Appendix E: GRADE profiles

E.1 Question 4.1: Signs and symptoms of Coeliac disease

Intussusception in adults

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With coeliac	Without coeliac	OR (95% CI) Absolute (95% CI)	
Ludvigsson (2013)	Case-control	Serious[1]	N/A	None[2]	Serious[3]	29096	144522	1.17 (0.84, 2.05)	LOW

Low BMI (<18.5) in adults

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With coeliac	Without coeliac	OR (95% CI)	Absolute (95% CI)	
Olen (2009)	Case-control	Serious[1]	N/A	None[2]	Serious[3]	174	787986	2.2 (1. 0, 4.8)		LOW

Visual acuity defects in adults

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With coeliac	Without coeliac	OR (95% CI) Absolute (95% CI)	
Mollazadegan (2009)	Case-control	Very serious[4]	N/A	None[2]	Serious[3]	69	6850	1.04 (0.63, 1.70)	VERY LOW

[1] Serious risk of bias as assessed by CASP cohort study quality appraisal checklist
[2] No serious indirectness, population were as specified in protocol
[3] Serious imprecision, confidence intervals are wide and cross line of no effect
[4] Very serious risk of bias as assessed by CASP cohort study quality appraisal checklist

Migraine in children

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With coeliac	Without coeliac	OR (95% CI)	Absolute (95% CI)	
Alehan (2008)	Case-control	Serious[1]	Serious[2]	None[3]	Very Serious[4]	5	215	8.46 (0.92, 77.15)		VERY LOW
Inaloo (2009)	Case-control	Serious[1]	Serious[2]	None[3]	Serious[5]	32	1558	1.00 (0.23, 4.24)		LOW

Apthous ulcers in children

Quality assessment						Number of patients		Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With coeliac	Without coeliac	OR (95% CI)	Absolute (95% CI)	
Campisi (2008)	Case-control	Serious[2]	N/A[6]	None[3]	Serious[5]	102	742	3.82 (2.49, 5.86)		LOW

Dental enamel defects

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	With coeliac	Without coeliac	OR (95% CI) Absolute (95% CI)	
El-Hodod (2012)	Case-control	Serious[2]	N/A[6]	None[3]	Serious[5]	32	828	22.4 (9.36, 52.37)	LOW

[1] serious risk of bias as assessed by CASP cohort study quality appraisal checklist
[2] Serious inconsistency, OR estimate and confidence intervals around OR do not overlap
[3] No serious indirectness, population was as defined in protocol
[4] Very serious imprecision, confidence intervals are very wide
[5] Serious imprecision - confidence intervals are wide
[6] N/A - Not applicable, only one study contributed to this analysis

E.2 Question 4.2

Modified GRADE profile for prevalence of coeliac disease in coexisting conditions and first-degree relatives

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Percentage with Coeliac disease (95% CI)	Quality
Addison's disease									
1 (Fichna et al., 2010)	Cross-sectional	No serious [1]	NA [2]	No serious [3]	Serious [4]	NA	85	1.2% (0.0, 6.4%)	VERY LOW
Arthritis									
3 (Atzeni et al., 2008; Coacciloli et al., 2010; Francis et al., 2002)	Case-series (1) and Cross-sectional (2)	No serious [1]	No serious [5]	No serious [3]	Serious [6]	NA	231	3.0% (0.8, 11.0%)	VERY LOW
Juvenile arthritis									
3 (George et al., 1996; Lepore et al., 1996; Robazzi et al., 2013)	Cross-sectional	No serious [1]	No serious [5]	No serious [3]	Serious [4]	NA	224	2.3% (0.9%, 5.3%)	VERY LOW
Cardiomyopathy (Adults)									

[1] Study at medium risk of bias but this is not expected to impact of findings
[2] Single study analysis
[3] Population and tests as specified in the review protocol
[4] Confidence intervals around point estimate cross the MID of prevalence (1%) of coeliac disease in the general population
[5] Low heterogeneity (I-squared less than 33%)
[6] Confidence intervals around point estimate cross the MID of prevalence (1%) of coeliac disease in the general population

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Percentage with Coeliac disease (95% CI)	Quality
3 (Chicco et al., 2010; Frustaci et al., 2002; Vizzardi et al., 2008)	Case-control (1) and Cross-sectional (2)	No serious [1]	Serious[1]	No serious[3]	Serious[4]	NA	641	2.2% (0.7%, 6.4%)	VERY LOW
Cardiomyopathy (Children)									
1 (De Menzes et al., 2012)	Cross-sectional	No serious [1]	NA[2]	No serious[3]	Serious[4]	NA	56	1.8% (0.3%, 9.5%)	VERY LOW
Down syndrome									
5 (Bonamico et al., 2001; Cerqueria et al., 2010; Goldacre et al., 2004; Pavlović et al., 2012; Wouters et al., 2009)	Case-control (1) and Cross-sectional (4)	No serious [1]	Very serious[2]	No serious[3]	No serious[6]	NA	2999	3.2% (1.3%, 7.4%)	LOW
Epilepsy									
4 (Cronin et al., 1998; Djurić et al., 2010; Peltola et al., 2009; Pratesi et al., 2003)	Case-control (3) and Cross-sectional (1)	No serious [1]	No serious[5]	No serious[3]	No serious[6]	NA	605	3.6% (1.9%, 6.7%)	LOW
Gastrointestinal conditions - Dyspepsia									

[1] Moderate heterogeneity (I-squared between 34% and 66%)
[2] High heterogeneity (I-squared greater than 67%)
[3] Population and tests as specified in the review protocol
[4] Confidence intervals around point estimate cross the MID of prevalence (1%) of coeliac disease in the general population
[5] Low heterogeneity (I-squared less than 33%)
[6] Confidence intervals around point estimate cross the MID of prevalence (1%) of coeliac disease in the general population

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Percentage with Coeliac disease (95% CI)	Quality
1 (Giangreco et al., 2008)	Cross-sectional	No serious [1]	NA[2]	No serious[3]	No serious[6]	NA	726	2.1% (1.3%, 3.4%)	LOW
Gastrointestinal conditions – Irritable bowel syndrome									
5 (Cash et al., 2011; Cristori et a., 2014; El-Salhy et al., 2011; Sanders et al., 2001; Sanders et al., 2003)	Case-control (1) and Cross-sectional (3)	No serious [1]	Very serious[8]	No serious[3]	Serious[4]	NA	2153	1.8% (0.7, 4.7%)	VERY LOW
Gastrointestinal conditions – Other									
5 (Aziz et al., 2010; Casella et al., 2010; Leeds et al., 2007; Lynch et al., 1995; Simondi et al., 2010)	Case-control (1) and Cross-sectional (4)	No serious [1]	Very serious[8]	No serious[3]	Serious[4]	NA	2220	2.9% (0.5, 16.6%)	VERY LOW
Liver disease									
9 (Bardella et al., 1997; Chatzicostas et al., 2002; Dickey et al., 1997; Drastich et al., 2012; Eapen et al., 2011; Gatselis et al., 2012; Germenis et al., 2005; Olsson et al., 1982; Thevenot et al., 2007)	Case-control (1), case series (2) and Cross-sectional (6)	No serious [1]	Very serious[8]	No serious[3]	Serious[4]	NA	3233	2.0% (0.7, 5.8%)	VERY LOW
Neurological disease									
1 (Ruggieri et al.,	Case control	No	NA[2]	No serious[3]	Serious[4]	NA	650	1.1% (0.5%, 2.3%)	VERY LOW

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Percentage with Coeliac disease (95% CI)	Quality
2008)		serious [1]							
Sarcidosis									
1 (Papadopoulos et al., 1999)	Cross-sectional	No serious [1]	NA [2]	No serious [3]	No serious [6]	NA	78	0%	LOW
Sjogren syndrome									
1 (Szodoray et al., 2004)	Cross-sectional	No serious [1]	NA [2]	No serious [3]	No serious [6]	NA	111	4.5% (1.9%, 10.1%)	LOW
Systemic sclerosis									
1 (Forbess et al., 2013)	Cross-sectional	No serious [1]	NA [2]	No serious [3]	No serious [6]	NA	72	0%	LOW
Autoimmune thyroid disease									
3 (Saatar et al., 2011; Sategna-Spadaccino et al., 2008)	Case-control (1) and Cross-sectional (2)	No serious [1]	Very serious [8]	No serious [3]	Serious [4]	NA	730	1.1% (0.2%, 6.2%)	VERY LOW
Turner syndrome									
4 (Bonamico et al., 2002; Dias et al., 2010; Frost et al., 2009; Mortensen et al., 2009))	Cross-sectional	No serious [1]	No serious [5]	No serious [3]	No serious [6]	NA	807	5.5% (4.1, 7.4%),	LOW
Type I diabetes									
12 (Adlercreutz et al., 2014; Barbato et al.,	Case-control (1), case	No serious	Very serious [8]	No serious [3]	No serious [6]	NA	9114	6.0% (4.0, 8.9%)	LOW

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Percentage with Coeliac disease (95% CI)	Quality
1998; Cev et al., 2010; Djurić et al., 2010; Galván et al., 2008; Kakleas et al., 2010; Leeds et al., 2010; Pham-Short et al., 2010; Picarelli et al., 2005; Salardi et al., 2008; Smith et al., 2000; Uibo et al., 2010)	series (1) and Cross-sectional (9)	[1]							
First-degree relatives									
9 (Almeida et al., 2008; Ascher et al., 1997; Biagi et al. 2008; da Silva Kotze et al., 2013; Estev et al., 2006; Oliveira et al., 2012; Rubio-Tapia et al., 2008; Szaflarska-Szczepanik et al., 2001; Vaquero et al., 2014)	Cohort (1), and Cross-sectional (8)	No serious [1]	Very serious[8]	No serious[3]	No serious[6]	NA	2094	9.0% (4.7%, 16.6%)	LOW

E.3 Question 4.3

Modified GRADE profile for risk of long-term consequences of undiagnosed or untreated biopsy-confirmed coeliac disease

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Estimation of risk (95% CI)	Quality
Osteoporosis – reported as any fracture									
1 (Jafri et al., 2008)	Case-control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	83	Adj HR 2.0 (1.0, 3.9)	VERY LOW
Osteoporosis – reported as risk of peripheral fracture									
1 (Jafri et al., 2008)	Case-control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	83	Adj HR 2.0 (1.0, 3.9)	VERY LOW
Osteoporosis – reported as risk of axial fracture									
1 (Jafri et al., 2008)	Case-control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	83	Adj HR 1.7 (0.7, 4.2)	VERY LOW
Osteoporosis – reported as risk of osteoporotic fracture									
1 (Jafri et al., 2008)	Case-control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	83	Adj HR 6.9 (0.7, 7.65)	VERY LOW
Malignancy – reported as non-Hodgkin's lymphoma, Hodgkin's lymphoma, small bowel, colon, oesophageal, melanoma, breast , stomach or other cancer									
1 (Silano et al.,	Retrospectiv	No	No serious[2]	No serious[3]	Serious[4]	None	1968	SIR 1.3 (1.0–	VERY

[1] No concerns over study design
[2] Single study analysis
[3] Population and outcome as specified in the review protocol
[4] Confidence intervals around point estimate cross MID

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Estimation of risk (95% CI)	Quality
2007)	e case series	serious[1]						1.7)	LOW
Malignancy – reported as small bowel cancer									
1 (Silano et al., 2007)	Retrospective case series	No serious[1]	No serious[2]	No serious[3]	No serious[1]	None	1968	SIR 25 (8.5–51.4)	LOW
Malignancy – reported as non-Hodgkin's lymphoma									
1 (Silano et al., 2007)	Retrospective case series	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	1968	SIR 4.7 (2.9–7.3)	LOW
Malignancy – reported as Hodgkin's lymphoma									
1 (Silano et al., 2007)	Retrospective case series	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	1968	SIR 10 (2.7–25)	LOW
Malignancy – reported as stomach cancer									
1 (Silano et al., 2007)	Retrospective case series	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	1968	SIR 3 (1.3–4.9)	LOW
Malignancy – reported as colon cancer									
1 (Silano et al., 2007)	Retrospective case series	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	1968	SIR 1.1 (0.68–1.56)	VERY LOW
Mortality – reported as Child mortality rate									
1 (Zugna et al., 2013)	Case-control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	12919	Adj HR 1.08 (0.94, 1.25)	VERY LOW
Mortality – reported as Risk of non-accidental death									
1 (Zugna et al.,	Case-control	No seriou	No serious[2]	No serious[3]	Serious[4]	None	12919	Adj HR 1.30	VERY

[5] Confidence intervals around point estimate do not cross MID

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Estimation of risk (95% CI)	Quality
2013)		s[1]						(0.65, 2.58)	LOW

Modified GRADE profile for risk of long-term consequences of undiagnosed or untreated serology-confirmed coeliac disease

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Estimation of risk (95% CI)	Quality
Osteoporosis									
1 (Godfrey et al., 2010)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	127	OR 2.59 (1.32, 5.09)	LOW
Osteoporosis reported as fracture risk									
1 (Sanchez et al., 2011)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	265	HR 1.53 (1.05, 2.14)	LOW
Osteoporosis – reported at T score less than -2.5									
1 (Duerksen et al., 2010)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	376	OR 2.67 (1.17, 2.02)	LOW

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Estimation of risk (95% CI)	Quality
Osteoporosis – reported as low Bone Mineral Density (osteoporosis or osteopenia)									
1 (LeBoff et al., 2013)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	208	OR 0.97 (0.10, 95.8)	VERY LOW
Malignancy – reported as CD related cancer									
1 (Godfrey et al., 2010)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	127	OR 2.02 (0.29, 14.38)	VERY LOW
Malignancy – reported as lymphoproliferative cancer									
1 (Lohi et al., 2009)	Cross-sectional	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	73	Adj RR 5.94 (1.41, 25.04)	LOW
Malignancy reported as breast cancer									
1 (Lohi et al., 2009)	Cross-sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	73	Adj RR 0.71 (0.10, 5.07)	VERY LOW
Malignancy – reported as risk of all cancer									
1 (Lohi et al., 2009)	Cross-sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	73	Adj RR 0.67 (0.28, 1.61)	VERY LOW
Malignancy – reported as risk of mortality due to cancer									
1 (Canavan et al., 2011	Case control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	87	Adj HR 1.18 (0.53, 2.65)	VERY LOW
Malignancy – reported as risk of mortality due to cancer									
1 (Lohi et al., 2009)	Cross-sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	73	Adj RR 0.91 (0.59, 1.38)	VERY LOW
Fertility – reported as risk of undiagnosed CD in those with infertility due to ovulation disorder									
1 (Hogen-Esch et	Case control	No	No serious[2]	No serious[3]	Serious[4]	None	1038	OR 5.36	VERY

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Estimation of risk (95% CI)	Quality
al., 2011)		serious[1]						(0.89, 32.57)	LOW
Fertility – reported as risk of undiagnosed CD in those with male factor infertility									
1 (Hogen-Esch et al., 2011)	Case control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	1038	OR 5.36 (0.89, 32.57)	VERY LOW
Fertility – reported as risk of undiagnosed CD in infertile (any cause) women									
1 (Hogen-Esch et al., 2011)	Case control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	1038	OR 2.43 (0.49, 12.09)	VERY LOW
Fertility – reported as risk of undiagnosed CD in infertile (any cause) men									
1 (Hogen-Esch et al., 2011)	Case control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	1038	OR 0.92 (0.21, 4.12)	VERY LOW
Fertility –- reported as unexplained fertility (women)									
1 (Hogen-Esch et al., 2011)	Case control	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	1038	OR 4.51 (1.36, 19.19)	LOW

Modified GRADE profile for risk of long-term consequences of undiagnosed or untreated serology-confirmed coeliac disease

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Estimation of risk (95% CI)	Quality
Osteoporosis									
1 (Godfrey et al., 2010)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	127	OR 2.59 (1.32, 5.09)	LOW
Osteoporosis reported as fracture risk									
1 (Sanchez et al., 2011)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	265	HR 1.53 (1.05, 2.14)	LOW
Osteoporosis – reported at T score less than -2.5									
1 (Duerksen et al., 2010)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	376	OR 2.67 (1.17, 2.02)	LOW
Osteoporosis – reported as low Bone Mineral Density (osteoporosis or osteopenia)									
1 (LeBoff et al., 2013)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	208	OR 0.97 (0.10, 95.8)	VERY LOW
Malignancy – reported as CD related cancer									
1 (Godfrey et al., 2010)	Cross sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	127	OR 2.02 (0.29, 14.38)	VERY LOW
Malignancy – reported as lymphoproliferative cancer									
1 (Lohi et al., 2009)	Cross-sectional	No serious	No serious[2]	No serious[3]	No serious[5]	None	73	Adj RR 5.94 (1.41, 25.04)	LOW

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Estimation of risk (95% CI)	Quality
		s[1]							
Malignancy reported as breast cancer									
1 (Lohi et al., 2009)	Cross-sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	73	Adj RR 0.71 (0.10, 5.07)	VERY LOW
Malignancy – reported as risk of all cancer									
1 (Lohi et al., 2009)	Cross-sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	73	Adj RR 0.67 (0.28, 1.61)	VERY LOW
Malignancy – reported as risk of mortality due to cancer									
1 (Canavan et al., 2011	Case control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	87	Adj HR 1.18 (0.53, 2.65)	VERY LOW
Malignancy – reported as risk of mortality due to cancer									
1 (Lohi et al., 2009)	Cross-sectional	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	73	Adj RR 0.91 (0.59, 1.38)	VERY LOW
Fertility – reported as risk of undiagnosed CD in those with infertility due to ovulation disorder									
1 (Hogen-Esch et al., 2011)	Case control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	1038	OR 5.36 (0.89, 32.57)	VERY LOW
Fertility – reported as risk of undiagnosed CD in those with male factor infertility									
1 (Hogen-Esch et al., 2011)	Case control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	1038	OR 5.36 (0.89, 32.57)	VERY LOW
Fertility – reported as risk of undiagnosed CD in infertile (any cause) women									
1 (Hogen-Esch et al., 2011)	Case control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	1038	OR 2.43 (0.49, 12.09)	VERY LOW
Fertility – reported as risk of undiagnosed CD in infertile (any cause) men									

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Estimation of risk (95% CI)	Quality
1 (Hogen-Esch et al., 2011)	Case control	No serious[1]	No serious[2]	No serious[3]	Serious[4]	None	1038	OR 0.92 (0.21, 4.12)	VERY LOW
Fertility –- reported as unexplained fertility (women)									
1 (Hogen-Esch et al., 2011)	Case control	No serious[1]	No serious[2]	No serious[3]	No serious[5]	None	1038	OR 4.51 (1.36, 19.19)	LOW

E.4 Question 4.4: Active case finding

Active case finding for coeliac disease

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			Relative (95% CI)	Absolute (95% CI)	Quality
No studies were identified										

E.5 Question 5.1: Serological testing - accuracy

IgA transglutaminase (IgA tTG)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for IgA tTG in coeliac disease in children									
2 studies: Mubarak (2011)[A] Panetta (2011)	cohort	Low[1,]	None[2,]	None[3]	Serious[4,]	None	376	96 (93 – 99)	MODERATE
Specificity for IgA tTG in coeliac disease in children									
2 studies: Mubarak (2011)[A] Panetta (2011)	Cohort	Low[1]	serious[5]	None[26]	None[6]	None	376	86 (78 – 91)	MODERATE
Sensitivity for IgA tTG in coeliac disease in adults									

[A] Mubarak (2001): The data presented here represents that collected in children ≥ 2 years old.
[1] Low risk of overall bias as assessed by QUADAS 2 tool
[2] No serious inconsistency: I^2 is < 33%
[3] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol.
[4] Serious imprecision: confidence intervals around each of the point estimates crosses 95%
[5] Serious inconsistency: I^2 is >33% and < 66 %
[6] No serious imprecision: confidence intervals do not cross 95%
[7] No serious imprecision: Confidence intervals do not cross the 95% point estimate

3 studies: Hopper (2008) Volta (2010) Swallow (2012)	Cohort	Low[1]	None[25]	None 26	None 29	None	2900	91% (85 – 95)	HIGH
Specificity for IgA tTG in coeliac disease in adults									
3 studies: Hopper (2008) Volta (2010) Swallow (2012)	Cohort	Low[1]	Serious[28]	None 26	None 29	None	2900	91% (90 – 92)	MODERATE
Sensitivity for IgA tTG in coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None 26	Serious[4]	None	268	97% (94 – 99)	MODERATE
Specificity for IgA tTG in coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None 26	None 29	None	268	87% (80 – 92)	HIGH

[1]Low risk of overall bias as assessed by QUADAS 2 tool
[2] No serious inconsistency: I^2 is < 33%
[3] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol.
[4] Serious imprecision: confidence intervals around each of the point estimates crosses 95%
[5] Serious inconsistency: I^2 is >33% and < 66 %
[6] No serious imprecision: confidence intervals do not cross 95%

IgA endomysial antibodies (IgA EMA)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for IgA EMA in coeliac disease in children									
2 studies: Mubarak (2011)[A] Panetta (2011)	Cohort	Low[1]	None[2]	None[26]	Serious[27]	None	376	97 (94-99)	MODERATE
Specificity for IgA EMA in coeliac disease in children (Marsh ≥3 criteria									
2 studies: Mubarak (2011)[A] Panetta (2011)	Cohort	Low[1]	Very Serious[28]	None[3]	None[29]	None	376	76 (67 – 83)	LOW
Sensitivity for IgA EMA in coeliac disease in adults									
3 studies: Hopper (2008) Volta (2010) Swallow (2012)	Cohort	Low[1]	None[2]	None[3]	None[7]	None	2900	85 (78-90)	HIGH
Specificity for IgA EMA in coeliac disease in adults									
3 studies: Hopper (2008) Volta (2010) Swallow (2012)		Low[1]	Serious[28]	None[3]	None[6]	None	2900	98 (98 – 99)	HIGH
Sensitivity for IgA EMA in coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None[3]	None[6]	None	268	98 (96-100)	HIGH
Specificity for IgA EMA in coeliac disease in children and adults									
Burgin-Wolff(2013)	Cohort	Low[1]	N/A	None	None	None	268	85 (78-91)	MODERATE

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
				[3]	29				

[1]Low risk of overall bias as assessed by QUADAS 2 tool
[2] No serious inconsistency: I^2 is < 33%
[3] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol.
[4] Serious imprecision: confidence intervals around each of the point estimates crosses 95%
[5] Serious inconsistency: I^2 is >33% and < 66 %
[6] No serious imprecision: confidence intervals do not cross 95%

IgA deamidated gliadin peptide (IgA DGP)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for IgA DGP in coeliac disease in children									
Mubarak (2011)	Cohort	Low[1]	N/A	None [26]	None [6]	None	212	82 (72 – 89)	HIGH
Specificity for IgA DGP in coeliac disease in children									
Mubarak (2011)	Cohort	Low[1]	N/A	None [26]	None [6]	None	212	86 (77 – 92)	HIGH
Sensitivity for IgA DGP in coeliac disease in adults									
Volta (2010)	Cohort	Low[1]	N/A	None 26	None [6]	None	144	83 (73 - 93)	HIGH
Specificity for IgA DGP in coeliac disease in adults									
Volta (2010)	Cohort	Low[1]	N/A	None 26	None [6]	None	144	80 (71 – 88)	HIGH
Sensitivity for IgA DGP in coeliac disease in children and adults									

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for IgA DGP in coeliac disease in children									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None 26	None [6]	None	268	78 (71 – 85)	HIGH
Specificity for IgA DGP in coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None 26	Serious[4]	None	268	97 (93-99)	MODERATE

[1]Low risk of overall bias as assessed by QUADAS 2 tool
[2] No serious inconsistency: I^2 is < 33%
[3] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol.
[4] Serious imprecision: confidence intervals around each of the point estimates crosses 95%
[5] Serious inconsistency: I^2 is >33% and < 66 %
[6] No serious imprecision: confidence intervals do not cross 95%

IgG deamidated gliadin peptide (IgG DGP)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for IgG DGP in coeliac disease in children									
Mubarak (2011)	Cohort	Low[1]	N/A	None [26]	None [6]	None	212	89 (80 – 95)	HIGH
Specificity for IgG DGP in coeliac disease in children									
Mubarak (2011)	Cohort	Low[1]	N/A	None [3]	None [6]	None	212	81 (71 – 88)	HIGH
Sensitivity for IgG DGP in coeliac disease in adults									
Volta (2010)	Cohort	Low[1]	N/A	None [26]	None [6]		144	83 (73 – 94)	HIGH
Specificity for IgG DGP in coeliac disease in adults									

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Volta (2010)	Cohort	Low[1]	N/A	None[3]	None[6]		144	97 (95 – 100)	HIGH
Sensitivity for IgG DGP in coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None[3]	None[6]	None	268	85 (80 – 90)	HIGH
Specificity for IgG DGP in coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None[3]	Serious27	None	268	92 (86 – 97)	MODERATE

[1]Low risk of overall bias as assessed by QUADAS 2 tool
[2] No serious inconsistency: I^2 is < 33%
[3] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol.
[4] Serious imprecision: confidence intervals around each of the point estimates crosses 95%
[5] Serious inconsistency: I^2 is >33% and < 66 %
[6] No serious imprecision: confidence intervals do not cross 95%

Human leucocyte antigen DQ2/DQ8 (HLA DQ2/DQ8)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for HLA DQ2/DQ8 genotyping in coeliac disease in children									
Clouzeau-Girard (2011)	Cohort	Low[1]	N/A	None[3]	None[6]	None	170	99 (96 – 100)	HIGH
Specificity for HLA DQ2/DQ8 genotyping in coeliac disease in children									
Clouzeau-Girard (2011)	Cohort	Low[1]	N/A	None[3]	None[6]	None	170	69 (59 – 79)	HIGH

[1]Low risk of overall bias as assessed by QUADAS 2 tool
[2] No serious inconsistency: I^2 is < 33%
[3] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol.
[4] Serious imprecision: confidence intervals around each of the point estimates crosses 95%
[5] Serious inconsistency: I^2 is >33% and < 66 %
[6] No serious imprecision: confidence intervals do not cross 95%

E.6 Question 5.2: Serological testing – Sequencing

IgG DGP + IgA EMA

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None [2]	None [3]	None	268	73 (66 – 80)	HIGH
Specificity for coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	Serious[4]	None	268	95 (91 – 98)	MODERATE

[1] Low risk of bias as assessed by the QUADAS 2 tool
[2] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol
[3] No serious imprecision: confidence intervals do not cross 95%
[4] Serious imprecision: confidence intervals around one of the point estimates crosses 95%
[5] Serious risk of bias, as assessed by CASP cohort study checklist

IgG DGP + IgA tTG

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in children and adults									
Burgin-Wolf (2013)	Cohort	Low[1]	N/A	None[2]	None **Error! Bookmark not defined.**	None	268	72 (65 – 80)	HIGH
Specificity for coeliac disease in children and adults									
Burgin-Wolf (2013)	Cohort	Low[1]	N/A	None[2]	None **Error! Bookmark not defined.**	None	268	96 (92 – 99)	MODERATE

[1] Low risk of bias as assessed by the QUADAS 2 tool
[2] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol
[3] No serious imprecision: confidence intervals do not cross 95%
[4] Serious imprecision: confidence intervals around one of the point estimates crosses 95%
[5] Serious risk of bias, as assessed by CASP cohort study checklist

IgA DGP + IgG DGP + IgA tTG

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None [2]	None [3]	None	268	73 (66 – 80)	HIGH
Specificity for coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	268	99 (98 – 100)	HIGH

IgA DGP + IgG DGP + IgA EMA

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in children and adults									

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	268	58 (50 – 66)	HIGH
Specificity for coeliac disease in children and adults									
Burgin-Wolff (2013)		Low[1]	N/A	None [2]	None [3]	None	268	99 (98 – 100)	HIGH

[1] Low risk of bias as assessed by the QUADAS 2 tool
[2] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol
[3] No serious imprecision: confidence intervals do not cross 95%
[4] Serious imprecision: confidence intervals around one of the point estimates crosses 95%
[5] Serious risk of bias, as assessed by CASP cohort study checklist

IgG DGP + IgA EMA+ IgA tTG

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in children and adults									

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	268	56 (48 – 64)	HIGH
Specificity for coeliac disease in children and adults									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	268	99 (98 – 100)	HIGH

[1] Low risk of bias as assessed by the QUADAS 2 tool
[2] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol
[3] No serious imprecision: confidence intervals do not cross 95%
[4] Serious imprecision: confidence intervals around one of the point estimates crosses 95
[5] Serious risk of bias, as assessed by CASP cohort study checklist

IgG DGP + IgA DGP + IgA EMA + IgA tTG

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in children and adults (Marsh ≥3 criteria)									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None [2]	None [3]	None	268	56 (48 – 64)	HIGH
Specificity for coeliac disease in children and adults (Marsh ≥3 criteria									
Burgin-Wolff (2013)	Cohort	Low[1]	N/A	None [2]	None [3]	None	268	99 (98 - 100)	HIGH

IgA tTG + IgA EMA + HLA Q2/DQ8 genotyping

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in children									
Clouzeau Girard (2011)	cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	170	99 (96 – 100)	HIGH
Specificity for coeliac disease in children									

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Clouzau Girard (2011)	cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	170	96 (92 – 100)	HIGH

[1] Low risk of bias as assessed by the QUADAS 2 tool
[2] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol
[3] No serious imprecision: confidence intervals do not cross 95%

IgA + IgG hTTG/DGP

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in children									
Mubarak (2011)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	Serious[4]	None	144	98 (93 -100)	HIGH
Specificity for coeliac disease in children									
Mubarak (2011)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	144	61 (45 – 66)	HIGH
Sensitivity for coeliac disease in adults									
Porcelli (2011)	Case-control	Serious[5]	N/A	None[2]	None **Error ! Bookmark not defined.**	None	201	100 (100-100)	MODERATE
Specificity for coeliac disease in adults									
Porcelli (2011)	Case-control	Serious[5]	N/A	None[2]	None[4]	None	201	90 (86 – 95)	MODERATE

[1] Low risk of bias as assessed by the QUADAS 2 tool
[2] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol
[3] No serious imprecision: confidence intervals do not cross 95%
[4] Serious imprecision: confidence intervals around one of the point estimates crosses 95 %
[5] Serious risk of bias, as assessed by CASP cohort study checklist

Algorithm: 2 step - If IgA tTG (+), and then IgA EMA (+)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in adults									
Hopper (2008)	Cohort	Low[1]	N/A	None [2]	None [3]	None	2000	86 (76 – 92)	HIGH
specificity for coeliac disease in adults									
Hopper (2008)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	2000	99 (98 – 99)	HIGH

Algorithm: 2 step - If IgA tTG (+) or equivocal, and then IgA EMA (+)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in adults									
Swallow (2012)	Cohort	Low[1]	N/A	None [2]	None [3]	None	756	87 (65-97)	HIGH
specificity for coeliac disease in adults									

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Swallow (2012)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	756	97 (95 – 98)	HIGH

[1] Low risk of bias as assessed by the QUADAS 2 tool
[2] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol
[3] No serious imprecision: confidence intervals do not cross 95%

Algorithm: If both IgA tTG (+) and IgA EMA (+)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in adults									
Hopper (2008) Swallow (2012)	Cohort	Low[1]	None [16]	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	2756	85 (68 – 93)	HIGH

[1] Low risk of bias as assessed by the QUADAS 2 tool
[2] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol
[3] No serious imprecision: confidence intervals do not cross 95%

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
specificity for coeliac disease in adults									
Hopper (2008) Swallow (2012)	Cohort	Low[1]	None[6]	None **Error ! Bookmark not defined.**	None **Error ! Bookmark not defined.**	None	2756	99 (98 – 100)	HIGH

Algorithm: If either IgA tTG (+) OR IgA EMA (+)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Sensitivity for coeliac disease in adults									
Hopper (2008)	Cohort	Low[1]	N/A	None **Error ! Bookmark not defined.**	serious[4]	None	2000	92 (84 – 96)	MODERATE
specificity for coeliac disease in adults									

[4] Serious imprecision: confidence intervals around one of the point estimates crosses 95
[5] Serious risk of bias, as assessed by CASP cohort study checklist

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of participants	Summary of findings (95% confidence interval)	Quality
Hopper (2008)	Cohort	Low[1]	N/A	None **Error! Bookmark not defined.**	None **Error! Bookmark not defined.**	None	2000	90 (89 – 92)	HIGH

[1] Low risk of bias as assessed by the QUADAS 2 tool
[2] No serious indirectness: Population, index test, and reference standard were as specified in the review protocol
[3] No serious imprecision: confidence intervals do not cross 95%
[4] Serious imprecision: confidence intervals around one of the point estimates crosses 95
[5] Serious risk of bias, as assessed by CASP cohort study checklist

E.7 Question 5.3

No studies were identified in the information searches for this question. Please see appropriate GRADE tables from question 7 and question 8 that contributed to answering this review question.

E.8 Question 5.4

GRADE profile for resolution of gastrointestinal and non-gastrointestinal symptoms

Quality assessment						Summary of findings		
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of patients**	**Effect (95%CI)**	**Quality**
Proportion of patients in clinical remission at 12 months follow up								
2 (Dickey 2000, MIdhagen 2004)	Cohort	Very serious[1]	No serious[2]	No serious[3]	No serious[4]	71	90.1% (80.7, 95.2%)	VERY LOW

[1] Unclear if consecutive participants recruited and no explanation given for exclusions
[2] Low heterogeneity (I-squared less than 33%)
[3] Population and outcome as specified in the review protocol
[4] Confidence intervals around point estimate above GDG agreed MID of 80% responders

Table xx: Summary GRADE profile for dietary non-adherence on GFD

Quality assessment						Summary of findings		
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of patients**	**Effect (95%CI)**	**Quality**
3 (Monzani 2001, Trigoni 2014, Zanchi 2013)	Cohort	Serious[1]	Very serious[2]	No serious[3]	Serious[4]	393	23.6% (9.2%, 48.5%)	VERY LOW

[1] Unclear if consecutive samples used in all three studies; 10% of sample did not reported on adherence in one study
[2] High heterogeneity (I-squared greater than 67%)
[3] Population and outcomes as specified in review protocol
[4] Confidence intervals around point estimate cross20% GDG estimate of non-adherence

GRADE profile for diagnostic accuracy of DGP IgA to monitor adherence to GFD

Quality assessment						Summary of findings		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Number of patients	Effect (95%CI%)	Quality
Sensitivity of A DGP IgA to discriminate between partially adherent and strictly adherent in children and young people at between 2 and 4 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	21	62% (32%, 85%)	LOW
Specificity of A DGP IgA to discriminate between partially adherent and strictly adherent in children and young people at between 2 and 4 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	21	13 (1, 53)	LOW
Sensitivity of A DGP IgA to discriminate between partially adherent and strictly adherent in children and young people at between 6 and 8 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	21	13 (1, 53)	LOW
Specificity of A DGP IgA to discriminate between partially adherent and strictly adherent in children and young people at between 6 and 8 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	21	20 (1, 70)	LOW
Sensitivity of A DGP IgA to discriminate between partially adherent and strictly adherent in children and young people at between 9 and 12 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	21	10 (1, 43)	LOW
Specificity of A DGP IgA to discriminate between partially adherent and strictly adherent in children and young people at between 9 and 12 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	21	43 (12, 80)	LOW

[1] No concerns over study design
[2] Single study analysis
[3] Population and outcomes as specified in the review protocol
[4] Point estimate and confidence intervals do not cross 95% threshold

GRADE profile for diagnostic accuracy of anti tTG IgA to monitor adherence to GFD

Quality assessment						Summary of findings		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Number of patients	Effect (%)	Quality
Sensitivity of anti tTG IgA to discriminate between partially adherent and strictly adherent in children and young people at between 2 and 4 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	21	100	LOW
Specificity of anti tTG IgA to discriminate between partially adherent and strictly adherent in children and young people at between 2 and 4 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	21	46 (20 – 74)	LOW
Sensitivity of anti tTG IgA to discriminate between partially adherent and strictly adherent in children and young people at between 6 and 8 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	Serious[5]	13	80 (30 – 99)	VERY LOW
Specificity of anti tTG IgA to discriminate between partially adherent and strictly adherent in children and young people at between 6 and 8 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	13	38 (10 – 74)	LOW
Sensitivity of anti tTG IgA to discriminate between partially adherent and strictly adherent in children and young people at between 9 and 12 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	20	56 (23 – 85)	LOW
Specificity of anti tTG IgA to discriminate between partially adherent and strictly adherent in children and young people at between 9 and 12 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	20	55 (25 – 82)	LOW

[1] No concern over study design
[2] Single study analysis
[3] Population and test as specified in review protocol
[4] Confidence intervals around point estimate do not cross 95% threshold
[5] Confidence interval around point estimate cross 95% threshold

GRADE profile for diagnostic accuracy of AGA IgA to monitor adherence to GFD

Quality assessment						Summary of findings		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Number of patients	Effect (%)	Quality
Sensitivity of AGA IgA to discriminate between partially adherent and strictly adherent in children and young people at between 2 and 4 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	21	38 (10 – 74)	LOW
Specificity of AGA IgA to discriminate between partially adherent and strictly adherent in children and young people at between 2 and 4 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	Serious[5]	21	96 (62 – 100)	VERY LOW
Sensitivity of AGA IgA to discriminate between partially adherent and strictly adherent in children and young people at between 6 and 8 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[6]	13	Not calculable	LOW
Specificity of AGA IgA to discriminate between partially adherent and strictly adherent in children and young people at between 6 and 8 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	13	80 (30 – 90)	LOW
Sensitivity of AGA IgA to discriminate between partially adherent and strictly adherent in children and young people at between 9 and 12 months of GFD								
1 (Monzani 2011)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[6]	20	Not calculable	LOW
Specificity of AGA IgA to discriminate between partially adherent and strictly adherent in children and young people at between 9 and 12 months of GFD								
1 (Monzani 2011)	Cohort	No serious[9]	NA[10]	No serious[11]	Serious[5]	20	91 (67 – 99)	VERY LOW

[1] No concerns over study design
[2] Single study analysis
[3] Population and test as specified in the review protocol
[4] Confidence intervals around point estimate do not cross 95% threshold
[5] Confidence intervals around point estimate cross 95% threshold
[6] Not enough data available to calculate effect size

GRADE profile for IgA anti-tTG ELISAto discriminate between partially adherent and strictly adherent at 24 months

Quality assessment						Summary of findings		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Number of patients	Effect (%)	Quality
Sensitivity of AGA IgA to discriminate between partially adherent and strictly adherent in children, young people and adults after GFD for 24 months								
1 (Zanchi 2013)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	315	44 (29 – 60)	LOW
Specificity of AGA IgA to discriminate between partially adherent and strictly adherent in children and young people at between 2 and 4 months of GFD								
1 (Zanchi 2013)	Cohort	No serious[1]	NA[2]	No serious[3]	Serious[5]	315	98 (96 – 99)	VERY LOW

[1] No apparent risk of bias
[2] Single study analysis
[3] Population and test as specified in the review protocol
[4] Confidence intervals around the point estimate do not cross 95% threshold
[5] Confidence intervals around the point estimate cross 95% threshold

GRADE profile for response to GFD defined by negative IgA EMA

Quality assessment						Number of patients	Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		Proportion testing negative ((%% CI)	Quality
At 3 months								
3 (Dickey 2000, Fotoulaki 1999; Midhagen 2004)	Cohort	No serious[1]	No serious[2]	No serious[3]	No serious[4]	100	31.2% (23.4%, 40.2%)	LOW
At 12 months								
4 (Dickey 2000, Midhagen 2004, Fotoulaki 1999; Trigoni 2014)	Cohort	No serious	Serious[5]	No serious[3]	No serious[4]	150	90.5% (83.1%, 94.9%)	VERY LOW

[1] No concerns over study design
[2] Low heterogeneity (I-squared less than 33%)
[3] Population and test as specified in the review protocol
[4] GDG did not agree a MID for this outcome
[5] Moderate heterogeneity (I-squared between 34% and 67%)

GRADE profile for response to GFD defined by negative IgA ARA

Quality assessment						Number of patients	Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		Proportion testing negative (95% CI)	Quality
At 3 months								
1 (Fotoulaki 1999)	Cohort	No serious[1]	No serious[2]	No serious[3]	No serious[4]	30	76.7% (57.3%, 89.4%)	LOW
At 12 months								
1 (Fotoulaki 1999)	Cohort	No serious[1]	No serious[2]	No serious[3]	No serious[4]	30	100% (No CI)	LOW

[1] No apparent risk of bias
[2] Single study analysis
[3] Population and test as specified in the review protocol
[4] GDG did not agree a MID for this outcome

GRADE profile for response to GFD defined by negative IgA tTG

Quality assessment						Number of patients	Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		Proportion testing negative	Quality
At 3 months								
2 (Midhagen 2004; Samasca 2011)	Cohort	Serious[1]	No serious[2]	No serious[3]	No serious[4]	64	65.5% (53.1%, 76.1%)	VERLOW
At 12 months								
3 (Midhagen 2004, Samasca 2011; Trigoni 2014)	Cohort	Serious[1]	Very serious[5]	No serious[3]	No serious[4]	115	76.5% (42.2%, 93.6%)	VERY LOW

[1] Unclear about population age range
[2] Low heterogeneity (I-squared below 33%)
[3] Population and test as specified in the review protocol
[4] GDG did not agree a MID for this outcome
[5] High heterogeneity (I-squared over 67%)

GRADE profile for response to GFD defined by negative IgG anti-tTG

Quality assessment						Number of patients	Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		Proportion testing negative (95%CI)	Quality
At 6 months								
1 Martin-Pagola 2007)	Cohort	No serious[1]	No serious[2]	No serious[3]	No serious[4]	93	63.4% (52.8%, 73.0%)	LOW
At 24 months								
1 Martin-Pagola 2007)	Cohort	No serious[1]	No serious[2]	No serious[3]	No serious[4]	93	96.7% (90.2%, 99.2%)	LOW

[1] No apparent risk of bias
[2] Single study analysis
[3] GDG did not agree a MID for this outcome
[4] Confidence intervals around point estimate do not cross 95% threshold

Table xx: GRADE profile for response to GFD defined by negative IgA AGA

Quality assessment						Number of patients	Effect	
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**		**Proportion testing negative (95% CI)**	**Quality**
At 3 months								
1 (MIdhagen 2004)	Cohort	No serious[1]	Not serious[2]	No serious[3]	No serious[4]	15	60.0% (32.9%, 82.5%)	LOW
At 12 months								
1 (MIdhagen 2004)	Cohort	No serious[1]	Not serious[2]	No serious[3]	No serious[4]	15	100% (No CI)	LOW

[1] No apparent risk of bias
[2] Single study analysis
[3] Population and test as specified in the review protocol
[4] GDG did not agree a MID for this outcome

GRADE profile for response to GFD defined histology at 12 months

Quality assessment						Number of patients	Effect	
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**		**Proportion testing negative**	**Quality**
Mucosal recovery								
3 (Dickey 2000, Martini 2002, Midhagen 2004)	Cohort	No serious[1]	NA[2]	No serious[3]	Serious[4]	172	12% to 89%	VERY LOW
Improvement (mucosal recovery or change in March criteria by a least 1 level)								
3 (Dickey 2000, Martini 2002)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[5]	154	58% to 62%	LOW
No change								
3 (Dickey 2000, Martini 2002, Midhagen 2004)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	172	11% to 38%	LOW

[1] No concerns over study design
[2] No test for heterogeneity carried out
[3] Population and test as specified in the review protocol
[4] Concerns over wide range in effect size
[5] No concern over range in effect size

GRADE profile for nutrition status at 12 months while on GFD

Quality assessment						Number of patients	Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		Proportion with nutritional inadequacies	Quality
1 (Shepherd 2012)	Cohort	No serious[1]	NA[2]	No serious[3]	No serious[4]	50	10%	LOW

[1] No concerns over study design
[2] Single study analysis
[3] Population and test as specified in the review protocol
[4] No concern over effect size

GRADE profile for healthcare involvement in follow-up monitoring

Quality assessment						Number of patients	Effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision		Proportion		
							Before	After	
Dietary adherence									
1 (Wylie 2005)	Before and after	Serious[1]	NA[2]	No serious[3]	No serious[4]	99	54%	66%	VERY LOW
Satisfaction with clinic									
1 (Wylie 2005)	Before and after	Serious[1]	NA[2]	No serious[3]	No serious[4]	99	42%	100%	VERY LOW

[1] Convenience sample used; unclear of number of participants In both phases
[2] Single study analysis
[3] Population and test as specified in the review protocol
[4] No concern over effect size

E.9 Question 6.1

Potential causes of non-responsive coeliac disease (NRCD)

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Proportion of patients (%)**	**Quality**
Incorrect diagnosis (%)								
3 studies: Dewar (2012) Leffler (2007) Abdulkarim (2002)	Cohort study	Low [1]	NA[2]	None[3]	None[4]	248	10 % - 12%	HIGH
Gluten ingestion (%)								
4 studies: Dewar (2012) Leffler (2007) Abdulkarim Van Weyenberg (2013)	Cohort study	Low [98]	NA[99]	None[100]	None[4]	265	36% - 82%	HIGH
Microscopic colitis (%)								

[1] Low risk of bias, as determined by QUADAS tool
[2] NA; Not applicable as no measure of heterogeneity was used
[3] No serious indirectness as study population and outcome of interest was uniform between studies
[4] No Serious imprecision as the GDG believed the range of estimates to accurately reflect clinical practice

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Proportion of patients (%)**	**Quality**
3 studies: Dewar (2012) Leffler (2007) Abdulkarim (2002)	Cohort study	Low [98]	NA[99]	None[100]	None[4]	248	6 %- 11%	HIGH
Bacterial overgrowth (%)								
3 studies: Dewar (2012) Leffler (2007) Abdulkarim (2002)	Cohort study	Low [1]	NA[2]	None[3]	None[4]	248	6 %-14 %	HIGH
Lactose intolerance (%)								
3 studies: Dewar (2012) Leffler (2007) Van Weyenberg (2002)	Cohort study	Low [98]	NA[99]	None[3]	None[4]	216	7 %-12 %	HIGH
Inflammatory colitis (%)								

[1] Low risk of bias as determined by QUADAS tool.
[2] NA; not applicable as no measure of heterogeneity was used in these estimates
[3] No serious indirectness as study population and outcome of interest was uniform between studies
[4] No Serious imprecision as the GDG believed the range of estimates to accurately reflect clinical practice

2 studies: Dewar (2012) Van Weyenberg (2002)	Cohort study	Low [98]	NA[99]	None[3]	None[4]	117	6 % - 7 %	HIGH
Irritable bowel syndrome (IBS) (%)								
3 studies: Dewar (2012) Leffler (2007) Abdulkarim (2002)	Cohort study	Low [98]	NA[99]	None[3]	None[4]	248	8% - 22%	HIGH
Refractory coeliac disease (%)								
3 studies: Dewar (2012) Leffler (2007) Abdulkarim (2002)	Cohort study	Low [1]	NA[2]	None[3]	None[4]	248	9% - 18%	HIGH

[1] Low risk of bias as assed by QUADAS tool
[2] NA; not applicable as no measure of heterogeneity was undertaken for these studies
[3] No serious indirectness
[4] No Serious imprecision as the GDG believed the range of estimates to accurately reflect clinical practice

Change to clinical management: Detection of RCD type I and RCD type II

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Summary of findings**	**Quality**
Aberrant T-cell receptor gene rearrangement (TCR)– sensitivity to diagnose RCD type II								
3 studies: Daum (2009) Arguelles-Grande (2013) Malamut (2009)	Cohort	Low[1]	NA[2]	None[3]	None[4]	N=146	97% - 100%	HIGH
Aberrant T-cell receptor gene rearrangement (TCR)– specificity to diagnose RCD type II								
3 studies: Daum (2009) Arguelles-Grande (2013) Malamut (2009)	Cohort	Low[1]	NA[2]	None[3]	None[4]	N=146	100%	HIGH
Immunohistochemistry to detect aberrant CD3(+) CD8(-) IEL phenotype - sensitivity to diagnose RCD type II								
3 studies: Daum (2009) Arguelles-Grande (2013) Malamut (2009)	Cohort	Low[1]	NA[2]	None[3]	None[4]	N=152	56% – 100%	HIGH
Immunohistochemistry to detect aberrant CD3(+) CD8(-) IEL phenotype - specificity to diagnose RCD type II								

[1] Low risk of bias as assessed by QUADAS tool
[2] NA; Measure of inconsistency not applicable as heterogeneity of data was not assessed
[3] No serious indirectness, populations of interest matched those outlined in the protocol
[4] No Serious imprecision as the GDG believed the range of estimates to accurately reflect clinical practice

3 studies: Daum (2009) Arguelles-Grande (2013) Malamut (2009)	Cohort	bias[1]	NA[2]	None[3]	None[4]	N=152	100%	HIGH

[1] Low risk of bias as assed by QUADAS tool
[2] NA; not applicable as no measure of heterogeneity was undertaken for these studies
[3] No serious indirectness
[4] No Serious imprecision as the GDG believed the range of estimates to accurately reflect clinical practice

Patient outcomes at follow-up: Detection of enteropathy associated T-cell lymphoma (EATL)

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Summary of findings**	**Quality**
18F-FDG PET - sensitivity to detect EATL								
Hadithi (2006)	Cohort	Low[1]	NA[2]	None[3]	None[4]	N= 30	100% (100-100)	HIGH
18F-FDG PET - specificity to detect EATL								
Hadithi (2006)	Cohort	Low[1]	NA[2]	None[3]	serious[5]	N= 30	90% (79 – 100)	MODERATE
Abdominal CT - sensitivity to detect EATL								
Hadithi(2006) Daum (2009)	Cohort	low[1]	None[6]	None[3]	Serious[5]	N=37	50% (36 – 69)	LOW
Abdominal CT - specificity to detect EATL								

[1] Low risk of bias, as assessed by QUADAS tool
[2] No measure of inconsistency as only one study was considered for this analysis
[3] No serious indirectness, population as specified within procotol
[4] No serious imprecision, confidence intervals do not cross 95%
[5] Serious imprecision – confidence intervals cross 95%
[6] No serious inconsistency – confidence intervals overlap

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Summary of findings**	**Quality**
Hadithi(2006) Daum (2009)	Cohort	low[1]	Serious[2]	None[3]	None[4]	N=37	76% (36-100)	
Magnetic resonance (MR) enteroclysis – sensitivity to detect EATL								
Van Weyenberg (2011)	Cohort	Low[1]	NA[5]	Serious[6]	Serious[7]	Total N = 28 (test group)	88% (47 – 99)	LOW
Magnetic resonance (MR) enteroclysis – specificity to detect EATL								
Van Weyenberg (2011)	Cohort	Low[1]	NA[5]	Serious[6]	Serious[7]	Total N = 28 (test group)	97% (87 – 99)	LOW
Double balloon enteroscopy – sensitivity to diagnose EATL and ulcerative jejunitis								
Hadithi (2007)	Cohort	Low[1]	NA[5]	None[3]	None[8]	N =21	100% (100- 100)	HIGH
Double balloon enteroscopy – specificity to detect EATL and ulcerative jejunitis								
Hadithi (2007)	Cohort	Low[1]	NA[5]	None[3]	None[8]	N =21	100% (100 – 100)	HIGH
Capsule endoscopy – sensitivity to detect EATL								
Daum (2007)	Cohort	Low[1]	Serious[2]	Serious[6]	serious[7]	N= 9 * Not possible in	50% (19 – 100)	VERY LOW

[1] Low risk of bias, as assessed by QUADAS tool
[2] Serious inconsistency between study estimates of effect
[3] No serious indirectness, population as specified within protocol
[4] No serious imprecision, confidence intervals do not cross 95%
[5] NA, not applicable, single study
[6] Serious indirectness, study participants were only those who scored<2 on MR enteroclysis. This group was composed of RCD I and 'uncomplicated CD' patients
[7] Serious imprecision, confidence intervals are wide
[8] No serious imprecision, confidence intervals are tight

Van Weyenberg (2013)						1/7 RCD I and 4/7 RCD II N=26* only data from RCD and EATL	0%* capsule unable to visualise distal small intestine in these patients.	
Capsule endoscopy – specificity to detect EATL								
Daum (2007) Van Weyenberg (2013)	Cohort	Low[1]	Serious[2]	Serious[6]	serious[7]	N= 9 * Not possible in 1/7 RCD I and 4/7 RCD II N=26* only data from RCD and EATL	100% (100-100)	VERY LOW

[1] Low risk of bias, as assessed by QUADAS tool
[2] Serious inconsistency between study estimates of effect
[3] No serious indirectness, population as specified within protocol
[4] No serious imprecision, confidence intervals do not cross 95%
[5] NA, not applicable, single study
[6] Serious indirectness, study participants were only those who scored<2 on MR enteroclysis. This group was composed of RCD I and 'uncomplicated CD' patients
[7] Serious imprecision, confidence intervals are wide
[8] No serious imprecision, confidence intervals are tight

Patient outcome at follow-up: cumulative survival at 5 year follow-up for RCD type I and RCD type II

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Summary of findings Percentage of patient population survival**	**Quality**
Cumulative survival at 5 years RCD type I								
4 studies: Daum (2009) Van Weyenberg (2013) [1] Arguelles Grande (2013) Malamut (2009)	Cohort	Low[2]	None[3]	None[4]	None[5]	112	90% (76 – 100)	MODERATE
Cumulative survival at 5 years RCD type II								
4 studies: Daum (2009) Van Weyenberg (2013) Arguelles Grande (2013) Malamut (2009)	Cohort	Low[2]	None[3]	None[4]	serious[6]	68	53% (12 – 94)	HIGH

[1] Corresponds to score on MR enteroclysis scoring system of <2. This group was composed of RCD I and 'uncomplicated CD' patients
[2] Low risk of bias as assed by QUADAS tool
[3] No Serious inconsistency; confidence intervals overlap
[4] No serious indirectness, population as specified within protocol
[5] No serious imprecision, confidence intervals are tight
[6] Serious imprecision, confidence intervals are wide

Predictive factors of EATL development in patients with RCD

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Odds ratio**	**Quality**
Aberrant Immunophenotype								
Liu (2009) Malamut (2009)	cohort	Low[1]	None[2]	None[3]	serious[4]	41	4.18 (0.8 – 20.7) – 5.00 (0.51 – 49)	MODERATE
Age								
Liu (2009) Malamut (2009)	cohort	Low[1]	Serious[5]	None[3]	Serious[4]	41	0.97 (0.92 - 1.04) - 1.3 (1.1 – 1.7)	MODERATE
Ulcerative jejunitis								
Liu (2009)	cohort	Low[1]	NA[6]	None[3]	Serious[4]	41	1.8 (0.7 – 4.7)	MODERATE
Gender								
Liu (2009)	cohort	Low[1]	NA	None[3]	Serious[4]	41	2.17 (0.45 – 10.44)	MODERATE
Persistent monoclonality								
Liu (2009)	cohort	Low[1]	NA	None[3]	Serious[4]	41	3.6 (0.6 – 21.6)	MODERATE
Persistent concurrent aberrant immunophenotype and monoclonality								
Liu (2009)	cohort	Low[1]	NA	None[3]	Serious[4]	41	9 (0.51 – 48.75)	MODERATE
Persistent >80% CD3+ CD8- IEL's								
Liu (2009)	cohort	Low[1]	NA[2]	None[3]	serious[4]	41	21.33 (2.94 – 154.6)	MODERATE
Persistent concurrent >80% CD3+ CD8- IEL's and monoclonality								
Liu (2009)	cohort	Low[7]	NA[8]	None[9]	serious[4]	41	45.33% (4.05 – 506.86)	MODERATE

[1] Low risk of bias as assessed by QUADAS tool
[2] No serious inconsistency as confidence intervals around estimates overlap
[3] No serious indirectness; population of interest matched study protocol
[4] Serious imprecision ;confidence intervals are wide
[5] Serious inconsistency, confidence intervals around estimates do not overlap
[6] NA; measure of inconsistency not applicable as only one study contributed to this analysis
[7] Low risk of bias as assessed by QUADAS tool
[8] NA; measure of inconsistency not applicable as only one study contributed to this analysis

[9] No serious indirectness; population of interest matched study protocol

Predictive factors for clinical worsening in RCD

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Hazard Ratio**	**Quality**
Age ≥ 50 years								
Arguelles Grande (2012)	Cohort	Low[1]	NA[2]	None[3]	serious[4]	73	1.55 (0.8 – 3)	HIGH
Monoclonality								
Arguelles Grande (2012)	Cohort	Low[1]	NA[2]	None[3]	Serious[4]	73	4.33 (1.7 – 10.98)	HIGH
Severe VA								
Arguelles Grande (2012)	Cohort	Low[1]	NA[2]	None[3]	None[5]	73	1.54 (0.25 - 0.8)	HIGH
Aberrant IEL immunophenotype								
Arguelles Grande (2012)	Cohort	Low[1]	NA[2]	None[3]	Serious[4]	73	3.01 (1.5 – 6.01)	MODERATE
Presence of non-EATL lymphoma								
Arguelles Grande (2012)	Cohort	Low[1]	NA[2]	None[3]	Serious[4]	73	2.76 (0.8 – 9.19)	MODERATE
Presence of proximal focal erythema on capsule endoscopy								
Van Weyenberg (2013)	Cohort	Low[1]	NA[2]	Serious[6]	Serious[4]	N=48	6.7 (1.2 – 38.7)	LOW
Absence of progression of capsule to distal intestina during capsule endoscopy								
Van Weyenberg (2013)	Cohort	Low	NA[2]	Serious[6]	Serious[4]	N=48	16.5 (1.2 – 224.9)	LOW

[1] Low risk of bias as assessed by QUADAS tool
[2] NA; measure of inconsistency not applicable as only one study contributed to this analysis
[3] No serious indirectness; population of interest matched study protocol
[4] Serious imprecision, wide confidence intervals
[5] No serious imprecision, tight confidence intervals
[6] Serious indirectness, results only reported in small proportion of patients in whom the capsule was tolerated

E.10 Question 6.2

Pharmacological management of refractory coeliac disease – summary of results

	Clinical improvement				Complications of CD	Adverse events	Serological response EMA/Anti-tTG Haemoglobin (mmol/L) Albumin (g/L)	Immunological response	Clinical and histological response	Histological improvement
	Proportion achieved	Specific symptom improvement	BMI (kg/m^2)	Body weight (kg)				Proportion of IELs per 100 epithelial cells		Achieved improvement Change in Marsh status
Aminosalicylates										
Mesalazine[9]	3/4 (75%) had total symptom reduction (1 had < 50%)[9]					1/10 had headache leading to withdrawal (unclear if they had budesonide)[9]				
Corticosteroids										
Budesonide[5,14]	12/15 (80%) complete 3/15 (20%) poor at 7m[5] 1/2 at 28m[14]		From mean 20.6 to 20.75 at 24m[14] (n=2)					From mean 64% (n=2) to 87% at 24m in one (not measured in the second)[14]	4/9 (44%) at mean 26 months[5]	Of 2 patients, one remained MIIIB at 24m (not measured in the other patient)[14]
Corticosteroids	32/40 (77%) at unclear follow-up[1]									7/40 (20%) had partial and 7/30 had complete

	Clinical improvement				Complications of CD	Adverse events	Serological response EMA/Anti-tTG Haemoglobin (mmol/L) Albumin (g/L)	Immunological response	Clinical and histological response	Histological improvement
	Proportion achieved	Specific symptom improvement	BMI (kg/m^2)	Body weight (kg)				Proportion of IELs per 100 epithelial cells		Achieved improvement Change in Marsh status
(unspecified)[1]										villous recovery[1]
Prednisone[3,4]					25% 5-year survival (Kaplan Meier) in one study (n=47)[3] Of the 5 discovered to have persistent aberrant clone in another, 5 developed EATL 18 and 24m later[4].			5 had aberrant clone at average 24m follow-up (unclear which 11 patients had biopsy to determine this)[4]	2/11 (18%) at mean 26m[4]	
Prednisolone[10,11]	11/15 (73%) at 41m[10]									3/14 (21%) at 41m[10] Of 5 patients who had subtotal villous atrophy, 1 had normal histology and 4 had partial villous atrophy at average 6 weeks[11]
Cytokine modulators										

	Clinical improvement				Complications of CD	Adverse events	Serological response EMA/Anti-tTG Haemoglobin (mmol/L) Albumin (g/L)	Immunological response	Clinical and histological response	Histological improvement
	Proportion achieved	**Specific symptom improvement**	**BMI (kg/m^2)**	**Body weight (kg)**				**Proportion of IELs per 100 epithelial cells**		**Achieved improvement Change in Marsh status**
Anti-TNF α (unspecified)[1]	3/4 (75%) at unclear follow-up[1]									1/4 (25%) had partial villous recovery at unclear follow-up[1]
Immunosupressants										
Azathioprine[1, 13]	3/5 (60%) at unclear follow-up[1]	Of 5 patients, abdominal pain resolved in all, fever resolved in both who presented with it and diarrhoea resolved in all 4 of 5 patients who presented with it 12m[13]	From median 17 (12-21) to 26 (19-30) at 12m (n=5)[13]	From median 46 (42-54) to 60 (49-77) at 12m (n=5)[13]	0/5 developed EATL after end of 12m trial (mean 11m follow-up after trial)[13]	1/5 had leukopenia and an opportunistic infection causing withdrawal at 7m and then death; 1/5 had pneumonia controlled without withdrawal; 1/5 had sepsis after small intestinal	Of 5 patients, 2 who were EMA positive and 2 who were anti-tTG negative were no longer positive/negative at 12m[13] From median 10 (8-12) to 13 (12-14) at 12m (n=5)[13] From median 2 (1-3) to 4 (3-4) at 12m (n=5)[13]	From median 48 (12-55) to 12 (7-16) at 12m (n=5)[13]		1/5 (20%) had partial and none had complete villous recovery at unclear follow-up[1] 5 had MIIIC and 2 had MIIIB at baseline but 2 had MII and 3 had M0 at 12m[13]

	Clinical improvement				Complications of CD	Adverse events	Serological response EMA/Anti-tTG Haemoglobin (mmol/L) Albumin (g/L)	Immunological response	Clinical and histological response	Histological improvement
	Proportion achieved	Specific symptom improvement	BMI (kg/m^2)	Body weight (kg)				Proportion of IELs per 100 epithelial cells		Achieved improvement Change in Marsh status
						perforation after laparotomy; 1/5 died 3 months after 12m trial from superior mesenteric artery infarction and 1/5[13]				
Cladribine[1,4,6,7]	From 6/17 (35%)[6] to 1/2 (50%)[1] 21/22 (95%)[7] at mean 22m, unclear follow-up, and 24m.		From mean 20.6 kg/m2 (SD 2.12) to 21.20 kg/m2 (SD 3.14) at 48d (n-17)[6] From mean 20.9 to 23 at 31m (n=32;		Ulcerative jejunitis resolved in all at 22 months[6] In same study, 7/17 (41%) developed and died from EATL within 56d; 2/17	Nausea & vomiting in 3 (17%), diarrhoea and bronchitis each in 1 (6%) in a study of 17 patients[6]	From mean 7.65 (SD1.35) to 7.69 (SD1.29) at 48d in one study (n-17)[6] From mean 7.8 to 7.9 at 31m in another study (n=32; includes 10 patients pre-treated with	35% (6/17) to 58% (13/22) had ≥20% decrease in aberrant IELs in two studies at 22m at 31m months (from average 73% and 61% to 58% and 56%)[6,7] 1 of 2 had aberrant clone at average 24m[4]		59% (10/17) to 58%(13/22) had improvement at 48d and 24m in 2 studies[6,7] Another study reported that 1 of 2 patients had partial histological response at unclear follow-up[1] Of 17 patients, 6 had MIIIC, 1 MIIIB, and 5 MIIIA at baseline but 4 had MIIIC, 3 had MIIIB, 8 had MIIIA, I each had MII and MI at

	Clinical improvement				Complications of CD	Adverse events	Serological response EMA/Anti-tTG Haemoglobin (mmol/L) Albumin (g/L)	Immunological response	Clinical and histological response	Histological improvement
	Proportion achieved	Specific symptom improvement	BMI (kg/m^2)	Body weight (kg)				Proportion of IELs per 100 epithelial cells		Achieved improvement Change in Marsh status
			includes 10 patients pre-treated with azathioprine or prednisone)[7]		(12%) died from bronchiectasis In another study, 5/32 (16%) developed and died from EATL[7]		azathioprine or prednisone)[7] From mean 30 (SD7.2) to 33.7 (SD7.49) at 48d in one study (n=17)[6] From mean 36 to 39 at 31m (n=32; includes 10 patients pre-treated with azathioprine or prednisone)[7]			mean 22m[6]
Cyclosporin[1,8]	1/2 (50%)[1] to 8/13 (62%)[8] in 2 studies					Nausea and abdominal cramps in 2/13 (15%) and gingivitis in 1/13 (8%)[8]				From 0/2[1] to 6/13 (46%)[8] Of 13 patients, 1 changed from MIIIA to MI at 2 m, 3 from MIIIA to MII after 2 m (1) and 6-12 m (2), 2 changed from MIIIB to MIIA at 2 or 6-12 m, 2 changed from MIIIC to MII or MIIIA at 6-12

	Clinical improvement				Complications of CD	Adverse events	Serological response EMA/Anti-tTG Haemoglobin (mmol/L) Albumin (g/L)	Immunological response	Clinical and histological response	Histological improvement
	Proportion achieved	Specific symptom improvement	BMI (kg/m^2)	Body weight (kg)				Proportion of IELs per 100 epithelial cells		Achieved improvement Change in Marsh status
										m, 4 were unchanged after 2 m (2 with MIIIA + 2 with MIIIC), and 1 with MIIIA was unchanged after 6-12m[8]
Methotrexate[1]	5/7 (71%) at unclear follow-up[1]									2/7 (28.5%) had partial villous recovery at unclear follow-up[1]
Tioguanine[12]	10/12 (83%) at average 12m[12]		Median 19.5 (16.7-27.8) to 22.4 (19.7-27.1) at average 12m[12]	Median 56.5 (46-86) to 65 (53-84) at average 12m[12]	1 death from septic shock &multi-organ failure[12]	Muscle spasm requiring withdrawal and liver test abnormality in 1/12 each (8%)[12]	Median 7.7 (6.5-9.7) to 8 (7.3-9.9) at average 12m[12] Median 38 (27-44) to 40 (32-45) at average 12m[12]			7/9 (78%) achieved at average 18 months (one beyond 48m); not determined in 3[12] n=6 from MIIIA/IIIB to M0 by 12m (5) or 24m (1) n=2 with MIIIA unchanged at 24 m but 1 had M0 at 48m Of 4 who died, 1 had reduced from MIIIC to MIIIB by 12m[12]
Drug combinations										
Aminosalicylates + corticosteroids										
Mesalamine+ budesonide[9]	2/6 (33%) had total symptom reduction, 1 had at least 50% and 3 had < 50%[9]					1/10 had headache leading				

	Clinical improvement				Complications of CD	Adverse events	Serological response EMA/Anti-tTG Haemoglobin (mmol/L) Albumin (g/L)	Immunological response	Clinical and histological response	Histological improvement
	Proportion achieved	Specific symptom improvement	BMI (kg/m^2)	Body weight (kg)				Proportion of IELs per 100 epithelial cells		Achieved improvement Change in Marsh status
						to withdrawal (unclear if they had budesonide) [9]				
Multiple corticosteroids										
Budesonide+ prednisone[5]	1/3 (33%) for each complete, moderate and poor at 7m[5]									
Corticosteroids + Immunosupressants										
Azathioprine+ prednisone[2,3,4]	17/18 (95%) at 52 weeks[2]				6/18 (33%) developed EATL; 7/18 (39%) died within follow-up of 52 weeks [2] 36% 5-year survival in one study				1/2 (50%) at mean 26m[4]	

	Clinical improvement				Complications of CD	Adverse events	Serological response EMA/Anti-tTG Haemoglobin (mmol/L) Albumin (g/L)	Immunological response	Clinical and histological response	Histological improvement
	Proportion achieved	Specific symptom improvement	BMI (kg/m^2)	Body weight (kg)				Proportion of IELs per 100 epithelial cells		Achieved improvement Change in Marsh status
					(n=46)[3]					
Azathioprine+ prednisone+ cladribine[3]					22% 5-year survival in one study (n=23)[3]					
Azathioprine+ budesonide[5,14]	3/4 (75%) complete and 1/4 (25%) poor at 7m[5] 2/2 (100%) at 28.5m[5]		From mean 19.7 to mean 21.5 at 24m[14] (n=2)			One of 2 had skin fragility at 14m; the same patient had postprandial abdominal pain and weight loss after budesonide[14]		1 had 90% at baseline but, in one, this was not measured; value were 50 and 67% in each patient at 24m[14]		Of 2 patients, one remained MIIIC at 24m; another with MIIIA was MIIIB at 24m[14]
Azathioprine+ budesonide+ prednisone[5]	5/7 (71%) moderate 2/7 (29%) poor at 7m[5]									

	Clinical improvement				Complications of CD	Adverse events	Serological response EMA/Anti-tTG Haemoglobin (mmol/L) Albumin (g/L)	Immunological response	Clinical and histological response	Histological improvement
	Proportion achieved	Specific symptom improvement	BMI (kg/m^2)	Body weight (kg)				Proportion of IELs per 100 epithelial cells		Achieved improvement Change in Marsh status
Multiple immunosupressants										
Cladribine + pre-treatment with azathioprine or prednisone[7]	7/10 (70%) at 24m[7]							1/10 (10%) at 24m[7]		2/10 (20%) at 24m[7]

1 Malamut 2009 – clinical response defined as reduction in diarrhoea with a decrease of 50% of the number of stools or of weight stools per day and/or the recovering of 50% of weight loss; partial histological response was defined as villous architecture improvement by at least one grade and was complete if villous architecture was restored to normal
2 Goerres 2003 – clinical response defined as disappearance of diarrhoea, or loss of fatigue or weakness; histological improvement was improvement in small intestinal histology which may or may not have had a decrease of intra-epithelial lymphocytosis
3 Al-Toma 2007
4 Rubio-Tapia – clinical response defined as disappearance of diarrhoea and at least 2 of the following: increase of BMI > 1 point, increase in albumin > 10% of baseline, increase of haemoglobin > 1 point, and/or reversion > or = to 1 stage of modified Marsh classification after treatment; clinical and histological response if clinical response and normal intestinal biopsy during follow-up
5 Brar 2007
6 Al-Toma 2006 – clinical response defined as disappearance of diarrhoea, improvement in performance status according to WHO scale or at least 2 of the following: increase of BMI >1 point, increase albumin 10% or more from baseline, or increase in haemoglobin >1 point
7 Tack 2011 – clinical response defined as improvement in diarrhoea, abdominal discomfort and/or signs of malapsorption, combined with at least 2 out of the following parameters of intestinal integrity within the normal range or an improvement of 1 or more points in haemoglobin, BMI and albumin; histological response (or complete histological remission) defined as normalisation of architecture of duodenum, classified as Marsh 0 or 1 lesion according to Modified Marsh classification
8 Wahab 2000 – clinical response defined as improved patient symptoms like fatigue, abdominal complaints, diarrhoea; histological response if normalisation of villi (to Marsh I or II); histological response was normalisation of villi (to Marsh I or II)
9 Jamma 2011
10 Cellier 2000 (3 required extended steroid therapy to maintain improvement) – clinical response defined as regression of diarrhoea and improvement in nutritional status

11 Peters 1978
12 Tack 2012 (four of the 10 patients who tolerated treatment for at least 6 months had been using corticosteroids at baseline& 2 were corticosteroid-dependent) – clinical response defined as amelioration of GI symptoms, combined with at least 2 of BMI, albumin, haemoglobin improving within reference range or by ≥1 point; histological response characterised by normalisation of the small mucosal architecture as Marsh 0 or 1 (partial was improvement in Marsh by 2 or more steps)
13 Mauriño 2002
14 Daum 2006 – clinical response was defined as increase of BMI by at least 10% or more OR a clinically significant decrease in bowel movements and an at least stable BMI

Quality appraisal of individual studies – Modified GRADE

Study	Risk of bias (Study design limitations)	Indirectness	Inconsistency	Imprecision	Overall quality
Al-Toma 2006	Serious[1]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Al-Toma 2007	Very serious[5]	None[2]	N/A[3]	Serious[6]	VERY LOW
Brar 2007	Very serious[5]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Cellier 2000	Very serious[5]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Daum 2006	Very serious[5]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Goerres 2003	Serious[1]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Jamma 2011	Very serious [5]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Malamut 2009	Very serious[5]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Mauriño 2002	Serious[1]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Peters 1978	Very serious[5]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Rubio-Tapia 2009	Very serious[5]	None[2]	N/A[3]	Serious[6]	VERY LOW
Tack 2011	Very serious[5]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Tack 2012	Very serious[5]	None[2]	N/A[3]	Very serious[4]	VERY LOW
Wahab 2000	Serious[1]	None[2]	N/A[3]	Very serious[4]	VERY LOW

[1] Serious risk of bias, prospective study design, however patient recruitment and treatment allocation methods were unclear
[2] No serious indirectness, Study sample represents the population of interest with regard to key characteristics sufficiently. Outcome of interest is adequately measured. The interventions used have sufficiently similar administration and dosage of those used in clinical practice
[3] N/A = not applicable for single study
[4] Very serious imprecision - very small sample size (N<10)
[5] Very serious risk of bias, retrospective study design, unclear if patients were consecutively recruited and unclear treatment allocation details
[6] Serious imprecision - small sample size (N <20)

E.11 Question 6.3

What is the effectiveness of nutritional management or nutritional support for people with refractory coeliac disease?

Quality assessment						Number of patients	Effect estimates	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision			
Resolution of gastrointestinal and non-gastrointestinal symptoms								
1[1]	Case series	Very serious[2]	No serious inconsistency[4]	No serious indirectness[5]	Very serious[3]	10	Good response = 6/10 (one discontinued) No improvement = 2/10 Inconclusive effect = 2/10 No patients needed total parenteral nutrition (1.5–2.0 years after the diet).	Very low

[1] Olaussen (2005)
[2] very serious risk of bias, no randomisation
[3] very serious imprecision, study very underpowered
[4] No serious inconsistency detected
[5] No serious indirectness, population as described in protocol

E.12 Question 6.4

GRADE profiles for the effectiveness of autologous stem cell transplant for people with refractory coeliac disease

Quality assessment						Number of patients		Absolute effect		Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	ASCT	No ASCT	ASCT	No ASCT	
Outcome: Mortality at the end of follow-up: Median in months (range) = 26 (10–67)										
1[1]	Case series	Very serious[2]	N/A[3]	Serious[4]	CBA[5]	13	5	23% (3/13)	100% (5/5)	Very low

[1] One study with 2 published papers: Tack (2011) & Al-toma (2007)
[2] Non-randomised study, prone to selection bias, unclear whether it was retrospective or prospective case series, unclear whether it was consecutive or non-consecutive recruitment, 3 of 5 patient in the No ASCT group had progressed into EATL before stem cells could be collected (and therefore treated as comparison), very small number of cases.
[3] N/A: Non-applicable, single study
[4] Specific subgroup of RCD: RCD type II who were unresponsive to cladribine therapy.

[5] CBA: Cannot be assessed

E.13 Question 7.1

Carers experience of diagnosis

Quality assessment								
Example	**Studies**	**Design**	**Risk of bias**	**inconsistency**	**indirectness**	**N**	**Supporting statement**	**Quality**
Understanding the diagnosis								
Difficulty getting a diagnosis	Cederborg (2011)	Qualitative	Low[1]	NA[2]	None[3]	20	*"I felt everything was not as it should be…many months before the diagnosis was made"*	HIGH
Curiosity about CD	Cederborg (2011)	Qualitative	Low[1]	NA[2]	None[3]	20	"*she never showed any symptoms, she had never been sick*"	HIGH
Lack of knowledge, anxiety	Rosen (2011)	Qualitative	Low[1]	NA[2]	None[3]	43	*"I wasn't totally sure…I had a little hope that maybe it wasn't so, but what was it then? Something even worse…I was scared…I got nightmares"*	HIGH
Relief at being given a diagnosis	Rosen (2011)	Qualitative	Low[1]	NA[2]	None[3]	43	"*We'd been to the paediatric clinic earlier for different diffuse problems, so when we found out about this, it was as if it suddenly dawned on me"*	HIGH
Transforming to a gluten-free diet (GFD)								
Getting used to the GFD	Cederborg (2011)	Qualitative	Low[1]	NA[2]	None[3]	20	*"I panicked about everything…the first two months were a mess"*	HIGH
Social impact of GFD	Cederborg (2011)	Qualitative	Low[1]	NA[2]	None[3]	20	*"he cannot spontaneously be with his peers.. he fears his peers with think his a bother"*	HIGH

[1] Low risk of bias as assessed by CASP qualitative studies checklist
[2] Not applicable, single study
[3] No serious indirectness, population and outcomes as specified in protocol

Adolescents' experience of diagnosis

Quality assessment								
Example	**Studies**	**Design**	**Risk of bias**	**inconsistency**	**indirectness**	**N**	**Supporting statement**	**Quality**
Understanding the diagnosis								
Resentment at not being involved in the decision to undergo testing	Rosen (2011)	Qualitative Cross-sectional	Low[1]	NA[2]	None[3]		Used to describe receiving the diagnosis: *"getting caught"* *"being stuck"*	HIGH
Anger at diagnosis	Rosen (2011)	Qualitative Cross-sectional	Low[1]	NA[2]	None[3]		*"I got very annoyed when my doctor called to say that I was gluten intolerant…because I had no symptoms"*	HIGH
Transforming to a gluten-free diet (GFD)								
Motivation to follow GFD	Rosen (2011)	Qualitative Cross-sectional	Low[1]	NA[2]	None[3]		*"To eat gluten-free is like, it's just best for me"* *"it sort of feels important"*	HIGH

Health related quality of life post diagnosis in asymptomatic individuals

Quality assessment			Quality

[1] Low risk of bias as assessed by CASP qualitative studies checklist
[2] NA: Not applicable, single study
[3] No serious indirectness, population and outcome as specified in protocol

Example	Studies	Design	Risk of bias	inconsistency	indirectness	N	Odds ratio (95% CI)	
Difference in mobility EQ-5D scores between cases and controls at follow-up 1 year post screening								
EQ5D - mobility	Nordyke (2011)	Cohort	Low[1]	NA[2]	None[3]	586	0.78 (0.21 - 2.84)	MODERATE
Difference in anxiety/depression EQ-5D scores between cases and controls at follow-up 1 year post screening								
EQ5D - anxiety	Nordyke (2011)	Cohort	Low[1]	NA[2]	None[3]	586	0.70 (0.39 - 1.26)	MODERATE
Difference in activity EQ-5D scores between cases and controls at follow-up 1 year post screening								
EQ5D-anxiety/depression	Nordyke (2011)	Cohort	Low[1]	NA[2]	None[3]	586	0.85 (0.19 - 3.89)	MODERATE
Difference in pain EQ-5D scores between cases and controls at follow-up 1 year post screening								
EQ5D HRQoL - pain	Nordyke (2011)	Cohort	Low[1]	NA[2]	None[3]	586	**0.51 (0.28 - 0.96)**	MODERATE
Difference in VAS general health dimension EQ-5D scores between cases and controls at follow-up 1 year post screening								
EQ5D HRQoL - general wellbeing	Nordyke (2011)	Cohort	Low[1]	NA[2]	None[3]	586	0 (0.00)	MODERATE

Gastrointestinal symptoms and health related quality of life in those following a GFD compared to gluten-containing diet in asymptomatic seropositive adults

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Mean differnce (95% CI)**	**Quality**
Improvement in GI symptoms in GSRS score between GFD and 'normal diet' groups								

[1] Low risk of bias as assessed by CASP qualitative studies checklist
[2] NA: not applicable, single study
[3] No serious indirectness, population and outcome as specified within protocol

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Mean differnce (95% CI)**	**Quality**
Kurppa (2014)	RCT	Low[1]	NA[2]	None[3]	None[4]	40	-0.14 (0.7 to - 0.1)	HIGH

Histological recovery in those following a GFD compared to gluten-containing diet in asymptomatic seropositive adults

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Mean differnce (95% CI)**	**Quality**
Improvement in expression of CD3+ intraepithelial lymphocytes between GFD and 'normal diet' groups								
Kurppa (2014)	RCT	Low[5]	NA[6]	None[7]	serious[8]	40	-0.12.5 (-39.5 to 14.4)	MODERATE

[1] Low risk of Bias as assessed by NICE RCT quality checklist.
[2] NA; not applicable, single study contributed to this data
[3] No serious indirectness; all participants were assumed to have CD on the basis of seropositivity to EMA
[4] No serious imprecision, confidence intervals are tight
[5] Low risk of Bias as assessed by NICE RCT quality checklist.
[6] NA; not applicable, single study contributed to this data
[7] No serious indirectness; all participants were assumed to have CD on the basis of seropositivity to EMA
[8] serious imprecision, confidence intervals are wide and cross the line of no effect

E.14 Question 7.2

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Mean difference score and 95% CI**	**Quality**
Mean difference in GSRS from baseline to 10 weeks between intervention and control groups								
Jacobssen (2007)	RCT	Serious[1]	NA[2]	None[3]	None[4]	105	-0.19 (-0.21, -0.17)	LOW

Specialised education, behavioural modification, and cognitive behavioural therapy intervention to improve GFD adherence

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Relative risk (95% CI)**	**Quality**
Improvement in adherence in CDAQ score between intervention and wait-list control groups								
Sainsbury (2012)	RCT	Serious[1]	NA[2]	serious[5]	None[4]	189	1.51 (0.82 - 2.78)	LOW

Specialised psychological support counselling to improve GFD adherence

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Relative risk (95% CI)**	**Quality**
Noncompliance to GFD in intervention compared to control group - post intervention follow-up								
Addolorato (2004)	RCT	Serious[1]	NA[2]	None[3]	None[4]	66	**0.23 (0.07 - 0.73)**	MODERATE

[1] Serious risk of bias as assed by NICE RCT quality checklist. Method of randomisation unclear; Statistical methodology not optimal - main group x treatment interaction not reported. Statistically significant difference between groups at baseline, where control participants reported fewer GI symptoms
[2] Not applicable; only study contributed to the analyses
[3] No serious indirectness, population was specified as in protocol
[4] No serious imprecision, tight confidence intervals
[5] Serious indirectness, all participants presented with anxiety

Useful sources of information about coeliac disease and the GFD

Quality assessment								
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Number of participants**	**Summary of findings Percentage of patient population survival**	**Quality**
Coeliac support association								
2 studies: Zarkadas (2012) Leffler (2008)	Cross-sectional	Low[1]	None[2]	None[3]	None[4]	6066	88% – 90 %	MODERATE
Another patient								
Zarkadas (2012)	Cross-sectional	Low[1]	NA[5]	None[3]	None[4]	5914	67%	MOERATE
GP								
2 studies: Zarkadas (2012) Leffler (2008)	Cross-sectional	Low[1]	None[2]	None[3]	None[4]	6066	25% - 36%	MODERATE
Gastroenterologist								
2 studies: Zarkadas (2012) Leffler (2008)	Cross-sectional	Low[1]	None[2]	None[3]	None[4]	6066	43% - 57%	MODERATE
Dietician								
2 studies: Zarkadas (2012) Leffler (2008)	Cross-sectional	Low[1]	None[2]	None[3]	None[4]	6066	52%-63%	MODERATE
Cookbook								
Zarkadas (2012)	Cross-sectional	Low[1]	NA[5]	None[3]	None[4]	5914	62%	MODERATE
Internet								

[1] Low risk of bias, as assessed by CASP qualitative study checklist
[2] No serious inconsistency, confidence intervals overlap
[3] No serious indirectness, population as specified in protocol
[4] No serious imprecision - estimates are consistent between studies
[5] NA = not applicable, only one study contributed to this analyses

Quality assessment								
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Number of participants	Summary of findings Percentage of patient population survival	Quality
Zarkadas(2012)	Cross-sectional	Low	NA[5]	None[3]	None[4]	5914	52%	MODERATE

Patient experience of the GFD

Quality assessment Number of patients								
Example	Studies	Design	Risk of bias	Inconsistency	Indirectness	N	Supporting statement	Quality
Emotional experience of elf management								
Embarassment of eating in social situation	3 studies: Olsson (2008) Rashid (2005) Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	6127	*"I felt embarrassed, thinking that the whole school knew I had coeliac disease...that was hard"*	MODERATE
Feeling a burden to family and friends	2 studies: Olsson (2008) Rashid (2005) Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	5959	*"…I ate normal food because my family thinks it's so awfully hard to explain about my diet"*	MODERATE
Avoid social situations because of food	Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	5912	*Not available*	MODERATE
Do not like others to feel sorry	Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	5912	*Not available*	MODERATE

[1] Low risk of bias as assessed by CASP qualitative research quality checklist
[2] NA, not appropriate for qualitative research for subjective personal experience outcomes
[3] No serious indirectness, population and outcomes were as specified within protocol

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**
for them								
Social experience of self-management								
Limited availability and palatability of GF foods	4 studies: Olsson (2008) Rashid (2005) Zarkadas (2012) Leffler (2008)	Qualitative	Low[1]	NA[2]	None[3]	6281	"*You know how different foods taste, so you choose the ones that taste the best…"*	MODERATE
Difficulty eating out	4 studies: Olsson (2008) Rashid (2005) Zarkadas (2012) Leffler (2008)	Qualitative	Low[1]	NA[2]	None[3]	6281	*"at a café…when the only pastries you can choose from are the ones you aren't allowed to eat"*	MODERATE
Difficulty travelling	4 studies: Olsson (2008) Rashid (2005) Zarkadas (2012) Leffler (2008)	Qualitative	Low[1]	NA[2]	None[3]	6281	*" When I was [in Vietnam], I ate normal food because my family thinks it's so awfully hard to explain about my diet"*	MODERATE
Feeling excluded from social activities	5 studies: Olsson (2008) Rashid (2005) Zarkadas (2012) Leffler (2008) Erichiello	Qualitative	Low[1]	NA[2]	None[3]	6485	*"I want to try things…I want to try things in life. I will never let a disease force me to not"*	MODERATE

Quality assessment Number of patients								
Example	Studies	Design	Risk of bias	Inconsistency	Indirectness	N	Supporting statement	Quality
	(2010)							

[1] Low risk of bias as assessed by CASP qualitative research quality checklist
[2] NA, not appropriate for qualitative research for subjective personal experience outcomes
[3] No serious indirectness, population and outcomes were as specified within protocol

Patient experience: Factors that positively influence adherence

Quality assessment								
Example	**Studies**	**Design**	**Risk of bias**	**inconsistency**	**indirectness**	**N**	**Outcome on GFD adherence**	**Quality**
Knowledge of CD and the GFD								
Good knowledge of GFD	**2 Studies:** Leffler (2008) Zarkadas (2012)	cross-sectional	Low[1]	None[2]	None[3]	6066	*Increased knowledge about the GFD and gluten containing foods was associated with better adherence*	MODERATE
Time spent following GFD								
Length of time on GFD	**2 Studies:** Leffler (2008) Zarkadas (2012)	cross-sectional	Low[1]	None[2]	None[3]	6066	*Participants who had followed the GFD for a longer amount of time were more likely to adhere to GFD*	MODERATE
Social and emotional factors								
Social relationships	**2 Studies:** Erichiello Leffler (2008)	cross-sectional	Low[1]	None[2]	None[3]	358	*Those with good social relationships adhered to GFD better than those with poor social relationships.*	MODERATE
School integration	Erichiello	cross-sectional	Low[1]	NA[4]	None[3]	204	*Those with excellent school integration adhered to GFD better than those with poor or sufficient integration*	MODERATE
Self-constraint	Erichiello	cross-sectional	Low[1]	NA[4]	None[3]	204	*People without feelings of self-constraint adhered better to GFD than those without feelings of self-constraint*	MODERATE
Membership of CD specialist organisation	Leffler (2008)	cross-sectional	Low[1]	NA[4]	None[3]	154	*A large proportion of participants felt that being a member of their local coeliac sociality was beneficial in improving adherence*	MODERATE
Prevention of adverse consequences								

[1] Low risk of bias as assessed by CASP qualitative research quality checklist
[2] No serious inconsistency, studies reflected similar qualitative outcomes
[3] No serious indirectness, population and outcomes of interest as identified within protocol
[4] NA, not applicable, single study

Quality assessment								
Example	**Studies**	**Design**	**Risk of bias**	**inconsistency**	**indirectness**	**N**	**Outcome on GFD adherence**	**Quality**
Prevention of symptoms	**2 Studies:** Zarkadas (2012) Leffler (2008)	cross-sectional	Low[1]	None[2]	None[3]	6066	*A large proportion of participants adhered to the GFD to avoid gluten-associated gastrointestinal symptoms*	MODERATE
Prevention of serious long term health complication	**2 Studies:** Zarkadas (2012) Leffler (2008)	cross-sectional	Low[1]	None[2]	None[3]	6066	*A large proportion of participants adhered to the GFD to avoid serious long term health consequences*	MODERATE

[1] Low risk of bias as assessed by CASP qualitative research quality checklist
[2] No serious inconsistency, studies reflected similar qualitative outcomes
[3] No serious indirectness, population and outcomes of interest as identified within protocol

Patient experience: Strategies to improve adherence

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**
Adaptive strategies for improving adherence								
Bringing own GF foods out	2 studies**:** Olsson (2008) Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	5959	Adaptive strategy: '*have snacks on hand'*	MODERATE
Seeking emotional support from family	2 studies**:** Olsson (2008) Zarkadas	Qualitative	Low[1]	NA[2]	None[3]	5959	Adaptive strategy: '*Invite friends/family to eat at my home"*	MODERATE

[1] Low risk of bias as assessed by CASP qualitative research quality checklist
[2] NA, not applicable, not appropriate for qualitative research subjective personal experience outcome
[3] No serious indirectness, population and outcomes of interest as specified in protocol

Quality assessment Number of patients								
Example	**Studies**	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**N**	**Supporting statement**	**Quality**
and friends	(2012)							
Avoid exposure to sensory aspects of gluten-containing foods	Olsson (2008)	Qualitative	Low[1]	NA[2]	None[3]	47	*"I never eat normal food…because I will only be tempted to continue and eat more"*	MODERATE
Reading every ingredient on food labels	Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	5912	Adaptive strategy: '*read every ingredient list'*	MODERATE
Labelling all GF foods and storing GF food in a separate area	Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	5912	Adaptive strategy: '*label all GF flours'* *Adaptive strategy: 'store GF foods in a separate area'*	MODERATE
Enquiring about gluten content of foods in restaurants	Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	5912	Adaptive strategy: '*call ahead to enquire about GF menu choices'*	MODERATE
Talking to others with CD	Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	5912	*'talk to others about coeliac disease and the GF diet'*	MODERATE
Reminding hosts of GFD if event involves food	Zarkadas (2012)	Qualitative	Low[1]	NA[2]	None[3]	5912	Adaptive strategy: *'Offer to bring a GF dish to events involving food'*	MODERATE

[1] Low risk of bias as assessed by CASP qualitative research quality checklist

[2] NA, not applicable, not appropriate for qualitative research subjective personal experience outcome
[3] No serious indirectness, population and outcomes of interest as specified in protocol

E.15 Question 7.3

E.15.1 SECTION 1: The role of oats in children

GRADE profiles for the role of oats (children) – Serological outcomes: IgA EMA, TGA and nitric oxide

Quality assessment						Number of patients		Effect		
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-oats[1]	Standard GFD	Relative (95% CI)	Absolute (95% CI)	Quality
CHILDREN (newly diagnosed): Serological outcome: IgA EMA positive at 12-month										
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	14/42 (33.3%)	12/50 (24%)	RR 1.39 (0.72 to 2.67)	9 more per 100 (from 7 fewer to 40 more)	LOW
CHILDREN (newly diagnosed): Serological outcome: TGA positive at 12-month										
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	7/42 (16.7%)	5/50 (10%)	RR 1.67 (0.57 to 4.87)	7 more per 100 (from 4 fewer to 39 more)	LOW
CHILDREN (newly diagnosed): Serological outcome: Nitric oxide (NO)[5] metabolites at 12-month (the cut-off value = 1406 µM):										
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	9/34 (26.5%)	8/41 (19.5%)	RR 1.36 (0.59 to 3.13)	7 more per 100 (from 8 fewer to 42 more)	LOW
CHILDREN (newly diagnosed): Serological outcome: IgA EMA positive at 12-month (SUBGROUP: GFD-oats ≥8g daily)										
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	12/34 (35.3%)	12/50 (24%)	RR 1.47 (0.75 to 2.88)	11 more per 100 (from 6 fewer to 45 more)	LOW
CHILDREN (newly diagnosed): Serological outcome: IgA EMA titres (1:10-1:20) at 12-month (SUBGROUP: GFD-oats ≥8g daily)										
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	5/34 (14.7%)	8/50 (16%)	RR 0.92 (0.33 to 2.57)	1 fewer per 100 (from 11 fewer to 25 more)	LOW
CHILDREN (newly diagnosed): Serological outcome: IgA EMA titres (1:40-1:80) at 12-month (SUBGROUP: GFD-oats ≥8g daily)										
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	7/34 (20.6%)	4/50 (8%)	RR 2.57 (0.82 to 8.12)	13 more per 100 (from 1 fewer to 57 more)	LOW
CHILDREN (newly diagnosed): Serological outcome: TGA positive at 12-month (SUBGROUP: GFD-oats ≥8g daily)										
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	7/34	5/50	RR 2.06	11 more per 100 (from 3 fewer to 49 more)	LOW

Quality assessment						Number of patients		Effect		Quality
						(20.6%)	(10%)	(0.71 to 5.95)		

GFD = gluten free diet; IgA = antigliadin antibody; EMA = antiendomysium antibody; TGA = antitissue transglutaminase
1 = GFD with oats (aimed at a daily oat intake of 25–50g)
2 = Three papers published from a single study (Hogberg 2004; Hollen 2006a; Hollen 2006b).
3 = Methods of randomisation not reported, high number of withdrawals from the GFD with oats group, no ITT analysis, only 12-month follow-up.
4 = Lower limit of 95%CI crosses over 1.25 and no effect.
5 = Nitric oxide (NO) metabolites in morning urine as indicator of ongoing inflammation in the small intestine (the cut-off value = 1406 µM).

GRADE profiles for the role of oats (children) – Histological outcome: IEL count

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-oats[1]	Standard GFD	Mean and SD	
CHILDREN (newly diagnosed): Histologicall outcome: IEL count (per 100 enterocytes) at 12-month									
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	42	50	GFD-oats = 16 (4.5) Standard GFD = 16 (5.0) P=0.84	LOW
CHILDREN (newly diagnosed): Histological outcome: IEL count (per 100 enterocytes) at 12-month (SUBGROUP: GFD-oats ≥8g daily)									
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	34	50	GFD-oats = 16 (4.0) Standard GFD = 16 (5.0) P=0.94	LOW

GFD = gluten free diet; IEL = intraepithelial lymphocytes; SD = standard deviation
1 = GFD with oats (aimed at a daily oat intake of 25–50g)
2 = Three papers published from a single study (Hogberg 2004; Hollen 2006a; Hollen 2006b).
3 = Methods of randomisation not reported, high number of withdrawals from the GFD with oats group, no ITT analysis, only 12-month follow-up.
4 = only p-value provided, unable to assess precision, small sample size, high uncertainty on the precision of effect estimate.

GRADE profiles for the role of oats (children) – Serological outcomes: IgA and IgG anti-avenin antibodies

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-oats[1]	Standard GFD	Median and range	

Quality assessment						Number of patients		Effect	Quality
CHILDREN (newly diagnosed): Serological outcome: IgA anti-avenin antibodies at 12-month									
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	38	43	GFD-oats = 0.24 (0.06 to 1.89) Standard GFD = 0.18 (0.01 to 1.05); P=0.13	LOW
CHILDREN (newly diagnosed): Serological outcome: IgG anti-avenin antibodies at 12-month									
3[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	38	45	GFD-oats = 0.93 (0.38 to 1.55) Standard GFD = 1.08 (0.51 to 1.62); P=0.26	LOW

GFD = gluten free diet
1 = GFD with oats (aimed at a daily oat intake of 25–50g)
2 = Three papers published from a single study (Hogberg 2004; Hollen 2006a; Hollen 2006b).
3 = Methods of randomisation not reported, high number of withdrawals from the GFD with oats group, no ITT analysis, only 12-month follow-up.
4 = only p-value provided, unable to assess precision, small sample size, high uncertainty on the precision of effect estimate.

E.15.2 SECTION 2: The role of oats in adults

GRADE profiles for the role of oats (adults) – Gastrointestinal symptoms: Cluster of flatulence, abdominal pain and distention, general well-being

Quality assessment						Number of patients		Effect	
Number of	**Design**	**Risk of**	**Inconsiste**	**Indirectnes**	**Imprecisio**	**GFD-**	**Standard**	**Mean change from baseline (SD)**	**Quality**

Quality assessment						Number of patients		Effect	Quality
studies		bias	ncy	s	n	oats[1]	GFD		
ADULTS (in remission): GI symptoms score (cluster)[2]: Mean change from baseline at 6-month (score 0 better)									
4[3]	RCT	Serious[4]	N/A	No serious	Serious[5]	26	26	GFD-oats = 6.7 (17.5) Std-GFD = 2.1 (10.8) *Mean change differences between groups = 4.6 (95%CI: -3.5 to 12.8)*	LOW
ADULTS (newly diagnosed): GI symptoms score (cluster)[2]: Mean change from baseline at 12-month (score 0 better)									
4[3]	RCT	Serious[4]	N/A	No serious	Serious[5]	19	21	GFD-oats = -8.2 (26.6) Std-GFD = -8.4 (22.7) *Mean change differences between groups = 0.2 (95%CI: -15.6 to 16.0)*	LOW

GFD = gluten free diet
1 = GFD with oats (the goal for the daily intake of oats was 50g to 70g).
2 = Average of the 4 variables (flatulence, abdominal pain and distention, general well-being), each variable measured on a 100-mm scale, ranging from 0 = no symptoms at all; to 100 = extremely severe symptoms.
3 = Four papers published from a single study (Janatuinen 1995; Janatuinen 2000; Janatuinen 2002; Kemppainen 2007).
4 = Inappropriate randomisation method (randomised by gender).
5 = Very small sample size (<400) and GDG unable to set MID.

GRADE profiles for the role of oats (adults) – Gastrointestinal symptoms: Gastrointestinal symptom rating scale (GSRS)

Quality assessment						Number of patients		Effect	Quality
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-oats[1]	Standard GFD	Mean, SD (with p-value for interaction between groups and time effects in ANOVA)	
ADULTS (in remission): GSRS total score: Mean score 12-month (lower score better)									
1[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	20	16	GFD-oats = 2.00 (0.50) Standard GFD = 1.94 (0.70)	LOW

Quality assessment						Number of patients		Effect	Quality
								P=0.094	
ADULTS (in remission): GSRS Diarrhoea score: Mean score 12-month (lower score better)									
1[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	20	16	GFD-oats = 2.03 (0.74) Standard GFD = 1.69 (0.91) P=0.010	LOW
ADULTS (in remission): GSRS Indigestion score: Mean score 12-month (lower score better)									
1[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	20	16	GFD-oats = 2.06 (0.59) Standard GFD = 2.13 (1.14) P=0.065	LOW
ADULTS (in remission): GSRS Abdominal pain score: Mean score 12-month (lower score better)									
1[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	20	16	GFD-oats = 1.56 (0.39) Standard GFD = 1.83 (0.58) P=0.297	LOW
ADULTS (in remission): GSRS Constipation score: Mean score 12-month (lower score better)									
1[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	20	16	GFD-oats = 2.24 (0.70) Standard GFD = 2.23 (1.23) P=0.297	LOW
ADULTS (in remission): GSRS Reflux score: Mean score 12-month (lower score better)									
1[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	20	16	GFD-oats = 2.07 (0.92) Standard GFD = 1.81 (0.87) P=0.781	LOW

GFD = gluten free diet
1 = GFD with oats (50g of oats-containing gluten-free products daily).
2 = Peraaho (2004).
3 = Blinding not reported, lack of baseline data (e.g. inclusion/exclusion criteria), unclear ITT was carried out.
4 = Very small sample size (<400) and GDG unable to set MID.

GRADE profiles for the role of oats (adults) – Histological outcomes: Villous atrophy and intraepithelial lymphocytes (IEL) count

Quality assessment						Number of patients		Effect	
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**GFD-oats[1]**	**Standard GFD**	**Mean change from baseline (SD)**	**Quality**
ADULTS (in remission): Villous atrophy (mean histopathological grade)[2]: Mean change from baseline at 6-month (score 0 better)									

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-oats[1]	Standard GFD	Mean change from baseline (SD)	Quality
4[3]	RCT	Serious[4]	N/A	No serious	Serious[5]	26	26	GFD-oats = 0.01 (0.36) Std-GFD = -0.06 (0.31) *Mean change differences between groups = 0.07 (95%CI: -0.12 to 0.26)*	LOW
ADULTS (newly diagnosed): Villous atrophy (mean histopathological grade)[2]: Mean change from baseline at 12-month (score 0 better)									
4[3]	RCT	Serious[4]	N/A	No serious	Serious[5]	19	21	GFD-oats = -1.07 (0.58) Std-GFD = -1.20 (0.42) *Mean change differences between groups = 0.13 (95%CI: -0.23 to 0.43)*	LOW
ADULTS (merged in remission group and newly diagnosed group): Villous atrophy (mean histopathological grade)[2]: Mean change from 6-12 months at 5-year (score 0 better)									
4[3]	RCT	Serious[4]	N/A	No serious	Serious[5]	35	28	GFD-oats = -0.55 (0.54) Std-GFD = -0.52 (0.45) *Mean change differences between groups = 0.03 (95%CI: -0.29 to 0.23)*	LOW
ADULTS (in remission): Intraepithelial lymphocytes (IEL) count/100 epithelial cells: Mean change from baseline at 6-month									
4[3]	RCT	Serious[4]	N/A	No serious	Serious[5]	26	26	GFD-oats = -0.6 (21.8) Std-GFD = 2.0 (11.7) *Mean change differences between groups = -2.6 (95%CI: -12.3 to 7.2)*	LOW
ADULTS (newly diagnosed): Intraepithelial lymphocytes (IEL) count/100 epithelial cells: Mean change from baseline at 12-month									
4[3]	RCT	Serious[4]	N/A	No serious	Serious[5]	19	21	GFD-oats = -23.8 (23.3) Std-GFD = -21.7 (14.5) *Mean change differences between groups = -2.1 (95%CI: -14.4 to 10.2)*	LOW

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-oats[1]	Standard GFD	Mean change from baseline (SD)	Quality
ADULTS (in remission): Intraepithelial lymphocytes (IEL) count/millimetre of epithelium: Mean at 12-month									

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-oats[1]	Standard GFD	Mean change from baseline (SD)	Quality
1[7]	RCT	Serious[8]	N/A	No serious	Serious[9]	18	13	GFD-oats = 44.6 (22.7) Std-GFD = 26.7 (21.0) P = 0.039	LOW

GFD = gluten free diet
1 = GFD with oats (the goal for the daily intake of oats was 50g to 70g).
2 = Villous atrophy was graded as 1 = partial; 2 = subtotal; or 3 = total. A grade of 0 indicates the absence of villous atrophy.
3 = Four papers published from a single study (Janatuinen 1995; Janatuinen 2000; Janatuinen 2002; Kemppainen 2007).
4 = Inappropriate randomisation method (randomised by gender).
5 = Very small sample size (<400) and GDG unable to set MID.
6 = GFD with oats (50g of oats-containing gluten-free products daily).
7 = Peraaho (2004).
8 = Blinding not reported, lack of baseline data (e.g. inclusion/exclusion criteria), unclear ITT was carried out.
9 = Very small sample size (<400) and GDG unable to set MID, no baseline data.

GRADE profiles for the role of oats (adults) – Serological outcomes: Anti-gliadin IgA, Anti-gliadin IgG and Anti-reticulin IgA

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-oats[1]	Standard GFD	Median change from baseline (range)	Quality
ADULTS (in remission): Anti-gliadin IgA (EU/ml): Median change from baseline at 6-month									
4[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	26	26	GFD-oats = 0.0 (-0.47 to 0.41) Std-GFD = 0.0 (0.0 to 0.39) P = 0.33	LOW
ADULTS (newly diagnosed): Anti-gliadin IgA (EU/ml): Median change from baseline at 12-month									
4[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	19	21	GFD-oats = -0.73 (-0.99 to 0.00) Std-GFD = -0.57 (-9.38 to 0.00) P = 0.69	LOW
ADULTS (in remission): Anti-gliadin IgG (EU/ml): Median change from baseline at 6-month									
4[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	26	26	GFD-oats = 0.0 (-1.21 to 2.02) Std-GFD = 0.0 (-2.63 to 0.86) P = 0.12	LOW
ADULTS (newly diagnosed): Anti-gliadin IgG (EU/ml): Median change from baseline at 12-month									

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-oats[1]	Standard GFD	Median change from baseline (range)	Quality
4[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	19	21	GFD-oats = -7.09 (-29.85 to 0.00) Std-GFD = -2.99 (-55.2 to 0.53) P = 0.99	LOW
ADULTS (in remission): Anti-reticulin IgA (EU/ml): Median change from baseline at 6-month									
4[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	26	26	GFD-oats = 0.0 (-50.0 to 0.00) Std-GFD = 0.0 (-50.0 to 0.00) P = 1.00	LOW
ADULTS (newly diagnosed): Anti-reticulin IgA (EU/ml): Median change from baseline at 12-month									
4[2]	RCT	Serious[3]	N/A	No serious	Serious[4]	19	21	GFD-oats = -200.0 (-2000.0 to 0.00) Std-GFD = -175.0 (-4000.0 to 5.00) P = 0.79	LOW

GFD = gluten free diet
1 = GFD with oats (the goal for the daily intake of oats was 50g to 70g).
2 = Four papers published from a single study (Janatuinen 1995; Janatuinen 2000; Janatuinen 2002; Kemppainen 2007).
3 = Inappropriate randomisation method (randomised by gender).
4= Only p-value provided for median, unable to assess precision, small sample size, high uncertainty on the precision of effect estimate.

E.15.3 SECTION 3: The role of kilned and unkilned oats in adults

GRADE profiles for the role of kilned oats and unkilned oats (adults) – Gastrointestinal symptoms: Abdominal pain, flatulence, abdominal distention, diarrhoea

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	GFD-kilned oats[1]	GFD-unkilned oats[2]	Categorical data: (a) Not at all (b) To some extent (c) Moderate (d) Extreme (Mann–Whitney U-test)	Quality
ADULTS (in remission): Abdominal pain at 6-month									
3[3]	RCT	Very serious[4]	N/A	No serious	Serious[5]	16	15	Kilned oats: (a)=11, (b)=5, (c)=0, (d)=0 Unkilned oats: (a)=11, (b)=4, (c)=0, (d)=0 p>0.05	Very low
ADULTS (in remission): Flatulence at 6-month									
3[3]	RCT	Very serious[4]	N/A	No serious	Serious[5]	16	15	Kilned oats: (a)=7, (b)=6, (c)=3, (d)=0 Unkilned oats: (a)=6, (b)=6, (c)=2, (d)=0 p>0.05	Very low
ADULTS (in remission): Abdominal distention at 6-month									
3[3]	RCT	Very serious[4]	N/A	No serious	Serious[5]	16	15	Kilned oats: (a)=10, (b)=4, (c)=2, (d)=0 Unkilned oats: (a)=11, (b)=3, (c)=1, (d)=0 p>0.05	Very low
ADULTS (in remission): Diarrhoea at 6-month									
3[3]	RCT	Very serious[4]	N/A	No serious	Serious[5]	16	15	Kilned oats: (a)=12, (b)=3, (c)=1, (d)=0	Very low

Quality assessment						Number of patients		Effect	Quality
								Unkilned oats: (a)=13, (b)=2, (c)=0, (d)=0 p>0.05	

GFD = gluten free diet
1 = The aim of the daily intake of kilned oats was 100g.
2 = The aim of the daily intake of unkilned oats was 100g.
3 = Three papers published from a single study.
4 = Downgraded 2 levels: methods of randomisation not reported, allocation concealment unclear, blinding unclear, potential reporting bias on some outcomes where there was a lack of details, no analysis of crossover effects.
5 = Only p-value provided, unable to assess precision, small sample size, high uncertainty on the precision of effect estimate.

GRADE profiles for the role of kilned oats and unkilned oats (adults) – Serological outcomes: Erythrocyte folate, serum vitamin B-12 and serum calcium

Quality assessment						Number of patients		Effect	Quality
Number of	Design	Risk of	Inconsiste	Indirectnes	Imprecisio	GFD-kilned	GFD-	Mean, SD (with p-value for	

Quality assessment						Number of patients		Effect	Quality
studies		bias	ncy	s	n	oats[1]	unkilned oats[2]	difference between groups at 6-month)	
ADULTS (in remission): Mean erythrocyte folate (nmol/L) at 6-month									
3[3]	RCT	Very serious[4]	N/A	No serious	Serious[5]	16	15	Kilned oats = 582 (185) Unkilned oats = 496 (102) P=0.18	Very low
ADULTS (in remission): Mean serum vitamin B-12 (pmol/L) at 6-month									
3[3]	RCT	Very serious[4]	N/A	No serious	Serious[5]	16	15	Kilned oats = 279 (109) Unkilned oats = 287 (93) P=0.68	Very low
ADULTS (in remission): Mean serum calcium (mmol/L) at 6-month									
3[3]	RCT	Very serious[4]	N/A	No serious	Serious[5]	16	15	Kilned oats = 2.30 (0.14) Unkilned oats = 2.30 (0.10) P=0.63	Very low

GFD = gluten free diet
1 = The aim of the daily intake of kilned oats was 100g.
2 = The aim of the daily intake of unkilned oats was 100g.
3 = Three papers published from a single study.
4 = Downgraded 2 levels: methods of randomisation not reported, allocation concealment unclear, blinding unclear, potential reporting bias on some outcomes where there was a lack of details, no analysis of crossover effects.
5 = Only p-value provided, unable to assess precision, small sample size, high uncertainty on the precision of effect estimate.
Note: Normal values for the general population: Erythrocyte folate: 315-850nmol/L; Serum vitamin B-12: 140-540pmol/L; Serum calcium: 2.2-2.65mmol/L.

E.15.4 SECTION 4: Nutritional supplements (adults)

GRADE profiles for the role of nutritional supplements (adults) – QoL outcome: Psychological general well-being (PGWB) scale

Quality assessment						Number of patients		Effect	
Number of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Supplements[1]	Placebo	Median and range	Quality
ADULTS (in remission): Median PGWB score at 6-month (higher score better)									

Quality assessment						Number of patients		Effect	Quality
1[2]	RCT	Very serious[3]	N/A	No serious	Serious[4]	11	12	Supplements = 105 (87 to 115) Placebo = 94 (40 to 121) p>0.05	Very low

1 = A daily dose of 0.8 mg folic acid, 0.5 mg cyanocobalamin (vitamin B-12) and 3 mg pyridoxine (vitamin B-6)
2 = Hallert (2009)
3 = Downgraded 2 levels: No mention of allocation concealment and not reported the method of randomisation, only conducted per-protocol analysis (no ITT), not clear whether the study sample has carried on their GFD or not during the trial.
4 = Very small sample size, only median with range were reported, unable to set MID, high uncertainty of the precision of the effect estimate.

GRADE profiles for the role of nutritional supplements (adults) – Serological outcome: Plasma total homocysteine (tHcy) level (marker of B vitamin status)

Quality assessment						Number of patients		Effect	
Number of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Supplements[1]**	**Placebo**	**Median and range**	**Quality**
ADULTS (in remission): Median P-tHcy (μmol／L) at 6-month									
1[2]	RCT	Very serious[3]	N/A	No serious	Serious[4]	11	12	Supplements = 7.9 (5.0 to 11.3) Placebo = 11.1 (5.3 to 22.4) P<0.001	Very low

1 = A daily dose of 0.8 mg folic acid, 0.5 mg cyanocobalamin (vitamin B-12) and 3 mg pyridoxine (vitamin B-6)
2 = Hallert (2009)
3 = Downgraded 2 levels: No mention of allocation concealment and not reported the method of randomisation, only conducted per-protocol analysis (no ITT), not clear whether the study sample has carried on their GFD or not during the trial.
4 = Very small sample size, only median with range were reported, unable to set MID, high uncertainty of the precision of the effect estimate.