawal of drug (without change to duration of treatment))												
1[10]	observational studies[11]	very serious[12]	serious[13]	serious[14]	very serious[6,9]	none	4/29 (13.8%)	2/18 (11.1%)	OR 1.28 (0.12 to 13.17)[8]	3 more per 100 (from 10 fewer to 51 more)	VERY LOW	
Adherence (any surgery compared with no surgery) (follow-up 9 to 60 months; assessed with: number of patients to default treatment)												
1[10]	observational studies[11]	very serious[12]	serious[13]	serious[14]	very serious[6,9]	none	1/74 (1.4%)	1/18 (5.6%)	OR 0.38 (0.02 to 6.34)[8]	3 fewer per 100 (from 5 fewer to 22 more)	VERY LOW	

[1] *Shin et al, 2002*
[2] *prospective*
[3] *allocation was based upon the time at which the patient was treated; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders*
[4] *unclear if groups were comparable at baseline; groups appeared to received the same 'other' care, although details provided are limited; unclear if groups were followed for an equal period*
[5] *antituberculosis regimens do not use all of or just the 4 standard recommended drugs; substitute for an outcome of interest*
[6] *GRADE rule of thumbs: <300 events*
[7] *unit of analysis is at the renal unit-, rather then the patient-, level*
[8] *odds raio and 95% confidence interval calculated by reviewer*
[9] *wide confidence interval*
[10] *Wong et al, 1984*
[11] *unclear if prospective or retrospective*
[12] *unclear if method of allocation to treatment groups unrelated to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; definition for 'default' not provided, and only a loose definition provided for 'treatment failure'*
[13] *groups were comparable at baseline, although only details of age and sex were provided; groups appeared to received the same 'other' care, although details provided are limited; follow-up had a wide range within each group, though the ranges appeared to be comparable*
[14] *intervention varies by more than the presence or absence of surgery (duration of antituberculosis chemotherapy is longer amongst those patients that received surgery)*

POST-OPERATIVE COMPLICATIONS

Shin et al, 2002

No details provided

Wong et al, 1984

Chest infection = 5 (6.8%)

Wound infection = 2 (2.7%)

Pneumothorax requiring chest drainage = 2 (2.7%)

Haemorrhage from anastomosis = 1 (1.4%)

Burst abdomen = 1 (1.4%)

Intestinal obstruction owing to adhesion (late complication) = 1 (1.4%)

A.11.8 Adjunctive surgery in the treatment of active DRUG RESISTANT tuberculosis

RANDOMISED CONTROLLED TRIALS

No randomised controlled trials identified

NON-RANDOMISED CONTROLLED TRIALS

No non-randomised controlled trials identified

OBSERVATIONAL STUDIES

Mortality

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Mortality - all-cause (follow-up unclear; assessed with: number of deaths)												
1[1]	observational studies	very serious[2]	serious[3]	very serious[4]	very serious[5,6]	none	1/3 (33.3%)	1/5 (20%)	OR 22.00 (0.08 to 51.60)[7]	65 more per 100 (from 18 fewer to 73 more)	VERY LOW	
Mortality - all-cause (follow-up unclear; assessed with: number of deaths)												
1[8]	observational studies[9]	very serious[10]	serious[11]	very serious[12]	serious[6]	plausible confounding would change effect[13]	2/35 (5.7%)	12/107 (11.2%)	OR 0.48 (0.10 to 2.26)[7]	5 fewer per 100 (from 10 fewer to 11 more)	VERY LOW	
Mortality - all-cause (follow-up unclear; assessed with: number of deaths)												
1[14]	observational studies[15]	very serious[16]	serious[17]	very serious[18]	serious[6]	none	1/35 (2.9%)	9/120 (7.5%)	OR 0.39 (0.05 to 3.21)[7]	4 fewer per 100 (from 7 fewer to 13 more)	VERY LOW	
Mortality - all-cause (follow-up unclear; assessed with: number of deaths)												
1[19]	observational studies[15]	serious[20]	serious[21]	very serious[22]	very serious[5,6]	none	1/19 (5.3%)	13/185 (7%)	OR 0.74 (0.09 to 5.95)[7]	2 fewer per 100 (from 6 fewer to 24 more)	VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Mortality - all-cause (follow-up unclear; assessed with: number of deaths)												
1[23]	observational studies[15]	serious[24]	serious[25]	very serious[26]	serious[6]	none	5/66 (7.6%)	13/186 (7%)	OR 1.09 (0.37 to 3.19)[7]	1 more per 100 (from 4 fewer to 12 more)	VERY LOW	
Mortality - all-cause (patients aged 40 years or younger) (follow-up 3 to 7 years after treatment initiation; assessed with: number of deaths of any cause among patients aged 40 years or younger)												
1[27]	observational studies[15]	very serious[28]	serious[29]	very serious[30]	serious[6]	none	-	-	OR 0.53 (0.17 to 1.67)	-	VERY LOW	
Mortality - TB-related (patients aged 40 years or younger) (follow-up 3 to 7 years after treatment initiation; assessed with: number of TB-related deaths among patients aged 40 years or younger)												
1[27]	observational studies[15]	very serious[28]	serious[29]	very serious[30]	serious[6]	none	-	-	OR 0.67 (0.21 to 2.14)	-	VERY LOW	

[1] *Cameron & Harrison, 1997*
[2] *unclear if method of allocation was related to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*
[3] *the mean age in the surgery group was significantly older than in the group that received antituberculosis chemotherapy alone (41 vs 27 years); unclear if groups were comparable for other baseline characteristics; groups appeared to receive the same 'other' care, although details provided are limited; unclear if groups were followed for an equal period*
[4] *2 patients, both in the surgery group, had comorbidities that might affect the choice or management of treatment; the interventions used varied by more than the presence or absence of surgery - the regimens of antituberculosis chemotherapy contained, on average, more drugs in the surgery group (3.7 vs 2)*
[5] *wide confidence intervals*
[6] *GRADE rule of thumb: <300 events*
[7] *odds ratio and 95% confidence intervals calculated by reviewer*
[8] *Karagöz et al, 2009*
[9] *prospective cohort*
[10] *allocation to surgery was based on specific criteria (drug resistance with high probability of failure or relapse, sufficiently localized disease with adequate cardiopulmonary reserve and the availability of drugs with adequate efficacy to cause rapid healing of the bronchial stump); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*
[11] *unclear if groups were comparable at baseline; unclear if groups received the same 'other' care; unclear if follow-up was comparable across the groups*
[12] *some patients had comorbidities that may affect the choice or management of treatment (12% had diabetes mellitus and 21.8% had COPD); no females; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[13] *those that received surgery were selected due to a high likelihood of treatment failure or relapse; therefore it is likely that the reduced incidence of treatment failure in this group would be even lower if this confounding factor were not present*
[14] *Kwon et al, 2008*
[15] *retrospective cohort*
[16] *method of allocation to treatment groups was related to potential confounding factors (criteria for surgery: MDR-TB refractory to at least 6 months of medical treatment with a primary localized lesion); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders*
[17] *unclear if groups were comparable at baseline; groups appeared to receive the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*
[18] *some patients had a comorbidity that might affect the choice or management of antituberculosis treatment (15% diabetes mellitus, 5% chronic liver disease, 3% malignancy); it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[19] *Leimane et al, 2005*
[20] *unclear if method of allocation to treatment groups unrelated to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
[21] *unclear if groups were comparable at baseline; groups appeared to receive the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*												
[22] *some patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*												
[23] *Törün et al, 2007*												
[24] *allocation to surgery was based on specific criteria (resistance to a high number of drugs and therefore a high possibility of relapse or treatment failure; continued localised cavitary disease; destroyed lung, and only if they had relatively robust cardiopulmonary functions); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders*												
[25] *unclear if groups were comparable at baseline; groups appeared to receive the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*												
[26] *18.7 % of patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*												
[27] *Kim et al, 2008*												
[28] *unclear if method of allocation to treatment groups unrelated to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made in the study design or analysis to balance confounders*												
[29] *unclear if groups were comparable at baseline; groups appeared to receive the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*												
[30] *22.6% of patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*												

Cure

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Cure (follow-up unclear; assessed with: number of patients to be considered a cure, defined as negative smear and culture throughout treatment for at least 18 months (or 24 months, in the absence of first line drugs) and if only 1 positive culture was reported during that time and there was no concomitant evidence of deterioration, a patient may still be considered cured, provided that this positive culture was followed by a minimum of 3 consecutive negative cultures)												
1[1]	observational studies[2]	very serious[3]	serious[4]	very serious[5]	very serious[6,7]	plausible confounding would change effect[8]	31/35 (88.6%)	71/107 (66.4%)	OR 3.93 (1.29 to 11.99)[9]	22 more per 100 (from 5 more to 30 more)	VERY LOW	
Cure (follow-up unclear; assessed with: number of patients to achieve a cure, defined as a patient who has completed treatment and consistently had negative culture results (with at least 5 negative results) during the final 12 months of treatment)												
1[10]	observational studies[11]	very serious[12]	serious[13]	very serious[14]	serious[6]	plausible confounding would change effect[15]	26/35 (74.3%)	60/120 (50%)	OR 2.89 (1.25 to 6.68)[9]	24 more per 100 (from 6 more to 37 more)	VERY LOW	
Cure (follow-up unclear; assessed with: number of patients to achieve a cure, defined as patients who completed treatment and were *M. tuberculosis* culture negative for the last 12 months of treatment)												
1[16]	observational studies[11]	serious[17]	serious[13]	very serious[18]	serious[6]	none	1/19 (5.3%)	113/185 (61.1%)	OR 1.78 (0.62 to 5.17)[9]	13 more per 100 (from 12	VERY LOW	

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
										fewer to 28 more)		
Cure (follow-up unclear; assessed with: of patients to achieve a cure, defined as completion of treatment and at least 5 consecutive negative cultures from samples collected at least 30 days apart in the final 12 months)												
1[19]	observational studies[11]	serious[20]	serious[13]	very serious[21]	serious[6]	none	55/66 (83.3%)	138/186 (74.2%)	OR 1.50 (0.64 to 3.46)[22]	7 more per 100 (from 9 fewer to 17 more)	VERY LOW	

[1] *Karagöz et al, 2009*
[2] *prospective cohort*
[3] *allocation to surgery was based on specific criteria (drug resistance with high probability of failure or relapse, sufficiently localized disease with adequate cardiopulmonary reserve and the availability of drugs with adequate efficacy to cause rapid healing of the bronchial stump); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*
[4] *unclear if groups were comparable at baseline; unclear if groups received the same 'other' care; unclear if follow-up was comparable across the groups*
[5] *some patients had comorbidities that may affect the choice or management of treatment (12% had diabetes mellitus and 21.8% had COPD); no females; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[6] *GRADE rule of thumb: <300 events*
[7] *wide confidence intervals*
[8] *those that received surgery were selected due to a high likelihood of treatment failure or relapse; therefore it is likely that the higher incidence of cure in this group would be even higher if this confounding factor were not present*
[9] *odds ratio and 95% confidence intervals calculated by reviewer*
[10] *Kwon et al, 2008*
[11] *retrospective cohort*
[12] *method of allocation to treatment groups was related to potential confounding factors (criteria for surgery: MDR-TB refractory to at least 6 months of medical treatment with a primary localized lesion); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders*
[13] *unclear if groups were comparable at baseline; groups appeared to received the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*
[14] *some patients had a comorbidity that might affect the choice or management of antituberculosis treatment (15% diabetes mellitus, 5% chronic liver disease, 3% malignancy); it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[15] *those that received surgery were selected due to a high likelihood of treatment failure (criteria for performing surgery: MDR-TB was refractory to chemotherapy after at least 6 months of treatment); therefore it is likely that the higher incidence of cure in this group would be even higher if this confounding factor were not present*
[16] *Leimane et al, 2005*
[17] *unclear if method of allocation to treatment groups unrelated to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*
[18] *some patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[19] *Törün et al, 2007*
[20] *allocation to surgery was based on specific criteria (resistance to a high number of drugs and therefore a high possibility of relapse or treatment failure; continued localised cavitary disease; destroyed lung, and only if they had relatively robust cardiopulmonary functions); blinding unclear, though unlikely; attempts appear to have been made to balance confounders in the multivariate analysis*
[21] *18.7 % of patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[22] *multivariate analysis*

Treatment failure

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
Treatment failure (follow-up unclear; assessed with: number of patients to experience microbiological failure, defined as patients who failed to achieve three consecutive negative sputum cultures over at least a 3-month period)												
1[1]	observational studies[2]	very serious[3]	serious[4]	serious[5]	serious[6]	plausible confounding would change effect[7]	9/108 (8.3%)	16/54 (29.6%)	OR 0.22 (0.09 to 0.53)[8]	21 fewer per 100 (from 11 fewer to 26 fewer)	VERY LOW	
Treatment failure (follow-up unclear; assessed with: number of patients to be considered a treatment failure, defined as persistence of positive smear and culture despite treatment for 18-24 months)												
1[9]	observational studies[10]	very serious[11]	serious[12]	very serious[13]	very serious[6,14]	none	1/35 (2.9%)	9/107 (8.4%)	OR 0.32 (0.04 to 2.62)[8]	6 fewer per 100 (from 8 fewer to 11 more)	VERY LOW	
Treatment failure (follow-up unclear; assessed with: number of patients to experience treatment failure, defined as ≥2 positive culture results recorded during the final 12 months or a positive result of any 1 of the final 3 cultures)												
1[15]	observational studies[2]	very serious[16]	serious[17]	very serious[18]	serious[6]	none	3/35 (8.6%)	19/120 (15.8%)	OR 0.50 (0.14 to 1.79)[8]	7 fewer per 100 (from 13 fewer to 9 more)	VERY LOW	
Treatment failure (follow-up unclear; assessed with: number of patients to experience treatment failure, defined as patients with more than 1 positive *M. tuberculosis* culture during the past 12 months of treatment, those with 1 of their last 3 *M. tuberculosis* cultures positive, or those remaining persistently *M. tuberculosis* culture positive with treatment being stopped by their physician)												
1[19]	observational studies[2]	serious[20]	serious[17]	very serious[21]	serious[6]	none	1/19 (5.3%)	28/185 (15.1%)	OR 0.31 (0.04 to 2.43)[8]	10 fewer per 100 (from 14 fewer to 15 more)	VERY LOW	
Treatment failure (follow-up unclear; assessed with: number of patients to experience treatment failure, defined as 2 or more positive cultures amongst final 5 samples collected in the final 12 months of therapy, or if any 1 of the final 3 cultures were positive)												
1[22]	observational studies[2]	serious[23]	serious[17]	very serious[24]	serious[6]	none	2/66 (3%)	14/186 (7.5%)	OR 0.38 (0.08 to 1.74)[8]	5 fewer per 100 (from 7 fewer to 5 more)	VERY LOW	

[1] *Chan et al, 2004*
[2] *retrospective cohort*
[3] *allocation to surgery was broadly based on potential confounding factors (a high likelihood of medical failure based on extensive drug resistance, localized cavitary disease within a lobe or total destruction of one lung, and predictably adequate postoperative lung function), although the authors also state that because of the retrospective nature of the study, there were no rigid criteria for selection or exclusion for surgery; blinding unclear, though unlikely; attempts were not made within the design or analysis to balance the groups for potential confounders*
[4] *unclear if groups were comparable at baseline; unclear if the groups received the same care apart from the intervention(s) studied; unclear if follow-up was balanced between the groups; unclear of groups were comparable for treatment completion and availability of outcome data*
[5] *it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[6] *GRADE rule of thumb: <300 events*
[7] *those that received surgery were selected due to a high likelihood of treatment failure; therefore it is likely that the reduced incidence of treatment failure in this group would be even lower if this*

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		

confounding factor were not present
[8] *odds ratio and 95% confidence intervals calculated by reviewer*
[9] *Karagöz et al, 2009*
[10] *prospective cohort*
[11] *allocation to surgery was based on specific criteria (drug resistance with high probability of failure or relapse, sufficiently localized disease with adequate cardiopulmonary reserve and the availability of drugs with adequate efficacy to cause rapid healing of the bronchial stump); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*
[12] *unclear if groups were comparable at baseline; unclear if groups received the same 'other' care; unclear if follow-up was comparable across the groups*
[13] *some patients had comorbidities that may affect the choice or management of treatment (12% had diabetes mellitus and 21.8% had COPD); no females; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[14] *wide confidence intervals*
[15] *Kwon et al, 2008*
[16] *method of allocation to treatment groups was related to potential confounding factors (criteria for surgery: MDR-TB refractory to at least 6 months of medical treatment with a primary localized lesion); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders*
[17] *unclear if groups were comparable at baseline; groups appeared to received the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*
[18] *some patients had a comorbidity that might affect the choice or management of antituberculosis treatment (15% diabetes mellitus, 5% chronic liver disease, 3% malignancy); it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[19] *Leimane et al, 2005*
[20] *unclear if method of allocation to treatment groups unrelated to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*
[21] *some patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*
[22] *Törün et al, 2007*
[23] *allocation to surgery was based on specific criteria (resistance to a high number of drugs and therefore a high possibility of relapse or treatment failure; continued localised cavitary disease; destroyed lung, and only if they had relatively robust cardiopulmonary functions); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders*
[24] *18.7 % of patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*

Adherence

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Adherence (follow-up unclear; assessed with: number of patients to complete the intended course of therapy)												
1[1]	observational studies	very serious[2]	serious[3]	very serious[4]	very serious[5,6]	none	1/3 (33.3%)	2/5 (40%)	OR 0.75 (0.04 to 14.97)[7]	7 fewer per 100 (from 37 fewer to 51 more)	VERY LOW	

[1] *Cameron & Harrison, 1997*
[2] *unclear if method of allocation was related to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*
[3] *the mean age in the surgery group was significantly older than in the group that received antituberculosis chemotherapy alone (41 vs 27 years); unclear if groups were comparable for other*

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
baseline characteristics; groups appeared to received the same 'other' care, although details provided are limited; unclear if groups were followed for an equal period												
[4] *2 patients, both in the surgery group, had comorbidities that might effect the choice or management of treatment; the interventions used varied by more than the presence or absence of surgery - the regimens of antituberculosis chemotherapy contained, on average, more drugs in the surgery group (3.7 vs 2)*												
[5] *wide confidence intervals*												
[6] *GRADE rule of thumb: <300 events*												
[7] *odds ratio and 95% confidence intervals calculated by reviewer*												

Treatment default

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Adherence - default (follow-up unclear; assessed with: number of patients to be considered a defaulter, defined as failure to complete treatment for any reason)												
1[1]	observational studies[2]	very serious[3]	serious[4]	very serious[5]	serious[6]	none	1/35 (2.9%)	15/107 (14%)	OR 0.18 (0.02 to 1.42)[7]	11 fewer per 100 (from 14 fewer to 5 more)	VERY LOW	
Adherence - default (follow-up unclear; assessed with: number of patients to default on treatment, defined as patients who interrupted treatment for 2 or more consecutive months)												
1[8]	observational studies[9]	serious[10]	serious[11]	very serious[12]	serious[6]	none	1/19 (5.3%)	25/185 (13.5%)	OR 0.36 (0.05 to 2.78)[7]	8 fewer per 100 (from 13 fewer to 17 more)	VERY LOW	
Adherence - incomplete treatment (follow-up unclear; assessed with: number of patients to experience incomplete treatment, defined as treatment interrupted for 2 or more consecutive months for any reason)												
1[13]	observational studies[9]	serious[14]	serious[11]	very serious[15]	serious[6]	none	4/66 (6.1%)	21/186 (11.3%)	OR 0.51 (0.17 to 1.54)[7]	5 fewer per 100 (from 9 fewer to 5 more)	VERY LOW	
[1] *Karagöz et al, 2009*												
[2] *prospective cohort*												
[3] *allocation to surgery was based on specific criteria (drug resistance with high probability of failure or relapse, sufficiently localized disease with adequate cardiopulmonary reserve and the availability of drugs with adequate efficacy to cause rapid healing of the bronchial stump); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*												
[4] *unclear if groups were comparable at baseline; unclear if groups received the same 'other' care; unclear if follow-up was comparable across the groups*												
[5] *some patients had comorbidities that may affect the choice or management of treatment (12% had diabetes mellitus and 21.8% had COPD); no females; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*												
[6] *GRADE rule of thumb: <300 events*												
[7] *odds ratio and 95% confidence intervals calculated by reviewer*												
[8] *Leimane et al, 2005*												
[9] *retrospective cohort*												
[10] *unclear if method of allocation to treatment groups unrelated to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made to balance*												

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		

confounders; unclear if length of follow-up was appropriate

[11] *unclear if groups were comparable at baseline; groups appeared to received the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*

[12] *some patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*

[13] *Törün et al, 2007*

[14] *allocation to surgery was based on specific criteria (resistance to a high number of drugs and therefore a high possibility of relapse or treatment failure; continued localised cavitary disease; destroyed lung, and only if they had relatively robust cardiopulmonary functions); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders*

[15] *18.7 % of patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group*

Favourable response to treatment

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance
Favourable response to treatment (follow-up unclear; assessed with: number of patients to experience an initial favourable response, defined as patients with at least three consecutive negative sputum cultures over a period of at least 3 months while on treatment)												
1[1]	observational studies[2]	serious[3]	serious[4]	very serious[5]	serious[6]	plausible confounding would change effect[7]	99/108 (91.7%)	38/54 (70.4%)	OR 4.23 (1.28 to 13.93)[8]	21 more per 100 (from 5 more to 27 more)	VERY LOW	
Favourable response to treatment (follow-up unclear; assessed with: number of patients to experience a favourable outcome, defined as treatment completion or cure)												
1[9]	observational studies[2]	very serious[10]	serious[11]	very serious[12]	serious[6]	none	-	-	OR 1.24 (0.69 to 2.26)	-	VERY LOW	
Favourable response to treatment (follow-up unclear; assessed with: number of patients to achieve a favourable clinical response, defined as the disappearance of signs and symptoms associated with active tuberculosis, regression of chest radiograph shadowing, and 2 consecutive culture-negative sputum specimens collected 2 weeks apart)												
1[13]	observational studies	very serious[14]	serious[15]	very serious[16]	very serious[6,17]	none	2/3 (66.7%)	4/5 (80%)	OR 0.50 (0.02 to 12.90)[18]	13 fewer per 100 (from 73 fewer to 18 more)	VERY LOW	
Favourable response to treatment (follow-up 3 to 7 years after treatment initiation; assessed with: number of patients to experience treatment success, defined as the sum of cure, treatment completion, and short-term treatment completion[19])												
1[20]	observational studies[2]	very serious[21]	serious[22]	very serious[23]	very serious[6,17]	none	-	-	OR 3.87 (1.69 to 8.88)[24]	-	VERY LOW	
Favourable response to treatment (follow-up unclear; assessed with: number of patients to achieve a favourable outcome, defined as cure or treatment completion)												
1[25]	observational studies[2]	very serious[26]	serious[22]	very serious[27]	serious[6]	plausible confounding would change effect[28]	31/35 (88.6%)	71/120 (59.2%)	OR 11.35 (3.02 to 42.74)[24]	35 more per 100 (from 22 more to 39 more)	VERY LOW	

[1] *Chan et al, 2004*
[2] *retrospective cohort*
[3] *allocation to surgery was broadly based on potential confounding factors (a high likelihood of medical failure based on extensive drug resistance, localized cavitary disease within a lobe or total destruction of one lung, and predictably adequate postoperative lung function), although the authors also state that because of the retrospective nature of the study, there were no rigid criteria for selection or exclusion for surgery; blinding unclear, though unlikely; a stepwise selection procedure was used to create a multiple predictor model for the incidence of favourable response*
[4] *unclear if groups were comparable at baseline; unclear if the groups received the same care apart from the intervention(s) studied; unclear if follow-up was balanced between the groups; unclear of groups were comparable for treatment completion and availability of outcome data*
[5] *it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group; outcome is a substitute for outcomes of interest*
[6] *GRADE rule of thumb: <300 events*
[7] *those that received surgery were selected due to a high likelihood of treatment failure; therefore it is likely that the increased incidence of favourable response in this group would be even higher if this confounding factor were not present*
[8] *a stepwise selection procedure was used to create a multiple predictor model for the incidence of favourable response*
[9] *Keshajvee et al, 2008*
[10] *unclear if method of allocation to treatment groups unrelated to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate; 'favourable outcome' is defined as treatment completion or cure, but the definitions for treatment completion and cure are not provided*

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute	Quality	Importance

[11] *unclear if groups were comparable at baseline; groups appeared to received the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups comparable for treatment completion and availability of outcome data*
[12] *it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group; 'favourable outcome' is a composite of outcomes of interest*
[13] *Cameron & Harrison, 1997*
[14] *unclear if method of allocation was related to potential confounding factors; blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*
[15] *the mean age in the surgery group was significantly older than in the group that received antituberculosis chemotherapy alone (41 vs 27 years); unclear if groups were comparable for other baseline characteristics; groups appeared to received the same 'other' care, although details provided are limited; unclear if groups were followed for an equal period*
[16] *2 patients, both in the surgery group, had comorbidities that might effect the choice or management of treatment; the interventions used varied by more than the presence or absence of surgery - the regimens of antituberculosis chemotherapy contained, on average, more drugs in the surgery group (3.7 vs 2); outcome is a surrogate for outcomes of interest*
[17] *wide confidence intervals*
[18] *odds ratio and 95% confidence intervals calculated by reviewer*
[19] *see evidence table in the appendix for full definition*
[20] *Kim et al, 2008*
[21] *unclear if method of allocation to treatment groups unrelated to potential confounding factors; blinding unclear, though unlikely; attempts appear to have been made to balance confounders in the multivariate analysis*
[22] *unclear if groups were comparable at baseline; groups appeared to received the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*
[23] *22.6% of patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group; outcome is a substitute for outcomes of interest*
[24] *multivariate analysis*
[25] *Kwon et al, 2008*
[26] *method of allocation to treatment groups was related to potential confounding factors (criteria for surgery: MDR-TB refractory to at least 6 months of medical treatment with a primary localized lesion); blinding unclear, though unlikely; attempts appear to have been made to balance confounders in the multiple logistic regression*
[27] *some patients had a comorbidity that might affect the choice or management of antituberculosis treatment (15% diabetes mellitus, 5% chronic liver disease, 3% malignancy); it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group; substitute for outcomes of interest*
[28] *those that received surgery were selected due to a high likelihood of treatment failure (criteria for performing surgery: MDR-TB was refractory to chemotherapy after at least 6 months of treatment); therefore it is likely that the higher incidence of favourable outcomes in this group would be even higher if this confounding factor were not present*

Poor response to treatment

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		
Poor response to treatment (follow-up unclear; assessed with: number of patients to experience a poor outcome, defined as treatment failure, death during treatment or default)												
1[1]	observational studies[2]	serious[3]	serious[4]	very serious[5]	serious[6]	none	8/37 (21.6%)	171/343 (49.9%)	OR 0.28 (0.12 to 0.62)[7]	28 fewer per 100 (from 12 fewer to 39 fewer)	VERY LOW	
Poor response to treatment (follow-up unclear; assessed with: number of patients to experience a poor outcome, defined as treatment failure, death during treatment or default)												
1[8]	observational studies[9]	very serious[10]	serious[11]	very serious[12]	serious[6]	plausible confounding would change effect[13]	4/13 (30.8%)	110/129 (85.3%)	OR 0.18 (0.04 to 0.78)[14]	34 fewer per 100 (from 3 fewer to 66 fewer)	VERY LOW	
Poor response to treatment (follow-up unclear; assessed with: number of patients to experience treatment failure, defined as failure (defined as ≥2 of 5 positive culture results recorded during the final 12 months or any 1 of the final 3 cultures being positive), relapse (defined as a cured patient or a patient who completed therapy who resumed treatment 16 months after completion of the first treatment because of the emergence of MDR-tuberculous bacilli) or death)												
1[15]	observational studies[9]	very serious[16]	very serious[17]	very serious[18]	serious[6]	none	17/60 (28.3%)	48/137 (35%)	OR 0.73 (0.38 to 1.42)[19]	7 fewer per 100 (from 18 fewer to 8 more)	VERY LOW	
Poor response to treatment (follow-up unclear; assessed with: number of patients to experience a long-term poor outcome, defined as death, treatment failure or incomplete treatment)												
1[20]	observational studies[9]	serious[21]	serious[22]	very serious[23]	serious[6]	none	11/66 (16.7%)	48/186 (25.8%)	OR 0.58 (0.28 to 1.19)[19]	9 fewer per 100 (from 17 fewer to 3 more)	VERY LOW	

[1] *Geiga et al, 2012*
[2] *prospective cohort*
[3] *decision to perform surgical resection was made by the Georgian National TB Program's Drug Resistance Committee, and was dependent upon the presence sufficient pulmonary function to tolerate resection and a localised lesion amenable to resection were required; blinding unclear, though unlikely; a binary multivariable logistic regression model was used to evaluate the independent association of potential risk factors with poor outcome; unclear if the length of follow-up was appropriate*
[4] *unclear if groups were comparable at baseline; unclear if groups appeared to received the same 'other' care; unclear if the length of follow-up was comparable across the groups; unclear if groups comparable for treatment completion and availability of outcome data*
[5] *some patients had comorbidities that may affect the choice or management of treatment (e.g. 9% had diabetes mellitus); it is unclear if the 2 interventions varied by more than the presence or absence of surgery; in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group; 'poor outcome' is a substitute outcome*
[6] *GRADE rule of thumb: <300 events*
[7] *a binary multivariable logistic regression model was used to evaluate the independent association of potential risk factors with poor outcome*
[8] *Jeon et al, 2009*
[9] *retrospective cohort*
[10] *allocation to surgery was based on specific criteria (surgical resection was considered for patients with localised cavitary lesions and anticipated adequate postoperative lung function, and for selected patients with bilateral lesions if medical treatment had failed or was expected to fail); blinding unclear, though unlikely; binary logistic regression analysis was performed; unclear if the length of follow-up was appropriate*
[11] *unclear if groups were comparable at baseline; unclear if groups appeared to received the same 'other' care; unclear if follow-up was comparable across the groups; unclear if groups comparable for treatment completion and availability of outcome data*
[12] *some patients had comorbidities that may affect the choice or management of treatment (15% had diabetes mellitus); it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group; 'poor outcome' is a substitute for outcomes of*

Appendix E: GRADE profiles

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Antituberculosis chemotherapy plus surgery	Antituberculosis chemotherapy alone	Relative (95% CI)	Absolute		

interest

[13] *those that received surgery were selected due to a high likelihood of treatment failure or because they had already failed; therefore it is likely that the reduced incidence of poor outcome in this group would be even lower if this confounding factor were not present*

[14] *binary logistic regression analysis with the backward elimination method was performed for variables with p < 0.2 in the univariate analysis, which included the use of surgery, and the Hosmer-Lemeshow test was used for testing the goodness-of-fit of the models*

[15] *Kim et al, 2007*

[16] *method of allocation to treatment groups related to potential confounding factors (criteria for surgery: MDR-TB refractory to at least 6 months of medical treatment with a primary localized lesion); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders; unclear if length of follow-up was appropriate*

[17] *surgery was performed more frequently in patients with XDR-TB (p<0.001); unclear if groups received the same 'other' care; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*

[18] *34.1% of patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group; outcome is a substitute for outcomes of interest*

[19] *odds ratio and 95% confidence intervals calculated by reviewer*

[20] *Törün et al, 2007*

[21] *allocation to surgery was based on specific criteria (resistance to a high number of drugs and therefore a high possibility of relapse or treatment failure; continued localised cavitary disease; destroyed lung, and only if they had relatively robust cardiopulmonary functions); blinding unclear, though unlikely; attempts do not appear to have been made to balance confounders*

[22] *unclear if groups were comparable at baseline; groups appeared to received the same 'other' care, although details provided are limited; unclear if follow-up was comparable between the groups; unclear if groups were comparable for treatment completion and availability of outcome data*

[23] *18.7 % of patients had a comorbidity that might affect the choice or management of antituberculosis treatment; it is unclear if the 2 interventions varied by more than the presence or absence of surgery - in particular, there is insufficient detail around the precise regimens of antituberculosis chemotherapy used in each group; outcome is a substitute for outcomes of interest*

A.11.9 Adjunctive surgery in the treatment of active DRUG RESISTANT tuberculosis

Mortality – all-cause

Cure

Treatment failure

Poor response to treatment

A.12 RQ S

A.12.1 Any resistance

Age

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London Date: 2004	0-14 years	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	234	1.0 (0.3 to 3.4)	VERY LOW
	15-29 years					no serious imprecision		1.0 (0.8 to 1.6)	
	≥60 years					no serious imprecision		**0.6 (0.4 to 1.0)**	
	reference: 30-59 years					-		-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
[3] Analyses not reported for a number of variables recorded and reported in population characteristics [4] Unclear if loss to follow-up sufficiently unrelated to key characteristics Abbreviations: CI, confidence interval; OR, odds ratio									

Sex

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London Date: 2004	Male reference: female	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	234	1.0 (0.7 to 1.4) -	VERY LOW
Melzer, 2010 East London/ Essex Date: 2003-6	Female reference: male	Observational with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	380	0.70 (0.33 to 1.49) -	LOW
[1] Unclear if prognostic factor and outcome measurement blinded [2] Multivariate analysis used, but unclear which confounders were controlled for [3] Analyses not reported for a number of variables recorded and reported in population characteristics Abbreviations: CI, confidence interval; OR, odds ratio									

HIV status

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Melzer, 2010 East London/ Essex Date: 2003-6	HIV-positive reference: HIV-negative	Observational with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[5]	380	1.93 (0.70 to 5.23) -	VERY LOW
[1] Unclear if prognostic factor and outcome measurement blinded [2] Multivariate analysis used, but unclear which confounders were controlled for [3] Analyses not reported for a number of variables recorded and reported in population characteristics [4] Unclear if loss to follow-up sufficiently unrelated to key characteristics [5] Wide confidence interval Abbreviations: CI, confidence interval; OR, odds ratio									

Previous history of tuberculosis

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007	**Previous**	Observational with	very serious[1,2,3]	serious[4]	no serious	no serious	234	**3.0 (1.9 to 4.9)**	VERY LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
London Date: 2004	**history of tuberculosis**	multivariate analysis			indirectness	imprecision			
	reference: no history of disease							-	
Melzer, 2010 East London/ Essex Date: 2003-6	Previous treatment of tuberculosis	Observational with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[5]	380	1.53 (0.41 to 5.62)	VERY LOW
	reference: no history of treatment							-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

Exposure

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Melzer, 2010 East London/ Essex Date: 2003-6	Previous exposure to drug resistant tuberculosis	Observational with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[5]	380	12.84 (0.68 to 240.2)	VERY LOW
	reference: no previous exposure to drug resistant tuberculosis							-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

Place of birth

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Melzer, 2010 East London/ Essex Date: 2003-6	Country of origin with high incidence of drug resistance	Observational with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	380	0.61 (0.25 to 1.47)	LOW
	reference: country of origin without high incidence of drug resistance							-	
Melzer, 2010 East London/ Essex Date: 2003-6	Date of arrival in the UK ≥2000 i.e. less than 3-6 years in the UK	Observational with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	380	0.71 (0.27 to 1.87)	LOW
	reference: date of arrival in the UK <2000 i.e. more than 3-6 years in the UK							-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
Abbreviations: CI, confidence interval; OR, odds ratio

Ethnicity

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London Date: 2004	South Asian	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	234	1.0 (0.6 to 1.6)	VERY LOW
	Black African					no serious imprecision		1.3 (0.8 to 2.0)	
	Black Caribbean					serious[5]		**3.0 (1.2 to 7.7)**	
	Other					no serious imprecision		**1.9 (1.0 to 3.4)**	
	reference: white					-		-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Abbreviations: CI, confidence interval; OR, odds ratio									

Imprisonment

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London Date: 2004	**Prison** reference: not in prison	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	234	**3.0 (1.7 to 5.5)** -	VERY LOW
[1] Unclear if prognostic factor and outcome measurement blinded [2] Multivariate analysis used, but unclear which confounders were controlled for [3] Analyses not reported for a number of variables recorded and reported in population characteristics [4] Unclear if loss to follow-up sufficiently unrelated to key characteristics Abbreviations: CI, confidence interval; OR, odds ratio									

Homelessness

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London Date: 2004	**Homeless** reference: not homeless	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	234	**1.6 (1.1 to 2.2)** -	VERY LOW
[1] Unclear if prognostic factor and outcome measurement blinded [2] Multivariate analysis used, but unclear which confounders were controlled for [3] Analyses not reported for a number of variables recorded and reported in population characteristics [4] Unclear if loss to follow-up sufficiently unrelated to key characteristics Abbreviations: CI, confidence interval; OR, odds ratio									

.12.2 First-line drug resistance

Adherence

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Pritchard, 2003 Leicestershire	**Poor adherence** reference: no	Matched case-control[1] with multivariate	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	104	**4.8 (1.6 to 14.4)** -	VERY LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Data: 1993-8	evidence of poor adherence	analysis							

[1] Cases and controls were matched on ethnic group, gender and age group
[2] Unclear if prognostic factor and outcome measurement blinded
[3] Authors had to rely on others' notes (potential for recall bias)
[4] Multivariate analysis used, but unclear which confounders were controlled for
[5] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

Previous history of tuberculosis

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Pritchard, 2003 Leicestershire Data: 1993-8	**Previous history of tuberculosis**	Matched case-control[1] with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	104	**3.7 (1.2 to 11.8)**	VERY LOW
	reference: no history of tuberculosis							-	

[1] Cases and controls were matched on ethnic group, gender and age group
[2] Unclear if prognostic factor and outcome measurement blinded
[3] Authors had to rely on others' notes (potential for recall bias)
[4] Multivariate analysis used, but unclear which confounders were controlled for
[5] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

Site of disease

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Pritchard, 2003 Leicestershire Data: 1993-8	Extrapulmonary	Matched case-control[1] with multivariate analysis	very serious[1,2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	104	No statistic provided Authors state that the effect was not significant	VERY LOW
	reference: pulmonary							-	

[1] Cases and controls were matched on ethnic group, gender and age group
[2] Unclear if prognostic factor and outcome measurement blinded
[3] Authors had to rely on others' notes (potential for recall bias)
[4] Multivariate analysis used, but unclear which confounders were controlled for
[5] Insufficient data provided to assess imprecision
Abbreviations: CI, confidence interval; OR, odds ratio

Place of birth

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Pritchard, 2003 Leicestershire Data: 1993-8	Non-UK birth	Matched case-control[1] with multivariate analysis	very serious[1,2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	104	No statistic provided Authors state that the effect was not significant	VERY LOW
	reference: UK birth							-	

[1] Cases and controls were matched on ethnic group, gender and age group
[2] Unclear if prognostic factor and outcome measurement blinded
[3] Authors had to rely on others' notes (potential for recall bias)
[4] Multivariate analysis used, but unclear which confounders were controlled for
[5] Insufficient data provided to assess imprecision
Abbreviations: CI, confidence interval; OR, odds ratio

Foreign travel

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Pritchard, 2003 Leicestershire Data: 1993-8	Travel outside the UK	Matched case-control[1] with multivariate analysis	very serious[1,2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	104	No statistic provided Authors state that the effect was not significant	VERY LOW
	reference: no travel outside the UK							-	

[1] Cases and controls were matched on ethnic group, gender and age group
[2] Unclear if prognostic factor and outcome measurement blinded
[3] Authors had to rely on others' notes (potential for recall bias)
[4] Multivariate analysis used, but unclear which confounders were controlled for
[5] Insufficient data provided to assess imprecision
Abbreviations: CI, confidence interval; OR, odds ratio

Time in the UK

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Pritchard, 2003 Leicestershire Data: 1993-8	Recent immigration to the UK	Matched case-control[1] with multivariate analysis	very serious[1,2,3,4,5]	no serious inconsistency	no serious indirectness	serious[6]	104	No statistic provided Authors state that the effect was not significant	VERY LOW
	reference: no recent immigration to the UK							-	

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
[1] Cases and controls were matched on ethnic group, gender and age group [2] Unclear if prognostic factor and outcome measurement blinded [3] Authors had to rely on others' notes (potential for recall bias) [4] Multivariate analysis used, but unclear which confounders were controlled for [5] Effect estimate not reported [6] Insufficient data provided to assess imprecision Abbreviations: CI, confidence interval; OR, odds ratio									

A.12.3 Isoniazid resistance

Age

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Patients living in London									
Kruijshaar, 2008 London Data: 1998 and 2005	Age (linear)	Observational with multivariate analysis	serious[1]	serious[6,7]	no serious indirectness	no serious imprecision	11 848	**0.99 (0.98 to 0.99)**	LOW
Maguire, 2011 London Data: 1995 to the third quarter of 2006	0-14 years	Unmatched case-control with multivariate analysis	very serious[1,2,7,8]	no serious inconsistency	no serious indirectness	no serious imprecision	18040	0.30 (0.09 to 1.01)	LOW
	25-34 years					no serious imprecision		0.79 (0.52 to 1.20)	
	35-44 years					no serious imprecision		**0.64 (0.41 to 1.00)**	
	45-64 years					no serious imprecision		**0.45 (0.27 to 0.74)**	
	≥65 years					no serious imprecision		**0.23 (0.10 to 0.51)**	
	reference: 15-24 years					-		-	
Neely, 2009 London Data: 2004	≤24 years	Unmatched case-control with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	355	1.7 (0.5 to 6.3)	VERY LOW
	25-44 years					serious[5]		2.1 (0.6 to 7.7)	
	reference: ≥45 years					-		-	

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London (non-outbreak) Date: 2004	0-14 years	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	129	0.8 (0.2 to 4.6)	VERY LOW
	15-29 years					no serious imprecision		1.1 (0.7 to 1.7)	
	≥60 years					no serious imprecision		0.5 (0.3 to 1.2)	
	reference: 30-59 years					-		-	
Patients living outside of London									
Kruijshaar, 2008 England, Wales and Northern Ireland, excluding London Data: 1998 and 2005	**Age (linear)**	Observational with multivariate analysis	serious[1]	serious[6,7]	no serious indirectness	no serious imprecision	16 633	**0.99 (0.98 to 0.99)**	LOW
Patients with no previous tuberculosis									
French, 2008 England and Wales Data: 1999-2005	**45-64 years**	Unmatched case-control with multivariate analysis	very serious[1,2,8,9]	no serious inconsistency	no serious indirectness	no serious imprecision	18005	**0.70 (0.59 to 0.83)**	LOW
	≥65 years					no serious imprecision		**0.34 (0.26 to 0.44)**	
	reference: 15-44 years					-		-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for all variables recorded and reported in population characteristics (possible selective reporting)
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval
[6] Loss to follow-up, its reasons and the characteristics of those lost not reported
[7] Approach to drug susceptibility testing not reported
[8] Cases and controls unmatched
[9] A number of factors reported in the univariate analyses were not reported as multivariate analyses
Abbreviations: CI, confidence interval; OR, odds ratio

Sex

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

Appendix E: GRADE profiles

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Patients living in London									
Kruijshaar, 2008 London Data: 1998 and 2005	Female reference: male	Observational with multivariate analysis	serious[1]	serious[5,6]	no serious indirectness	no serious imprecision	11 848	0.92 (0.79 to 1.08) -	LOW
Maguire, 2011 London Data: 1995 to the third quarter of 2006	Male reference: female	Unmatched case-control with multivariate analysis	very serious[1,2,7,8]	no serious inconsistency	no serious indirectness	no serious imprecision	18040	1.34 (0.98 to 1.83) -	LOW
Neely, 2009 London Data: 2004	**Male** reference: female	Unmatched case-control with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	serious[9]	355	**2.7 (1.1 to 6.6)** -	VERY LOW
Story, 2007 London (non-outbreak) Date: 2004	Male reference: female	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	129	1.0 (0.7 to 1.6) -	VERY LOW
Patients living outside of London									
Kruijshaar, 2008 England, Wales and Northern Ireland, excluding London Data: 1998 and 2005	**Female** reference: male	Observational with multivariate analysis	serious[1]	serious[5,6]	no serious indirectness	no serious imprecision	16 633	**0.81 (0.69 to 0.96)** -	LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Loss to follow-up, its reasons and the characteristics of those lost not reported
[6] Approach to drug susceptibility testing not reported
[7] Cases and controls unmatched
[8] Some data collected by questionnaire (i.e. may be some reliance on recall)
[9] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

Exposure

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Degree of exposure to drug resistant tuberculosis									
Neely, 2009 London Data: 2004	**Close**	Unmatched case-control with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	355	**6.2 (1.7 to 21.8)**	VERY LOW
	reference: casual							-	
Neely, 2009 London Data: 2004	**Cases to whom contact was exposed: ≥2**	Unmatched case-control with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	355	**3.1 (1.1 to 8.4)**	VERY LOW
	reference: 1							-	
Exposure to smear-positive drug resistant tuberculosis									
Neely, 2009 London Data: 2004	Exposure to cases with smear-positive drug resistant tuberculosis	Unmatched case-control with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	355	2.2 (0.8 to 6.2)	VERY LOW
	reference: no exposure to smear-positive drug resistant tuberculosis							-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for number of drug-using cases to whom contact exposed, which was recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

Previous history of tuberculosis

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Patients living in London									
Kruijshaar, 2008 London Data: 1998 and 2005	**Previous history of tuberculosis**	Observational with multivariate analysis	serious[2]	serious[6,7]	no serious indirectness	no serious imprecision	11 848	**1.35 (1.02 to 1.78)**	LOW
	reference: no history of tuberculosis							-	
Patients living outside of London									
Kruijshaar,	**Previous**	Observational with	serious[2]	serious[6,7]	no serious	no serious	16 633	**1.80 (1.40 to 2.32)**	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
2008 England, Wales and Northern Ireland, excluding London Data: 1998 and 2005	**history of tuberculosis** reference: no history of tuberculosis	multivariate analysis			indirectness	imprecision		 -	

[1] Cases and controls were matched on ethnic group, gender and age group
[2] Unclear if prognostic factor and outcome measurement blinded
[3] Authors had to rely on others' notes (potential for recall bias)
[4] Multivariate analysis used, but unclear which confounders were controlled for
[5] Wide confidence interval
[6] Loss to follow-up, its reasons and the characteristics of those lost not reported
[7] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

Smear status

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Maguire, 2011 London Data: 1995 to the third quarter of 2006	Smear-positive reference: smear-negative	Unmatched case-control with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	no serious imprecision	18040	1.37 (0.98 to 1.93) -	LOW
Patients with previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	**Smear-positive** reference: smear-negative	Unmatched case-control with multivariate analysis	very serious[1,6,7]	no serious inconsistency	no serious indirectness	serious[5]	639	**3.2 (1.1 to 9.2)** -	LOW
Patients with no previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	Smear-positive reference: smear-negative	Unmatched case-control with multivariate analysis	very serious[1,6,7]	no serious inconsistency	no serious indirectness	no serious imprecision	8762	1.1 (0.8 to 1.4) -	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
[1] Unclear if prognostic factor and outcome measurement blinded [2] Cases and controls unmatched [3] Some data collected by questionnaire (i.e. may be some reliance on recall) [4] Multivariate analysis used, but unclear which confounders were controlled for [5] Wide confidence interval [6] Not all factors that underwent univariate analysis were entered into the multivariate analyses; unclear how factors were selected for the multivariate analyses [7] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993–1994 and 1998–2000) Abbreviations: CI, confidence interval; OR, odds ratio									

Site of disease

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Patients living in London									
Maguire, 2011 London Data: 1995 to the third quarter of 2006	Extrapulmonary tuberculosis	Unmatched case-control with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	no serious imprecision	18040	1.52 (0.98 to 2.36)	LOW
	reference: pulmonary tuberculosis							-	
Kruijshaar, 2008 London Data: 1998 and 2005	Pulmonary tuberculosis	Observational with multivariate analysis	serious[1]	serious[5,6]	no serious indirectness	no serious imprecision	11 848	1.06 (0.89 to 1.25)	LOW
	reference: extrapulmonary tuberculosis							-	
Patients living outside of London									
Kruijshaar, 2008 England, Wales and Northern Ireland, excluding London Data: 1998 and 2005	**Pulmonary tuberculosis**	Observational with multivariate analysis	serious[1]	serious[5,6]	no serious indirectness	no serious imprecision	16 633	**0.82 (0.69 to 0.98)**	LOW
	reference: extrapulmonary tuberculosis							-	
[1] Unclear if prognostic factor and outcome measurement blinded [2] Cases and controls unmatched [3] Some data collected by questionnaire (i.e. may be some reliance on recall) [4] Multivariate analysis used, but unclear which confounders were controlled for [5] Loss to follow-up, its reasons and the characteristics of those lost not reported [6] Approach to drug susceptibility testing not reported									

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Abbreviations: CI, confidence interval; OR, odds ratio									

HIV status

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Patients with no previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	HIV-positive reference: HIV-negative	Unmatched case-control with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	8762	1.3 (0.8 to 1.9) -	LOW
French, 2008 England and Wales Data: 1999-2005	HIV-positive reference: HIV-negative	Unmatched case-control with multivariate analysis	very serious[1,4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	18005	1.02 (0.80 to 1.30) -	LOW
Patients with previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	HIV-positive reference: HIV-negative	Unmatched case-control with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	639	0.6 (0.1 to 4.6) -	LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Not all factors that underwent univariate analysis were entered into the multivariate analyses; unclear how factors were selected for the multivariate analyses
[3] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993-4 and 1998-2000)
[4] Cases and controls unmatched
[5] Multivariate analysis used, although it was unclear which confounders were accounted for
[6] A number of factors reported in the univariate analyses were not reported as multivariate analyses
Abbreviations: CI, confidence interval; OR, odds ratio

Place of residence

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Patients with no previous tuberculosis									
Conaty, 2004 England and Wales	**London residence** reference: non-	Unmatched case-control with multivariate	very serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	8762	**1.4 (1.1 to 1.7)** -	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Data: 1993-4 and 1998-2000	London residence	analysis							
French, 2008 England and Wales Data: 1999-2005	**London residence**	Unmatched case-control with multivariate analysis	very serious[1,4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	18005	**1.52 (1.34 to 1.72)**	LOW
	reference: non-London residence							-	
Patients with previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	London residence	Unmatched case-control with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	639	1.8 (0.9 to 3.7)	LOW
	reference: non-London residence							-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Not all factors that underwent univariate analysis were entered into the multivariate analyses; unclear how factors were selected for the multivariate analyses
[3] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993-4 and 1998-2000)
[4] Cases and controls unmatched
[5] Multivariate analysis used, although it was unclear which confounders were accounted for
[6] A number of factors reported in the univariate analyses were not reported as multivariate analyses
Abbreviations: CI, confidence interval; OR, odds ratio

Place of birth

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Time in the UK in patients with previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	In the UK <5 years	Unmatched case-control with multivariate analysis	very serious[1,3,6]	no serious inconsistency	no serious indirectness	serious[5]	639	2.8 (0.8 to 9.7)	VERY LOW
	In the UK 5-9 years					serious[5]		**5.3 (1.2 to 23.5)**	VERY LOW
	In the UK ≥10 years					no serious imprecision		0.9 (0.3 to 3.8)	LOW
	reference: born in the UK					-		-	-
Time in the UK in patients with no previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4	In the UK <5 years	Unmatched case-control with multivariate analysis	very serious[1,3,6]	no serious inconsistency	no serious indirectness	no serious imprecision	8762	1.1 (0.8 to 1.5)	LOW
	In the UK 5-9 years					no serious imprecision		1.2 (0.8 to 1.7)	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
and 1998-2000	In the UK ≥10 years					no serious imprecision		0.9 (0.7 to 1.3)	LOW
	reference: born in the UK					-		-	-
Time in the UK in patients who have residence in London									
Kruijshaar, 2008 London Data: 1998 and 2005	**Years in the UK (linear)**	Observational with multivariate analysis	serious[1]	serious[7,8]	no serious indirectness	no serious imprecision	11 848	**1.04 (1.00 to 1.07)**	LOW
Time in the UK in patients who have residence outside of London									
Kruijshaar, 2008 England, Wales and Northern Ireland, excluding London Data: 1998 and 2005	Years in the UK (linear)	Observational with multivariate analysis	serious[1]	serious[7,8]	no serious indirectness	no serious imprecision	16 633	1.01 (0.98 to 1.05)	LOW
Place of birth in patients who have residence in London									
Kruijshaar, 2008 London Data: 1998 and 2005	**Born outside of the UK**	Observational with multivariate analysis	serious[1]	serious[7,8]	no serious indirectness	no serious imprecision	11 848	**0.76 (0.60 to 0.95)**	LOW
	reference: born in the UK							-	
Maguire, 2011 London Data: 1995 to the third quarter of 2006	**Born in the UK**	Unmatched case-control with multivariate analysis	very serious[1,2,9,10]	no serious inconsistency	no serious indirectness	no serious imprecision	18040	**2.40 (1.68 to 3.43)**	LOW
	reference: born outside of the UK							-	
Story, 2007 London (outbreak) Date: 2004	**Born in the UK**	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	38	**2.8 (1.1 to 7.0)**	VERY LOW
	reference: born outside of the UK							-	
Place of birth in patients who have residence outside of London									
Kruijshaar,	**Born outside of**	Observational with	serious[1]	serious[7,8]	no serious	no serious	16 633	**1.49 (1.16 to 1.92)**	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
2008 England, Wales and Northern Ireland, excluding London Data: 1998 and 2005	**the UK** reference: born in the UK	multivariate analysis			indirectness	imprecision		 -	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval
[6] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993–1994 and 1998–2000)
[7] Loss to follow-up, its reasons and the characteristics of those lost not reported
[8] Approach to drug susceptibility testing not reported
[9] Cases and controls unmatched
[10] Some data collected by questionnaire (i.e. may be some reliance on recall)
Abbreviations: CI, confidence interval; OR, odds ratio

Ethnicity

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 London Data: 1998 and 2005	**Black Caribbean**	Observational with multivariate analysis	serious[1]	serious[6,7]	no serious indirectness	no serious imprecision	11 848	**2.93 (2.11 to 4.09)**	LOW
	Black African					no serious imprecision		1.08 (0.80 to 1.45)	LOW
	Black other					no serious imprecision		1.38 (0.75 to 2.55)	LOW
	Indian, Pakistani, Bangladeshi					no serious imprecision		0.89 (0.66 to 1.19)	LOW
	Chinese					no serious imprecision		1.41 (0.75 to 2.64)	LOW
	Other					no serious imprecision		1.04 (0.74 to 1.46)	LOW
	reference: white					-		-	-
Kruijshaar, 2008	Black Caribbean	Observational with multivariate	serious[1]	serious[6,7]	no serious indirectness	no serious imprecision	16 633	1.35 (0.77 to 2.36)	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
England, Wales and Northern Ireland, excluding London Data: 1998 and 2005	Black African	analysis				no serious imprecision		0.99 (0.68 to 1.43)	LOW
	Black other					no serious imprecision		0.99 (0.30 to 3.28)	LOW
	Indian, Pakistani, Bangladeshi					no serious imprecision		1.26 (0.94 to 1.69)	LOW
	Chinese					no serious imprecision		1.71 (0.99 to 2.95)	LOW
	Other					no serious imprecision		**1.65 (1.11 to 2.44)**	LOW
	reference: white					-		-	-
Maguire, 2011 London Data: 1995 to the third quarter of 2006	**Black Caribbean**	Unmatched case-control with multivariate analysis	very serious[1,2,8,9]	no serious inconsistency	no serious indirectness	serious[5]	18040	**12.52 (7.69 to 20.37)**	VERY LOW
	Black (other)					serious[5]		**3.29 (1.35 to 8.02)**	VERY LOW
	White					no serious imprecision		**2.94 (1.79 to 4.83)**	LOW
	Indian subcontinent					no serious imprecision		**0.57 (0.30 to 1.10)**	LOW
	Chinese					serious[5]		0.68 (0.09 to 5.05)	VERY LOW
	Other					no serious imprecision		1.210 (0.67 to 2.19)	LOW
	reference: Black African					-		-	-
Patients with previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	Indian subcontinent	Unmatched case-control with multivariate analysis	very serious[1,2,10]	no serious inconsistency	no serious indirectness	no serious imprecision	639	1.2 (0.4 to 3.7)	LOW
	Black African					no serious imprecision		0.9 (0.2 to 3.8)	LOW
	Other					no serious imprecision		0.5 (0.1 to 2.6)	LOW
	reference: white					-		-	-
Story, 2007 London (outbreak) Date: 2004	South Asian	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	38	1.1 (0.2 to 6.7)	VERY LOW
	Black African							0.8 (0.1 to 7.2)	VERY LOW
	Black Caribbean							**9.7 (2.6 to (35.4)**	VERY LOW
	Other							**6.1 (1.6 to 23.3)**	VERY LOW
	reference: White							-	-

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London (non-outbreak) Date: 2004	South Asian	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	129	1.0 (0.5 to 2.1)	VERY LOW
	Black African							1.4 (0.7 to 2.6)	VERY LOW
	Black Caribbean							1.6 (0.3 to 10.2)	VERY LOW
	Other							2.5 (0.9 to 7.1)	VERY LOW
	reference: White							-	-
Patients with no previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	**Indian subcontinent**	Unmatched case-control with multivariate analysis	very serious[1,2,10]	no serious inconsistency	no serious indirectness	no serious imprecision	8762	**1.6 (1.2 to 2.1)**	LOW
	Black African					no serious imprecision		**1.7 (1.2 to 2.4)**	LOW
	Other					no serious imprecision		**1.9 (1.3 to 2.8)**	LOW
	reference: white					-		-	-
French, 2008 England and Wales Data: 1999-2005	**Black Caribbean**	Unmatched case-control with multivariate analysis	very serious[1,2,8,11]	no serious inconsistency	no serious indirectness	no serious imprecision	18005	**3.11 (2.36 to 4.08)**	LOW
	Black African					no serious imprecision		**1.22 (1.00 to 1.50)**	LOW
	Indian/ Pakistani/ Bangladeshi					no serious imprecision		**1.18 (0.99 to 1.42)**	LOW
	Other					no serious imprecision		**1.40 (1.12 to 1.76)**	LOW
	reference: white					-		-	-

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval
[6] Loss to follow-up, its reasons and the characteristics of those lost not reported
[7] Approach to drug susceptibility testing not reported
[8] Cases and controls unmatched
[9] Some data collected by questionnaire (i.e. may be some reliance on recall)
[10] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993–1994 and 1998–2000)
[11] A number of factors reported in the univariate analyses were not reported as multivariate analyses
Abbreviations: CI, confidence interval; OR, odds ratio

Employment

Study	Factor	Quality assessment	Number of	Summary of findings	Quality

		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	patients	Adjusted OR (95% CI)	
Healthcare									
Maguire, 2011 London Data: 1995 to the third quarter of 2006	Healthcare profession reference: other (not: prisoner, healthcare, unemployed, asylum seeker/ refugee, drug dealer/sex worker, educational, retired)	Unmatched case-control with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	no serious imprecision	18040	1.53 (0.67 to 3.51) -	LOW
Education									
Maguire, 2011 London Data: 1995 to the third quarter of 2006	Educational profession reference: other (not: prisoner, healthcare, unemployed, asylum seeker/ refugee, drug dealer/sex worker, educational, retired)	Unmatched case-control with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	no serious imprecision	18040	1.22 (0.67 to 2.23) -	LOW
Drug dealer/ sex worker									
Maguire, 2011 London Data: 1995 to the third quarter of 2006	**Drug dealer/ sex worker** reference: other (not: prisoner, healthcare, unemployed, asylum seeker/ refugee, drug dealer/sex worker, educational, retired)	Unmatched case-control with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	18040	**187.07 (28.40 to 1232.35)** -	VERY LOW
Unemployed									
Maguire, 2011 London Data: 1995 to the third quarter of 2006	**Unemployed** reference: other (not: prisoner, healthcare, unemployed, asylum seeker/	Unmatched case-control with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	no serious imprecision	18040	**4.09 (2.97 to 5.63)** -	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
	refugee, drug dealer/sex worker, educational, retired)								
Retired									
Maguire, 2011 London Data: 1995 to the third quarter of 2006	Retired reference: other (not: prisoner, healthcare, unemployed, asylum seeker/ refugee, drug dealer/sex worker, educational, retired)	Unmatched case-control with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	no serious imprecision	18040	1.69 (0.71 to 4.06) -	LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Cases and controls unmatched
[3] Some data collected by questionnaire (i.e. may be some reliance on recall)
[4] Multivariate analysis used, but unclear which confounders were controlled for
[5] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

Drug use

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London (outbreak) Date: 2004	**Problem drug use** reference: no problem drug use	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	38	**3.5 (1.6 to 7.7)** -	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

Asylum seekers/refugee

Study	Factor	Quality assessment	Number of	Summary of findings	Quality

		Design	Risk of bias	Inconsistency	Indirectness	Imprecision	patients	Adjusted OR (95% CI)	
Maguire, 2011 London Data: 1995 to the third quarter of 2006	**Asylum seeker/ refugee** reference: other (not: prisoner, healthcare, unemployed, asylum seeker/ refugee, drug dealer/sex worker, educational, retired)	Unmatched case-control with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	serious[5]	18040	**8.09 (1.02 to 64.41)** -	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Cases and controls unmatched
[3] Some data collected by questionnaire (i.e. may be some reliance on recall)
[4] Multivariate analysis used, but unclear which confounders were controlled for
[5] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

Imprisonment

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Maguire, 2011 London Data: 1995 to the third quarter of 2006	**Imprisonment** reference: other (not: prisoner, healthcare, unemployed, asylum seeker/ refugee, drug dealer/sex worker, educational, retired)	Unmatched case-control with multivariate analysis	very serious[1,2,6,7]	no serious inconsistency	no serious indirectness	serious[5]	18040	**20.21 (6.75 to 60.56)** -	VERY LOW
Story, 2007 London (outbreak) Date: 2004	**Imprisonment** reference: not being imprisoned	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	38	**10.3 (4.0 to 26.5)** -	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval
[6] Cases and controls unmatched

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
[7] Some data collected by questionnaire (i.e. may be some reliance on recall) Abbreviations: CI, confidence interval; OR, odds ratio									

Homelessness

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London (non-outbreak) Date: 2004	Hostel/street homeless	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	129	2.0 (0.9 to 4.5)	VERY LOW
	reference: not homeless							-	
[1] Unclear if prognostic factor and outcome measurement blinded [2] Multivariate analysis used, but unclear which confounders were controlled for [3] Analyses not reported for a number of variables recorded and reported in population characteristics [4] Unclear if loss to follow-up sufficiently unrelated to key characteristics Abbreviations: CI, confidence interval; OR, odds ratio									

A.12.4 Rifampicin resistance

Age

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	**Age (linear)**	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	28481	**0.98 (0.97 to 0.99)**	LOW
[1] Unclear if prognostic factor and outcome measurement blinded [2] Loss to follow-up, its reasons and the characteristics of those lost not reported [3] Approach to drug susceptibility testing not reported Abbreviations: CI, confidence interval; OR, odds ratio									

Sex

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	Female reference: male	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	28447	0.83 (0.64 to 1.08) -	LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Loss to follow-up, its reasons and the characteristics of those lost not reported
[3] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

Previous history of tuberculosis

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	**Previous history of tuberculosis** reference: no history of tuberculosis	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	22671	**4.72 (3.50 to 6.35)** -	LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Loss to follow-up, its reasons and the characteristics of those lost not reported
[3] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

Site of disease

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998	**Pulmonary** reference: extrapulmonary	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	28341	**1.48 (1.10 to 1.98)** -	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
and 2005									

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Loss to follow-up, its reasons and the characteristics of those lost not reported
[3] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

Place of residence

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	London reference: Outside London	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	28485	0.81 (0.62 to 1.05) -	LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Loss to follow-up, its reasons and the characteristics of those lost not reported
[3] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

Place of birth

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Place of birth									
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	**Not born in the UK** reference: born in the UK	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	25557	**1.88 (1.24 to 2.86)** -	LOW
Time in the UK									
Kruijshaar, 2008 England, Wales and Northern	Years in the UK (linear)	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	28485	1.03 (0.98 to 1.09)	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Ireland Data: 1998 and 2005									

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Loss to follow-up, its reasons and the characteristics of those lost not reported
[3] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

Ethnicity

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	Black Caribbean	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	27257	1.28 (0.59 to 2.79)	LOW
	Black African					no serious imprecision		0.98 (0.59 to 1.64)	LOW
	Black other					serious[4]		1.87 (0.69 to 5.06)	VERY LOW
	Indian, Pakistani, Bangladeshi					no serious imprecision		0.94 (0.59 to 1.50)	LOW
	Chinese					no serious imprecision		0.83 (0.28 to 2.45)	LOW
	Other					no serious imprecision		0.97 (0.54 to 1.75)	LOW
	reference: white					-		-	-

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Loss to follow-up, its reasons and the characteristics of those lost not reported
[3] Approach to drug susceptibility testing not reported
[4] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio

A.12.5 Multidrug resistance

Age

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	Age (linear)	Observational with multivariate analysis	serious[1]	serious[6,7]	no serious indirectness	no serious imprecision	28481	0.98 (0.59 to 1.08)	LOW
Patients with no previous tuberculosis									
French, 2008 England and Wales Data: 1999-2005	**45-64 years**	Unmatched case-control with multivariate analysis	very serious[1,2,3,4]	no serious inconsistency	no serious indirectness	no serious imprecision	16935	**0.52 (0.27 to 0.99)**	LOW
	≥65 years					serious[5]		**0.35 (0.14 to 0.90)**	VERY LOW
	reference: 15-44 years					-		-	-

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Cases and controls unmatched
[4] A number of factors reported in the univariate analyses were not reported as multivariate analyses
[5] Wide confidence interval
[6] Loss to follow-up, its reasons and the characteristics of those lost not reported
[7] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

Sex

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	Female	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	28447	0.80 (0.59 to 1.08)	LOW
	reference: male							-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Loss to follow-up, its reasons and the characteristics of those lost not reported
[3] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

Previous history of tuberculosis

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	Previous history of tuberculosis reference: no history of tuberculosis	Observational with multivariate analysis	serious[2]	serious[6,7]	no serious indirectness	no serious imprecision	28485	1.04 (0.76 to 1.42) -	LOW

[1] Cases and controls were matched on ethnic group, gender and age group
[2] Unclear if prognostic factor and outcome measurement blinded
[3] Authors had to rely on others' notes (potential for recall bias)
[4] Multivariate analysis used, but unclear which confounders were controlled for
[5] Wide confidence interval
[6] Loss to follow-up, its reasons and the characteristics of those lost not reported
[7] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

Smear status

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Patients with previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	**Smear-positive** reference: smear-negative	Unmatched case-control with multivariate analysis	very serious[1,3,4]	no serious inconsistency	no serious indirectness	serious[2]	630	**5.9 (1.8 to 19.0)** -	LOW
Patients with no previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	Smear-positive reference: smear-negative	Unmatched case-control with multivariate analysis	very serious[1,6,7]	no serious inconsistency	no serious indirectness	no serious imprecision	8210	1.4 (0.7 to 2.5) -	LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Wide confidence interval
[3] Not all factors that underwent univariate analysis were entered into the multivariate analyses; unclear how factors were selected for the multivariate analyses
[4] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993–1994 and 1998–2000)
Abbreviations: CI, confidence interval; OR, odds ratio

Site of disease

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	**Pulmonary tuberculosis** reference: extrapulmonary tuberculosis	Observational with multivariate analysis	serious[1]	serious[2,3]	no serious indirectness	no serious imprecision	28341	**1.40 (1.00 to 1.96)** -	LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Loss to follow-up, its reasons and the characteristics of those lost not reported
[3] Approach to drug susceptibility testing not reported
Abbreviations: CI, confidence interval; OR, odds ratio

HIV status

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Patients with no previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	**HIV-positive** reference: HIV-negative	Unmatched case-control with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[7]	662	**2.5 (1.2 to 5.2)** -	LOW
French, 2008 England and Wales Data: 1999-2005	HIV-positive reference: HIV-negative	Unmatched case-control with multivariate analysis	very serious[1,4,5,6]	no serious inconsistency	no serious indirectness	no serious imprecision	16935	0.91 (0.47 to 1.76) -	LOW
Patients with previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	HIV-positive reference: HIV-negative	Unmatched case-control with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[7]	8210	2.8 (0.6 to 11.9) -	LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Not all factors that underwent univariate analysis were entered into the multivariate analyses; unclear how factors were selected for the multivariate analyses
[3] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993-4 and 1998-2000)
[4] Cases and controls unmatched

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
[5] Multivariate analysis used, although it was unclear which confounders were accounted for [6] A number of factors reported in the univariate analyses were not reported as multivariate analyses Abbreviations: CI, confidence interval; OR, odds ratio									

Place of residence

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Patients with no previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	**London residence** reference: non-London residence	Unmatched case-control with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	662	**2.0 (1.2 to 3.3)** -	LOW
Patients with previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	London residence reference: non-London residence	Unmatched case-control with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	8210	1.2 (0.6 to 2.4) -	LOW
[1] Unclear if prognostic factor and outcome measurement blinded [2] Not all factors that underwent univariate analysis were entered into the multivariate analyses; unclear how factors were selected for the multivariate analyses [3] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993-4 and 1998-2000) Abbreviations: CI, confidence interval; OR, odds ratio									

Place of birth

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Time in the UK									
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	Years in the UK (linear)	Observational with multivariate analysis	serious[1]	serious[4,5]	no serious indirectness	no serious imprecision	25557	1.62 (0.99 to 2.66)	LOW
Time in the UK in patients with previous tuberculosis									

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	**In the UK <5 years**	Unmatched case-control with multivariate analysis	very serious[1,2,6]	no serious inconsistency	no serious indirectness	serious[3]	8210	**5.8 (1.8 to 18.5)**	VERY LOW
	In the UK 5-9 years					serious[3]		2.2 (0.4 to 11.6)	VERY LOW
	In the UK ≥10 years					serious[3]		1.7 (0.4 to 6.9)	LOW
	reference: born in the UK					-		-	-
Time in the UK in patients with no previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	**In the UK <5 years**	Unmatched case-control with multivariate analysis	very serious[1,2,6]	no serious inconsistency	no serious indirectness	serious[3]	630	**3.2 (1.4 to 7.4)**	LOW
	In the UK 5-9 years					serious[3]		**3.0 (1.1 to 8.5)**	LOW
	In the UK ≥10 years					no serious imprecision		1.2 (0.4 to 3.7)	LOW
	reference: born in the UK					-		-	-
Place of birth									
Kruijshaar, 2008 England, Wales and Northern Ireland Data: 1998 and 2005	Born outside of the UK	Observational with multivariate analysis	serious[1]	serious[4,5]	no serious indirectness	no serious imprecision	25557	1.01 (0.95 to 1.08)	LOW
	reference: born in the UK							-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Analyses not reported for a number of variables recorded and reported in population characteristics
[3] Wide confidence interval
[4] Loss to follow-up, its reasons and the characteristics of those lost not reported
[5] Approach to drug susceptibility testing not reported
[6] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993–1994 and 1998–2000)
Abbreviations: CI, confidence interval; OR, odds ratio

Ethnicity

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Kruijshaar, 2008	Black Caribbean	Observational with multivariate	serious[1]	serious[6,7]	no serious indirectness	no serious imprecision	27257	1.01 (0.30 to 3.43)	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
England, Wales and Northern Ireland Data: 1998 and 2005	Black African	analysis				no serious imprecision		1.77 (0.92 to 3.41)	LOW
	Black other					serious[5]		2.44 (0.68 to 8.81)	VERY LOW
	Indian, Pakistani, Bangladeshi					no serious imprecision		1.63 (0.91 to 2.95)	LOW
	Chinese					no serious imprecision		1.77 (0.56 to 5.54)	LOW
	Other					no serious imprecision		1.32 (0.62 to 2.84)	LOW
	reference: white					-		-	-
Patients with previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	**Indian subcontinent**	Unmatched case-control with multivariate analysis	very serious[1,2,10]	no serious inconsistency	no serious indirectness	serious[5]	8210	**5.8 (1.8 to 18.5)**	VERY LOW
	Black African					serious[5]		2.2 (0.4 to 11.6)	VERY LOW
	Other					serious[5]		1.7 (0.4 to 6.9)	VERY LOW
	reference: white					-		-	-
Story, 2007 London Date: 2004	South Asian	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	1540	1.6 (0.8 to 3.0)	LOW
	Black African					no serious imprecision		**2.5 (1.2 to 5.7)**	LOW
	Black Caribbean					serious[5]		1.6 (0.3 to 10.2)	VERY LOW
	Other					serious[5]		2.5 (0.9 to 7.1)	VERY LOW
	reference: White					-		-	-
Patients with no previous tuberculosis									
Conaty, 2004 England and Wales Data: 1993-4 and 1998-2000	Indian subcontinent	Unmatched case-control with multivariate analysis	very serious[1,2,10]	no serious inconsistency	no serious indirectness	no serious imprecision	630	0.8 (0.4 to 1.5)	LOW
	Black African					no serious imprecision		0.6 (0.3 to 1.2)	LOW
	Other					serious[5]		0.3 (0.1 to 0.9)	VERY LOW
	reference: white					-		-	-
French, 2008 England and Wales Data: 1999-2005	Black Caribbean	Unmatched case-control with multivariate analysis	very serious[1,2,8,11]	no serious inconsistency	no serious indirectness	no serious imprecision	16935	1.40 (0.39 to 5.01)	LOW
	Black African					no serious imprecision		2.02 (0.88 to 4.64)	LOW
	Indian/ Pakistani/ Bangladeshi					no serious imprecision		1.33 (0.61 to 2.90)	LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
	Other					no serious imprecision		1.39 (0.56 to 3.45)	LOW
	reference: white					-		-	-

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
[5] Wide confidence interval
[6] Loss to follow-up, its reasons and the characteristics of those lost not reported
[7] Approach to drug susceptibility testing not reported
[8] Cases and controls unmatched
[9] Some data collected by questionnaire (i.e. may be some reliance on recall)
[10] Multivariate analysis used, although effect estimates only adjusted for age and two periods of analysis (1993–1994 and 1998–2000)
[11] A number of factors reported in the univariate analyses were not reported as multivariate analyses
Abbreviations: CI, confidence interval; OR, odds ratio

Homelessness

Study	Factor	Quality assessment					Number of patients	Summary of findings Adjusted OR (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Story, 2007 London Date: 2004	**Ever homeless**	Observational with multivariate analysis	very serious[1,2,3]	serious[4]	no serious indirectness	no serious imprecision	1540	**2.1 (1.1 to 4.1)**	VERY LOW
	reference: not homeless							-	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Analyses not reported for a number of variables recorded and reported in population characteristics
[4] Unclear if loss to follow-up sufficiently unrelated to key characteristics
Abbreviations: CI, confidence interval; OR, odds ratio

A.12.6 International surveillance data

Countries with a high burden of multidrug resistant tuberculosis, according to the World Health Organisation[1]:

A.13 RQ U, V & W

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	7RE	4 RE	Relative (95% CI)	Absolute	
Response											
1	randomised trials	very serious1	no serious inconsistency2	no serious indirectness3	serious 4	none	110/113 (97.3%)	105/113 (92.9%)	RR 1.05 (0.99 to 1.11)	46 more per 1000 (from 9 more to 102 more)	VERY LOW
Relapse											
1	randomised trials	very serious1	no serious inconsistency2	no serious indirectness3	serious 4	none	2/92 (2.2%)	6/86 (7%)	RR 0.31 (0.06 to 1.5)	48 fewer per 1000 (from 66 fewer to 35 more)	VERY LOW
Adverse effects											
1	randomised trials	very serious1	no serious inconsistency2	no serious indirectness3	serious 4	none	1/113 (0.88%)	1/113 (0.88%)	RR 1 (0.06 to 15.79)	0 fewer per 1000 (from 8 fewer to 131 more)	VERY LOW

[1] Serious risk of bias due to concerns over trail methodology re blinding, allocation concealment, method of allocation
[2] Single study analysis
[3] Population and intervention as specified in the review protocol
[4] Confidence intervals around point estimate cross line of no effect

3RSZH or 3RSHZ + 2SHZ

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Response								
1	randomis	very	serious2	serious3	serious4	RCT but data not	32/35 (91%)	VERY

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
(Balasubramanian, 1990)	ed trials	serious1				stratified by Resistance status		LOW
Relapse at 5 years								
1 (Balasubramanian, 1990)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	6/32 (19%)	VERY LOW

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

6RSH

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Response to 6RSH								
1 (East African/ British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	19/20 (95%)	VERY LOW
1 (Hong Kong Chest Service/British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	34/40 (85%)	VERY LOW
Relapse at 24 – 30 months								
1 (East African/ British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	3/13 (23%)	VERY LOW

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
1 (Hong Kong Chest Service/British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	4/29 (14%)	VERY LOW

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

SHRZ/$S_2H_2Z_2$

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Response to SHRZ/S2H2Z2								
1 (Hong Kong Chest Service/British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	16/20 (80%)	VERY LOW
Relapse at 24 months								
1 (Hong Kong Chest Service/British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	3/14 (21%)	VERY LOW

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

SHRE/$S_2H_2Z_2$SHR

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Response to SHRE/S2H2Z2SHR								
1 (Hong Kong Chest Service/British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	22/22 (100%)	VERY LOW
Relapse at 24 months								
1 (Hong Kong Chest Service/British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	9/21 (43%)	VERY LOW

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

SHRE/$S_2H_2Z_2$SHR

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Response to SHRE/S2H2Z2SHR								
1 (Hong Kong Chest Service/British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	22/22 (100%)	VERY LOW
Relapse at 24 months								
1 (Hong Kong Chest	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	9/21 (43%)	VERY LOW

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Service/British MRC, 1977)								

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

$S_3H_3Z_3R_3$/ $S_2H_2Z_2$

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Response to S3H3Z3R3/ S2H2Z2								
1 (Hong Kong Chest Service/British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	20/21 (95%)	VERY LOW
Relapse at 24 months								
1 (Hong Kong Chest Service/British MRC, 1977)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	20/15 (13%)	VERY LOW

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

3RSZH

Quality assessment							No of patients	
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**AntiTB regimen**	**Quality**
Response to 3RSZH								
1 (Tuberculosis Research Centre, Madras and National Tuberculosis Institute, Bangalore, 1986)	randomised trials	very serious1	serious2	serious3	serious4	none	32/34 (94%)	VERY LOW
Relapse at 24 months								
1 (Tuberculosis Research Centre, Madras and National Tuberculosis Institute, Bangalore, 1986)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	7/33 (21%)	VERY LOW

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

6SRZH

Quality assessment							No of patients	
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**AntiTB regimen**	**Quality**
Response to 6SRZH								
1 (Tanzania/British MRC Collaborative Investigation,	randomised trials	very serious1	serious2	serious3	serious4	none	12/18 (67%)	VERY LOW

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
1997)								
Relapse								
1 (Tanzania/British MRC Collaborative Investigation, 1997)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	2/10 (20%)	VERY LOW

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

2EHRZ2/4EHR$_2$-

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Response to 2EHRZ2/4EHR2-								
1 (Tuberculosis Research Centre/Indian Council of Medical Research, 1997)	randomised trials	very serious1	serious2	serious3	serious4	none	47/59 (80%)	VERY LOW
Relapse (timepoint not stated)								
1 (Tuberculosis Research Centre/Indian Council of Medical	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	11/21 (54%)	VERY LOW

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Research, 1997)								

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

2EHRZ7/6EH7

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Response to 2EHRZ7/6EH7								
1 (Tuberculosis Research Centre/Indian Council of Medical Research, 1997)	randomised trials	very serious1	serious2	serious3	serious4	none	16/94 (83%)	VERY LOW
Relapse (timepoint not stated)								
1 (Tuberculosis Research Centre/Indian Council of Medical Research, 1997)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	6/21 (29%)	VERY LOW

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

2HRZ2/4HR2-

Quality assessment							No of patients	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	AntiTB regimen	Quality
Response to 2HRZ2/4HR2-								
1 (Tuberculosis Research Centre/Indian Council of Medical Research, 1997)	randomised trials	very serious1	serious2	serious3	serious4	none	28/74 (38%)	VERY LOW
Relapse (timepoint not stated)								
1 (Tuberculosis Research Centre/Indian Council of Medical Research, 1997)	randomised trials	very serious1	serious2	serious3	serious4	RCT but data not stratified by Resistance status	4/21 (19%)	VERY LOW

[1] Very serious risk of bias due to concerns over randomisation, allocation concealment, blinding
[2] Single study analysis
[3] Intervention not as specified in review protocol
[4] Descriptive only results used

•

A.14 RQ Z

A.14.1 Management of treatment interruptions

Sequential reintroduction without pyrazinamide SE→H→R compared to simultaneous reintroduction HRZE in patients receiving treatment for pulmonary or pleural tuberculosis who have experienced drug-induced hepatotoxicity[1]

Number of evaluations	Quality assessment					Number of patients		Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Sequential reintroduction without pyrazinamide SE→H→R	Simultaneous reintroduction HRZE		
Adverse events – recurrence of drug-induced hepatitis[1] (number of patients in whom drug-induced hepatitis[1] recurred following treatment reintroduction)									
1[2]	RCT	serious[4]	serious[5]	no serious indirectness	serious[6]	0/20	6/25	OR 0.07 (95% CI 0.00 to 1.39)	VERY LOW
Cure[3] (number of patients to achieve a cure[3])									
1[2]	RCT	serious[4]	serious[5]	no serious indirectness	very serious[6,7]	20/20	20/25	OR 1.24 (95% CI 0.02 to 65.4)	VERY LOW

[1] Drug-induced hepatitis was defined as normalisation of liver functions after withdrawal of all antituberculosis drugs, and at least one of the following criteria:
• a rise to five times the normal levels (40 U/L) of serum AST and/or ALT
• a rise in the level of serum total bilirubin over 1.5 mg/dl
• any increase in AST and/or ALT above pretreatment levels, together with anorexia, nausea, vomiting and jaundice

[2] Tahaoglu, 2001

[3] Cure was defined as a sputum smear-positive patient who is smear-negative at completion of treatment

[4] Unclear method of randomisation; unclear if allocation concealment used; unclear blinding

[5] Risk factors for hepatotoxicity (age, sex, alcohol consumption, hepatitis markers, radiological extension of the disease in the lungs, pretreatment serum albumin level, diabetes mellitus, additional hepatotoxic drug use, body weight and body mass index) were compared statistically to ensure that there was no increased susceptibility to hepatotoxicity in either group; however, reintroduction without pyrazinamide group had more individuals with extensive disease (P = 0.001) and more individuals with hypoalbuminemia (P = 0.053)

[6] GRADE rule of thumb: <300 events

[7] Wide confidence interval

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; CI, confidence interval; E, ethambutol; H, isoniazid; OR, odds ratio; R, rifampicin; S, streptomycin; Z, pyrazinamide

Sequential reintroduction R→H→Z compared to simultaneous reintroduction HRZ in patients receiving treatment for tuberculosis who have experienced drug-induced hepatotoxicity[1]

Number of evaluations	Quality assessment					Number of patients		Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Sequential reintroduction R→H→Z	Simultaneous reintroduction HRZ		
Adverse events – recurrence of drug-induced hepatitis[1] (number of patients in whom drug-induced hepatitis[1] recurred following treatment reintroduction)									
1[2]	RCT	serious[3]	serious[4]	no serious indirectness	serious[5]	6/59	4/29	OR 0.71 (95% CI 0.18 to 2.73)	VERY LOW

[1] Drug-induced hepatotoxicity was diagnosed if criteria 1, 2, or 3 were present in combination with criteria 4 and 5:

1) an increase ≥5 times the upper limit of the normal levels (50 IU/l) of serum AST and/or ALT on 1 occasion, or >3 times the upper limit of normal (>150 IU/l) on 3 consecutive occasions;

2) an increase in serum total bilirubin >1.5 mg/dl;

3) any increase in serum AST and or ALT level above pretreatment values together with anorexia, nausea, vomiting, and jaundice;

4) absence of serological evidence of infection with hepatitis A, B, C, or E virus; and

5) improvement in liver function test results (serum bilirubin level <1 mg/dl; AST and ALT level <100 IU/l) after withdrawal of antituberculosis drugs

[2] Sharma, 2010

[3] Unclear blinding; unclear length of follow-up

[4] Unclear if length of follow-up equal in each group

[5] GRADE rule of thumb: <300 events

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; CI, confidence interval; H, isoniazid; OR, odds ratio; R, rifampicin; Z, pyrazinamide

Sequential reintroduction H→R→Z compared to simultaneous reintroduction HRZ in patients receiving treatment for tuberculosis who have experienced drug-induced hepatotoxicity[1]

Number of evaluations	Quality assessment					Number of patients		Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Sequential reintroduction H→R→Z	Simultaneous reintroduction HRZ		
Adverse events – recurrence of drug-induced hepatitis[1] (number of patients in whom drug-induced hepatitis[1] recurred following treatment reintroduction)									
1[2]	RCT	serious[3]	serious[4]	no serious	serious[5]	5/58	4/29	OR 0.59 (95%	VERY

Number of evaluations	Quality assessment					Number of patients		Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Sequential reintroduction H→R→Z	Simultaneous reintroduction HRZ		
				indirectness				CI 0.15 to 2.39)	LOW

[1] Drug-induced hepatotoxicity was diagnosed if criteria 1, 2, or 3 were present in combination with criteria 4 and 5:

1) an increase ≥5 times the upper limit of the normal levels (50 IU/l) of serum AST and/or ALT on 1 occasion, or >3 times the upper limit of normal (>150 IU/l) on 3 consecutive occasions;

2) an increase in serum total bilirubin >1.5 mg/dl;

3) any increase in serum AST and or ALT level above pretreatment values together with anorexia, nausea, vomiting, and jaundice;

4) absence of serological evidence of infection with hepatitis A, B, C, or E virus; and

5) improvement in liver function test results (serum bilirubin level <1 mg/dl; AST and ALT level <100 IU/l) after withdrawal of antituberculosis drugs

[2] Sharma, 2010

[3] Unclear blinding; unclear length of follow-up

[4] Unclear if length of follow-up equal in each group

[5] GRADE rule of thumb: <300 events

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; CI, confidence interval; H, isoniazid; OR, odds ratio; R, rifampicin; Z, pyrazinamide

Sequential reintroduction compared to simultaneous reintroduction in patients receiving treatment for tuberculosis who have experienced drug-induced hepatotoxicity[1,2]

Number of evaluations	Quality assessment					Number of patients		Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Sequential reintroduction	Simultaneous reintroduction		
Adverse events – recurrence of drug-induced hepatitis[1,2] (number of patients in whom drug-induced hepatitis[1,2] recurred following treatment reintroduction)									
2[3,4]	RCT	serious[5,6]	serious[7,8]	no serious indirectness	serious[9]	11/137	14/83	OR 0.44 (95% CI 0.18 to 1.03)	VERY LOW

[1] Drug-induced hepatitis in Tahaoglu (2001) was defined as normalisation of liver functions after withdrawal of all antituberculosis drugs, and at least one of the following criteria:

• a rise to five times the normal levels (40 U/L) of serum AST and/or ALT

• a rise in the level of serum total bilirubin over 1.5 mg/dl

Number of evaluations	Quality assessment					Number of patients		Summary of findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Sequential reintroduction	Simultaneous reintroduction		
• any increase in AST and/or ALT above pretreatment levels, together with anorexia, nausea, vomiting and jaundice [2] Drug-induced hepatotoxicity in Sharma (2010) was diagnosed if criteria 1, 2, or 3 were present in combination with criteria 4 and 5: 1) an increase ≥5 times the upper limit of the normal levels (50 IU/l) of serum AST and/or ALT on 1 occasion, or >3 times the upper limit of normal (>150 IU/l) on 3 consecutive occasions; 2) an increase in serum total bilirubin >1.5 mg/dl; 3) any increase in serum AST and or ALT level above pretreatment values together with anorexia, nausea, vomiting, and jaundice; 4) absence of serological evidence of infection with hepatitis A, B, C, or E virus; and 5) improvement in liver function test results (serum bilirubin level <1 mg/dl; AST and ALT level <100 IU/l) after withdrawal of antituberculosis drugs [3] Tahaoglu, 2001 [4] Sharma, 2010 [5] Tahaoglu, 2001: unclear method of randomisation; unclear if allocation concealment used; unclear blinding [6] Sharma, 2010: unclear blinding; unclear length of follow-up [7] Tahaoglu, 2001: risk factors for hepatotoxicity (age, sex, alcohol consumption, hepatitis markers, radiological extension of the disease in the lungs, pretreatment serum albumin level, diabetes mellitus, additional hepatotoxic drug use, body weight and body mass index) were compared statistically to ensure that there was no increased susceptibility to hepatotoxicity in either group; however, reintroduction without pyrazinamide group had more individuals with extensive disease (P = 0.001) and more individuals with hypoalbuminemia (P = 0.053) [8] Sharma, 2010: unclear if length of follow-up equal in each group [9] GRADE rule of thumb: <300 events [10] Forest plot: Abbreviations: CI, confidence interval; OR, odds ratio									

A.15 RQs AA and BB

A.15.1 Behrman 1998. Tuberculosis control in an urban emergency department

Number of evaluations	Quality assessment					Number of patients		Summary of findings		Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision					
Phase I	Prospective cohort	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	ED[2] 6/50	OHEs[3] 51/2514	RR[4] 5.9 (95% CI 2.7-13.1)[5]	Absolute difference 10% (1-19%)	
Phase II	Prospective cohort	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision	ED[2] 0/64	OHEs[3] 36/3000	NC	1.2% (1-2%)	

[1] Unclear blinding participants, personnel and investigators, and how authors addressed potential confounders
[2] ED emergency department employees except physicians
[3] OHEs Other hospital employees
[4] RR risk ratio
[5] Wide confidence interval
Abbreviations: ED: emergency department; OHEs other health employees; CI, confidence interval; NC, not calculable, RR, risk ratio;

A.15.2 Blumerg et al. 1995. Preventing the nosocomial transmission of tuberculosis

Evaluation	Quality assessment					Number of individuals	Summary of Findings	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
TB exposure episode	'descriptive case series' observational	Very Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision[2,]	35/103 18/358	OR 95% CI 9.72 (4.99 to 19.25)[2]	VERY LOW
TST conversion in HCWs evaluated every 6 months x 2.5 years	'descriptive case series' observational	Very Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision		Jan 1992 3.3% to June 1994 0.04%	VERY LOW
Cumulative number of exposure per month	'descriptive case series' observational	Very Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision		35.4/month at 8 month 3.3/month at 28 month (p < 0.001)	VERY LOW

[1] Limitations in design; unequal length of follow up periods; lack of allocation; blinding, other limitations

[2] Wide confidence interval

Abbreviations: HCW; health care workers, RCT randomized control trial; CI, confidence interval; OR, odds ratio

A.15.3 Chamie et al 2013. Household ventilation and tuberculosis transmission in Kampala, Uganda

Outcome of Interest	Quality assessment					Homes (households)	Summary of findings Co-prevalent (n) vs no-co-prevalent (n) households median ACH [IQR] p = 0.05	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Index case sleeping room ventilation rates	Nested case control	Very Serious[1]	No serious inconsistency	No serious indirectness	Serious imprecision[2]	61 (208)	12 [8-15] (12) vs 15 [11-18] (49) *P* = 0.12 (12)	VERY LOW
AFB smear-positive index cases	Nested case control	Very Serious[1]	No serious inconsistency	No serious inconsistency	Serious imprecision[2]	61 (208)	11 [8-14] (11) vs 15 [11-19] (48) *P* = 0.06	VERY LOW
AFB smear positive index case, non-HIV infected	Nested case control	Very Serious[1]	No serious inconsistency	No serious inconsistency	Serious imprecision[2]	61 (208)	11 [8-14] (11) vs 17 [10-20] (12) *p* = 0.1	VERY LOW

[1] Limitations in study design, unclear/lack of blinding, potential recruitment bias

[2] Uncertainty about the results due to low number of households participating

Abbreviations: ACH: air changes per hour; AFB: acid fast bacilli; Co-prevalent: IQR: interquartile range; TB: presence of TB in household acquired from or not from index case;

A.15.4 Da Costa 2009

Outcome and evaluations	Quality assessment					Number of participants	Number of conversions observed /months Conversions/1000 person-month; 95% CI	Summary of Findings Adjusted[a] HR (95% CI)	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision				
TST conversion									
Period I (1999-2001)	Prospective Cohort	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	406	25/4307 5.8; 4.9-6.7		VERY LOW
Period II (2002-2003)	Prospective Cohort	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	193	15/3858 3.7;2.8-4.6 *P* = 0.006	0.24 (0.10-0.54)	VERY LOW
Exposure to pulmonary TB case in hospital (yes)									
Period I (1999-2001)	Prospective Cohort	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	406	11/1661 6.6;5.1-8.1		VERY LOW
Period II (2002-2003)	Prospective Cohort	Serious[1]	No serious inconsistency	No serious indirectness	Serious[2]	193	8/1997 4;2.7-5.3	0.31 (0.13-0.73)	VERY LOW

[a] Adjusted for exposure to pulmonary TB case and professional category (i.e., admin clerk, physician, nurse, social worker, lab & technician, housekeeper)

[1] Unclear inclusion and exclusion of participants; unclear/lack blinding; unclear reasons and characteristics of individuals lost at follow up.

[2] GRADE rule of thumb <300 events

Abbreviations: CI, confidence interval; HR: hazard ratios, TST: Tuberculin skin test

A.15.5 Gonzalez-Angulo et al 2013. Knowledge and acceptability of patient specific infection control measures for TB

Outcome of Interest	Quality assessment					Infection Control Measure	Acceptability of IC measure (TB treatment only). % of Absolute difference (CI) *p* value	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Acceptability of infection control measures	Prospective (questionnaire) cohort	Very Serious[1]	No serious inconsistency	No serious indirectness	Serious imprecision[2] [100 participants (50 diagnosed, 50 suspects)]	**Hospital** Use of face mask Cough hygiene Complete a course of TB treatment Isolation from other patients **Home** Cough hygiene Use of mask Cosleeping Ventilation (natural) Ventilation (mechanical) Isolation **Workplace** Stop working-2 wks Cough hygiene Use of mask	 5 (-3.34-4.88) p 0 .5 - - 5 (-15.71-23.61) p 0.804 2 (-.32-2.44) p 1 22 (2.89-30.49) p 0.23 12 (-4339-20.72) p 0.18 24 (2.07-38.27) p 0.31 22 (4.68-26.71) p 0.12 5 (-5.97-9.63) p 0.625 (-4.39-20.72) p 0.18 - 15 (-1.03-19.39) p 0.70 Spearman correlation coefficient 0.5288 p = .0033	VERY LOW

[1] Limitations in study design, unclear/lack of blinding, unclear exclusion of participants or lost to follow up

[2] GRADE rule of thumb <300 events

Abbreviations: CI: confidence interval TB: tuberculosis; '-' no difference between baseline and end of treatment: wks: weeks

A.15.6 Hubad et al 2012. Inadequate hospital ventilation system increases the risk of nosocomial TB

Outcome of Interest / Location	Quality assessment					Area (IS6110 copy per m^3 of air/ calculated TB cell equivalent per m^3 of air)	Calculated time (hrs)[3]	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Risk of exposure / TB Ward	Prospective interventional study	Very Serious[1]	No serious inconsistency	No serious indirectness	Serious imprecision[2]	Patient room (<10) /- Corridor 177 ±32 / 19±3 Collection room (<10) /-	- 1 -	VERY LOW
Risk of exposure – Diagnostic Laboratory		Very Serious[1]	No serious inconsistency	No serious inconsistency	Serious imprecision[2]	Incubation room 187±49 / 20±5 Corridor 55±22 / 6±2 Lab room (culture) (<10) /-	1 3 -	VERY LOW
Risk of exposure –non TB areas		Very Serious[1]	No serious inconsistency	No serious inconsistency	Serious imprecision[2]	Corridor 98±30 / 10±3 Bioch Lab (<10) /-	2 -	VERY LOW

[1] Limitations in study design, unclear/lack of blinding, unclear how authors address confounders
[2] Uncertainty about the results due to low number of measurements and locations observed
[3] Time after which it is believed that a person would have been exposed to an *M tuberculosis* infectious dose
Abbreviations: hrs: hours, m^3: cubic meter, TB: tuberculosis

A.15.7 Lygizos et al 2013. Natural ventilation reduces high TB transmission risk in traditional homes in rural KwaZulu-natal, SA

Outcome of Interest	Quality assessment					Summary of Findings % Risk of TB, SD % (p value)	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		
TB risk estimation after 10 hours of exposure	Prospective Interventional Cohort	Very Serious[1]	No serious inconsistency	No serious indirectness	Serious imprecision[2]	a) windows and door closed was 55.4%, 27.8, b) upon opening windows 21.5%, SD 14.1 (p <0.001) c) upon opening windows and door together was 9.6%, SD 4.7 (p <0.001) Estimated risk of TB infection increased in parallel to exposure time (p <0.001)	VERY LOW

[1] Limitations in study design, unclear/lack of blinding, potential recruitment bias

[2] Uncertainty about the results due to low number of households participating (n=24)

Abbreviations: SD: standard deviation, TB: tuberculosis

A.15.8 Nardell et al 2008. Safety of upper-room ultraviolet germicidal air disinfection for room occupants: results from the TB UV shelter study

Outcome of Interest	Quality assessment					Summary of Findings		Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
TST conversion	Double blind placebo/control field trial	Serious[1]	No serious inconsistency	No serious indirectness	Serious imprecision[2]	33611 staff and homeless residents	"inconclusive results"	VERY LOW
Adverse Effects	Double blind placebo/control field trial	Serious[1]	No serious inconsistency	No serious indirectness	Serious imprecision[2]	223/3,611 interviews (6%) included a report of a *skin or eye symptom* *Skin or eye symptom* 95/223 (43%) occurred entirely in active UV periods 92/223 (42%) occurred entirely in placebo UV periods 36/223 (16%) uncertain when symptoms occurred Pearson Chi-square value of 0.066 (not statistically significant) One instance of UV-related keratoconjunctivitis occurred, caused by human error		VERY LOW

[1] Limitations in study design, unclear how participant and shelter staff blinding was achieved, potential bias due to loss to follow up, unclear how confounding factors were addressed

[2] Uncertainty about the results due to low number of shelters participating (n=14), <300 events as per GRADE rule of thumb

Abbreviations: SD: standard deviation, TB: tuberculosis; UV: ultraviolet

A.15.9 Richardson 2014. Shared air: a renewed focus on ventilation for the prevention of tuberculosis transmission

Outcome of Interest	Quality assessment					Summary of Findings Rudnick and Milton TB transmission risk	Quality
	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		
TB risk or transmission	Prospective Interventional Cohort	Serious[1]	No serious inconsistency	No serious indirectness	Serious imprecision[2]	Classrooms had 5 to 6 air changes per hour (average sizes of 31 students and class volume of 180,000 litters or 180 m^3) Ventilation rate: 60.2% of students time was spent above the recommended threshold	VERY LOW

[1] Limitations in study design, unclear/lack of blinding, lack on information on confounders and how they were addressed, loss to follow up
[2] Uncertainty about the results due to low number of students participating (n=64)
Abbreviations: TB: tuberculosis

A.16 RQ CC and DD

A.16.1 Duration of isolation to minimise risk of infection to others

Length of Isolation

Study	Factor	Quality assessment					Number of patients	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Ritchie 2007 NZ	Length of isolation	observational	serious[1,]	Serious[2]	No serious indirectness[3]	serious imprecision[4]	143	1516 days saved	VERY LOW
Kalamuddin 2014 Singapore	Time spent in isolation	observational	serious[1,]	Serious[2]	No serious indirectness[3]	serious imprecision[4]	121	3 days vs 5 days *p* , 0.01	VERY LOW

[1] Unclear if outcome measurement blinded
[2] Heterogeneity in populations,
[3] Does not directly asses infectivity, and does not directly measures the outcome of interest
[4] Small sample size according to GRADE rule of thumb >300 events
Abbreviations: CI, confidence interval; OR, odds ratio; HR, hazard ratio

Number of sputum samples

Study	Factor	Quality assessment					Number of patients	Summary of findings	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Lippincott 2014 US	Xpert MTB/RIF strategy	observational	serious[1,]	Serious[2]	serious indirectness[3]	serious imprecision[4]	207 180 148	68hrs(IQR 47.1-97.5) smear 3-samples vs 2-samples 41.2 (IQR 26.6-54.8) and 3-samples 54.0 (IQR 43.3-80)	VERY LOW
Wilmer 2011 Canada	Third AFB smear	observational	serious[1,]	Serious[2]	serious indirectness[3]	serious imprecision[4]	116	Average delay for third specimen 0.95 days/patient	VERY LOW

[1] Unclear if outcome measurement blinded
[2] Heterogeneity in populations,
[3] Does not directly asses infectivity, and does not directly measures the outcome of interest
[4] Small sample size according to GRADE rule of thumb >300 events
Abbreviations: CI, confidence interval; OR, odds ratio; HR, hazard ratio

Appendix E: GRADE profiles

A.16.2 Determining level of infectiousness – time to sputum smear conversion

Age

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Rekha 2007 India	Age >45 yr	observational with multivariate analysis	very serious[1,2]	No serious	no serious indirectness	serious imprecision[3]	86	OR 1.8 (1.02 – 3.16)	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Small sample size and wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio; HR, hazard ratio

Sputum smear grade

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Bouti 2013 Morocco	Grade 3+	observational with multivariate analysis	very serious[1,2]	serious[3]	no serious indirectness	serious imprecision[4]	37	OR: 7.1 (2.5-11.2) [4]	VERY LOW
Horne 2010 USA	Grades 1+ to 4+	observational with multivariate analysis	very serious[1,2]	serious[3]	no serious indirectness	no serious imprecision[4]	98	HR: 0.45 (0.35-0.57)	VERY LOW
Rekha 2007 India	Higher pre-treatment grade (grades 2+ to 3+)	observational with multivariate analysis	very serious[1,2]	serious[3]	no serious indirectness	no serious imprecision[4]	157	OR 2.64 (1.76-3.96)	VERY LOW
Wang 2009 Taiwan	Grade 2+	observational with multivariate analysis	very serious[1,2]	serious[3]	no serious indirectness	no serious imprecision[4]	75	HR: 0.6 (0.43-0.84)	VERY LOW
	Grade 3+						72	HR: 0.47 (0.33-0.66)	
	Grade 4+ Reference: Grade 1+						82	HR: 0.5 (0.35-0.71)	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Heterogeneity in populations,
[4] Wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio; HR, hazard ratio

Miliary

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Bouti 2013 Morocco	Miliary	Observational with multivariate analysis	very serious[1,2]	No serious inconsistency	no serious indirectness	no serious imprecision[3]	ns	Adjusted OR: 8.8 (2.3-19.4)[3]	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Wide confidence interval
Abbreviations: CI, confidence interval; ns, no statistically significant (value no reported) ; OR, odds ratio

Two zones involved in X-ray

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Rekha 2007 India	>2 zones involved	observational with multivariate analysis	very serious[1,2]	No serious	no serious indirectness	serious imprecision[3]	179	1.31 (1.09 - 1.57)	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Small sample size and wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio; HR, hazard ratio

Bilateral radiological lesions

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Bouti 2013 Morocco	Bilateral radiological lesions	Observational with multivariate analysis	very serious[1,2,]	no serious inconsistency	no serious indirectness	Serious[3]	68	OR (95% CI) 13.4 (1.8-55.6) [3]	VERY LOW
Horne 2010 USA	Bilateral radiological lesions	Observational with multivariate analysis	very serious[1,2,]	no serious inconsistency	no serious indirectness	Serious[3]	43	Ns (values not reported)	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Sample size, and wide confidence interval
Abbreviations: CI, confidence interval; ns, no statistically significant; OR, odds ratio

Cavitation

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Bouti 2013 Morocco	Cavitation	Observational with multivariate analysis	very serious[1,2,]	serious inconsistency	no serious indirectness	Serious[3]	42	ns	VERY LOW
Horne 2010 USA	Cavitation	Observational with multivariate analysis	very serious[1,2,]	serious inconsistency	no serious indirectness	Serious[3]	44	ns	VERY LOW
Wang 2009 Taiwan	Cavitation	Observational with multivariate analysis	very serious[1,2,]	serious inconsistency	no serious indirectness	Serious[3]	85	HR 95% CI 0.26 (0.18-0.38)	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Sample size and wide confidence interval
Abbreviations: CI, confidence interval; ns, no statistically significant (values not reported) ; HR, hazard ratio

First two month regimen

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Wang 2009 Taiwan	Treatment interruption	Observational with multivariate analysis	very serious[1,2,3]	no serious inconsistency	no serious indirectness	Serious[3]	15	HR: 0.46 (0.27-0.79)	VERY LOW
	Other than HERZ Reference: HERZ						99	HR: 0.63 (0.53-0.87)	

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[5] Small sample size and wide confidence interval
Abbreviations: CI, confidence interval; HR, hazard ratio; HERZ, isoniazid, rifampicin, ethambutol and pyrazinamide

Drug Resistance

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Horne 2010 USA	Drug resistance	Observational with multivariate analysis	very serious[1,2,]	no serious inconsistency	no serious indirectness	serious [3]	22	HR: 2.30 (1.08-4.89)	VERY LOW
Wang 2009 Taiwan	Drug resistance	Observational with multivariate analysis	very serious[1,2,]	no serious inconsistency	no serious indirectness	serious [3]	48	ns	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Sample size and wide confidence interval
Abbreviations: CI, confidence interval; ns, no statistically significant (values not reported) ; HR, hazard ratio

A.16.3 Risk factors for continued risk of infection – time to culture conversion

Age

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Rekha 2007 India	Age >45 yr	observational with multivariate analysis	very serious[1,2]	No serious	no serious indirectness	serious imprecision[3]	67	OR 3.5 (1.56 – 7.84)	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Small sample size and wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio; HR, hazard ratio

2 zones involved in X-ray

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Rekha 2007	>2 zones involved	observational with multivariate	very serious[1,2]	No serious	no serious indirectness	serious imprecision[3]	152	OR 1.41 (1.04-1.90)	VERY LOW

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
India		analysis							

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Small sample size and wide confidence interval
Abbreviations: CI, confidence interval; OR, odds ratio; HR, hazard ratio

Culture grade

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Horne 2010 USA	Grades 1+ to 4+ scale	Observational with multivariate analysis	very serious[1,2,]	serious inconsistency[3]	no serious indirectness	serious imprecision[4]	98	HR: 0.52 (0.40-0.67)	VERY LOW
Rekha 2007 India	Higher pre-treatment grade (grades 2+ to 3+)	observational with multivariate analysis	very serious[1,2]	serious[3]	no serious indirectness	serious imprecision[4]	205	OR 3.5 (1.35-9.26)	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] measurement and sample heterogeneity
[4] Small sample size and wide confidence interval
Abbreviations: CI, confidence interval; HR, hazard ratio

Drug resistance

Study	Factor	Quality assessment					Number of patients	Summary of findings (95% CI)	Quality
		Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Horne 2010 USA	Drug resistance	Observational with multivariate analysis	very serious[1,2,]	No serious inconsistency	no serious indirectness	serious imprecision[3]	22	HR: 2.30 (1.02-5.21)	VERY LOW

[1] Unclear if prognostic factor and outcome measurement blinded
[2] Multivariate analysis used, but unclear which confounders were controlled for
[3] Small sample size and wide confidence interval
Abbreviations: CI, confidence interval; HR, hazard ratio

Appendix E: GRADE profiles

A.17 RQ HH

GRADE tables for outcome of risk of developing active tuberculosis in those diagnosed with latent tuberculosis

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of tuberculosis (Hazard ratio)	Relative (95% CI)	Quality	Importance
Radhakrishnan et al (assessed with: clinical and biochemical diagnosis) follow up adjusted for person years (follow up period 15 years)										
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	253,186 participant Infected= 3118 Hazard ratios • Not infected female child- 1.0 • Infected female child- 8.3 • Infected male child- 12.2 • Infected female adult- 15.8 • Infected male adult- 50.6 • No TB case at home- 1.0 • INH susceptible contact- 1.8 • INH resistant contact- 2.2	• Not infected female child: • Infected female child: (5.6-12.3) • Infected male child: (8.4-17.6) • Infected female adult: (11.0-22.7) • Infected male adult: (34.2-74.8) • No TB case at home: • INH susceptible contact: (1.4-2.2) • INH resistant contact: (1.5-3.3)	⊙○○○ VERY LOW	CRITICAL
Casado (follow-up median 43 months)										
1	observational studies	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	131 participants Hazard ratios • CD4 cell count (per each unit of increase)- 0.995 (P=0.06) • Persistence of predisposing factors for TB- 3.17 (P=0.0002)	• CD4 cell count (per each unit of increase)- (0.992-1.003) • Persistence of predisposing factors for TB- (1.56-17)	⊙○○○ VERY LOW	CRITICAL
Mori et al (case control)										
1	observational studies	very serious[4]	no serious inconsistency	serious[5]	Serious[6]	none[7]	Case n= 46, Control n=46 • Adjusted odds ratio • 6 or more months of isoniazid therapy- 0.02 • Alcohol abuse- 3.8 • Diabetes- 5.2	• Adjusted odds ratio • 6 or more months of isoniazid therapy- (0.002-0.16) • Alcohol abuse- (1.15-12.3) • Diabetes- (1.22-22.1)	⊙○○○ VERY LOW	CRITICAL
Leung et al (cohort)										
1	observational studies	serious[8]	no serious inconsistency	serious indirectness[9]	no serious imprecision	none	N=435 • Adjusted odds ratio • Number currently smoked per day • <10- 1.00 • 10-<20- 1.89 • ≥ 20- 2.54 Non-significant findings included age, past/current regular alcohol use, body mass	• Adjusted odds ratio • Number currently smoked per day • <10- reference • 10-<20- (1.19-5.05) • ≥ 20- (1.63-8.16)	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of tuberculosis (Hazard ratio)	Relative (95% CI)	Quality	Importance
							index, presence of other co-morbidities, BCG scar, tuberculin status/treatment of latent tuberculosis infection, principle job type, duration of silica dust exposure, profusion, size and shape of lung nodules and progressive massive fibrosis			
Martinez-Pino et al (cohort)										
1	observational studies	very serious[10]	no serious inconsistency	no serious indirectness	no serious imprecision	none	N=7902, 428 participants with latent TB • Adjusted odds ratio • Age • <35- reference • ≥35 years- 6.1 • Nadir CD4 • ≥200 cells/µl- reference • <200 cells/µl- 5.6 Non-significant variables included, gender, known date of HIV diagnosis, known start date of HAART[2], HAART[2] at TST[1], HAART[2] at TB diagnosis, ethnicity, education, socio-economic strata, previous incarceration, anti-HCV antibodies, HbsAg, CD4 cell count at enrolment, CD4 <200 cells/µl at enrolment, HIV viral load at enrolment. massive fibrosis.	• Adjusted odds ratio • Age • <35- reference • ≥35 years- (1.1-33.7) • Nadir CD4 • ≥200 cells/µl- reference • <200 cells/µl- (1.3-23.7)	⊙○○○ VERY LOW	CRITICAL
Di Perri et al (cohort)										
1	observational studies	serious[11]	no serious inconsistency	no serious indirectness	no serious imprecision	none	N=44 • Adjusted odds ratio • After multivariate analysis only • CD4 cell count and β-2 microglobulin serum levels retained statistical significance in the prognosis of developing active tuberculosis. Non-significant variables included, total lymphocytes	-	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of tuberculosis (Hazard ratio)	Relative (95% CI)	Quality	Importance
Antonucci et al (cohort)										
1	observational studies	no Serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	N=197 • Hazard ratio • CD4 >0.35 x 109/L- 5.49 • CD4 0.20–0.35 x 109/L- 14.78 • CD4 <0.20 x 109/L- 31.18	• Hazard ratio • CD4 >0.35 x 109/L-(1.32-27.09) • CD4 0.20–0.35 x 109/L- (3.49-62.63) • CD4 <0.20 x 109/L- (7.62-127.50)	⊙⊙○○ LOW	CRITICAL
Gessner et al (cohort)										
1	observational studies	very serious risk of bias[12]	no serious inconsistency	no serious indirectness	no serious imprecision	none	N=282 • Odds ratio • Left upper lobe lesion in adult- 12 • Alaska native child- 8.9 • Adult is parent of child- 8.3 • Age of child- 1.5 Non-significant variables included, 3 or 4+ culture positive adults, 3 or 4+ smear positive adults, gender	• Odds ratio • Left upper lobe lesion in adult- (2.2–65) • Alaska native child- (1.1–73) • Adult is parent of child- (1.6–44) • Age of child- (1.1–2.0)	⊙○○○ VERY LOW	CRITICAL

[1] Unclear if cohorts were matched for the amount that recieved BCG vaccination or placebo in the initial randomised clinical trial. Cohort was significantly older in persons in households without a TB case. Isoniazid susceptible cohort had the lowest proportion of males. Isoniazid resistant cohort had the highest proportion infected. Therefore variance was not spread evenly between groups.

[2] Unclear if all patients recieved the same standard of care. Definition of outcome was unclear: persistence of predisposing conditions for TB infection was highlighted as the main risk factor with no attempt to break down the data any further. Unclear if valid and reliable method used to determine outcome.

[3] Low number of participants (n=131)

[4] The study does not ask a clearly focused question: It attempts to illicit the benefit of isoniazid preventive therapy in those that are tuberculin reactors however some non-reactors were also included in the analysis thereby confounding the study data. Also since documented TST[1] reactors are more likely to be offered chemoprophylaxis, the control group is likely to overestimate the proportion of latently infected people in the population who receive preventive therapy. The data on risk factors for developing tuberculosis is more useful but still confounded by the presence of non-TST[1] reactors in the case group. The cases and controls are taken from comparable populations, however, control patients were found to be more compliant to treatment when compared to tuberculosis cases. No measures appear to have been taken to prevent knowledge of primary exposure(s) from influencing case ascertainment. Exposure to diabetes may have not been measured in a standard and reliable fashion since patients with high random or fasting blood glucose recordings were listed as being diabetic, however British guidelines require more than just one isolated raised blood glucose level. Chart documentation supplied many of the other diagnosis such as notation of alcohol abuse or admissions related to alcoholism which may not have been accurate. Unclear how long participant's histories were tracked for.

[5] Population does not exactly match population of interest: Native American people were enrolled; these people have an incidence of TB two to three times that of the surrounding populations. Not all patients in the active tuberculosis group had a documented positive TST[1] test prior to TB diagnosis. 1 had a negative TST[1] and 8 had an unknown infection status.

[6] Number of participants was small (n=92)

[7] funding was unclear

[8] The sample included those who had an induration less than 10 mm however tuberculin status was later adjusted for in multivariate analysis. Patients did not receive the same standard of care for latent TB as some were treated and others were not. Information on the number treated and on which treatment regimen is provided. Patients were also seen in differing clinics with potential for variance in standard of care. Adjustments for treatment of latent tuberculosis were attempted in multivariate analysis. Data was recorded by questionnaire which is vulnerable to recall bias.

[9] The population was amongst male high risk silicotic patients in Hong Kong, there may be some generalizability issues here

[10] The patients may not have received same standard of care since participants spread over 20 different hospitals. Comparisons in baseline characteristics were not made between those that accepted treatment and those who refused to initiate therapy. Unclear why CD4 count at registration<200 vs. ≥200 cells/µl was not included in final multivariate analysis when it was significant at the univariate level. There were clear differences in populations at baseline between those who had no TB, prevalent TB and incident TB. Information on TST[1] was not available for 4848 patients. Compared with patients with available TST[1] results, these patients were more likely to have had no education or only primary education (61.8% vs 49.1%), to be of lower socio-economic status (50.5% vs 40.2%) and to have a CD4 cell count of <200 cells/µl at enrolment (18.4% vs 14.3%, P=<0.001). No information on treatment adherence was provided either for those who received isoniazid or those who received

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of tuberculosis (Hazard ratio)	Relative (95% CI)	Quality	Importance

HAART[2] therapy.

[11] *Participants received the same standard of care in regard to monitoring however immunological evaluation was performed at baseline and subsequently at 3-6 month intervals, leaving some uncertainty about consistency of monitoring tests. Multivariate analysis was performed using the Cox model. However the study has failed to adjust for external risk factors that may be relevant such as malnutrition, alcoholism, homelessness and drug dependence.*

[12] *It was stated that once diagnosed infected children were treated however it is unclear under what regimen they were treated for latent tuberculosis and whether all received the same standard of care. Unclear if the 30 villages in the area performed the same level of monitoring or care for the children and the infected adults. Few baseline characteristics are reported. The methods used to observe risk factors are unlikely to be reliable as data was recorded retrospectively. Definition of diagnosis of active and latent tuberculosis was not stated in full and the methods used to observe risk factors are unlikely to be reliable as data was recorded retrospectively. Observation period was for 7 years between 1987 and 1994. Unclear if length of observation was the same for all children (or if adjustments were made).*

Risk of developing hepatotoxicity for those receiving treatment for latent tuberculosis

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of Hepatotoxicity	Relative (95% CI)	Quality	Importance
Tedla et al (n=1,995)										
1	observational studies	no serious risk of bias	no serious inconsistency	no serious risk of indirectness1	no serious imprecision	none	Relative risk • CD4 lymphocyte count • CD4 <200 cells/mm³- 2.80 • CD4 ≥200 cells/mm³- 1.00 Non-significant variables Age, sex, BMI, antiretroviral therapy, efavirenz, nevirapine, NNRTI, co-trimoxazole, alcohol, alcohol-dependence, Hepatitis B viral serological testing	Relative risk • CD4 lymphocyte count • CD4 <200 cells/mm³- 2.80 (1.14-6.84) • CD4 ≥200 cells/mm³- 1.00	⊙⊙○○ LOW	CRITICAL
Fountain et al (assessed with: AST levels greater than 5 times upper limit of normal) (n=3,377)										
1	observational studies	Serious[6]	no serious inconsistency	no serious indirectness	no serious imprecision	none[7]	Adjusted odds ratio • Baseline AST > upper limit of normal- 5.398 • Age ≥ 50 years- 3.699	Adjusted odds ratio • Baseline AST > upper limit of normal- (2.081-13.999) • Age ≥ 50 years- (1.428-9.584)	⊙○○○ VERY LOW	CRITICAL
Fernandez-Villar (n=415)										
1	observational studies	serious[8]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Adjusted odds ratios • Excessive alcohol consumption- 4.2 • Baseline abnormal ALT- 4.3 (odds ratios calculated by comparing to those who did not have any of the above)	Adjusted odds ratios • Excessive alcohol consumption- (1.6-10.8) • Baseline abnormal ALT- (1.6-11.4)	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of Hepatotoxicity	Relative (95% CI)	Quality	Importance
Nolan et al (n=11,141)										
1	observational studies	serious[9]	no serious inconsistency	no serious indirectness	no serious imprecision	none[10]	Adjusted odds ratios • Non-significant variables were: • Sex, Age and Race	-	⊙○○○ VERY LOW	CRITICAL
Dickinson et al, 1981										
1	observational studies	serious[11]	no serious inconsistency	no serious indirectness	serious[12]	none	Only age was found to be significantly correlated with liver dysfunction after adjustment for all other factors- (P= 0.034) • Non-significant variables were: • Rapid/slow acetylation phenotype, sex and race	-	⊙○○○ VERY LOW	CRITICAL
Lee et al (n=3788)										
1	observational studies	serious[13]	no serious inconsistency	no serious indirectness	serious[12]	none[7]	Odds ratio • Gender • Female- 4.1 • Male- reference Non-significant variables were: Race, age, alcohol use, illicit drug use, pyrazinamide dose, presumed recent infection	• Gender • Female- (1.2-14.3) • Male- reference	⊙○○○ VERY LOW	CRITICAL
Lobato et al (N= 1,246)										
1	observational studies	no serious	no serious inconsistency	no serious indirectness	no serious risk	none[7]	Adjusted odds ratio • Age- 0.97 • Unemployed within past 24 months- 0.51 • Elevated AST before therapy- 0.72 Non-significant findings included: Sex, US birth, race, homelessness, prior positive tuberculin skin test, previous incarceration, injection drug use, non-injection drug use, excess alcohol	• Age- (0.95-0.99) • Unemployed- (0.27-0.97) • Elevated AST- (0.54-0.95)	⊙⊙○○ LOW	CRITICAL
Vinnard et al (N= 219)										
1	observational studies	very serious[14]	no serious inconsistency	no serious indirectness	no serious risk	none[7]	Hazard ratio • Hepatitis C- 3.03 Age was not associated with treatment discontinuation due to suspected toxicity.	Hazard ratio • Hepatitis C- 1.08-8.52	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of Hepatotoxicity	Relative (95% CI)	Quality	Importance
Smith et al (retrospective cohort) (n=9145)										
1	observational studies	serious[15]	no serious inconsistency	no serious indirectness	no serious	none	Independent variables associated with subsequent hepatic events following treatment for latent tuberculosis infection include: • Hospital admission • Any physician visits for liver disease • High Charlson comorbidity score during the 6 months before treatment initiation • Age stratified adjusted odds ratio • ≤ 35- 1.00 (reference) • 36-50- 2.7 • 51-65- 5.7 • >65- 34.2	• Age stratified adjusted odds ratio • ≤ 35- (reference) • 36-50- (0.5-16.0) • 51-65- (1.0-33.7) • >65- (7.6-153.8)	⊙○○○ VERY LOW	CRITICAL

[1] *Participants were from Botswana, however immigration into England from Africa may produce similar TB cohorts in the UK. Participants were HIV infected.*

[2] *The study does not ask a clearly focused question: It attempts to illicit the benefit of isoniazid preventive therapy in those that are tuberculin reactors however some non-reactors were also included in the analysis thereby confounding the study data. Also since documented TST[1] reactors are more likely to be offered chemoprophylaxis, the control group is likely to overestimate the proportion of latently infected people in the population who receive preventive therapy. The data on risk factors for developing tuberculosis is more useful but still confounded by the presence of non-TST[1] reactors in the case group. The cases and controls are taken from comparable populations, however, control patients were found to be more compliant to treatment when compared to tuberculosis cases. No measures appear to have been taken to prevent knowledge of primary exposure(s) from influencing case ascertainment. Exposure to diabetes may have not been measured in a standard and reliable fashion since patients with high random or fasting blood glucose recordings were listed as being diabetic, however British guidelines require more than just one isolated raised blood glucose level. Chart documentation supplied many of the other diagnosis such as notation of alcohol abuse or admissions related to alcoholism which may not have been accurate. Unclear how long participant's histories were tracked for.*

[3] *Population does not exactly match population of interest: Native American people were enrolled; these people have an incidence of TB two to three times that of the surrounding populations. Not all patients in the active tuberculosis group had a documented positive TST[1] test prior to TB diagnosis. 1 had a negative TST[1] and 8 had an unknown infection status.*

[4] *Number of participants was small (n=92)*

[5] *funding was unclear*

[6] *Patients did not receive the same level of care as participants within the last three years of the study were given 9 months of isoniazid instead of 6 months. Follow up did not include the last 3 months of treatment. Treatment completion was poor across the board with only 43.13 % of patients completing 3 months of therapy.*

[7] *unclear who provided funding for this study*

[8] *Patients did not receive the same level of care as rules regarding monitoring adherence; some of the participants were enrolled in a methadone maintenance therapy programme where isoniazid was administered alongside. Others had their adherence monitored by means of pill count, urine samples and family supervision. Follow up did not appear to continue beyond treatment period.Treatment completion was fairly low with 76.9% of patients completing.*

[9] *Treatment completion was fairly low with 64% of patients completing 6 months of therapy. Attempts to find the systematic differences between those who did or did not complete treatment were not made. 84% of patients on the multidrug therapy arm completed therapy. Dose and length of treatment was unclear and may vary. Method of diagnosis was based on the assumption that all hepatotoxic patients would be symptomatic, subclinical cases would have been missed.*

[10] *unclear source of funding*

[11] *Population does not exactly match population of interest. Participants included 36 who were PPD[1] negative and therefore potentially not latently infected. These patients recieved a shorter duration of treatment. Follow up did not extend beyond treatment period. Treatment completion was low.*

[12] *low population*

[13] *Patients did not receive the same standard of care as rifabutin was substituted for rifampicin in HIV positive patients on protease inhibitors or non-nucleoside reverse transcriptase inhibitors. Doses of*

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of Hepatotoxicity	Relative (95% CI)	Quality	Importance
rifampicin and pyrazinamide initially followed guidelines established for HIV infected patients and those with active tuberculosis but dose of pyrazinamide was subsequently limited based on an expert opinion published in the American Thoracic Society guidelines. Follow up did not appear to continue beyond treatment period. Treatment completion was low [14] *Patients appear to have received a great variety of different standards of care. Variability included testing for comorbidities, number of isoniazid tablets provided per prescription and frequency of follow up visits. The proportion of patients in the cohort without testing for important comorbidities was not determined. Women were under-represented in this study. Treatment completion was low: 46% of veterans who initiated latent tuberculosis therapy completed treatment satisfactorily. Comparisons were not made between those that accepted treatment and those who refused to initiate therapy. Data was not available for why 46% of patients discontinued treatment. Uncertain how many variables were included in multivariate analysis. data was gathered by retrospectively examining clinical charts which is unlikely to be reliable. Definition of treatment completion outcome was unclear. Also ALT[1] levels were available for only 84% of the participants at baseline and 71% of the participants during therapy which meant diagnosis of hepatotoxicity was reliant upon the clinician reporting this is both unclear and unreliable. Baseline characteristics were not provided for all patients* [15] *They took patients receiving the treatment for latent for tuberculosis as having had latent tuberculosis when this may not have been the case. This is an indirect definition of latent tuberculosis. Definition of risk factors was clear but unlikely to be reliable since this was a retrospective study and data was retrieved from administrative health data.*										

GRADE summary for those at risk of developing adverse events as a result of treatment for latent tuberculosis

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of adverse events	Relative (95% CI) Absolute	Quality	Importance
Lobue et al (n=3,788)										
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Odds ratio • Gender • M- reference • F- 1.6 Age • 0-14- reference • 15-34- 1.3 • 35-49- 1.8 • 50-64- 2.2 • 65+- 1.5 Homeless • N- reference • Y- 2.2 Correctional facility • N- reference • Y- 2.6 The occurrence of hepatotoxicity was also associated with self-reported intravenous drug use	Odds ratio • Gender • M- • F- (1.4-2.0) Age • 0-14- reference • 15-34- (1.0-1.6) • 35-49- (1.3-2.5) • 50-64- (1.3-3.8) • 65+- (0.6-3.2) Homeless • N- reference • Y- (1.2-4.2) Correctional facility • N- reference • Y- (1.5-4.5)	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Risk of adverse events	Relative (95% CI) Absolute	Quality	Importance
Pettit et al (cohort) (n=1323)										
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	None	Isoniazid discontinuation due to adverse events Adjusted relative risk • Female sex- 1.67 • Current alcohol use- 1.14	 • Female sex- (1.32-2.10) • Current alcohol use- (1.13-1.77)	⊙○○○ VERY LOW	CRITICAL

[1] Patients did not receive the same level of care as rules regarding monitoring were altered during the study due to changes in American Thoracic Society Guidelines. Initially all patients over 35 were monitored with monthly transaminase levels as well as those at higher risk of hepatotoxicity; later this was changed to only those at higher risk. Follow up did not exceed treatment period.Treatment completion was poor with only 64% of patients completing 6 months of therapy. The paper does not provide the exact doses and lengths of regimens used.

[2] Adjusted relative risk was adjusted for study site, sex and current alcohol use. No other significant factors appear to have been adjusted for. Methods used to record the risk were generally reliable and valid although taken from in-person interviews which may have been subject to recall bias especially the factors of alcohol and substance use. Reasons for treatment default were taken second hand from medical charts which may not have been reliable. 15% of participants were lost to follow up.

GRADE summary for those at risk of non-completion of treatment for latent tuberculosis

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Completion of treatment	Relative (95% CI) Absolute	Quality	Importance
Gilroy (assessed with: completion of 6 months isoniazid therapy) (n=335)										
1	observational studies	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None[3]	Only ALT level at baseline was statistically significant for non-completion after adjustment for other variables Non-significant variables recorded included: Gender, ethnicity, alcohol use and number of medications regularly taken.	-	⊙○○○ VERY LOW	CRITICAL
Lobue et al (at risk for lower completion rates of anti-tuberculosis regimen) (3788)										
1	observational studies	serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Odds ratio Risk of lower completion rates Self-reported excess alcohol use • N- reference • Y- 0.1 Homelessness • N- reference • Y- 0.2	Odds ratio Risk of lower completion rates Self-reported excess alcohol use • N- reference • Y- (0.0-0.6) Homelessness • N- reference	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Completion of treatment	Relative (95% CI) Absolute	Quality	Importance
							Any other adverse event (not hep tox) • N- reference • Y- 0.8 Higher completion rates were associated with female sex, younger age groups, white/Hispanic race and non-USA country of birth.	• Y- (0.1-0.5) Any other adverse event • N- reference • Y- (0.7-0.9)		
Lobato et al (at risk for lower completion rates of anti-tuberculosis regimen) (n=1,246)										
1	observational studies	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	Adjusted odds ratio • Female sex- 0.35 • Hispanic ethnicity- 0.59 • Unemployed- 1.43 • Injection drug use within past 12 months- 0.54 • Excess alcohol- 1.35	Adjusted odds ratio • Female sex- (0.23-0.54) • Hispanic ethnicity- (0.46-0.75) • Unemployed- (1.07-1.90) • Injection drug use- (0.31-0.95) • Excess alcohol- (1.04-1.76)	⊙⊙○○ LOW	CRITICAL
Oni et al (cohort) (n=164)										
1	observational studies	serious risk of bias[5]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Adjusted odds ratio • Time since HIV diagnosis: 0.81; 0.68-0.98 • Alcohol drinkers: OR 4.05; 1.89-9.06	• Time since HIV diagnosis: (0.68-0.98) • Alcohol drinkers: (1.89-9.06)	⊙○○○ VERY LOW	CRITICAL
Goswami et al (cohort) (n=496)										
1	observational studies	serious risk of bias[6]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Relative risk Risk for initiating therapy: • Close contact to a TB case- 2.5 • Non-employment reason for screening- 1.6 • Lower educational level- 1.3 • Having a regular physician- 1.4 • Fear of getting sick with TB without medicine- 1.7 • Prior incarceration- 1.7 • Risk for treatment completion: • Plan to tell friends or family about latent tuberculosis diagnosis.- 2.0	Relative risk Risk for initiating therapy: • Close contact to a TB case- 1.8-3.6 • Non-employment reason for screening- 1.0-2.5 • Lower educational level- 1.1-1.6 • Having a regular physician- 1.0-2.0 • Fear of getting sick with TB without medicine- 1.2-2.6 • Prior incarceration- 1.1-2.8 • Risk for treatment completion: • Plan to tell friends or family about latent tuberculosis diagnosis.- 1.0-3.9	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Completion of treatment	Relative (95% CI) Absolute	Quality	Importance
Anibarro et al (retrospective cohort) (n=599)										
1	observational studies	serious risk of bias[6]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Adjusted odds ratios Risk for treatment completion: • <36- 0.33 • ≥36-1 Sex • M-0.58 • F- 1 Immigrant (<5 years residence) • Y- 0.21 • N- 1 Social risk factors (unemployment, alcohol abuse, illegal drug abuse or residence in a correctional facility) • Y- 0.21 • N- 1	Adjusted odds ratios Risk for treatment completion: • <36- (0.30-0.76) • ≥36- 1 Sex • M- 0.37-0.92) • F Immigrant (<5 years residence) • Y- (0.12-0.37) • N- 1 Social risk factors • Y- (0.11-0.39) • N -1	⊙○○○ VERY LOW	CRITICAL
Li et al (retrospective cohort) (n=15,035)										
1	observational studies	serious risk of bias[8]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Adjusted odds ratios Risk for treatment completion: Age, years • <18- NS • 18-24- NS • 25-35- reference • ≥35- 1.16 Race/ethnicity • Asian- 1.20 • Non-Hispanic black- 1.11 • Non-Hispanic white- reference • Hispanic- 1.10 • Other/unknown- NS Country of birth • Non-US-born- 1.08 • US-born- reference	Adjusted odds ratios Risk for treatment completion: Age, years • <18- NS • 18-24- NS • 25-35- reference • ≥35- (1.11-1.22) Race/ethnicity • Asian- (1.10-1.30) • Non-Hispanic black- (1.02-1.19) • Non-Hispanic white- reference • Hispanic- (1.02-1.19) • Other/unknown- (0.92-1.11) Country of birth • Non-US-born (1.03-1.13) • US-born- reference	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Completion of treatment	Relative (95% CI) Absolute	Quality	Importance
							Risk group • Contact- 1.51 • Medical risk- 1.45 • Population risk- 1.16 • Low risk- reference Ever on directly observed preventive therapy • Yes- 1.26 • No- reference Treatment regimen Isoniazid alone- reference • Rifamycin alone- 1.20	Risk group • Contact- (1.38-1.66) • Medical risk- (1.32-1.60) • Population risk- (1.07-1.27) • Low risk- reference Ever on directly observed preventive therapy • Yes- (1.18-1.34) • No- reference Treatment regimen Isoniazid alone- reference • Rifamycin alone- 1.20 (1.14-1.26)		
Machado et al (cohort) (n=101)										
1	observational studies	serious risk of bias[9]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Relative risk Report of adverse effect- 2.69 Distance to health centre • 0-5- reference • 5.1-10- NS • >10- 0.39 Number of buses required to commute • 1- reference • 2- 1.84	Adjusted odds ratios Report of adverse effect- (1.3-5.8) Distance to health centre • 0-5- • 5.1- • >10- (0.2-0.8) Number of buses required to commute • 1- reference • 2- (1.0-3.3)	⊙○○○ VERY LOW	CRITICAL
Kwara et al (retrospective cohort) (n=672)										
1	observational studies	serious risk of bias[10]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Odds ratio Report of adverse effect • N- reference • Y- 3.6 Medical insurance • Y-reference • N- 1.7 Non-significant variables included age, and being postpartum.	Odds ratios Report of adverse effect • N- reference • Y- (2.2-6.2) Medical insurance • Y-reference • N- (1.1-2.7)	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Completion of treatment	Relative (95% CI) Absolute	Quality	Importance
Haley et al (retrospective cohort) (n=749)										
1	observational studies	serious risk of bias[11]	no serious inconsistency	no serious indirectness	no serious imprecision	none	Adjusted odds ratio • Hispanic subjects (n=534) • Contact with an infectious TB case- 3.7 • Alcohol use reported at baseline- 1.7 • Other medications reported at baseline- 2.2 • Non-Hispanic subjects (n=215) • Black race- 2.6 • Age- 0.97 • Foreign birth- 0.5 Non-significant findings included work or residence in a correctional facility in past year	Adjusted odds ratio • Hispanic subjects (n=534) • Contact with an infectious TB case- (1.8-7.4) • Alcohol use reported at baseline- (1.1-2.8) • Other medications reported at baseline- (1.3-3.8) • Non-Hispanic subjects (n=215) • Black race- (1.5-4.7) • Age- 0.94-0.99) • Foreign birth- (0.2-0.9)	⊙○○○ VERY LOW	CRITICAL

[1] *follow up did not appear to go beyond treatment period. Treatment completion was low 76% of participants completing. This was a retrospective study taken from the charts of patients and therefore likely to incurr recording bias. Definition of treatment completion was unclear as patients were assumed to be compliant if they kept monthly appointments at clinic.*

[2] *population was low at 335 patients*

[3] *source of funding was unclear*

[4] *Patients did not receive the same level of care as rules regarding monitoring were altered during the study due to changes in American Thoracic Society Guidelines. Initially all patients over 35 were monitored with monthly transaminase levels as well as those at higher risk of hepatotoxicity; later this was changed to only those at higher risk. Follow up did not exceed treatment period.Treatment completion was poor with only 64% of patients completing 6 months of therapy. The paper does not provide the exact doses and lengths of regimens used.*

[5] *Multivariate analysis was used however the significant factor of smoking was not included in the multivariate analysis model as the alcohol variable provided a better fit of the model instead. It is unclear why all significant factors could not have been included. Definition of risk factors was unlikely to be valid or reliable since alcohol use and smoking was self-reported, as were other important factors.A valid and reliable method of measurement was not used as patients were assumed to be compliant if they kept monthly appointments at the clinic and self-reported adherence.*

[6] *Unclear if the type of preventive therapy used was included in multivariate analysis. Some patients were taking 4 months of rifampicin, some were taking 9 months of isoniazid. Completion rate of isoniazid participants was 52%, completion rate in those treated with rifampicin was 61% (p=0.3). At least six months of isoniazid was completed by 63% of participants. Definition of risk factors was clear but unlikely to be valid or reliable since all risk factors were self-reported at baseline.*

[7] *Definition of risk factors was clear but unlikely to be reliable since this was a retrospective study and data was retrieved from administrative health data. Definition of treatment completion outcome was clear but also reliant upon retrospective data. Due to differences in the methods of evaluating adherence on the different hospital sites treatment completion was chosen as an endpoint instead. Participants did not receive the same level of care apart from intervention studied as different participants were taking different drugs in various combinations with different durations. Patients on one hospital site also received urine tests at every visit which may have improved adherence as patients knew they would be tested*

[8] *There was no attempt to adjust for the differing types of dosing schedules in the isoniazid group, or for the patients taking rifabutin or rifampicin in the rifamycin group. Definition of treatment completion outcome was clear but also reliant upon retrospective data. Different methods of evaluating adherence was used depending on the age and regimen of the participant: patients aged >18 years were considered to have completed treatment if they took 6-9 months of isoniazid daily or twice weekly within a 9-12 month period; or > 4 months of daily rifamycin doses within 6 months. Patients younger than 18 years were considered to have completed treatment if they had taken 9 or more months of daily or twice weekly isoniazid therapy within a 12 month period, or 6 or more months of daily rifamycin therapy within 9 months. Outcome measure was not reliable as there was no guarantee that patients were taking their medication despite regular attendance at clinic to pick up their monthly supply of medications.*

[9] *Definition of risk factors was clear but unlikely to be reliable since number of buses required to commute was discovered by asking the transportation agency rather than the patients themselves who may have another means of transport. Data was gathered by questionnaire. Definition of treatment completion outcome was clear but may be unreliable since the patient was judged to be adherent on the basis of attending monthly appointments and picking up pills; it is uncertain if patients were actually taking the pills.*

Quality assessment							No of patients	Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Completion of treatment	Relative (95% CI) Absolute	Quality	Importance

[10] *Definition of risk factors was mostly clear however the definition of "medical risk factor," wasn't. Data is unlikely to be reliable since it was obtained by looking retrospectively at medical records. Definition of treatment completion outcome was clear but may be unreliable since the patient was judged to be adherent on the basis of attending monthly appointments and picking up pills; it is uncertain if patients were actually taking the pills. Patients did not necessarily receive the same standard of care due to increased clinical monitoring and blood tests in certain age groups of patients in accordance to guidelines.*

[11] *Definition of risk factors was clear however data is unlikely to be reliable since it was obtained by looking retrospectively at medical records. Definition of treatment completion outcome was clear but may be unreliable since the patient was judged to be adherent on the basis of attending monthly appointments and picking up pills; it is uncertain if patients were actually taking the pills. Data was also retrospective.*

A.18 RQ II

Author(s):
Date: 2014-03-03
Question: Should 3 months isoniazid vs 3 months placebo be used for latent tuberculosis?
Settings: Czechoslovakia, Finland, Germany, Hungary, Poland, Romania, Yugoslavia
Bibliography:

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months isoniazid	3 months placebo	Relative (95% CI)	Absolute	Quality	Importance
Incidence of active tuberculosis (follow-up median 5 years[1]; assessed with: Clinical diagnosis and biomedical testing)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	75/6956 (1.1%)	97/6990 (1.4%) 0%	-	14 fewer per 1000 (from 14 fewer to 14 fewer)	⊙⊙○○ LOW	CRITICAL

[1] *No average provided, however five year follow up was complete for 97.2% of the population.*
[2] *Unclear for how many participants in each group were no outcome data available or whether groups were comparable with respect to systematic differences between groups in terms of those for whom no outcome data was available. Also more patients were lost in the longer duration treatment regimens.*
[3] *Number of events less than 300*

Author(s):
Date: 2014-03-03
Question: Should 6 months isoniazid vs 6 months placebo be used for latent TB?
Settings:
Bibliography:

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	6 months isoniazid	6 months placebo	Relative (95% CI)	Absolute	Quality	Importance
Incidence of active tuberculosis (follow-up median 5 years[1]; assessed with: clinical and biomedical diagnosis)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	34/6965 (0.49%)	97/6990 (1.4%) 0%	-	14 fewer per 1000 (from 14 fewer to 14 fewer)	⊙⊙○○ LOW	CRITICAL

[1] *No average provided however follow up was completed for 97.2% of participants*
[2] *unclear if groups were comparable for availability of outcome data*
[3] *Number of events below 300*

Author(s):
Date: 2014-03-04
Question: Should 12 months Isoniazid vs 12 months placebo be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	12 months Isoniazid	12 months placebo	Relative (95% CI)	Absolute		
Incidence of active TB (follow-up median 5 years[1]; assessed with: Clinical and biomedical diagnosis)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	None	24/6919 (0.35%)	97/6990 (1.4%) 0%	-	14 fewer per 1000 (from 14 fewer to 14 fewer)	⊙⊙○○ LOW	CRITICAL

[1] *No average provided however follow up was completed for 97.2% of participants.*
[2] *Unclear is groups were comparable for availability of outcome data*
[3] *Event number below 300*

Author(s):
Date: 2014-03-04
Question: Should 3 months isoniazid vs no treatment be used for latent TB?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months isoniazid	No treatment	Relative (95% CI)	Absolute		
Incidence of active TB (follow-up 8 years; assessed with: Clinical and biochemical diagnosis)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none[3]	10/82 (12.2%)	17/85 (20%) 0%	-	200 fewer per 1000 (from 200 fewer to 200 fewer)	⊙○○○ VERY LOW	CRITICAL

[1] *Unclear if an appropriate method of randomisation or allocation concealment was used. No blinding was employed. Unclear how groups were comparable for length of follow up or availability of outcome data. No precise definition of outcome. Unclear if a valid and reliable method was used to determine outcome.*
[2] *events less than 300*
[3] *However no information given on funding*

Author(s):
Date: 2014-03-04
Question: Should 1 month isoniazid and rifampicin vs no treatment be used for latent TB?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	1 month isoniazid and rifampicin	No treatment	Relative (95% CI)	Absolute		
Incidence of active tuberculosis (follow-up 8 years; assessed with: clinical and biochemical diagnosis)												
1	randomised trials	very serious	no serious inconsistency	no serious indirectness	serious[1]	none[2]	9/83 (10.8%)	17/85 (20%) 0%	-	200 fewer per 1000 (from 200 fewer to 200 fewer)	⊙○○○ VERY LOW	CRITICAL

[1] *event number less than 300*
[2] *no information on funding provided*

Author(s):
Date: 2014-03-04
Question: Should 3 months isoniazid and rifampicin vs no treatment be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months isoniazid and rifampicin	No treatment	Relative (95% CI)	Absolute		
Incidence of Active Tuberculosis (follow-up 8 years; assessed with: clinical and biochemical diagnosis)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none[3]	4/85 (4.7%)	17/85 (20%) 0%	-	200 fewer per 1000 (from 200 fewer to 200 fewer)	⊙○○○ VERY LOW	CRITICAL

[1] *Unclear if an appropriate method of randomisation or allocation concealment was used. No blinding was employed. Unclear how groups were comparable for length of follow up or availability of outcome data. No precise definition of outcome. Unclear if a valid and reliable method was used to determine outcome.*
[2] *Event number less than 300*
[3] *no information on funding provided*

Author(s):
Date: 2014-03-04
Question: Should 1 month isoniazid, pyrazinamide and rifampicin vs no treatment be used for latent TB?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	1 month isoniazid, pyrazinamide and rifampicin	No treatment	Relative (95% CI)	Absolute		
Incidence of active tuberculosis (follow-up 8 years; assessed with: clinical and biochemical diagnosis)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none[3]	0/80 (0%)	17/85 (20%) 0%	-	200 fewer per 1000 (from 200 fewer to 200 fewer)	⊙○○○ VERY LOW	CRITICAL

[1] Unclear if an appropriate method of randomisation or allocation concealment was used. No blinding was employed. Unclear how groups were comparable for length of follow up or availability of outcome data. No precise definition of outcome. Unclear if a valid and reliable method was used to determine outcome.
[2] event rate less than 300
[3] no information provided on funding

Author(s):
Date: 2014-03-04
Question: Should 36 months isoniazid vs 6 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	36 months isoniazid	6 months isoniazid	Relative (95% CI)	Absolute		
Incidence of active tuberculosis (follow-up 3 years[1]; assessed with: clinical and biochemical diagnosis)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	4/252 (1.6%)	12/216 (5.6%) 0%	-	56 fewer per 1000 (from 56 fewer to 56 fewer)	⊙⊙⊙○ MODERATE	CRITICAL
Mortality (follow-up 3 years[1]; assessed with: number of deaths)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	5/252 (2%)	13/216 (6%) 0%	-	60 fewer per 1000 (from 60 fewer to 60 fewer)	⊙⊙⊙○ MODERATE	CRITICAL

[1] No follow up beyond 3 year treatment period
[2] event rates less than 300

Author(s):
Date: 2014-03-04
Question: Should 4 months rifampicin vs 6 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	4 months rifampicin	6 months isoniazid	Relative (95% CI)	Absolute		
Incidence of adverse events leading to discontinuation (follow-up 1 months; assessed with: Any adverse event leading to permanent discontinuation of treatment.)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	4/190 (2.1%)	22/183 (12%) 0%	-	120 fewer per 1000 (from 120 fewer to 120 fewer)	⊙⊙○○ LOW	CRITICAL
Treatment completion (follow-up 1 months; assessed with: Number of patients who completed treatment)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	163/190 (85.8%)	142/183 (77.6%) 0%	-	776 fewer per 1000 (from 776 fewer to 776 fewer)	⊙⊙○○ LOW	CRITICAL

[1] *Neither participants nor clinicians were blinded to treatment allocation. Unclear if there was adequate concealment of allocation. Unclear if groups were comparable for numbers that did not complete treatment.*
[2] *event numbers less than 300*

Author(s):
Date: 2014-03-04
Question: Should 3 months rifapentine and isoniazid vs 6 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months rifapentine and isoniazid	6 months isoniazid	Relative (95% CI)	Absolute		
Incidence of active tuberculosis (follow-up 3-6 years; assessed with: clinical and biochemical presentation)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	24/328 (7.3%)	22/327 (6.7%) 0%	-	67 fewer per 1000 (from 67 fewer to 67 fewer)	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months rifapentine and isoniazid	6 months isoniazid	Relative (95% CI)	Absolute		
Mortality (follow-up 3-6 years; assessed with: number of deaths)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	17/328 (5.2%)	25/327 (7.6%) 0%	-	76 fewer per 1000 (from 76 fewer to 76 fewer)	⊙○○○ VERY LOW	CRITICAL
Hepatotoxicity (follow-up 3-6 years; assessed with: a grade 3 or 4 elevation in the aminotransferase levels)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None	5/328 (1.5%)[3]	18/327 (5.5%)[3] 0%	-	55 fewer per 1000 (from 55 fewer to 55 fewer)	⊙○○○ VERY LOW	CRITICAL

[1] Neither participants nor clinicians were kept blinded to treatment regimen. Isoniazid alone treatment was self administered while other treatments were directly observed therapy.
[2] event number less than 300
[3] calculated from percentages

Author(s):
Date: 2014-03-04
Question: Should 3 months rifampicin and isoniazid vs 6 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months rifampicin and isoniazid	6 months isoniazid	Relative (95% CI)	Absolute		
Incidence of active tuberculosis (follow-up 3-6 years; assessed with: clinical and biochemical diagnosis)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None	24/329 (7.3%)	22/327 (6.7%) 0%	-	67 fewer per 1000 (from 67 fewer to 67 fewer)	⊙○○○ VERY LOW	CRITICAL
Mortality (follow-up 3-6 years; assessed with: number of deaths)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None	16/329 (4.9%)	25/327 (7.6%) 0%	-	76 fewer per 1000 (from 76 fewer to 76 fewer)	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months rifampicin and isoniazid	6 months isoniazid	Relative (95% CI)	Absolute		
Hepatotoxicity (follow-up 3-6 years; assessed with: Grade 3 or 4 raised aminotransferases)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None	8/329 (2.4%)[3]	18/327 (5.5%)[3] 0%	-	55 fewer per 1000 (from 55 fewer to 55 fewer)	⊙○○○ VERY LOW	CRITICAL

[1] Neither participants nor clinicians were kept blinded to treatment regimen. Isoniazid alone treatment was self administered while other treatments were directly observed therapy.
[2] event rate less than 300
[3] calculated from percentages provided

Author(s):
Date: 2014-03-04
Question: Should continous isoniazid vs 6 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Continous isoniazid	6 months isoniazid	Relative (95% CI)	Absolute		
Incidence of active tuberculosis (follow-up 3-6 years; assessed with: clinical and biochemical diagnosis)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None	8/164 (4.9%)	22/327 (6.7%) 0%	-	67 fewer per 1000 (from 67 fewer to 67 fewer)	⊙○○○ VERY LOW	CRITICAL
Mortality (follow-up 3-6 years; assessed with: number of deaths)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None	8/164 (4.9%)	25/327 (7.6%) 0%	-	76 fewer per 1000 (from 76 fewer to 76 fewer)	⊙○○○ VERY LOW	CRITICAL
Hepatotoxicity (follow-up 3-6 years; assessed with: grade 3 or 4 raised aminotransferases)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None	46/164 (28%)[3]	18/327 (5.5%)[3] 0%	-	55 fewer per 1000 (from 55 fewer to 55 fewer)	⊙○○○ VERY LOW	CRITICAL

[1] Neither participants nor clinicians were kept blinded to treatment regimen. Isoniazid alone treatment was self administered while other treatments were directly observed therapy.
[2] event number less than 300
[3] calculated from percentages provided

Author(s):
Date: 2014-03-04
Question: Should 3 months rifabutin dose 300 mg and isoniazid vs 6 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months rifabutin dose 300 mg and isoniazid	6 months isoniazid	Relative (95% CI)	Absolute	Quality	Importance
Treatment completion (follow-up mean 18 months; assessed with: number achieving 80% adherence to drugs taken)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	reporting bias[3]	13/16 (81.3%)	10/14 (71.4%) 0%	-	714 fewer per 1000 (from 714 fewer to 714 fewer)	⊙○○○ VERY LOW	CRITICAL

[1] Unclear if an appropriate method of randomisation was used. Unclear if adequate concealment of allocation. Unclear if groups were comparable at baseline. Neither participants nor clinicians were blinded to treatment group. Unclear if groups were comparable for treatment completion. No precise definition of outcome.
[2] event rate less than 300
[3] Pharmacy funded with poor information about methods and trial terminated early

Author(s):
Date: 2014-03-04
Question: Should 3 months rifabutin 600 mg and isoniazid vs 6 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect			
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months rifabutin 600 mg and isoniazid	6 months isoniazid	Relative (95% CI)	Absolute	Quality	Importance
Treatment completion (follow-up mean 17-19 months; assessed with: adherence to drug regimen >80%)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None	13/14 (92.9%)	10/14 (71.4%) 0%	-	714 fewer per 1000 (from 714 fewer to 714 fewer)	⊙○○○ VERY LOW	CRITICAL

[1] Unclear if an appropriate method of randomisation was used. Unclear if adequate concealment of allocation. Unclear if groups were comparable at baseline. Neither participants nor clinicians were blinded to treatment group. Unclear if groups were comparable for treatment completion. No precise definition of outcome.
[2] event number less than 300

Author(s):
Date: 2014-03-04
Question: Should 3 months rifampicin and isoniazid vs 6 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months rifampicin and isoniazid	6 months isoniazid	Relative (95% CI)	Absolute		
Treatment completion (follow-up 5 years; assessed with: adhering to >80% of prescribed dose)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious	None	213/296 (72%)	154/294 (52.4%) 0%	-	524 fewer per 1000 (from 524 fewer to 524 fewer)	⊙⊙○○ LOW	CRITICAL
Hepatotoxicity (follow-up 5 years; assessed with: Liver enzymes > 3 times the normal level)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	none	4/296 (1.4%)	10/294 (3.4%) 0%	-	34 fewer per 1000 (from 34 fewer to 34 fewer)	⊙⊙○○ LOW	CRITICAL
Nausea or vomiting (follow-up 5 years; assessed with: without hepatotoxicity)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	none	23/296 (7.8%)	24/294 (8.2%) 0%	-	82 fewer per 1000 (from 82 fewer to 82 fewer)	⊙⊙○○ LOW	IMPORTANT
Cutaneous toxicity (follow-up 5 years; assessed with: Rash, pruritis, photosensitivity)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	none	8/296 (2.7%)	5/294 (1.7%) 0%	-	17 fewer per 1000 (from 17 fewer to 17 fewer)	⊙⊙○○ LOW	IMPORTANT
Headache (follow-up 5 years)												
1	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	None	5/296 (1.7%)	8/294 (2.7%) 0%	-	27 fewer per 1000 (from 27 fewer to 27 fewer)	⊙⊙○○ LOW	IMPORTANT

[1] *event number less than 300*
[2] *Neither clinicians nor participants were blinded to treatment group. Groups were not comparable at baseline for sex and number of illegal immigrants. Groups were not comparable for treatment completion and there was a high loss to follow up.*

Author(s):
Date: 2014-03-04
Question: Should 3 months rifapentine and isoniazid vs 9 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months rifapentine and isoniazid	9 months isoniazid	Relative (95% CI)	Absolute		
incidence of active tuberculosis (follow-up 33 months; assessed with: clinical and biochemical diagnosis)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	4/3273 (0.12%)[3]	8/2585 (0.31%)[3]	-	3 fewer per 1000 (from 3 fewer to 3 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		
Completion of therapy (follow-up 33 months; assessed with: patients who completed therapy)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	None	713/3273 (21.8%)[4]	2585/3745 (69%)[4]	-	690 fewer per 1000 (from 690 fewer to 690 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		
Discontinuation of treatment due to adverse events (follow-up 33 months; assessed with: Number who discontinued treatment due to adverse events)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	196/3986 (4.9%)	139/3745 (3.7%)	-	37 fewer per 1000 (from 37 fewer to 37 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		
Mortality (follow-up 33 months; assessed with: Number of deaths)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	31/3986 (0.78%)	39/3745 (1%)	-	10 fewer per 1000 (from 10 fewer to 10 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		
Hepatotoxicity (follow-up 33 months months; assessed with: unclear definition)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18/4040 (0.45%)	103/3759 (2.7%)	-	27 fewer per 1000 (from 27 fewer to 27 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months rifapentine and isoniazid	9 months isoniazid	Relative (95% CI)	Absolute		
Rash (follow-up 33 months; assessed with: Unclear)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	31/4040 (0.77%)	21/3759 (0.56%)	-	6 fewer per 1000 (from 6 fewer to 6 fewer)	⊙○○○ VERY LOW	IMPORTANT
								0%		-		
Possible Hypersensitivity (follow-up 33 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	152/4040 (3.8%)	17/3759 (0.45%)	-	5 fewer per 1000 (from 5 fewer to 5 fewer)	⊙○○○ VERY LOW	IMPORTANT
								0%		-		
Adverse event (follow-up 33 months; assessed with: grade 3 or 4)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	229/4040 (5.7%)	244/3759 (6.5%)	-	65 fewer per 1000 (from 65 fewer to 65 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		

[1] *Unclear if adequate concealment of allocation. Neither clinican nor participant were blinded to treatment group. Treatment group did not recieve the same care appart from intervention studied combination therapy was directly observed, isoniazid was self administered. Unclear if groups were comparable for treatment completion or availability of outcome data.*
[2] *event number less than 300*
[3] *Data available in the evidence table for results adjusted per patient-year*
[4] *Calculated from number that discontinued treatment*

Author(s):
Date: 2014-03-04
Question: Should 9 months isoniazid vs 3 months placebo be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	9 months isoniazid	3 months placebo	Relative (95% CI)	Absolute		
Hepatotoxicity (assessed with: raised aminotransferases)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious	none[2]	8/60 (13.3%)	1/60 (1.7%)	-	17 fewer per 1000 (from 17 fewer to 17 fewer)	⊙⊙○○ LOW	CRITICAL
								0%		-		
Rash												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none[2]	7/60 (11.7%)	6/60 (10%)	-	100 fewer per 1000 (from 100 fewer to 100 fewer)	⊙⊙○○ LOW	
								0%		-		
Nausea												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none[2]	2/60 (3.3%)	1/60 (1.7%)	-	17 fewer per 1000 (from 17 fewer to 17 fewer)	⊙⊙○○ LOW	
								0%		-		

[1] *results were taken from 3 months into the trial; no data from beyond this point. Unclear if groups were comparable for treatment completion or availability of outcome data.*
[2] *unclear source of funding*
[3] *event number less than 300*

Author(s):
Date: 2014-03-04
Question: Should 12 months isoniazid vs No treatment be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	12 months isoniazid	No treatment	Relative (95% CI)	Absolute		
incidence of active tuberculosis (follow-up mean 33-39 months; assessed with: clinical and biochemical diagnosis)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	2/58 (3.4%)[3]	6/60 (10%)[3]	-	100 fewer per 1000 (from 100 fewer to 100 fewer)	⊙⊙○○ LOW	CRITICAL
								0%		-		

[1] *No blinding of participants or clinicians.*
[2] *event number less than 300*
[3] *number adjusted for person-years available in evidence table*

Author(s):
Date: 2014-03-04
Question: Should 2 months rifampicin and pyrazinamide vs 12 months isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	2 months rifampicin and pyrazinamide	12 months isoniazid	Relative (95% CI)	Absolute		
Incidence of active tuberculosis (follow-up mean 36-37 months; assessed with: clinical and biochemical diagnosis)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	28/791 (3.5%)	29/792 (3.7%)	-	37 fewer per 1000 (from 37 fewer to 37 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		
Mortality (follow-up mean 36-37 months; assessed with: Number of deaths)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	139/791 (17.6%)	159/792 (20.1%)	-	201 fewer per 1000 (from 201 fewer to 201 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		

[1] *Unclear if appropriate method of randomisation was used. Unclear if adequate concealment of treatment groups. Unclear if patients recieved same care apart from intervention. Neither participants nor clinicians were blinded to treatment group. Groups were not comparable for treatment completion.*
[2] *Number of events less than 300*

Author(s):
Date: 2014-03-05
Question: Should 12 months isoniazid vs 12 months placebo be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	12 months isoniazid	12 months placebo	Relative (95% CI)	Absolute		
incidence of active tuberculosis (assessed with: broad review of history and chest xray)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	0/6403 (0%)	7/6484 (0.11%)	-	1 fewer per 1000 (from 1 fewer to 1 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		

[1] *Unclear method of randomisation and concealment of allocation. Groups were not similar in terms of mortality, weight and abnormal xrays at baseline. Unclear if groups were comparable for treatment completion or availability of outcome data. Follow up appears only to be during treatment period and outcome was deter by a broad review of history and chest xray possibly leading to high diagnostic bias.*
[2] *event number less than 300*

Author(s):
Date: 2014-03-05
Question: Should 6 months of isoniazid and ethambutol vs 36 months of isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	6 months of isoniazid and ethambutol	36 months of isoniazid	Relative (95% CI)	Absolute		
Incidence of active tuberculosis (assessed with: clinical and biochemical diagnosis)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	6 months, isoniazid and ethambutol n=141 TB incidence/100 personyears (95% CI²) 3.18 (1.38-4.97)) TB incidence/100 personyears	36 months isoniazid, n=132 TB incidence/100 personyears (95% CI²) 1.81 (0.69-3.04) TB incidence/100 personyears (95% CI²) per protocol analysis	6 months, isoniazid and ethambutol n=141 Adjusted incidence rate ratio (95% CI²) 1.48 (0.55, 3.96) Adjusted incidence rate ratio (95% CI²), per protocol analysis 1.57 (0.50, 4.9) 36 months isoniazid, n=132 Adjusted incidence		⊙⊙○○ LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	6 months of isoniazid and ethambutol	36 months of isoniazid	Relative (95% CI)	Absolute		
							(95% CI²) per protocol analysis 2.80 (1.06-4.70)	1.84 (0.37-3.32)	rate ratio (95% CI²) Reference Adjusted incidence rate ratio (95% CI²), per protocol analysis reference			
Mortality												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	6 months, isoniazid and ethambutol n=141 Mortality/100 personyears (95% CI²) 2.91 (1.19-4.63) Mortality/100 personyears (95% CI²) per protocol analysis 3.08 (1.26-4.89)	36 months isoniazid, n=132 Mortality/100 personyears (95% CI²) 2.53 (1.21-3.85) Mortality/100 personyears (95% CI²) per protocol analysis 2.15 (0.56-3.74)	6 months, isoniazid and ethambutol n=141 Adjusted incidence rate ratio (95% CI²) 1.51 (0.56, 4.02) Adjusted incidence rate ratio (95% CI²), per protocol analysis 1.43 (0.53, 3.8) 36 months isoniazid, n=132 Adjusted incidence rate ratio (95% CI²) Reference Adjusted incidence rate ratio (95% CI²), per protocol analysis reference			

[1] *Neither participants nor clinicans were blinded to treatment groups. Unclear if groups were comparable for treatment completion and availability of outcome data.*
[2] *event number less than 300*

Author(s):
Date: 2014-03-05
Question: Should 6 months of isoniazid and ethambutol vs 36 months of isoniazid be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	5-9 months of isoniazid therapy	No treatment	Relative (95% CI)	Absolute		
Incidence of active tuberculosis (follow-up 10 years; assessed with: clinical or bacteriological diagnosis)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	24/1451 (1.7%)[3]	10/1519 (0.66%)[3]	-	7 fewer per 1000 (from 7 fewer to 7 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		
Mortality (follow-up 10 years; assessed with: number of deaths)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	7/1451 (0.48%)[3]	7/1519 (0.46%)[3]	-	5 fewer per 1000 (from 5 fewer to 5 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		

[1] randomisation by date of birth was used, unclear if adequate concealment. Patients in the treatment grouo were younger. Unclear if comparison groups recieved the same care apart from treatment. Neither participants nor clinicians were blinded to treatment allocation. Unclear if groups were comparable in terms of treatment completion or availability of outcome data.
[2] event number less than 300
[3] follow up data available on a per year basis

Author(s):
Date: 2014-03-05
Question: Should 3 months of rifapentine and isoniazid vs 2 months of rifampicin and pyrazinamide be used for latent tuberculosis?
Settings:
Bibliography:

Quality assessment							No of patients		Effect	Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months of rifapentine and isoniazid	2 months of rifampicin and pyrazinamide	Relative (95% CI)		
Incidence of active tuberculosis (follow-up at least 2 years; assessed with: clinical and bacteriological diagnosis)											
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	Rifapentine and isoniazid group: • 3 cases in 564 person years of follow up (0.5/100 person-years)	Rifampicin and pyrazinamide group: • 1 case in 522 person- years of follow up (0.2/100 person-years)	Relative risk, 2.8; 95% CI, 0.3-26.8; p=0.66 i.e. non significant	⊙○○○ VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	3 months of rifapentine and isoniazid	2 months of rifampicin and pyrazinamide		Relative (95% CI)		
Mortality (follow-up at least 2 years; assessed with: number of deaths)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	1/206 (0.49%)	3/193 (1.6%)	-	16 fewer per 1000 (from 16 fewer to 16 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		
Hepaxicity (follow-up at least 2 years; assessed with: Grade 3 or 4)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	2/206 (0.97%)	20/193 (10.4%)	-	104 fewer per 1000 (from 104 fewer to 104 fewer)	⊙○○○ VERY LOW	CRITICAL
								0%		-		

[1] *Groups did not recieve the same care apart from intervention studied; one group was mostly self administered, the other entirely directly observed. Neither participants nor clinicians were blinded to treatment groups. Groups were not comparable for availability of outcome data. TB diagnoses were confirmed from medical records and health department data bases. Trial was stopped early.*

[2] *event number less than 300*

Appendix E: GRADE profiles