Appendix F:Full GRADE profiles [2014 update]

1.1 Full GRADE profiles (review question 2)

Review question 2:

Which risk factors indicate endoscopy in order to exclude Barrett's oesophagus?

1.1.1 Patients who had undergone endoscopy due to various indications (compared those with confirmed BO with no BO)

		Gender (Male)		Age (various thresholds)		Smoking (Smoker)		Alcohol consumption		BMI (various thresholds)		Hiatal hernia		GORD symptoms		Oesophagitis (endo)		H pylori (diff. ref.)	
		Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI
1	Abrams (2008)	***1.86***	***(1.20 to 2.87)***	***2.35***	***(1.16 to 4.76)[a]***							***3.53***	***(2.17 to 5.72)***	***2.87***	***(1.84 to 4.45)[p]***				
2	Ford (2005)	***2.70***	***(2.18 to 3.35)***	***1.03***	***(1.02 to 1.03)[b]***														
3	Johansson (2007)	1.80	(0.70 to 5.20)	***1.05***	***(1.01 to 10.9)[b]***	1.80	(0.70 to 4.40)[h]	0.60	(0.20 to 1.70)	1.10	(0.30 to 3.30)[l]			2.00	(0.80 to 5.00)[r]			1.70	(0.70 to 4.60)[s]
4	Voutilainen (2000)	***3.20***	***(1.27 to 8.12)***	***1.03***	***(1.00 to 1.06)[b]***											***6.57***	***(2.69 to 16.06)[u]***		
5	Jonaitis (2011)	1.56	(0.26 to 1.22)	***1.06***	***(1.01 to 1.20)[c]***	***4.62***	***(1.01 to 12.51)[i]***			1.11	(0.92 to 1.33)[m]	***5.22***	***(1.86 to 14.7)***					***5.60***	***(1.38 to 22.72)[t]***
6	Omer (2012)	***3.20***	***(2.30 to 4.40)***	0.97	(0.68 to 1.40)[c]	1.20	(0.84 to 1.60)	1.10	(0.59 to 1.90)[j]	1.20	(0.84 to 1.7)[n]								
7	Lam (2008)	***2.68***	***(1.32 to 5.45)***	1.01	(0.99 to 1.04)[d]	1.71	(0.78 to 3.76)	1.29	(0.58 to 2.86)										
8	Menon (2011)	***1.07***	***(1.01 to 1.07)***	***1.02***	***(1.02 to 1.02)[e]***							***1.22***	***(1.17 to 1.27)***			***3.46***	***(3.33 to 3.59)***		
9	Thrift (2012)**	***2.17***	***(1.50 to 3.14)***	***1.14***	***(1.06 to 1.23)[f]***	***1.93***	***(1.15 to 3.24)***			1.41	(0.90 to 2.22)[o]								
10	Khoury (2012)	***0.30***	***(0.20 to 0.44)[A]***																
11	Nelsen (2012)									2.08	(0.81 to 4.96)[n]								
12	Rubenstein (2010)																		
13	Bu (2006)									***3.30***	***(1.60 to 6.70)[k]***								
14	Conio (2002)					0.70	(0.40 to 1.40)[g]	1.30	(0.90 to 2.00)			***3.90***	***(2.50 to 6.00)***	***5.80***	***(4.00 to 8.40)[q]***				
15	Fan (2009)																		

GRADE

Risk of bias	Serious[1]	Serious[3]	Serious[5]	Serious[8]	Serious[10]	Serious[12]	Serious[14]	Serious[17]	Serious[19]
Indirectness	No serious	No serious	No serious	No serious	No serious	No serious	No serious	No serious	No serious
Inconsistency	Serious[22]	Serious[22]	Serious[6]	No serious	No serious	No serious	Serious[15]	No serious	Serious[20]
Imprecision	Serious[2]	Serious[4]	Serious[7]	Very serious[9]	Serious[11]	Very serious[13]	Very serious[16]	Very serious[18]	Very serious[21]
Other considerations	No serious	No serious	No serious	No serious	No serious	No serious	No serious	No serious	No serious
CONFIDENCE	**Very low**	**Low**	**Very low**	**Very low**	**Low**	**Very low**	**Very low**	**Very low**	**Very low**

Footnote:

A = Reference: Male
a = 60-69 yrs (Reference: <40 yrs); [Other age thresholds vs. Reference]: 40-49 yrs (Adj OR = 0.86, 95%CI: 0.34 to 2.18); 50-59 yrs (Adj OR = 1.49, 95%CI: 0.69 to 3.20); >70 yrs (Adj OR = 1.55, 95%CI: 0.75 to 3.23)
b = Each additional year
c = >60 yrs
d = Age threshold not reported
e = >50 yrs
f = Every 5 additional years
g = Smoking >20 per day (Reference: Non-smoker) [Other thresholds vs. Reference]: Smoking 1-20 per day (Adj OR = 1.0, 95%CI: 0.6 to 1.7)
h = Smoking everyday
i = Smoking >10 per day (Reference: Smoking <10 per day)
j = >14 drinks per week (Reference: Non-drinker) [Other thresholds vs. Reference]: <2 drinks per week (Adj OR = 1.0, 95%CI: 0.65 to 1.50); 2-14 drinks per week (Adj OR = 0.83, 95%CI: 0.55 to 1.30)
k = >30kg/m2 (Reference: <22kg/m2); [Other BMI thresholds vs. Reference]: 22-24.9kg/m2 (Adj OR = 1.2, 95%CI: 0.6 to 2.5); 25-29.9kg/m2 (Adj OR = 1.6, 95%CI: 0.9 to 3.1)
l = >26.6kg/m2 (Reference: <23.6kg/m2); [Other BMI thresholds vs. Reference]: 23.6-26.6kg/m2 (Adj OR = 0.9, 95%CI: 0.3 to 2.9)
m = Reference and threshold were not reported
n = >30kg/m2 (Reference: <30kg/m2)
o = >30kg/m2 (Reference: <25kg/m2); [Other BMI thresholds vs. Reference]: 25-30kg/m2 (Adj OR = 0.96, 95%CI: 0.64 to 1.44)
p = Reflux indication (Reference: No reflux)
q = Weekly GORD symptoms (Reference: No weekly GORD symptoms)
r = Reflux symptoms >50 times per year (Reference: <50 times per year)
s = Reference: H pylori negative
t = Reference: H pylori positive
u = Also reported oesophagitis confirmed by biopsies: Adj OR = 1.84 (95%CI: 0.75 to 4.50)

Footnote for GRADE:
1 = Downgraded 1 level: 7 out of 10 studies are retrospective; all 10 studies did not control for potential confounding factors.
2 = Downgraded 1 level: only 1 out of 10 studies had carried out model diagnostics and validation.
3 = Downgraded 1 level: 6 out of 9 studies are retrospective; all 9 studies did not control for potential confounding factors.
4 = Downgraded 1 level: only 1 out of 9 studies had carried out model diagnostics and validation.
5 = Downgraded 1 level: 3 out of 5 studies are retrospective; all 5 studies did not control for potential confounding factors.
6 = Downgraded 1 level: inconsistency among the effect estimates.
7 = Downgraded 1 level: only 1 out of 5 studies had carried out model diagnostics and validation.
8 = Downgraded 1 level: 2 out of 4 studies are retrospective; only 1 out of 4 studies controlled for potential confounding factors.
9 = Downgraded 2 levels: no studies had carried out model diagnostics and validation.
10 = Downgraded 1 level: 2 out of 6 studies are retrospective; only 2 out of 6 studies controlled for potential confounding factors.
11 = Downgraded 1 level: only 1 out of 6 studies had carried out model diagnostics and validation.
12 = Downgraded 1 level: 2 out of 4 studies are retrospective; only 1 out of 4 studies controlled for potential confounding factors.
13 = Downgraded 2 levels: no studies had carried out model diagnostics and validation.
14 = Downgraded 1 level: 1 out of 3 studies are retrospective; only 1 out of 3 studies controlled for potential confounding factors.
15 = Downgraded 1 level: inconsistency among the effect estimates.
16 = Downgraded 2 levels: no studies had carried out model diagnostics and validation.
17 = Downgraded 1 level: 1 out of 2 studies are retrospective; both studies did not control for potential confounding factors.
18 = Downgraded 2 levels: no studies had carried out model diagnostics and validation.

19 = Downgraded 1 level: both studies did not control for potential confounding factors.
20 = Downgraded 1 level: inconsistency among the effect estimates.
21 = Downgraded 2 levels: no studies had carried out model diagnostics and validation.
22 = Downgrade 1 level: inconsistent directions of effect estimate across different studies

1.1.2 Patients who had undergone endoscopy due to various indications (compared those with confirmed BO with no BO) [ETHNICITY]

		Blacks[a]		Hispanic[a]		Others[a]		White		Non-Asian		Afro-Carribean	
		Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI
1	Abrams (2008)	***0.34***	***(0.12 to 0.97)***	***0.38***	***(0.18 to 0.84)***	0.91	(0.56 to 1.58)						
2	Ford (2005)							***6.03***	***(3.56 to 10.2)***[c]			0.49	(0.11 to 2.17)[f]
3	Johansson (2007)												
4	Voutilainen (2000)												
5	Jonaitis (2011)												
6	Omer (2012)							1.00	(0.56 to 1.9)[d]				
7	Lam (2008)									***3.55***	***(1.85 to 6.85)***[e]		
8	Menon (2011)												
9	Thrift (2012)**												
10	Khoury (2012)	***0.28***	***(0.16 to 0.48)***[b]			0.37	(0.14 to 1.02)						
11	Nelsen (2012)												
12	Rubenstein (2010)	***0.26***	***(0.13 to 0.54)***										
13	Bu (2006)												
14	Conio (2002)												
15	Fan (2009)	0.56	(0.28 to 1.09)[b]	0.94	(0.46 to 1.92)	0.40	(0.06 to 2.93)						

GRADE

Risk of bias	Serious[1]	Serious[3]	Serious[6]	Serious[8]	Serious[11]	Serious[11]
Indirectness	No serious	No serious	No serious	No serious	No serious	No serious
Inconsistency	No serious	Serious[4]	No serious	Serious[9]	NA	NA
Imprecision	Very serious[2]	Very serious[5]	Very serious[7]	Very serious[10]	Very serious[12]	Very serious[12]
Other considerations	No serious	No serious	No serious	No serious	No serious	No serious
CONFIDENCE	**Very low**	**Very low**	**Very low**	**Very low**	**Very low**	**Very low**

Footnote:
a = Reference: White
b = African American
c = Reference: South Asian
d = Reference: Others
e = Reference: Asian
f = Reference: South Asian

Footnote for GRADE:
1 = Downgraded 1 level: 1 out of 2 studies are retrospective; both studies did not control for potential confounding factors.
2 = Downgraded 2 levels: no studies had carried out model diagnostics and validation.
3 = Downgraded 1 level: both studies are retrospective; only 1 study controlled for potential confounding factors.
4 = Downgraded 1 level: inconsistency among the effect estimates.
5 = Downgraded 2 levels: no studies had carried out model diagnostics and validation.
6 = Downgraded 1 level: both studies are retrospective; only 1 study controlled for potential confounding factors.
7 = Downgraded 2 levels: no studies had carried out model diagnostics and validation.
8 = Downgraded 1 level: both studies are retrospective and did not control for potential confounding factors.
9 = Downgraded 1 level: inconsistency among the effect estimates.
10 = Downgraded 2 levels: no studies had carried out model diagnostics and validation.
11 = Downgraded 1 level: retrospective study and did not control for potential confounding factors.
12 = Downgraded 2 levels: did not carry out model diagnostics and validation.
NA = Cannot be assessed.

1.1.3 Patients who had undergone endoscopy due to various indications (compared those with confirmed BO with no BO) [OTHER RISK FACTORS]

		Other risk factors	Adj OR	95%CI	Other risk factors	Adj OR	95%CI	Other risk factors	Adj OR	95%CI	Other risk factors	Adj OR	95%CI
1	Abrams (2008)												
2	Ford (2005)	Middle status[a]	***1.98***	***(1.48 to 2.65)***	High status[a]	***1.58***	***(1.16 to 2.15)***						
3	Johansson (2007)												
4	Voutilainen (2000)												
5	Jonaitis (2011)	Ulcer/stricture present	***11.95***	***(2.51 to 41.4)***									
6	Omer (2012)	PPI[c]	0.91	(0.64 to 1.30)	H2RA[c]	0.71	(0.39 to 1.30)	Aspirin[e]	***0.56***	***(0.39 to 0.80)***	NSAID[e]	0.92	(0.53 to 1.60)
7	Lam (2008)												
8	Menon (2011)	Stricture present	***1.20***	***(1.07 to 1.35)***									
9	Thrift (2012)**	Education School[b]	***2.08***	***(1.23 to 3.50)***	PPI or H2RA in last 5 yrs	***2.07***	***(1.46 to 2.93)***						
10	Khoury (2012)												
12	Nelsen (2012)	Waist circumference ≥97.8cm[d]	***4.05***	***(1.45 to 57.2)***	GE junction fat[f] ≥6.1cm^2	***5.97***	***(1.28 to 27.7)***	Subcutaneous fat[g] ≥97cm^2	3.20	(0.58 to 10.3)	Visceral fat[g] ≥97cm^2	***3.51***	***(1.04 to 22.9)***
13	Rubenstein (2010)												
14	Bu (2006)												
15	Conio (2002)	Ulcer present	***2.20***	***(1.30 to 3.50)***									
16	Fan (2009)												

GRADE

		Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	CONFIDENCE
2	Ford (2005)	Serious[1]	No serious	NA	Very serious[2]	No serious	**Very low**
5	Jonaitis (2011)	Serious[3]	No serious	NA	Very serious[2]	No serious	**Very low**
6	Omer (2012)	Serious[1]	No serious	NA	Very serious[2]	No serious	**Very low**
8	Menon (2011)	Serious[1]	No serious	NA	Very serious[2]	No serious	**Very low**
9	Thrift (2012)**	Serious[1]	No serious	NA	No serious	No serious	**Moderate**
12	Nelsen (2012)	No serious	No serious	NA	Very serious[2]	No serious	**Low**
15	Conio (2002)	No serious	No serious	NA	Very serious[2]	No serious	**Low**

Footnote:
a = Social status (Reference: Low status)
b = Reference: University level
c = Reference: No acid suppressant
d = Reference: <97.8cm (adjusted for BMI)
e = Reference: No medication
f = Reference: <6.1cm2 (adjusted for BMI)
g = Reference: <97cm2 (adjusted for BMI)

Footnote for GRADE:

NA = Cannot be assessed.
1 = Downgraded 1 level: retrospective study, did not control for potential confounding factors.
2 = Downgraded 2 levels: did not carry out any model diagnostics and validation.
3 = Downgraded 1 level: did not control for potential confounding factors.

1.1.4 Patients who had undergone endoscopy due to GORD symptoms (compared those with confirmed BO with no BO)

		Gender (Male)		Age (various thresholds)		Smoking (Smoker)		Alcohol consumption		African-American		Duration of GORD		Heartburn/regurgitation		Nocturnal heartburn		Hiatal hernia	
		Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI
1	Campos (2001)	***2.60***	***(1.60 to 4.30)***									***2.10***	***(1.40 to 3.20)[d]***					***4.10***	***(2.10 to 8.00)[c]***
2	Eloubeidi (2001)			***4.86***	***(1.50 to 15.80)[a]***									***4.38***	***(1.26 to 17.00)***	***0.36***	***(0.14 to 0.91)***		
3	Gerson (2001)	***3.70***	***(2.04 to 6.67)***	0.93	(0.63 to 1.37)[b]					0.39	(0.11 to 1.37)[g]			***1.80***	***(1.06 to 3.06)***	***1.73***	***(1.05 to 2.84)[i]***		
4	Gerson (2007)	***3.27***	***(1.81 to 5.90)***	1.01	(1.00 to 1.03)[b]	1.33	(0.90 to 1.98)	1.06	(0.71 to 1.58)			***1.39***	***(1.15 to 1.69)[f]***						
5	Koek (2008)	***2.77***	***(1.17 to 6.53)***																

GRADE									
Risk of bias	Serious[1]	Serious[1]	Serious[1]	Serious[1]	Serious[1]	Serious[1]	Serious[1]	Serious[1]	Serious[1]
Indirectness	No serious	No serious	No serious	No serious	No serious	No serious	No serious	No serious	No serious
Inconsistency	No serious	Serious[3]	NA	NA	NA	No serious	No serious	Serious[3]	NA
Imprecision	Very serious[2]	Very serious[2]	Very serious[2]	Very serious[2]	Very serious[2]	Very serious[2]	Very serious[2]	Very serious[2]	Very serious[2]
Other considerations	No serious	No serious	No serious	No serious	No serious	No serious	No serious	No serious	No serious
CONFIDENCE	**Very low**	**Very low**	**Very low**	**Very low**	**Very low**	**Very low**	**Very low**	**Very low**	**Very low**

Footnote:
a = >40 yrs (Reference: <40 yrs)
b = Age threshold or reference threshold not reported.
c = >4cm long (Reference: No hiatal hernia); for 2-4cm (Adj OR = 2.4, 95%CI: 1.4 to 4.6)
d = Duration >5 yrs
e = Each additional year
f = Duration of each additional year
g = Reference: White [Other ethnicity: Asian Adj OR = 0.72, 95%CI: 0.28 to 1.83; Hispanic Adj OR = 0.49, 95%CI: 0.18 to 1.38]
i = Nocturnal pain
NR = Not reported

Footnote for GRADE:
1 = Downgraded 1 level: all studies did not control potential confounding factors.
2 = Downgraded 2 levels: did not carry out any model diagnostics and validation.
3 = Downgraded 1 level: inconsistency among the effect estimates.
NA = Cannot be assessed.

1.1.5 Patients who had undergone endoscopy due to GORD symptoms (compared those with confirmed BO with no BO) [OTHER RISK FACTORS]

		Risk factors	Adj OR	95%CI			Risk factors	Adj OR	95%CI
2	Eloubeidi (2001)	Severe hearburn	***0.13***	***(0.04 to 0.42)***	5	Koek (2008)	Acid exp (7.5% of time)	***5.11***	***(2.66 to 9.83)***[j]
2	Eloubeidi (2001)	Heartburn >1 per wk	***3.01***	***(1.35 to 6.73)***	5	Koek (2008)	No. acid episodes >5min (7.5% of time)	***6.78***	***(1.81 to25.42)***[k]
1	Campos (2001)	Ab. bilirubin exp	***4.20***	***1.90 to 9.70***	5	Koek (2008)	DGOR exp (20.1% of time)	***4.18***	***(1.89 to 9.24)***[l]
1	Campos (2001)	Defective LES	***2.70***	***1.40 to 5.40***					
1	Campos (2001)	Defective DCA	***2.20***	***1.40 to 3.05***					

Note: Ab = Abnormal; exp = exposure; LES = lower oesophageal sphincter; DCA = distal contraction amplitude; DGOR = duodeno-gastro-oesophageal reflux

GRADE

		Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	CONFIDENCE
2	Eloubeidi (2001)	Serious[1]	No serious	NA	Very serious[2]	No serious	**Very low**
1	Campos (2001)	Serious[1]	No serious	NA	Very serious[2]	No serious	**Very low**
5	Koek (2008)	Serious[1]	No serious	NA	Very serious[2]	No serious	**Very low**

Footnote:
Ab = Abnormal; exp = exposure; LES = lower oesophageal sphincter; DCA = distal contraction amplitude; DGOR = duodeno-gastro-oesophageal reflux
j = For other thresholds: 0.6% of time Adj OR = 3.54 (95%CI: 1.23 to 10.17); 2.4% of time Adj OR = 3.69 (95%CI: 1.77 to 7.69)
k = For other thresholds: 0.6% of time Adj OR = 4.05 (95%CI: 1.51 to 10.87); 2.4% of time Adj OR = 4.42 (95%CI: 1.27 to 15.41)
l = For other thresholds: 0.6% of time Adj OR = 3.04 (95%CI: 0.09 to 10.25); 4.9% of time Adj OR = 3.74 (95%CI: 1.48 to 9.46)

Footnote for GRADE:
1 = Downgraded 1 level: all studies did not control potential confounding factors.
2 = Downgraded 2 levels: did not carry out any model diagnostics and validation.
NA = Cannot be assessed.

1.1.6 Patients who had undergone endoscopy because of suspected BO (compared those with confirmed BO with no BO)

		Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI
1	Wang (2008)	**Gender (Male)**									
		1.82	***(1.49 to 2.22)***								
		Age (50-59 yrs)		**Age (60-69 yrs)**		**Age (70-79 yrs)**		**Age (>80 yrs)**			
		1.72	***(1.36 to 2.17)***	***1.85***	***(1.44 to 2.37)***	***2.33***	***(1.75 to 3.10)***	***1.96***	***(1.25 to 3.08)***		
		Blacks		**Hispanic**		**Asian/Pasific Islander**		**Native American**		**Multiracial**	
		0.24	***(0.14 to 0.41)***	0.82	(0.42 to 1.60)	0.48	(0.11 to 2.08)	1.04	(0.62 to 1.75)	1.83	(0.14 to 24.6)
		Hiatal hernia									
		1.46	***(1.22 to 1.74)***								
		Length of BO >3cm									
		4.61	***(3.73 to 5.69)***								

Age = Reference: 18-49 yrs; Ethnicity = Reference: White; Length of BO = Reference: <3cm

GRADE

		Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	CONFIDENCE
1	Wang (2008)	Serious[1]	No serious	No serious	Serious[2]	No serious	**Low**

Footnote for GRADE:
1 = Downgraded 1 level: the study did not control potential confounding factors.
2 = Downgraded 1 level: the study lacks reproducibility (no validation).

1.1.7 SHORT BO: Patients who had undergone endoscopy due to various indications (compared those with SHORT BO with no BO)

	Reflux symptoms		Presence of tongues[a]		Age (per decade)		Oesophagitis[b]		Inflammation GO[c]	
	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI
De Mas (1999)	***4.70***	***(2.2 to 10.2)***	***2.80***	***(1.2 to 6.4)***						
Nandurkar (1997)					***1.03***	***(1.01 to 1.06)***	***3.20***	***(1.4 to 7.2)***	***5.90***	***(2.2 to 15.6)***

Footnote:

a = Tongue-like changes of the columnar epithelium

b = Histologically confirmed

c = Inflammation at the gastro-oesophageal junction

GRADE

	Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	CONFIDENCE
De Mas (1999)	Serious[1]	No serious	NA	Very serious[2]	No serious	**Very low**
Nandurkar (1997)	Serious[1]	No serious	NA	Very serious[2]	No serious	**Very low**

Footnote for GRADE:

1 = Downgraded 1 level: all studies did not control potential confounding factors.

2 = Downgraded 2 levels: did not carry out any model diagnostics and validation.

NA = Cannot be assessed.

1.1.8 Patients with short (<3cm) segment columnar-appearing mucosa in the oesophagus (compared those with intestinal metaplasia vs. no intestinal metaplasia)

	Gender (Male)		Age[a]		GORD symptoms		H. pylori infection		Corpus/antrum[b]	
	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI	Adj OR	95%CI
Dietz (2006)	0.93	(0.40 to 2.15)	***2.87***	***(1.14 to 7.24)***	0.63	(0.26 to 1.54)	1.79	(0.74 to 4.35)	***5.71***	***(2.09 to 15.6)***

Footnote:

a = Age thresholds and reference not reported.

b = Presence of Corpus/antrum gastric intestinal metaplasia

GRADE

	Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	CONFIDENCE
Dietz (2006)	Serious[1]	Serious[2]	NA	Very serious[3]	No serious	**Very low**

Footnote for GRADE:

1 = Downgraded 1 level: the study did not control potential confounding factors.

2 = Downgraded 1 level: indirect population = only included those aged 40 yrs or above.

3 = Downgraded 2 levels: did not carry out any model diagnostics and validation.

NA = Cannot be assessed.

1.1.9 Patients with GORD who have relatives of BO compared with matched controls with GORD but have no relatives of BO

	Have relatives of BO	
	Adj OR	95%CI
Romero (2002)	1.58	(0.46 to 5.45)

GRADE

	Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	CONFIDENCE
Romero (2002)	No serious	No serious	NA	Very serious[1]	No serious	**Low**

Footnote for GRADE:

1 = Downgraded 2 levels: did not carry out any model diagnostics and validation.

NA = Cannot be assessed.

1.1.10 Vegetable and fruit intake to predict BO (patients with BO compared with matched controls with no BO)

	Vegetables[a]		Fruit[b]		Vegetables & fruit[c]	
	Adj OR	**95%CI**	**Adj OR**	**95%CI**	**Adj OR**	**95%CI**
Thompson (2009)	***0.33***	***(0.17 to 0.63)***	0.76	(0.42 to 1.36)	***0.39***	***(0.21 to 0.75)***

Footnote:

a = >1.24 Servings/1000kcal/day (Reference: <0.67 servings) [Other thresholds vs reference]: 0.67-1.23 servings (Adj OR = 0.40, 95%CI: 0.23 to 0.71)

b = >1.00 Servings/1000kcal/day (Reference: <0.44 servings) [Other thresholds vs reference]: 0.44-0.99 servings (Adj OR = 0.73, 95%CI: 0.42 to 1.26)

c = >2.31 Servings/1000kcal/day (Reference: <1.24 servings) [Other thresholds vs reference]: 1.24-2.30 servings (Adj OR = 0.49, 95%CI: 0.28 to 0.86)

GRADE

	Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	CONFIDENCE
Thompson (2009)	No serious	No serious	NA	Very serious[1]	No serious	**Low**

Footnote for GRADE:

1 = Downgraded 2 levels: did not carry out any model diagnostics and validation.

NA = Cannot be assessed.

1.1.11 Risk factors to predict BO length (different populations with different indications for endoscopy)

1) Patients with confirmed BO (to predict long-segment BO ≥3cm)

Dickman (2005)	Adj OR	95%CI	Adj OR	95%CI
Age[a]	0.70	(0.40 to 1.30)		
Hiatal hernia	***1.90***	***(1.00 to 3.40)***		
BMI[b]	1.40	(0.80 to 2.50)[1]	***1.60***	***(1.00 to 2.80)[2]***
Ethnicity (White)[c]	1.60	(0.60 to 4.00)		
PPI	0.60	(0.30 to 1.20)		
Actively smoking[d]	***0.60***	***(0.30 to 0.96)***		
Dysplasia	***2.20***	***(1.02 to 4.60)***		
H2RA	1.56	(0.88 to 2.80)		

Footnote:

a = age >50 yrs old (Reference: >50 yrs old); b = Reference: <25kg/m^2; [1 = BMI >25kg/m^2 (overweight), 2 = BMI >30kg/m^2 (obese)]
c = Reference: other racial groups
d = Reference: not actively smoking

2) Patients who had undergone endoscopy due to various indications (to predict long-segment BO ≥3cm)

	Gender (male)		Hiatal hernia	
Abrams (2008)*	**Adj OR**	**95%CI**	**Adj OR**	**95%CI**
	6.37	***(1.29 to 31.4)***	***12.81***	***(2.61 to 63.0)***

3)Patients who had undergone endoscopy due to GORD (to predict long-segment BO ≥3cm)

	Longest reflux epi[a]		Hiatal hernia[d]		Defective LES[g]	
Campos (2001)*	**Adj OR**	**95%CI**	**Adj OR**	**95%CI**	**Adj OR**	**95%CI**
	8.10	***(2.80 to 24.0)[b]***	***17.80***	***(4.10 to 76.6)[e]***	***16.90***	***(1.60 to 181.4)***
	6.80	***(2.30 to 20.1)[c]***	***8.50***	***(2.30 to 31.7)[f]***		

Footnote:

a = Longest reflux episode (Reference: <19.9 min); b = >31.7 min; c = 19.9-31.7 min.
d = Hiatal hernia (Reference: <2cm); e = >4cm; e = 2-4cm.
g = Defective lower oesophageal sphincter.
* = Sub-analysis (also included in other overall multivariate analysis).

GRADE

	Risk of bias	Indirectness	Inconsistency
1) Dickman (2005)	Serious[1]	No serious	NA
2) Abrams (2008)*	Serious[3]	No serious	NA
3) Campos (2001)*	Serious[1]	No serious	NA

Footnote for GRADE:

1 = Downgraded 1 level: the study did not control potential confounders.
2 = Downgraded 2 levels: the study did not carry out model diagnostics and validation.
3 = Downgraded 1 level: retrospective study and did not control potential confounders.

1.2 Full GRADE profiles (review question 4)

Review question 4:

What is the clinical effectiveness of PPIs in patients with severe erosive reflux disease:

- to control/reduce oesophagitis?
- as maintenance therapy?

1.2.1 Outcome: Healing

1.2.1.1 Network meta-analysis for healing phase

Number of Studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
18 RCTs[a]	not serious[1]	serious[2]	serious[3]	very serious[4]	Very low

[1] No serious limitations. All studies used an appropriate method of randomisation (limited selection bias) and the majority of studies had some level of blinding. Healing measured according to well defined criteria (unlikely to lead to detection bias).

[2] I^2 was estimable for 6 links in the network: it was >50% for 3 links in the network (pantoprazole 40 mg/d v. ranitidine 300 mg/d, esomeprazole 40 mg/d v. lansoprazole 30 mg/d, esomeprazole 40 mg/d v rabeprazole 50 mg/d [ER]) and it was 0% for the 3 others (pantoprazole 40 mg/d v. nizatidine 300 mg/d, esomeprazole 20 mg/d v. omeprazole 20 mg/d, esomeprazole 40 mg/d v omeprazole 20 mg/d). There was fair agreement between direct and indirect estimates in the network loop.

[3] The majority of the evidence came from trials that were not designed or powered to focus on people with severe oesophagitis only, and effectiveness evidence was only available where subgroups of interest were reported.

[4] Wide confidence intervals for effect estimates which are likely due to small study sizes and/or reliance on subgroup results from trials that were not powered to detect differences between treatments in people with severe oesophagitis only. Most (12/18) of the 'links' in network include only 1 trial. As a consequence, there is substantial uncertainty of the ranking within the network.

[a] Fennerty (2005), Laine (2011), Richter (2000), Armstrong (2001), Kovacs (2002), Koop (1995), Meneghelli (2002), Jansen (1999), Robinson (1995), Mee (1996), Castell (2002), Gillessen (2004), Kahrilas (2000), Mossner (1995), Pace (2005), Richter (2001),Schmitt (2006),Lightdale (2006)

1.2.1.2 PPI versus placebo: no trials identified met the inclusion criteria

1.2.1.3 PPI versus H2RA: no trials identified met the inclusion criteria

1.2.1.4 Double-dose PPI versus full-dose PPI

Laine 2011 (1)

Rabeprazole-ER 50 mg compared to Esomeprazole 40 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rabeprazole-ER 50 mg	Esomeprazole 40 mg	Relative (95% CI)	Absolute		
Healing after 4 weeks in Grades C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
2[a]	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	serious imprecision[4]	none	556/1052 (52.9%)	539/1068 (50.5%)	RR 1.05 (0.96 to 1.14)	25 more per 1000 (from 20 fewer to 71 more)	Low	IMPORTANT[3]
Healing after 8 weeks in Grades C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
2[a]	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	serious imprecision[4]	none	828/1052 (78.7%)	819/1068 (76.7%)	RR 1.03 (0.98 to 1.07)	23 more per 1000 (from 15 fewer to 54 more)	Low	IMPORTANT[3]

a Laine (2011): 2 RCTs reported in one paper.
1 Blinding of the assessment of baseline endoscopy data was described, but assessment of endoscopy results for outcomes was not blinded.
2 Greater loss to follow up in the intervention group than the control
3 Endoscopic healing rather than a true patient-oriented outcome
4 The lower limit of the 95%CI crosses over 1.25.

1.2.1.5 Full-dose PPI versus low-dose PPI

1.2.1.5.1 Individual PPIs

Bibliography: Jaspersen 1998 (2)

Lansoprazole 30 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lansoprazole 30 mg	Omeprazole 20 mg	Relative (95% CI)	Absolute		

Healing after 4 weeks treatment (follow-up 4 weeks; assessed with: Endoscopy)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	2/10 (20%)	9/10 (90%)	RR 0.222 (0.06 to 0.78)	700 fewer per 1000 (from 198 fewer to 846 fewer)	⊕⊕OO LOW	IMPORTANT[2]

1 Upper limit of the 95%CI crosses over 0.75, and very low event rate.
2 Endoscopic healing rather than a true patient-oriented outcome

Bibliography: Jaspersen 1998 (2)

Pantoprazole 40 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Pantoprazole 40 mg**	**Omeprazole 20 mg**	**Relative (95% CI)**	**Absolute**		
Healing after 4 weeks treatment (follow-up 4 weeks; assessed with: Endoscopy)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1]	none	3/10 (30%)	9/10 (90%)	RR 0.333 (0.13 to 0.88)	600 fewer per 1000 (from 108 fewer to 783 fewer)	⊕⊕OO LOW	IMPORTANT[2]

1 Upper limit of the 95%CI crosses over 0.75, and very low event rate.
2 Endoscopic healing rather than a true patient-oriented outcome

Mee 1996 (9)

Lansoprazole 30 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Lansoprazole 30 mg**	**Omeprazole 20 mg**	**Relative (95% CI)**	**Absolute**		
Healing after 4 weeks in Grade 3 and 4 patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[4]	none	18/40 (45%)	24/42 (57.1%)	RR 0.79 (0.51 to 1.21)	120 fewer per 1000 (from 280 fewer to 120 more)	⊕⊕OO LOW	IMPORTANT[5]
Healing after 8 weeks in Grade 3 and 4 patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1,2,3]	no serious inconsistency	no serious indirectness	Serious[4]	none	26/37 (70.3%)	27/38 (71.1%)	RR 0.99 (0.74 to 1.32)	7 fewer per 1000 (from 185 fewer to 227 more)	⊕⊕OO LOW	IMPORTANT[5]

1 Concealment of treatment allocation not described.
2 Blinding of outcome assessment not described
3 Imbalance between treatment groups: significantly more smokers in lansoprazole group than omeprazole (28% vs 19%)

4 Upper limit of the 95%CI crosses over 0.75, and very low event rate.
5 Endoscopic healing rather than a true patient oriented outcome

Mossner 1995 (10)

Pantoprazole 40 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pantoprazole 40 mg	Omeprazole 20 mg	Relative (95% CI)	Absolute		
Healing after 4 weeks in Grade 3 patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	serious[3]	none	21/36 (58.3%)	12/22 (54.5%)	RR 1.069 (0.668 to 1.713)	38 more per 1000 (from 181 fewer to 389 more)	⊕⊕OO LOW	IMPORTANT[4]

1 Concealment of treatment allocation not described
2 Blinding of outcome assessment not described
Upper limit of the 95%CI crosses over 0.75, and very low event rate.
4 Endoscopic healing rather than a true patient oriented outcome

Richter 2001 (12)

Esomeprazole 40 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Esomeprazole 40 mg	Omeprazole 20 mg	Relative (95% CI)	Absolute		
Healing after 4 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious imprecision[3]	none	215/317 (67.8%)	152/320 (47.5%)	RR 1.43 (1.24 to 1.64)	204 more per 1000 (from 114 more to 304 more)	LOW	IMPORTANT[2]
Healing after 8 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious imprecision[3]	none	268/317 (84.5%)	217/320 (67.8%)	RR 1.25 (1.14 to 1.36)	170 more per 1000 (from 95 more to 244 more)	LOW	IMPORTANT[2]

1 Blinding of outcome assessment not described
2 Endoscopic healing rather than a true patient oriented outcome
3 The lower limit of 95%CI crosses over 1.25

Schmitt 2006 (13)

Esomeprazole 40 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Esomeprazole 40 mg	Omeprazole 20 mg	Relative (95% CI)	Absolute		
Healing after 4 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious imprecision[2]	none	115/189 (60.8%)	81/169 (47.9%)	RR 1.269 (1.045 to 1.542)	129 more per 1000 (from 22 more to 260 more)	Moderate	IMPORTANT[1]
Healing after 8 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious imprecision[2]	none	167/189 (88.4%)	131/169 (77.5%)	RR 1.140 (1.035 to 1.255)	109 more per 1000 (from 27 more to 198 more)	Moderate	IMPORTANT[1]

1 Endoscopic healing rather than a true patient oriented outcome
2 The lower limit of 95%CI crosses over 1.25

Pace 2005 (11)

Rabeprazole 20 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Rabeprazole 20 mg	Omeprazole 20 mg	Relative (95% CI)	Absolute		
Healing after 8 weeks in Grade 3 patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1,2,3]	no serious inconsistency	no serious indirectness	serious[4]	none	14/15 (93.3%)	13/15 (86.7%)	RR 1.077 (0.847 to 1.369)	67 more per 1000 (from 133 fewer to 320 more)	⊕⊕OO LOW	IMPORTANT[5]

1 Concealment of treatment allocation not described
2 Outcome: 'healing' not clearly defined
3 Blinding of outcome assessment not described
4 Low number of events, the lower limit of 95%CI crosses over 1.25
5 Endoscopic healing rather than a true patient oriented outcome

Kahrilas 2000 (8)

Esomeprazole 40 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Esomeprazole 40 mg	Omeprazole 20 mg	Relative (95% CI)	Absolute		
Healing after 8 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious imprecision[3]	none	136/166 (81.9%)	133/182 (73.1%)	RR 1.121 (1.001 to 1.256)	88 more per 1000 (from 1 more to 187 more)	Low	IMPORTANT[2]

1 Blinding of outcome assessment was not described
2 Endoscopic healing rather than a true patient oriented outcome
3 The effect estimate does not reach MID and the lower limit of 95%CI crosses over 1.25

1.2.1.5.2 Pooled full-dose PPIs vs. low-dose PPIs

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Full-dose PPIs	Low-dose PPIs	Relative (95% CI)	Absolute		
Healing after 4 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
5[a]	randomised trials	Serious[1]	no serious inconsistency	no serious indirectness	serious imprecision[3]	none	374/602 (62.1%)	287/573 (50.1%)	RR 1.24 (1.12 to 1.38)	120 more per 1000 (from 60 more to 190 more)	LOW	IMPORTANT
Healing after 8 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
5[b]	randomised trials	Serious[2]	no serious inconsistency	no serious indirectness	serious imprecision[3]	none	611/724 (84.4%)	521/724 (72%)	RR 1.17 (1.11 to 1.24)	122 more per 1000 (from 79 more to 173 more)	LOW	IMPORTANT

Full-dose PPIs: Lansoprazole 30mg; pantoprazole 40mg; esomeprazole 40mg; rabeprazole 20mg
Low-dose PPIs: Omeprazole 20mg
a Jaspersen (1998); Mee (1996); Mossner (1995); Richter (2001); Schmitt (2006)
b Mee (1996); Kahrilas (2000); Richter (2001); Schmitt (2006); Pace (2005)
1 Three out of the 5 RCTs were downgraded in risk of bias – overall downgraded 1-level
2 Four out of the 5 RCTs were downgraded in risk of bias – overall downgraded 1-level
3 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

1.2.1.6 Double-dose PPI versus low-dose PPI: no trials identified that met the inclusion criteria

1.2.1.7 Full-dose PPI versus full-dose PPI

1.2.1.7.1 Individual PPIs

Fennerty 2005 (3)

Esomeprazole 40 mg compared to Lansoprazole 30 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Esomeprazole 40 mg	Lansoprazole 30 mg	Relative (95% CI)	Absolute		
Healing after 4 weeks in Grades C and D patients (follow-up 4 weeks; assessed with: Endoscopy)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious imprecision[2]	none	278/498 (55.8%)	238/501 (47.5%)	RR 1.175 (1.041 to 1.326)	83 more per 1000 (from 19 more to 155 more)	Moderate	IMPORTANT[1]
Healing after 8 weeks in Grades C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious imprecision[2]	none	386/498 (77.5%)	367/501 (73.3%)	RR 1.058 (0.986 to 1.136)	42 more per 1000 (from 10 fewer to 100 more)	Moderate	IMPORTANT[1]

1 Endoscopic healing rather than a true patient-oriented outcome
2 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

Castell 2002 (14)

Esomeprazole 40 mg compared to Lansoprazole 30 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Esomeprazole 40 mg	Lansoprazole 30 mg	Relative (95% CI)	Absolute		
Healing after 8 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious imprecision[3]	none	552/640 (86.3%)	477/646 (73.8%)	RR 1.17 (1.11 to 1.23)	126 more per 1000 (from 81 fewer to 170 more)	Low	IMPORTANT[2]

1 Blinding of outcome assessment was not described, but study described concealment of treatment allocation
2 Endoscopic healing rather than a true patient oriented outcome
3 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

1.2.1.7.2 Pooled full-dose PPIs vs. full-dose PPIs

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Full-dose PPIs (1)	Full-dose PPIs (2)	Relative (95% CI)	Absolute		
Healing after 4 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1[a]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious imprecision[2]	none	374/602 (62.1%)	287/573 (50.1%)	RR 1.24 (1.12 to 1.38)	120 more per 1000 (from 60 more to 190 more)	Moderate	IMPORTANT
Healing after 8 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
2[b]	randomised trials	Serious[1]	no serious inconsistency	no serious indirectness	serious imprecision[2]	none	611/724 (84.4%)	521/724 (72%)	RR 1.17 (1.11 to 1.26)	122 more per 1000 (from 79 more to 173 more)	Low	IMPORTANT

Full-dose PPIs (1): Esomeprazole 40mg
Full-dose PPIs (2): lansoprazole 30mg
a Fennerty (2005)
b Fennerty (2005); Castell (2002)
1 One out of the 2 RCTs was downgraded in risk of bias – overall downgraded 1-level
2 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

1.2.1.8 Low-dose PPI versus low-dose PPI

1.2.1.8.1 Individual PPIs

Kahrilas 2000 (8)

Esomeprazole 20 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Esomeprazole 20 mg	Omeprazole 20 mg	Relative (95% CI)	Absolute		
Healing after 8 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious imprecision[3]	none	124/165 (75.2%)	133/182 (73.1%)	RR 1.028 (0.908 to 1.165)	20 more per 1000 (from 67 fewer to 121 more)	Low	IMPORTANT[2]

1 Blinding of outcome assessment was not described, but study described concealment of treatment allocation
2 Endoscopic healing rather than a true patient oriented outcome
3 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

Lightdale 2006 (16)

Esomeprazole 20 mg compared to Omeprazole 20 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Esomeprazole 20 mg	Omeprazole 20 mg	Relative (95% CI)	Absolute		
Healing after 8 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious imprecision[3]	none	122/158 (77.2%)	110/154 (71.4%)	RR 1.03 (0.91 to 1.16)	21 more per 1000 (from 64 fewer to 114 more)	Low	IMPORTANT[2]

1 Blinding of outcome assessment not described, but study described concealment of treatment allocation
2 Endoscopic healing rather than a true patient oriented outcome
3 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

1.2.1.8.2 Pooled low-dose vs. low-dose

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Low-dose PPIs (1)	Low-dose PPIs (2)	Relative (95% CI)	Absolute		
Healing after 8 weeks in Grade C and D patients (follow-up 8 weeks; assessed with: Endoscopy)												
2[a]	randomised trials	Serious[1]	no serious inconsistency	no serious indirectness	serious imprecision[2]	none	246/323 (76.2%)	243/336 (72.3%)	RR 1.05 (0.96 to 1.15)	36 more per 1000 (from 29 fewer to 108 more)	Low	IMPORTANT

Low-dose PPIs (1): Esomeprazole 20mg
Low-dose PPIs (2): Omeprazole 20mg
a Kahrilas (2000); Lightdale (2006)
1 One out of the 2 RCTs was downgraded in risk of bias – overall downgraded 1-level
2 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

1.2.2 Outcome – Maintenance

1.2.2.1 Network meta-analysis for maintenance phase

Number of Studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
5 RCTs[a]	not serious[1]	serious[2]	serious[3]	very serious[4]	Very low

[1] No serious limitations. All studies used an appropriate method of randomisation (limited selection bias) and the majority of studies had some level of blinding. Relapse measured according to well defined criteria (unlikely to lead to detection bias).

[2] I^2 not calculated for pairwise comparisons due to model used (cloglog-link hazard ratio model – no direct frequentist equivalent); however, in the odds-ratio-based model that was also explored with these data (see appendix E), I^2 was >50% in 3 links and <50% in 4 others. There was some inconsistency between direct and indirect estimates in the network loop. Definitions of relapse were inconsistent or unclear between trials.

[3] The majority of the evidence came from trials that were not designed or powered to focus on people with severe oesophagitis only, and effectiveness evidence was only available where subgroups of interest were reported.

[4] Wide confidence intervals for effect estimates which are likely due to small study sizes and/or reliance on subgroup results from trials that were not powered to detect differences between treatments in people with severe oesophagitis only. Some (3/10) of the 'links' in network include only 1 trial. As a consequence, there is substantial uncertainty of the ranking within the network.

[a] Robinson (1996); Richter (2004); Metz (2003);De Vault (2006);Lauritsen (2003)

1.2.2.2 PPI vs. placebo

Robinson 1996 (1)

Lansoprazole 15 mg and 30 mg compared to Placebo in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lansoprazole 15 mg and 30 mg	Placebo	Relative (95% CI)	Absolute		
Grade 3 and 4 patients remaining in remission after 12 months (follow-up 12 months; assessed with: Endoscopy)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious	none	52/67 (77.6%)	8/35 (22.9%)	RR 3.40 (1.82 to 6.33)	549 more per 1000 (from 187 more to 1000 more)	High	IMPORTANT[1]

1 Endoscopic healing rather than true patient oriented outcome

1.2.2.3 PPI vs. H2RA

1.2.2.3.1 Individual PPIs and H2RAs

Metz 2003 (2)

Pantoprazole 10 mg compared to Ranitidine 300 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pantoprazole 10 mg	Ranitidine 300 mg	Relative (95% CI)	Absolute		
Grade 3 and 4 patients remaining in remission after 12 months (follow-up 12 months; assessed with: Endoscopy)												
1	randomised trials	serious[1,2,3]	no serious inconsistency	no serious indirectness	very serious[4]	none	0/34 (0%)	3/34 (8.8%)	-	88 fewer per 1000 (from 88 fewer to 88 fewer)	⊕OOO VERY LOW	IMPORTANT[5]

1 Method of randomisation and concealment of treatment allocation not described
2 Blinding of outcome assessment not described
3 Significantly greater drop out rates in rantidine-treated patients compared with pantoprazole
4 Low number of events, RR not calculable, very imprecise.
5 Endoscopic healing rather than true patient oriented outcome

Metz 2003 (2)

Pantoprazole 20 mg compared to Ranitidine 300 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pantoprazole 20 mg	Ranitidine 300 mg	Relative (95% CI)	Absolute		
Grade 3 and 4 patients remaining in remission after 12 months (follow-up 12 months; assessed with: Endoscopy)												
1	randomised trials	serious[1,2,3]	no serious inconsistency	no serious indirectness	very serious[4]	none	15/23 (65.2%)	3/34 (8.8%)	RR 7.391 (2.409 to 22.675)	564 more per 1000 (from 124 more to 1000 more)	⊕OOO VERY LOW	IMPORTANT[5]

1 Method of randomisation and concealment of treatment allocation not described
2 Blinding of outcome assessment not described
3 Significantly greater drop out rates in rantidine-treated patients compared with pantoprazole
4 Although the effect estimate reaches the MID but the low number of events and wide CIs have raised concern on certainty, GDG agreed to downgrade.
5 Endoscopic healing rather than true patient oriented outcome

Metz 2003 (2)

Pantoprazole 40 mg compared to Rantidine 300 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pantoprazole 40 mg	Rantidine 300 mg	Relative (95% CI)	Absolute		
Grade 3 and 4 patients remaining in remission after 12 months (follow-up 12 months)												
1	randomised trials	serious[1,2,3]	no serious inconsistency	no serious indirectness	very serious[4]	none	16/26 (61.5%)	3/34 (8.8%)	RR 6.974 (2.27 to 21.427)	527 more per 1000 (from 112 more to 1000 more)	⊕OOO VERY LOW	IMPORTANT[5]

1 Method of randomisation and concealment of treatment allocation not described
2 Blinding of outcome assessment not described
3 Significantly greater drop out rates in rantidine-treated patients compared with pantoprazole
4 Although the effect estimate reaches the MID but the low number of events and wide CIs have raised concern on certainty, GDG agreed to downgrade.
5 Endoscopic healing rather than true patient oriented outcome

Richter 2004 (3)

Pantoprazole 10 mg compared to Ranitidine 300 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pantoprazole 10 mg	Ranitidine 300 mg	Relative (95% CI)	Absolute		
Grade 3 and 4 patients remaining in remission after 12 months (follow-up 12 months; assessed with: Endoscopy)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	Very serious[2]	none	8/30 (26.7%)	5/26 (19.2%)	RR 1.387 (0.517 to 3.718)	74 more per 1000 (from 93 fewer to 523 more)	Very low	IMPORTANT[3]

1 No description of concealment of treatment allocation
2 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25; and the low number of events and wide CIs have raised concern on certainty, GDG agreed to downgrade 2-level.
3 Endoscopic healing rather than true patient oriented outcome

Richter 2004 (3)

Pantoprazole 20 mg compared to Ranitidine 300 mg in severe erosive esophagitis

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pantoprazole 20 mg	Ranitidine 300 mg	Relative (95% CI)	Absolute		
Grade 3 and 4 patients remaining in remission after 12 months (follow-up 12 months; assessed with: Endoscopy)												
1	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	serious[3]	none	17/31 (54.8%)	5/26 (19.2%)	RR 2.852 (1.219 to 6.672)	356 more per 1000 (from 42 more to 1000 more)	⊕⊕OO LOW	IMPORTANT[4]

1 No description of concealment of treatment allocation
2 Blinding of outcome assessment was not described
3 Although the effect estimate reaches the MID but the low number of events and wide CIs have raised concern on certainty, GDG agreed to downgrade.
4 Endoscopic healing rather than true patient oriented outcome

Richter 2004 (3)

Pantoprazole 40 mg compared to Ranitidine 300 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pantoprazole 40 mg	Ranitidine 300 mg	Relative (95% CI)	Absolute		
Grade 3 and 4 patients remaining in remission after 12 months (follow-up 12 months; assessed with: Endoscopy)												
1	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	serious[3]	none	14/19 (73.7%)	5/26 (19.2%)	RR 3.832 (1.667 to 8.807)	545 more per 1000 (from 128 more to 1000 more)	⊕⊕OO LOW	IMPORTANT[4]

1 No description of concealment of treatment allocation
2 Blinding of outcome assessment was not described
3 Although the effect estimate reaches the MID but the low number of events and wide CIs have raised concern on certainty, GDG agreed to downgrade.
4 Endoscopic healing rather than true patient oriented outcome

1.2.2.3.2 pooled PPIs vs. H2RAs

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PPIs	H2RAs	Relative (95% CI)	Absolute		
Grade 3 and 4 patients remaining in remission after 12 months (follow-up 12 months; assessed with: Endoscopy)												
2[a]	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	No serious	none	70/163 (42.9%)	24/180 (13.3%)	RR 3.21 (2.17 to 4.76)	295 more per 1000 (from 156 more to 501 more)	Moderate	IMPORTANT

PPIs: Pantoprazole 10mg, 20mg, 40mg
H2RAs: Ranitidine 300mg
a Richter (2004); Metz (2003)
1 Both RCTs were downgraded in risk of bias – overall downgraded 1-level.

.

1.2.2.4 Double-dose PPI versus full-dose PPI: no trials identified met the inclusion criteria

1.2.2.5 Full-dose PPI versus low-dose PPI: no trials identified met the inclusion criteria

1.2.2.6 Double-dose PPI versus low-dose PPI: no trials identified met the inclusion criteria

1.2.2.7 Full-dose PPI versus full-dose PPI: no trials identified met the inclusion criteria

1.2.2.8 Low-dose PPI versus low-dose PPI

1.2.2.8.1 Individual PPIs

Lauritsen 2003 (4)

Esomeprazole 20 mg compared to Lansoprazole 15 mg in severe erosive esophagitis

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Esomeprazole 20 mg	Lansoprazole 15 mg	Relative (95% CI)	Absolute		
Grade C and D patients remaining in remission after 6 months (follow-up 6 months; assessed with: Endoscopy)												
1	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	serious imprecision[4]	none	87/114 (76.3%)	60/102 (58.8%)	RR 1.297 (1.071 to 1.572)	175 more per 1000 (from 42 more to 336 more)	Low	IMPORTANT[3]

1 Concealment of treatment allocation not described
2 Blinding of outcome assessment not described
3 Endoscopic healing rather than true patient oriented outcome
4 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

DeVault 2006 (5)

Esomeprazole 20 mg compared to Lansoprazole 15 mg in severe erosive esophagitis

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Esomeprazole 20 mg	Lansoprazole 15 mg	Relative (95% CI)	Absolute		
Grade C and D patients remaining in remission after 6 months (follow-up 6 months; assessed with: Endoscopy)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious imprecision[2]	none	96/121 (79.3%)	91/131 (69.5%)	RR 1.142 (0.987 to 1.321)	99 more per 1000 (from 9 fewer to 223 more)	Moderate	IMPORTANT[1]

1 Endoscopic healing rather than true patient oriented outcome
2 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

1.2.2.8.2 Pooled low-dose PPI versus low-dose PPI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Low-dose PPIs (1)	Low-dose PPIs (2)	Relative (95% CI)	Absolute		
Grade C and D patients remaining in remission after 6 months (follow-up 6 months; assessed with: Endoscopy)												
2[a]	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious imprecision[1]	none	183/235 (77.9%)	151/233 (64.8%)	RR 1.21 (1.07 to 1.36)	136 more per 1000 (from 45 more to 233 more)	Moderate	IMPORTANT

Low-dose PPIs (1): Esomeprazole 20mg
Low-dose PPIs (2): Lansoprazole 15mg
a DeVault (2006); Lauritsen (2003)
1 The effect estimates do not reach MID and the lower limit of 95%CI crosses over 1.25

1.3 Full GRADE profiles (review question 5)

Review question 5i:

In patients with symptoms of dyspepsia who are positive for Helicobacter pylori, which eradication regimens are the most clinically effective in the eradication of H pylori?

1.3.1 Network meta-analysis for *H pylori* eradication

Number of Studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
16 RCTs[a]	not serious[1]	very serious[2]	not serious[3]	very serious[4]	Very low

[1] No serious limitations. All studies used an appropriate method of randomisation (limited selection bias) and the majority of studies had some level of blinding. Eradication was measured using a biological test in all instances (very unlikely to lead to detection bias).

[2] I^2 was 84% for PPI/CLA/NIT vs PPI/AMO/NIT which may indicate considerable heterogeneity; I^2 was 61.3% for PPI vs PPI/AMO/CLA which may indicate considerable heterogeneity; I^2 was 0% for all other comparisons which may indicate that any inconsistency might not be important. There was some inconsistency between direct and indirect estimates in the network loop.

[3] All aspects of PICO conform to review protocol.

[4] Very wide confidence intervals for effect estimates which are likely due to very small study sizes causing uncertainty of the ranking within the network (most were ranked from 1 to 14); many of the 'links' in network include only 1 trial; limited head-to-head trials.

[a] Antos (2006); Arkkila (2005); Basu (2011); Chiba (1999); Ecclissato (2002); Hsu (2001); Katelaris (2000); Katelaris (2002); Koivisto (2005); Laine (2000); Laine (2003); Lee (1999); Lerang (1997)a; Lerang (1997)b; Ohlin (2002); van Zanten (2003)

[all compared to PPI/AMO/CLA]

Abbreviations: RCT, randomised controlled trial.

1.3.2 Eradication (pair-wise comparisons)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Regimen 1	Regimen 2	Relative (95% CI)	Absolute		
Eradication – Regimen 1: PPI/BIS/AMO/AZI (10 days); Regimen 2: PPI/BIS/AMO/CLA (10 days); (assessed with: rapid urease test and histology on repeat endoscopy)												
1	randomised trials[1]	no serious risk of bias	no serious inconsistency	serious[2]	serious[3]	none	15/29 (51.7%)	22/26 (84.6%)	RR 1.64 (1.11 to 2.41)	331 more per 1000 (from 57 more to 729 more)	LOW	CRITICAL
Eradication – Regimen 1: PPI/CLA/NIT (7 days, Nitroimidazole - metronidazole); Regimen 2: PPI/CLA/NIT (7 days, Nitroimidazole - tinidazole); (assessed with: C14 urea breath test)												
1	randomised trials[4]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5]	none	36/41 (87.8%)	44/44 (100%)	RR 0.88 (0.78 to 0.99)	120 fewer per 1000 (from 10 fewer to 220 fewer)	MODERATE	CRITICAL
Eradication – Regimen 1: PPI/AMO/NIT (7 days); Regimen 2: PPI/AMO/NIT (7 days, triple dose); (assessed with: culture, histology and C14 urea breath test)												

1	randomised trials[6]	serious[7]	no serious inconsistency	no serious indirectness	serious[8]	none	32/38 (84.2%)	29/35 (82.9%)	RR 1.02 (0.83 to 1.25)	17 more per 1000 (from 141 fewer to 207 more)	LOW	CRITICAL
Eradication – Regimen 1: PPI/BIS/NIT/TET (10 days); Regimen 2: PPI/BIS/NIT/TET (14 days); (assessed with: histology and C14 urea breath test)												
1	randomised trials[9]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[10]	none	199/215 (92.6%)	185/202 (91.6%)	RR 1.01 (0.96 to 1.07)	9 more per 1000 (from 37 fewer to 64 more)	MODERATE	CRITICAL
Eradication – Regimen 1: PPI/CLA/NIT (7 days, 250mg CLA); Regimen 2: PPI/CLA/NIT (7 days, 500mg CLA); (assessed with: C14 urea breath test)												
1	randomised trials[11]	serious[12]	no serious inconsistency	no serious indirectness	serious[8]	none	62/82 (75.6%)	63/80 (78.8%)	RR 0.96 (0.81 to 1.14)	32 fewer per 1000 (from 150 fewer to 110 more)	LOW	CRITICAL
Eradication – Regimen 1: PPI/AMO/CLA (3 days); Regimen 2: PPI/AMO/CLA (7 days); (assessed with: C14 urea breath test)												
1	randomised trials[13]	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious	none	51/187 (27.3%)	150/194 (77.3%)	RR 0.35 (0.28 to 0.45)	503 fewer per 1000 (from 425 fewer to 557 fewer)	HIGH	CRITICAL
Eradication – Regimen 1: PPI/AMO/CLA (3 days); Regimen 2: PPI/AMO/CLA (10 days); (assessed with: C14 urea breath test)												
1	randomised trials[13]	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious	none	51/187 (27.3%)	304/402 (75.6%)	RR 0.36 (0.28 to 0.46)	484 fewer per 1000 (from 408 fewer to 544 fewer)	HIGH	CRITICAL
Eradication – PPI/AMO/CLA (7 days); Regimen 2: PPI/AMO/CLA (10 days); (assessed with: C14 urea breath test)												
1	randomised trials[13]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[10]	none	150/194 (77.3%)	304/402 (75.6%)	RR 1.02 (0.93 to 1.12)	15 more per 1000 (from 53 fewer to 91 more)	MODERATE	CRITICAL

1 Sullivan (2002)
2 Study included a population with numerous varied conditions including gastric associated lymphoid tissue or intestinal metaplasia
3 95% CI crosses MID
4 Abbas (2003)
5 95% CI borderline to no effect
6 Bayerdorffer (1999)
7 Multi-centre trial (German data was extracted only) but could not determine any of the baseline characteristics by country
8 95% CI crosses MID and 95% CIs cross the line of no effect
9 Dore (2011)
10 95% CI crosses MID and 95% CIs cross the line of no effect
11 Ellenreider (1998)
12 Randomisation protocol used may result in bias
13 Vakil (2004)

1.3.3 Adherence to medication (pairwise comparison)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Regimen 1	Regimen 2	Relative (95% CI)	Absolute		
Adherence to medication – Regimen 1: PPI/AMO/CLA (7 days); Regimen 2: H2RA/BIS/CLA (7 days); (assessed with: tablet/capsule counts)												
1	randomised trials[1]	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	128/152 (84.2%)	143/153 (93.5%)	RR 0.90 (0.83 to 0.98)	93 fewer per 1000 (from 19 fewer to 159 fewer)	MODERATE	CRITICAL
Adherence to medication – Regimen 1: PPI/AMO/CLA (14 days); Regimen 2: PPI/AMO (14 days); (assessed with: tablet/capsule counts)												
1	randomised trials[3]	serious[2]	no serious inconsistency	no serious indirectness	No serious	none	33/34 (97.1%)	30/31 (96.8%)	RR 1.00 (0.92 to 1.09)	0 fewer per 1000 (from 77 fewer to 87 more)	MODERATE	CRITICAL
Adherence to medication – Regimen 1: PPI/AMO/CLA (7 days); Regimen 2: BIS/NIT/TET (14 days); (assessed with: tablet/capsule counts)												
1	randomised trials[5]	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious	none	130/134 (97%)	116/137 (84.7%)	RR 1.15 (1.06 to 1.24)	127 more per 1000 (from 51 more to 203 more)	HIGH	CRITICAL
Adherence to medication – Regimen 1: PPI/AMO/CLA (7 days[5]/10 days[6]); Regimen 2: PPI/BIS/NIT/TET (7 days[5]/10 days[6]); (assessed with: tablet/capsule counts)												
2	randomised trials[5,6]	no serious risk of bias	serious[7]	no serious indirectness	no serious imprecision	none	259/271 (95.6%)	252/272 (92.6%)	RR 1.03 (0.99 to 1.08)	28 more per 1000 (from 9 fewer to 74 more)	MODERATE	CRITICAL
Adherence to medication – Regimen 1: PPI/BIS/NIT/TET (7 days); Regimen 2: BIS/NIT/TET (14 days); (assessed with: tablet/capsule counts)												
1	randomised trials[5]	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious	none	126/134 (94%)	116/137 (84.7%)	RR 1.11 (1.02 to 1.21)	93 more per 1000 (from 17 more to 178 more)	HIGH	CRITICAL
Adherence to medication – Regimen 1: PPI/AMO/CLA (10 days); Regimen 2: PPI/TET/QUI/NTZ (7 DAYS); (assessed with: patient interview during course of therapy)												
1	randomised trials[8]	very serious[9]	no serious inconsistency	no serious indirectness	No serious	none	85/90 (94.4%)	87/90 (96.7%)	RR 0.98 (0.92 to 1.04)	19 fewer per 1000 (from 77 fewer to 39 more)	LOW	CRITICAL
Adherence to medication – Regimen 1: PPI/AMO/CLA (10 days); Regimen 2: PPI/TET/QUI/NTZ (10 days); (assessed with: patient interview during course of therapy)												
1	randomised trials[8]	very serious[9]	no serious inconsistency	no serious indirectness	No serious	none	85/90 (94.4%)	85/90 (94.4%)	RR 1.00 (0.93 to 1.07)	0 fewer per 1000 (from 66 fewer to 66 more)	LOW	CRITICAL
Adherence to medication – Regimen 1: PPI/BIS/NIT/TET (10 days); Regimen 2: PPI/BIS/NIT/TET (14 days); (assessed with: patient interview at completion of therapy)												

1	randomised trials[10]	serious[11]	no serious inconsistency	no serious indirectness	no serious imprecision	none	207/209 (99%)	187/192 (97.4%)	RR 1.02 (0.99 to 1.04)	19 more per 1000 (from 10 fewer to 39 more)	MODERATE	CRITICAL

1 van Zanten (2003)
2 Patients and investigators not blinded
3 Chiba (1996)
5 Katelaris (2002)
6 Laine (2003)
7 Laine (2003) population was active duodenal ulcer patients; Katelaris (2002) population was ulcer negative dyspepsia patients
8 Basu (2011)
9 Limited methodology for compliance measurement given and no allocation blinding following randomisation
10 Dore (2011)
11 Allocation not blinded following randomisation

1.3.4 Adverse events (pairwise comparison)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Regimen 1**	**Regimen 2**	**Relative (95% CI)**	**Absolute**		
Abnormal liver function test – Regimen 1: PPI/CLA/NIT (7 days); Regime 2: PPI/AMO/NIT (7 days); (assessed with: patient interview at 1, 2 and 6 weeks)												
1	randomised trials[1]	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	7/113 (6.2%)	6/114 (5.3%)	RR 0.85 (0.29 to 2.45)	8 fewer per 1000 (from 37 fewer to 76 more)	LOW	IMPORTANT
Dermatitis – Regimen 1: PPI/AMO/CLA (7 days); Regimen 2: PPI/AMO/QUI (7 days); (assessed with: patient interview at completion of treatment)												
1	randomised trials[3]	serious[4]	no serious inconsistency	no serious indirectness	very serious[2]	none	0/31 (0%)	2/30 (6.7%)	RR 0.19 (0.01 to 3.88)	54 fewer per 1000 (from 66 fewer to 192 more)	VERY LOW	IMPORTANT
Rash – Regimen 1: PPI/AMO/NIT (14 days); Regimen 2: BIS/NIT/TET (14 days); (assessed with: patient questionnaire at completion of treatment)												
1	randomised trials[5]	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	9/46 (19.6%)	9/54 (16.7%)	RR 1.17 (0.51 to 2.71)	28 more per 1000 (from 82 fewer to 285 more)	LOW	IMPORTANT
Rash – Regimen 1: PPI/BIS/NIT/TET (7 days); Regimen 2: PPI/AMO/CLA (7 days); (assessed with: patient reported at 2 and 8 weeks)												
1	randomised trials[6]	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	7/134 (5.2%)	4/134 (3%)	RR 1.75 (0.52 to 5.84)	22 more per 1000 (from 14 fewer to 144 more)	LOW	IMPORTANT
Rash – Regimen 1: PPI/BIS/NIT/TET (7 days); BIS/NIT/TET (14 days); (assessed with: patient reported at 2 and 8 weeks)												

1	randomised trials[6]	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	7/134 (5.2%)	16/137 (11.7%)	RR 0.45 (0.19 to 1.05)	64 fewer per 1000 (from 95 fewer to 6 more)	LOW	IMPORTANT
Rash – Regimen 1: PPI/AMO/CLA (7 days); BIS/NIT/TET (14 days); (assessed with: patient reported at 2 and 8 weeks)												
1	randomised trials[6]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[10]	none	4/134 (3%)	16/137 (11.7%)	RR 0.26 (0.09 to 0.74)	86 fewer per 1000 (from 30 fewer to 106 fewer)	MODERATE	IMPORTANT
Loose stools – Regimen 1: PPI/AMO/AZI/BIS (10 days); PPI/AMO/CLA/BIS (10 days); (assessed with: patient recording of side effects during treatment)												
1	randomised trials[7]	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	5/29 (17.2%)	6/27 (22.2%)	RR 0.78 (0.27 to 2.25)	49 fewer per 1000 (from 162 fewer to 278 more)	LOW	IMPORTANT
Loose stools – Regimen 1: PPI/AMO/CLA (7 days); Regimen 2: BIS/H2RA/CLA (7 days); (assessed with: patient checklist at completion of treatment)												
1	randomised trials[8]	serious[9]	no serious inconsistency	no serious indirectness	serious[10]	none	64/156 (41%)	45/156 (28.8%)	RR 1.42 (1.04 to 1.94)	121 more per 1000 (from 12 more to 271 more)	LOW	IMPORTANT
Loose stools – Regimen 1: PPI/AMO/QUI (7 days); Regimen 2: PPI/AMO/CLA (7 days); (assessed with: patient interview at completion of treatment)												
1	randomised trials[3]	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	9/30 (30%)	10/31 (32.3%)	RR 0.93 (0.44 to 1.96)	23 fewer per 1000 (from 181 fewer to 310 more)	LOW	IMPORTANT
Loose stools – Regimen 1: PPI/AMO/CLA (14 days); Regimen 2: PPI/AMO (14 days); (assessed with: patient reported during treatment[11] / completion of treatment[12])												
2	randomised trials[11,12]	serious[13]	no serious inconsistency	no serious indirectness	No serious	none	24/84 (28.6%)	15/129 (11.6%)	RR 2.47 (1.4 to 4.33)	171 more per 1000 (from 47 more to 387 more)	MODERATE	IMPORTANT
Loose stools – Regimen 1: BIS/NIT/TET (14 days); Regimen 2: PPI/AMO/CLA (7 days); (assessed with: patient reported at 2 and 8 weeks)												
1	randomised trials[6]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[10]	none	53/137 (38.7%)	34/134 (25.4%)	RR 1.52 (1.06 to 2.18)	132 more per 1000 (from 15 more to 299 more)	MODERATE	IMPORTANT
Loose stools – Regimen 1: BIS/NIT/TET (14 days); Regimen 2: PPI/BIS/NIT/TET (7 days); (assessed with: patient reported at 2 and 8 weeks)												
1	randomised trials[6]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[10]	none	53/137 (38.7%)	46/134 (34.3%)	RR 0.89 (0.65 to 1.22)	38 fewer per 1000 (from 120 fewer to 76 more)	MODERATE	IMPORTANT
Loose stools – Regimen 1: PPI/BIS/NT/TET (7 days[6] / 10 days[14]); Regimen 2: PPI/AMO/CLA (7 days[6] / 10 days[14]); (assessed with: patient reported at 2 and 8 weeks[6] / completion[14])												

2	randomised trials[6,14]	no serious risk of bias	serious[15]	no serious indirectness	serious[10]	none	69/286 (24.1%)	47/281 (16.7%)	RR 1.45 (1.05 to 2.01)	75 more per 1000 (from 8 more to 169 more)	LOW	IMPORTANT
Loose stools – Regimen 1: BIS/NIT/TET (14 days); Regimen 2: PPI/AMO/NIT (14 days); (assessed with: patient questionnaire at completion of treatment)												
1	randomised trials[5]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[10]	none	41/54 (75.9%)	30/46 (65.2%)	RR 1.16 (0.9 to 1.51)	104 more per 1000 (from 65 fewer to 333 more)	MODERATE	IMPORTANT
Loose stools – Regimen 1: PPI/AMO/NIT (7 days); Regimen 2: PPI/CLA/NIT (7 days); (assessed with: patient interview at 1, 2 and 6 weeks)												
1	randomised trials[1]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[10]	none	13/114 (11.4%)	6/113 (5.3%)	RR 2.15 (0.85 to 5.45)	61 more per 1000 (from 8 fewer to 236 more)	MODERATE	IMPORTANT
Loose stools – Regimen 1: PPI/AMO/NIT (14 days); Regimen 2: H2RA/AMO/NIT (14 days); (assessed with: patient interview at completion of treatment)												
1	randomised trials[18]	very serious[19]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/60 (5%)	4/60 (6.7%)	RR 0.75 (0.18 to 3.21)	17 fewer per 1000 (from 55 fewer to 147 more)	VERY LOW	IMPORTANT
Loose Stools – Regimen 1: PPI/BIS/NIT/TET (14 days); Regimen 2: PPI/BIS/NIT/TET (10 days); (assessed with: patient interview at completion of treatment)												
1	randomised trials[20]	very serious[13]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/202 (1.5%)	5/215 (2.3%)	RR 0.64 (0.15 to 2.64)	8 fewer per 1000 (from 20 fewer to 38 more)	VERY LOW	IMPORTANT
Loose stools – Regimen 1: PPI/CLA/NIT (500mg CLA / 7 days); Regimen 2: PPI/CLA/NIT (250mg CLA / 7 days); (assessed with: patient recorded in a diary during treatment)												
1	randomised trials[22]	serious[23]	no serious inconsistency	no serious indirectness	very serious[2]	none	5/72 (6.9%)	4/71 (5.6%)	RR 1.12 (0.13 to 4.02)	7 more per 1000 (from 49 fewer to 170 more)	VERY LOW	IMPORTANT
Loose stools – Regimen 1: PPI/CLA/NIT (NIT = TIN / 7 days); Regimen 2: PPI/CLA/NIT (NIT = MET / 7 days); (assessed with: questionnaire at completion of treatment)												
1	randomised trials[24]	serious[25]	no serious inconsistency	no serious indirectness	serious[10]	none	2/44 (4.5%)	8/41 (19.5%)	RR 4.29 (0.97 to 19.5)	642 more per 1000 (from 6 fewer to 1000 more)	LOW	IMPORTANT

1 Katelaris (2000)
2 95% CI crosses both MID (0.75 and 1.25)
3 Antos (2006)
4 outcome assessment not blinded
5 Lerang (1997)b
6 Katelaris (2002)
7 Sullivan (2002)

8 van Zanten (2003)
9 Patients and investigators not blinded
10 95% CI crosses one MID
11 Chiba (1996)
12 Ohlin (2002)
13 Methodology unclear for adverse event detection. No blinding following randomisation.
14 Laine (2003)
15 Laine population -active duodenal ulcer, Katelaris population ulcer negative dyspepsia
18 Hsu (2001)
19 Methodology unclear including the adbverse event and randomisation method . No allocation blinding.
20 Dore (2011)
22 Ellenreider (1998)
23 Randomisation protocol used may result in bias
24 Abbas (2003)
25 Methods of randomisation and allocation concealment not given

1.3.5 Antibiotic resistance (pairwise comparison)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Regimen 1	Regimen 2	Relative (95% CI)	Absolute		
Antibiotic resistance (to macrolides) – Regimen 1: PPI/AMO/CLA (14 days); Regimen 2: PPI/AMO (14 days); (assessed with: E-test sensitivity testing at 6 weeks)												
1	randomised trials[1]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	0/1 (0%)[3]	0/41 (0%)[3]	-	-	MODERATE	IMPORTANT
Antibiotic resistance (to penicillins) – Regimen 1: PPI/AMO/CLA (14 days); Regimen 2: PPI/AMO (14 days); (assessed with: E-test sensitivity testing at 6 weeks)												
1	randomised trials[1]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	0/1 (0%)[3]	0/41 (0%)[3]	-	-	MODERATE	IMPORTANT

1 Ohlin (2002)
2 zero event rate, precision not assessable.
3 After treatment *H pylori* was cultured in 42 patients (1 patient treated with PPI/AMO/CLA and 41 patients treated with PPI/AMO)

Review question 5ii:

What *H pylori* eradication regimens should be offered as second-line treatments when first-line treatments fail?

1.3.6 Eradication (pairwise comparison)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Regimen 1	Regimen 2	Relative (95% CI)	Absolute		
Eradication – Regimen 1: PPI/BIS/NIT/TET (7 days); Regimen 2: PPI/BIS/NIT/TET (14 days); (assessed with:)												
2	randomised trials[1,2]	no serious risk of bias	serious[3]	no serious indirectness	No serious	none	80/124 (64.5%)	96/131 (73.3%)	RR 0.88 (0.75 to 1.04)	88 fewer per 1000 (from 183 fewer to 29 more)	MODERATE	CRITICAL
Eradication – Regimen 1: PPI/AMO/NIT (7 days, low-dose); Regimen 2: PPI/AMO/NIT (7 days, high-dose); (assessed with:)												
1	randomised trials[5]	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious	none	106/121 (87.6%)	93/107 (86.9%)	RR 1.01 (0.91 to 1.11)	9 more per 1000 (from 78 fewer to 96 more)	HIGH	CRITICAL
Eradication – Regimen 1: PPI/BIS/AMO/TET (7 days); Regimen 2: PPI/BIS/AMO/TET (14 days); (assessed with:)												
1	randomised trials[6]	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious	none	75/92 (81.5%)	78/95 (82.1%)	RR 0.99 (0.87 to 1.14)	8 fewer per 1000 (from 107 fewer to 115 more)	HIGH	CRITICAL
Eradication – Regimen 1: PPI/AMO/QUI (7 days); Regimen 2: PPI/AMO/QUI (7 days); (assessed with:)												
1	randomised trials[7]	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious	none	50/62 (80.6%)	49/62 (79%)	RR 1.02 (0.85 to 1.22)	16 more per 1000 (from 119 fewer to 174 more)	HIGH	CRITICAL
Eradication – Regimen 1: PPI/AMO/QUI (7 days); Regimen 2: PPI/AMO/QUI (7 days; double-dose); (assessed with:)												
1	randomised trials[8]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[9]	none	26/40 (65%)	28/40 (70%)	RR 0.93 (0.68 to 1.26)	49 fewer per 1000 (from 224 fewer to 182 more)	LOW	CRITICAL
Eradication – Regimen 1: PPI/AMO/QUI (7 days); Regimen 2: PPI/AMO/QUI (10 days); (assessed with:)												
1	randomised trials[8]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	none	26/40 (65%)	36/40 (90%)	RR 0.72 (0.56 to 0.93)	252 fewer per 1000 (from 63 fewer to 396 fewer)	MODERATE	CRITICAL
Eradication – Regimen 1: PPI/AMO/QUI (10 days); Regimen 2: PPI/AMO/QUI (10 days, double-dose); (assessed with:)												
1	randomised trials[8]	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious[4]	none	36/40 (90%)	34/40 (85%)	RR 1.06 (0.9 to 1.25)	51 more per 1000 (from 85 fewer to 213 more)	HIGH	CRITICAL

Eradication – Regimen 1: PPI/AMO/QUI (7 days); Regimen 2: PPI/AMO/QUI (10 days, double-dose); (assessed with:)											
1	randomised trials[8]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	none	26/40 (65%)	34/40 (85%)	RR 0.76 (0.59 to 0.99)	204 fewer per 1000 (from 8 fewer to 349 fewer)	MODERATE	CRITICAL
Eradication – Regimen 1: PPI/AMO/QUI (7 days, double-dose); Regimen 2 - PPI/AMO/QUI (10 days, double-dose); (assessed with:)											
1	randomised trials[8]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	none	28/40 (70%)	34/40 (85%)	RR 0.82 (0.65 to 1.05)	153 fewer per 1000 (from 298 fewer to 42 more)	MODERATE	CRITICAL

1 Mantzaris (2005)
2 Nista (2003)
3 Mantzaris (2005) only included patients with inactive duodenal ulcer; Nista (2003) included non-ulcer dyspepsia patients
4 95% CI crosses one MID
5 Matsuhisa (2006)
6 Uygun (2008)
7 Cheng (2007)
8 Di Caro (2009)
9 95% CIs cross both MIDs

1.3.7 Network meta-analysis for *H pylori* eradication

Number of Studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
18 RCTs[a]	not serious[1]	very serious[2]	not serious[3]	very serious[4]	Very low

[1] No serious limitations. All studies used an appropriate method of randomisation (limited selection bias) and the majority of studies had some level of blinding. Eradication was measured using a biological test in all instances (very unlikely to lead to detection bias).

[2] I^2 was >44.4% for 5 comparisons which indicates inconsistency (between HH2RA/BIS/NIT/TET vs. PPI/BIS/NIT/TET; PPI/BIS/NIT/TET vs. PPI/BIS/AMO/TET; PPI/BIS/NIT/TET vs. PPI/QUI/NIT; PPI/BIS/NIT/TET vs. PPI/AMO/QUI and PPI/AMO/NIT vs. PPI/AMO/QUI) within the network. There was some inconsistency between direct and indirect estimates in the network loop.

[3] All aspects of PICO conform to review protocol.

[4] Very wide confidence intervals for effect estimates which are likely due to very small study sizes causing uncertainty of the ranking within the network; many of the 'links' in network include only 1 trial; limited head-to-head trials.

[a] Bago (2009); Cheon (2006a); Cheon (2006b); Chi (2003); Chuah (2012); Georgopoulos (2002); Gisbert (2007); Gisbert (1999); Hu (2011); Koksal (2005); Kuo (2009); Matsumoto (2006); Michopoulos (2000); Nista (2003); Ueki (2009); Uygun (2008); Wu (2006); Wu (2011);

[all compared to H2RA/BIS/AMO/CLA]

1.3.8 Adherence to medication (pairwise comparison)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Regimen 1	Regimen 2	Relative (95% CI)	Absolute		
Adherence to medication – Regimen 1: BIS/OME/NIT/TET (14 days); BIS/OME/NIT/TET (7 days); (assessed with:)												
1	randomised trials[1]	no serious risk of bias	no serious inconsistency	no serious indirectness	No serious	none	54/61 (88.5%)	51/54 (94.4%)	RR 0.94 (0.84 to 1.05)	57 fewer per 1000 (from 151 fewer to 47 more)	HIGH	IMPORTANT
Adherence to medication – Regimen 1: PPI/AMO/QUI; Regimen 2: PPI/AMO/QUI (double-dose); (assessed with:)												

1[3]	randomised trials[3]	very serious[4]	no serious inconsistency	no serious indirectness	No serious	none	57/60 (95%)	56/62 (90.3%)	RR 1.05 (0.95 to 1.16)	45 more per 1000 (from 45 fewer to 145 more)	LOW	IMPORTANT
Adherence to medication – Regimen 1: PPI/AMO/QUI (10 days); Regimen 2: PPI/AMO/QUI (7 days); (assessed with:)												
1[5]	randomised trials[5]	serious[6]	no serious inconsistency	no serious indirectness	No serious	none	33/40 (82.5%)	36/40 (90%)	RR 0.92 (0.77 to 1.09)	72 fewer per 1000 (from 207 fewer to 81 more)	MODERATE	IMPORTANT
Adherence to medication – Regimen 1: PPI/AMO/QUI (7 days, double-dose); Regimen 2: PPI/AMO/QUI (7 days); (assessed with:)												
1[5]	randomised trials[5]	serious[6]	no serious inconsistency	no serious indirectness	No serious	none	31/40 (77.5%)	36/40 (90%)	RR 0.86 (0.71 to 1.05)	126 fewer per 1000 (from 261 fewer to 45 more)	MODERATE	IMPORTANT
Adherence to medication – Regimen 1: PPI/AMO/QUI (10 days, double-dose); Regimen 2: PPI/AMO/QUI (7 days); (assessed with:)												
1[5]	randomised trials[5]	serious[6]	no serious inconsistency	no serious indirectness	No serious	none	36/40 (90%)	36/40 (90%)	RR 1 (0.86 to 1.16)	0 fewer per 1000 (from 126 fewer to 144 more)	MODERATE	IMPORTANT
Adherence to medication – Regimen 1: PPI/AMO/QUI (10 days); Regimen 2: PPI/AMO/QUI (10 days, double-dose); (assessed with:)												
1[5]	randomised trials[5]	serious[6]	no serious inconsistency	no serious indirectness	serious[2]	none	33/40 (82.5%)	36/40 (90%)	RR 1.09 (0.91 to 1.3)	81 more per 1000 (from 81 fewer to 270 more)	LOW	IMPORTANT
Adherence to medication – Regimen 1: PPI/AMO/QUI (7 days, double-dose); Regimen 2 – PPI/AMO/QUI (10 days, double-dose); (assessed with:)												
1[5]	randomised trials[5]	serious[6]	no serious inconsistency	no serious indirectness	serious[2]	none	31/40 (77.5%)	36/40 (90%)	RR 1.16 (0.95 to 1.41)	144 more per 1000 (from 45 fewer to 369 more)	LOW	IMPORTANT

1 Mantzaris 2005
2 95% CI crosses one MID
3 Cheng 2007
4 No methodology provided for adherence reporting and no blinding in the study
5 Di Caro 2009
6 Randomisation protocol used could potentially lead to bias

1.3.9 Network meta-analysis for adherence to medication

Number of Studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
12 RCTs[a]	not serious[1]	serious[2]	not serious[3]	very serious[4]	Very low

[1] No serious limitations.

[2] I^2 was 0% for all comparisons which may indicate that any inconsistency might not be important. There was some inconsistency between direct and indirect estimates in the network loop.

[3] All aspects of PICO conform to review protocol.

[4] Very wide confidence intervals for effect estimates which are likely due to very small study sizes causing uncertainty of the ranking within the network; many of the 'links' in network include only 1 trial; limited head-to-head trials.

[a] Bago (2009); Cheon (2006b); Chi (2003); Chuah (2012); Georgopoulos (2002); Gisbert (1999); Gisbert 92007); Hu (2011); Koksal (2005); Kuo (2009); Wu (2006); Wu (2011)

[all compared to H2RA/BIS/AMO/CLA]

1.3.10 Adverse events – loose stools (pairwise comparison)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Regimen 1	Regimen 2	Relative (95% CI)	Absolute		
Loose stools – Regimen 1: PPI/AMO/QUI (7 days); Regimen 2: PPI/AMO/QUI (7 days; high-dose); (assessed with:)												
1	randomised trials[1]	very serious[2]	no serious inconsistency	no serious indirectness	very serious[3]	none	3/62 (4.8%)	5/62 (8.1%)	RR 0.60 (0.15 to 2.4)	32 fewer per 1000 (from 69 fewer to 113 more)	VERY LOW	CRITICAL
Loose stools – Regimen 1: PPI/AMO/NIT (7 days; low-dose); Regimen 2: PPI/AMO/NIT (7 days; high-dose); (assessed with:)												
1	randomised trials[4]	very serious[2]	no serious inconsistency	no serious indirectness	No serious	none	9/118 (7.6%)	25/106 (23.6%)	RR 0.32 (0.16 to 0.66)	160 fewer per 1000 (from 80 fewer to 198 fewer)	LOW	CRITICAL
Loose stools – Regimen 1: PPI/BIS/NIT/TET (7 days); Regimen 2: PPI/BIS/NIT/TET (14 days); (assessed with:)												
1	randomised trials[6]	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[3]	none	1/70 (1.4%)	6/70 (8.6%)	RR 0.17 (0.02 to 1.35)	71 fewer per 1000 (from 84 fewer to 30 more)	LOW	CRITICAL

1 Cheng (2007)
2 No methodology provided for adverse event reporting and no blinding in the study
3 95% CIs cross both MIDs
4 Matsuhisa (2006)
6 Nista (2003)

1.3.11 Network meta-analysis for adverse events (loose stools)

Number of Studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
14 RCTs[a]	not serious[1]	serious[2]	not serious[3]	very serious[4]	Very low

[1] No serious limitations.

[2] I^2 was 64.7% for PPI/BIS/NIT/TET vs. PPI/BIS/AMO/TET which may indicate considerable level of heterogeneity; I^2 was 0% for all other comparisons which may indicate that any inconsistency might not be important.

[3] All aspects of PICO conform to review protocol.

[4] Very wide confidence intervals for effect estimates which are likely due to very small study sizes and rare events causing uncertainty of the ranking within the network; almost all of the 'links' in network include only 1 trial; limited head-to-head trials.

[a] Cheon (2006a); Cheon (2006b); Chi (2003); Chuah (2012); Gisbert (2007); Hu (2011); Koksal (2005); Kuo (2009); Matsumoto (2006); Michopoulos (2000); Nista (2003); Ueki (2009); Wu (2006); Wu (2011)

[all compared to H2RA/BIS/AMO/CLA]

1.3.12 Adverse events – mouth dryness (pairwise comparison)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Regimen 1	Regimen 2	Relative (95% CI)	Absolute		
Mouth dryness – Regimen 1: H2RA/BIS/NIT/TET; Regimen 2: H2RA/BIS/AMO/CLA; (assessed with:)												
1	randomised trials[1]	serious[2]	no serious inconsistency	no serious indirectness	very serious[3]	none	0/28 (0%)	2/28 (7.1%)	RR 0.20 (0.01 to 3.99)	57 fewer per 1000 (from 71 fewer to 214 more)	VERY LOW	CRITICAL

1 Koksal (2005)

2 Randomisation protocol used may lead to high risk of bias and lack of blinding was used in the study
3 95% CI crosses both MIDs

1.3.13 Adverse events – rash (pairwise comparison)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Regimen 1	Regimen 2	Relative (95% CI)	Absolute		
Rash – Regimen 1: PPI/BIS/NIT/TET (7 days); Regimen 2: PPI/BIS/NIT/TET (14 days); (assessed with:)												
1	randomised trials[1]	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	0/70 (0%)	1/70 (1.4%)	RR 0.33 (0.01 to 8.04)	10 fewer per 1000 (from 14 fewer to 101 more)	LOW	CRITICAL

1 Nista 2003
2 95% CIs cross both MIDs

1.3.14 Network meta-analysis for adverse events (rash)

Number of Studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
13 RCTs[a]	not serious[1]	not serious[2]	not serious[3]	very serious[4]	Low

[1] No serious limitations.

[2] I^2 was 33.5% for PPI/AMO/QUI vs PPI/AMO/NIT which may indicate low levels of heterogeneity; I^2 was 0% for all other comparisons which may indicate that any inconsistency might not be important.

[3] All aspects of PICO conform to review protocol.

[4] Very wide confidence intervals for effect estimates which are likely due to very small study sizes causing uncertainty of the ranking within the network; all of the 'links' in network include only 1 trial; limited head-to-head trials.

[a] Chuah (2012a); Chuah (2102b); Gisbert (2007); Hu (2011); Koksal (2005); Kuo (2009); Kuo (2013); Matsumoto (2006); Nista (2003) Ueki (2009); Wu (2006); Wu (2011); Michopoulos (2000)

[all compared to H2RA/BIS/AMO/CLA]

1.3.15 Recurrence (pairwise comparison)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Regimen 1	Regimen 2	Relative (95% CI)	Absolute		
Recurrence – Regimen 1:												
1	randomised trials[1]	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	0/36 (0%)	0/45 (0%)	-	-	MODERATE	IMPORTANT

1 Mantzaris (2005)
2 Zero event rate, precision not assessable.

1.4 Full GRADE profile (review question 6)

Review question 6:

What is the effectiveness of laparoscopic fundoplication compared to medical management in patients with GORD?

1.4.1 Health related QOL. SF-36 General (higher score denotes better outcome) 1 year follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lap fundoplication	PPI	Relative (95% CI)	Absolute		
Health related QOL (follow-up median 1 years; measured with: SF-36 general; Better indicated by higher values)												
1[1]	randomised trials	serious[2,4]	no serious inconsistency	no serious indirectness	serious[3]	none	52	52	-	MD 9 higher (0.19 lower to 18.19 higher) Favours lap fundoplication	LOW	CRITICAL

1 Anvari 2006 and Goeree 2011 (one study with two reports)
2 Lack of blinding of intervention - although impractical in this instance
3 Less than 400 patients in continuous outcome
4 Groups may have different prognostic factors at baseline

1.4.2 Health related QOL. REFLUX score (higher score denotes better outcome) 1 year follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lap fundoplication	PPI	Relative (95% CI)	Absolute		
Health related QOL (follow-up median 1 years; measured with: REFLUX score; Better indicated by higher values)												
1[1]	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	178	179	-	MD 11.2 higher (6.89 to 15.51 higher) Favours lap fundoplication	LOW	CRITICAL

1 Grant 2008 & 2012 REFLUX
2 Lack of blinding of intervention - although impractical in this instance
3 Less than 400 patients in continuous outcome

1.4.3 Health related QOL. GERSS score (lower score denotes better outcome) 1 year follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lap fundoplication	PPI	Relative (95% CI)	Absolute		
Health related QOL (follow-up median 1 years; measured with: GERSS score; Better indicated by lower values)												
1[1]	randomised trials	serious[2,4]	no serious inconsistency	no serious indirectness	serious[3]	none	52	52	-	MD 5.3 lower (8.75 to 1.85 lower) Favours lap fundoplication	LOW	CRITICAL

1 Anvari 2006 and Goeree 2011
2 Lack of blinding of intervention - although impractical in this instance
3 Less than 400 patients in continuous outcome
4 Groups may have different prognostic factors at baseline

1.4.4 All Health related QOL. GI wellbeing / REFLUX / GERSS score (higher score denotes better outcome) 1 year follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lap fundoplication	PPI	Relative (95% CI)	Absolute		
Health related QOL (follow-up median 1 years; measured with: GI wellbeing / REFLUX / GERSS score; Better indicated by higher values)												
3[1,4,5]	randomised trials	serious[2]	serious[3]	no serious indirectness	No serious	none	339	339	-	MD 0.45 higher (0.30 to 0.60 higher) Favours lap fundoplication	LOW	CRITICAL

1 Anvari 2006 and Goeree 2011
2 Lack of blinding of intervention - although impractical in this instance

3 Studies using different scales pooled
4 Grant 2008 & 2012 REFLUX
5 Mahon 2005

1.4.5 Health related QOL QOLRAD score (higher score denotes better outcome) 5 years FU

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lap fundoplication	PPI	Relative (95% CI)	Absolute		
Health related QOL (follow-up median 5 years; measured with: REFLUX score; Better indicated by higher values)												
1[1]	randomised trials	Serious[2]	serious inconsistency[3]	no serious indirectness	serious[3]	none	288	266	-	MD 0.37 higher (0.24 to 0.5 higher) Favours lap fundoplication	LOW	CRITICAL

1 Galmiche 2011 LOTUS
2 Lack of blinding of intervention - although impractical in this instance
3 Studies using different scales pooled

1.4.6 Health related QOL REFLUX score 5 years follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lap fundoplication	PPI	Relative (95% CI)	Absolute		
Health related QOL (follow-up median 5 years; measured with: REFLUX score; Better indicated by higher values)												
1[2]	randomised trials	Serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	178	179	-	MD 6.4 higher (1.6 to 11.2 higher) Favours lap fundoplication	LOW	CRITICAL

1 Lack of blinding of intervention - although impractical in this instance
2 Grant 2008 & 2012 REFLUX
3 less than 400 patients in continuous outcome

1.4.7 Health related QOL EQ-5D score 1 year follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Lap fundoplication	PPI	Relative (95% CI)	Absolute		
Health related QOL (follow-up median 1 years; measured with: EQ-5D score; Better indicated by higher values)												

2[1,4]	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	No serious	none	230	231	-	MD 2.16 higher (2.34 lower to 6.65 higher) Favours lap fundoplication	Moderate	CRITICAL

1 Anvari 2006 and Goeree 2011
2 Lack of blinding of intervention - although impractical in this instance
4 Grant 2008 & 2012 REFLUX

1.4.8 Health related QOL EQ-5D score 5 years follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Lap fundoplication**	**PPI**	**Relative (95% CI)**	**Absolute**		
Health related QOL (follow-up median 5 years; measured with: EQ-5D score; Better indicated by higher values)												
1[2]	randomised trials	Serious[1]	no serious inconsistency	no serious indirectness	Serious[3]	none	178	179	-	MD 0.047 higher (0.01 lower to 0.11 higher) Favours lap fundoplication	LOW	CRITICAL

1 Lack of blinding of intervention - although impractical in this instance
2 Grant 2008 & 2012 REFLUX
3 Less than 400 patients in continuous outcome

1.4.9 Health related QOL. SF-36 score 5 years follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Lap fundoplication**	**PPI**	**Relative (95% CI)**	**Absolute**		
Health related QOL (follow-up median 5 years; measured with: SF-36; Better indicated by lower values)												
1[2]	randomised trials	Serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	178	179	-	MD 2.76 higher (0.21 to 5.31 higher) Favours PPIs	LOW	CRITICAL

1 Lack of blinding of intervention - although impractical in this instance
2 Grant 2008 & 2012 REFLUX
3 Less than 400 patients in continuous outcome

1.4.10 Health related QOL. Visual Analogue Scale score1 year follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Lap fundoplication**	**PPI**	**Relative (95% CI)**	**Absolute**		

Health related QOL (follow-up median 1 years; measured with: Visual Analogue Scale; Better indicated by higher values)												
2[1,4]	randomised trials	serious[2,5]	no serious inconsistency	no serious indirectness	No serious	none	230	231	-	MD 2.67 higher (0.56 lower to 5.89 higher) Favours lap fundoplication	Moderate	CRITICAL

1 Anvari 2006 and Goeree 2011
2 Lack of blinding of intervention - although impractical in this instance
4 Grant 2008 & 2012 REFLUX
5 Groups may have different prognostic factors at baseline

1.4.11 Symptom Control. Proportion of patients in remission 5 years follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Lap fundoplication**	**PPI**	**Relative (95% CI)**	**Absolute**		
Symptom control (follow-up median 5 years; assessed with: Patients symptom free with no medication.												
1[1]	randomised trials	serious[2,3]	no serious inconsistency	no serious indirectness	No serious	none	245/288 (85.1%)	245/266 (92.1%)	RR 0.92 (0.87 to 0.98) (favours PPI medication group)	8 fewer per 1000 (from 2 fewer to 13 fewer)	Moderate	CRITICAL

1 Galmiche 2011 LOTUS
2 Lack of blinding of intervention - although impractical in this instance
3 Incomplete / inconsistent follow up of patients for certain outcomes without ITT analysis

1.4.12 Symptom Control. Patients with acid reflux 5 years follow-up (Dichotomous outcome)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Lap fundoplication**	**PPI**	**Relative (95% CI)**	**Absolute**		
Symptom control (follow-up median 5 years; assessed with: Acid regurgitation)												
1[1]	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	No serious	none	6/288 (2.1%)	35/266 (13.2%)	RR 0.16 (0.07 to 0.37) (favours lap fundoplication group)	84 fewer per 1000 (from 63 fewer to 93 fewer)	Moderate	IMPORTANT

1 Galmiche 2011 LOTUS
2 Lack of blinding of intervention - although impractical in this instance

1.4.13 Mortality. Overall mortality at 1 year follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Lap fundoplication**	**PPI**	**Relative**	**Absolute**		

									(95% CI)			
Mortality (follow-up median 1 years; assessed with: Absolute mortality)												
1[1]	randomised trials	serious[2,3]	no serious inconsistency	no serious indirectness	serious[4]	none	0/52 (0%)	0/52 (0%)	-	-	LOW	CRITICAL
										-		

1 Anvari 2006 and Goeree 2011
2 Lack of blinding of intervention - although impractical in this instance
3 Groups may have different prognostic factors at baseline
4 Zero event - unable to calculate relative risk, high uncertainty of the effect estimate.

1.4.14 Serious adverse event: Any serious event reported (either bleeding, perforation, pneumothorax, or dysphagia) at 1 year follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Lap fundoplication**	**PPI**	**Relative (95% CI)**	**Absolute**		
Serious adverse event (any of the following events reported)(bleeding, perforation, pneumothorax, dysphagia) (follow-up mean 1 years; assessed with)												
3[1,3,4]	randomised trials	serious[2,5]	no serious inconsistency	no serious indirectness	serious[6]	none	15/337 (4.5%)	0/338 (0%)	_	-	LOW	IMPORTANT

1 Anvari 2006 and Goeree 2011
2 Lack of blinding of intervention - although impractical in this instance
3 Grant 2008 & 2012 REFLUX
4 Mahon 2005
5 Differential drop out and no ITT analysis
6 Zero event in one arm - unable to calculate relative risk, high uncertainty of the effect estimate

1.4.15 Acid reflux – 24hr monitoring. % time <4pH 1 year follow-up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Lap fundoplication**	**PPI**	**Relative (95% CI)**	**Absolute**		
pH monitoring % time <4 1 year FU (follow-up median 1 years; Better indicated by higher values)												
1[1]	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	52	52	-	MD 3.63 higher (1.15 to 6.12 higher) Favours lap fundoplication	LOW	CRITICAL

1 Anvari 2006 and Goeree 2011
2 Lack of blinding of intervention - although impractical in this instance
3 Less than 400 patients in continuous outcome

1.5 Full GRADE profiles (review question 8)

Review question 8:

Should surveillance be used for patients with Barrett's oesophagus to detect progression to cancer, and improve survival?

1.5.1 Cancer incidence

1.5.1.1 Cohort studies – all studies

1 Patients in formal arm had only 1 year follow up

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Surveillance	No surveillance	Relative (95% CI)	Absolute		
Cancer incidence - Cohort studies (follow-up mean 4.9 years; measured with: Incidence per patient year follow up; Better indicated by lower values)												
3[2,3,9]	observational studies	serious[5]	serious[1,6]	serious[7]	not assessable	serious[8]	Range from 108 to 195	-	-	Incidence range from 0.37 to 1.85% (per patient year)	VERY LOW	CRITICAL

2 Fitzgerald (2001)
3 Gladman (2006)
5 Patients selected for surveillance based on age and fitness to undergo surgery
6 Control arm of trial was informal surveillance rather than no surveillance
7 Patients with a mixture of levels of dysplasia were included
8 Protocol excluded studies with n<100 patients
9 Macdonald (2000)

1.5.1.2 Case series - all studies

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Surveillance case series	Control	Relative (95% CI)	Absolute		
cancer incidence per patient year - overall (follow-up mean 6550 patient-years)												
20[4,16,17,18,20,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39]	observational studies[7]	no serious risk of bias	serious[12]	serious[3]	not assessable	none	Range from 101 to 16365	-	-	Incidence range from 0.00 to 2.03% (per	VERY LOW	CRITICAL

										patient year)		

For table notes please see end of document

1.5.1.3 Subgroup analysis by degree of dysplasia at baseline

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Surveillance case series	Control	Relative (95% CI)	Absolute		
cancer incidence per patient year - No HGD (follow-up mean 13465 patient-years)												
6[28,29,31,34,40,41]	observational studies[7]	no serious risk of bias	serious[12]	no serious indirectness	not assessable	none	Range from 102 to 16365	-	-	Incidence range from 0.36 to 0.65% (per patient year)	VERY LOW	CRITICAL
cancer incidence per patient year - No LGD or HGD (follow-up mean 3817 patient-years)												
2[4,38]	observational studies[7]	no serious risk of bias	serious[12]	no serious indirectness	not assessable	none	Range from 248 to 1204	-	-	Incidence range from 0.27 to 0.51% (per patient year)	VERY LOW	CRITICAL
cancer incidence per patient year - Mixed (follow-up mean 2764 patient-years)												
14[16,17,18,20,25,26,27,30,32,33,35,36,37,39]	observational studies[7]	no serious risk of bias	serious[42]	no serious indirectness	not assessable	none	Range from 101 to 1099	-	-	Incidence range from 0.00 to 2.03% (per patient year)	VERY LOW	CRITICAL

For table notes please see end of document

1.5.1.4 Case series – all studies - alternative analysis (studies with ≤5% HDG patients grouped as no HGD)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Surveillance case series	Control	Relative (95% CI)	Absolute		
Cancer incidence per patient year <5% HGD grouped as no HGD - overall (follow-up mean 6550 patient-years)												
20[4,16,17,18,20,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39]	observational studies[7]	no serious risk of bias	serious[12]	serious[3]	not assessable	none	Range from 101 to 16365	-	-	Incidence range from 0.00 to 2.03% (per patient year)	VERY LOW	CRITICAL

For table notes please see end of document

1.5.1.5 Case series – subgroup analysis by degree of dysplasia at baseline - alternative analysis (studies with ≤5% HDG patients grouped as no HGD)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Surveillance case series	Control	Relative (95% CI)	Absolute		
Cancer incidence per patient year <5% HGD grouped as no HGD - No HGD (follow-up mean 10249 patient-years)												
10[25,26,28,29,31,32,34,35,36,39]	observational studies[7]	serious[46]	serious[12]	no serious indirectness	not assessable	none	Range from 102 to 16365	-	-	Incidence range from 0.00 to 2.03% (per patient year)	VERY LOW	CRITICAL
Cancer incidence per patient year <5% HGD grouped as no HGD - No LGD or HGD (follow-up mean 3817 patient-years)												
2[4,38]	observational studies[7]	no serious risk of bias	serious[12]	no serious indirectness	not assessable	none	Range from 248 to 1204	-	-	Incidence range from 0.27 to 0.51% (per patient year)	VERY LOW	CRITICAL
Cancer incidence per patient year <5% HGD grouped as no HGD - Mixed (follow-up mean 2211 patient-years)												
7[16,17,20,27,30,33,37]	observational studies[7]	no serious risk of bias	serious[42]	no serious indirectness	not assessable	none	Range from 101 to 1099	-	-	Incidence range from 0.00 to 0.37% (per patient year)	VERY LOW	CRITICAL

For table notes please see end of document

1.5.2 HGD incidence

1.5.2.1 Cohort studies – all studies

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Surveillance	No surveillance	Relative (95% CI)	Absolute		
HGD incidence - Cohort studies (follow-up mean 4.9 years; measured with: Incidence per patient year follow up; Better indicated by lower values)												
2[2,3]	observational studies	serious[5]	serious[1,6]	serious[7]	not assessable	serious[8]	Range from 108 to 195	-	-	Incidence range from 0.19 to 0.27% (per patient year)	VERY LOW	CRITICAL

1 Patients in formal arm had only 1 year follow up
2 Fitzgerald (2001)
3 Gladman (2006)
5 Patients selected for surveillance based on age and fitness to undergo surgery
6 Control arm of trial was informal surveillance rather than no surveillance
7 Patients with a mixture of levels of dysplasia were included
8 Protocol excluded studies with n<100 patients

1.5.2.2 Case series – all studies

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Surveillance case series	Control	Relative (95% CI)	Absolute		
HGD incidence per patient year - overall (follow-up mean 7396 patient-years)												
17[4,16,20,25,26,28,29,31,32,34,35,36,37,38,39,40,43]	observational studies[7]	no serious risk of bias	serious[3]	serious[44]	not assessable	none	Range from 102 to 16365	-	-	Incidence range from 0.05 to 1.67% (per patient year)	VERY LOW	CRITICAL

For table notes please see end of document

1.5.2.3 Subgroup analysis by degree of dysplasia at baseline

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Surveillance case series	Control	Relative (95% CI)	Absolute		
HGD incidence per patient year - No HGD (follow-up mean 1272 patient-years;)												
6[28,29,31,34,40,43]	observational studies[7]	no serious risk of bias	serious[3]	serious[44]	not assessable	none	Range from 102 to 713	-	-	Incidence range from 0.21 to 1.03% (per patient year)	VERY LOW	CRITICAL
HGD incidence per patient year - No LGD or HGD (follow-up mean 3817 patient-years;)												
2[4,38]	observational studies[7]	no serious risk of bias	no serious inconsistency	serious[45]	not assessable	none	Range from 248 to 1204	-	-	Incidence range from 0.41 to 0.48% (per patient year)	VERY LOW	CRITICAL
HGD incidence per patient year - Mixed (follow-up mean 3865 patient-years)												
9[16,20,25,26,32,35,36,37,39]	observational studies[7]	no serious risk of bias	no serious inconsistency	serious[42]	not assessable	none	Range from 121 to 1099	-	-	Incidence range from 0.05 to 1.67% (per patient year)	VERY LOW	CRITICAL

For table notes please see end of document

1.5.2.4 Case series – all studies - alternative analysis (studies with ≤5% HDG patients grouped as no HGD)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Surveillance case series	Control	Relative (95% CI)	Absolute		

HGD incidence per patient year <5% HGD grouped as no HGD - overall (follow-up mean 7396 patient-years)												
17[4,16,20,25,26,28,29,31,32,34,35,36,37,38,39,40,43]	observational studies[7]	no serious risk of bias	serious[3]	serious[44]	not assessable	none	Range from 102 to 16365	-	-	Incidence range from 0.05 to 1.67% (per patient year)	VERY LOW	CRITICAL

For table notes please see end of document

1.5.2.5 Case series - subgroup analysis by degree of dysplasia at baseline - alternative analysis (studies with ≤5% HDG patients grouped as no HGD)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Surveillance case series**	**Control**	**Relative (95% CI)**	**Absolute**		
HGD incidence per patient year <5% HGD grouped as no HGD - No HGD (follow-up mean 8802 patient-years;)												
12[25,26,28,29,31,32,34,35,36,39,40,43]	observational studies[7]	no serious risk of bias	serious[3]	serious[44]	not assessable	none	Range from 102 to 16365	-	-	Incidence range from 0.21 to 1.67% (per patient year)	VERY LOW	CRITICAL
HGD incidence per patient year <5% HGD grouped as no HGD - No LGD or HGD (follow-up mean 3817 patient-years)												
2[4,38]	observational studies[7]	no serious risk of bias	no serious inconsistency	serious[45]	not assessable	none	Range from 248 to 1204	-	-	Incidence range from 0.41 to 0.48% (per patient year)	VERY LOW	CRITICAL
HGD incidence per patient year <5% HGD grouped as no HGD - Mixed (follow-up mean 4158 patient-years)												
3[16,20,37]	observational studies[7]	no serious risk of bias	no serious inconsistency	serious[42]	not assessable	none	Range from 123 to 1099	-	-	Incidence range from 0.40 to 0.56% (per patient year)	VERY LOW	CRITICAL

For table notes please see end of document

1.5.3 Oesophageal Cancer related Mortality

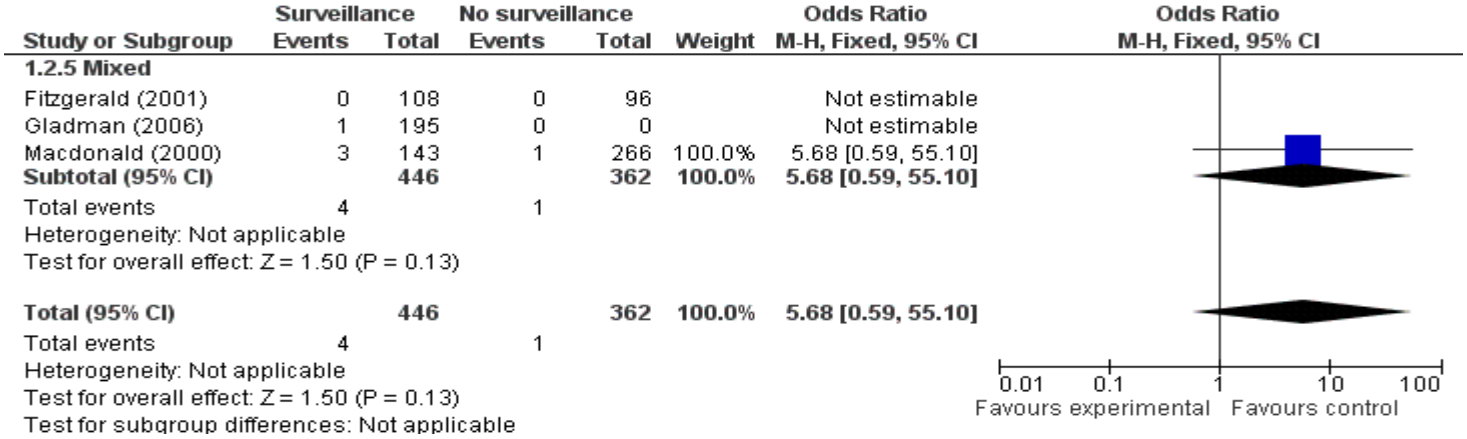
1.5.3.1 Cohort studies - all studies

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Surveillance**	**No surveillance**	**Relative (95% CI)**	**Absolute**		

Mortality - Mixed (follow-up mean 4.9 years; assessed with: Oesophageal cancer related mortality)												
3[2,3,4]	observational studies	serious[5]	no serious inconsistency	no serious indirectness	serious imprecision[11]	none	4/446 (0.9%)	1/362 (0.3%)	OR 5.68 (0.59 to 55.1)	13 more per 1000 (from 1 fewer to 130 more)	VERY LOW	CRITICAL

1 Patients in formal arm had only 1 year follow up
2 Fitzgerald (2001)
3 Gladman (2006)
4 Macdonald (2000)
5 Patients selected for surveillance based on age and fitness to undergo surgery

1.5.3.1.1 Forest plot Surveillance Vs No surveillance, outcome: Mortality



1.5.3.2 Case control study

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Cases in surveillance**	**Controls in surveillance**	**Relative (95% CI)**	**Absolute**		
Mortality: Case control study - (follow-up: 14 years; assessed with: Oesophageal cancer related mortality): adjusted for dysplasia status												
1[10]	observational study (case control)	No serious	no serious inconsistency	no serious indirectness	serious imprecision[13]	none	21/38 (55.3%)	61/101 (60.4%)	Adj OR 0.99 (0.36 to 2.75)	NR	VERY LOW	CRITICAL

Mortality: Case control study - (follow-up: 14 years; assessed with: Oesophageal cancer related mortality): adjusted for dysplasia status and length of BO											
1[10]	observational study (case control)	No serious	no serious inconsistency	no serious indirectness	serious imprecision[13]	none	21/38 (55.3%)	61/101 (60.4%)	Adj OR 1.14 (0.39 to 3.32)	NR	VERY LOW	CRITICAL

1.5.3.3 Case series – all studies

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Surveillance case series**	**Control**	**Relative (95% CI)**	**Absolute**		
Mortality (follow-up 3.8 to 7.3 years; assessed with: Oesophageal cancer related mortality)												
5[4,15,16,17,18]	observational studies[7]	no serious risk of bias	no serious inconsistency	serious[12]	not assessable	none	0/248 (0%)[4] 0/705 (0%)[15] 1/1099 (0.009%)[16] 1/136 (0.74%)[17] 2/212 (0.94%)[18]	-	-	-	VERY LOW	IMPORTANT

For table notes please see end of document

1.5.4 Quality of life

1.5.4.1 Case series – all studies

Quality assessment							No of patients		Effect	Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Surveillance case series**	**Control**	**Absolute**		
Quality of life Hospital anxiety and depression (HAD) Anxiety (0 to 21 lower scores better) (measured with: HAD anxiety scale; Better indicated by lower values)											
2[5,6]	observational studies[7]	no serious risk of bias	serious[8]	serious[9]	not assessable	none	151 and 192	-	Scores: 5.3 and 6.1	VERY LOW	IMPORTANT
Quality of life Hospital anxiety and depression (HAD) depression (0 to 21 lower scores better) (measured with: HAD depression scale; Better indicated by lower values)											
2[5,6]	observational studies[7]	no serious risk of bias	serious[12]	serious[9]	not assessable	none	151 and 192	-	Scores: 2.4 and 4.0	VERY LOW	IMPORTANT
Quality of life Trust in Physician score (TIPS) (11 to 55 points higher score better) (measured with: TIPS score; Better indicated by higher values)											
1[5]	observational studies[7]	no serious risk of bias	serious[12]	no serious indirectness	not assessable	none	151	-	Median score 44 points, range 27 to 55 points	VERY LOW	IMPORTANT
Quality of life - QOLRAD (measured with: Patient self reported scale; 0 to 7 points Better indicated by higher values)											
1[14]	observational	no serious	no serious	no serious	not	none	15	-	Mean score 6.8 points	VERY	IMPORTANT

	studies[7]	risk of bias	inconsistency	indirectness	assessable					LOW	
Preference for treatment of HGD Surveillance / oesophagectomy / PDT[21] (measured with: % choosing each scenario)											
1[22]	observational studies[7]	no serious risk of bias	serious[23]	no serious indirectness	not assessable	none	20	-	Significantly more patients chose Surveillance 70% (14/20) , than oesophagectomy 15% (3/20) , and PDT 15% (3/20) (p=0.0024) two tailed Chi-square	VERY LOW	IMPORTANT
Satisfaction score on 7 point likert scale[24] (measured with 0 to 7 points likert scale - higher scores better; Better indicated by higher values)											
1[20]	observational studies[7]	no serious risk of bias	no serious inconsistency	serious[12]	not assessable	none	123	-	88% of 102 patients who returned questionnaires were very satisfied (6+ on 0 to 6 scale) with their care	VERY LOW	IMPORTANT
Quality of life – SF-36 (measured with: SF-36 domains 0 to 100 points Better indicated by higher values)											
1[5]	observational studies[7]	no serious risk of bias	serious[12]	no serious indirectness	not assessable	none	151	-	Pain 57.2 points, General perception of health 53.9 points, mental health 72.4 points, physical functioning 57.0 points, role limitations emotional 63.0, role limitations physical 50.9, social functioning 88.1, energy 53.1. All SF-36 domains were significantly lower in the BO surveillance patients than in an age, sex, and socio-economic adjusted general population cohort except for mental health	VERY LOW	IMPORTANT

For table notes please see end of document

1.5.5 Adverse events

1.5.5.1 Case series – all studies

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Surveillance case series**	**Control**	**Relative (95% CI)**	**Absolute**		
Adverse events (follow-up 3.8 to 7.3; assessed with: Serious adverse event as defined in protocol)												
3[15,17,20]	observational studies[7]	no serious risk of bias	no serious inconsistency	serious[3]	not assessable	none	5/705 (0.5%)[15] 0/136 (0%)[17] 0/123 (0%)[20]	-	-	-	VERY LOW	CRITICAL
							Bleeding attributed to concomitant oesophageal stricture dilation (2 patients); cardiac dysrhythmias (2 patients); and one respiratory arrest			-		

For table notes please see end of document

1.5.6 Table notes

1 Control arm of trial was informal surveillance rather than no surveillance
2 Patients in formal arm had only 1 year follow up
3 Patients with a mixture of levels of dysplasia were included

4 Wong (2010)
5 Cooper (2009)
6 Kruijshaar (2006)
7 Case series
8 High lost to follow up
9 All SF-36 domains were significantly lower in the BO surveillance patients than in an age, sex, and socio-economic adjusted general population cohort except for mental health
10 Chorley (2013)
11 GDG unable to define MIDs, very low event rate, high uncertainty of the precision.
12 Patients selected for surveillance based on age and fitness to undergo surgery
13 No model diagnostics for the regression model, high uncertainty on precision.
14 Fisher (2002)
15 Levine (2000)
16 Schnell (2001)
17 Streitz (1998)
18 Switzer-Taylor (2008)
20 Schoenfeld (1998)
22 Hur (2005)
23 Patients instructed to imagine scenario where they had dysplasia. Profile of safety and efficacy of treatment options presented is questionable.
25 Abela (2008)
26 Ajumobi (2010)
27 Bani-Hani (2000)
28 Conio (2003)
29 de Jonge (2010)
30 Drewitz (1997)
31 Ferraris (1997)
32 Hillman (2003)
33 Horwhat (2007)
34 Katz (1998)
35 O'Connor (1999)
36 Olithselvan (2007)
37 Ramus (2009)
38 Wani (2011)
39 Weston (2004)
40 Murphy (2005)
41 Nilsson (2000)
42 Recall period varied during the study
43 Sikkema (2011)
44 Circumferential quad biopsy not used in all patients
45 Not all patients were on PPIs for acid suppression a proportion on H2RAs
46 Follow up was initially retrospective, and later prospective