Chapter 3. Results

Celiac 1: Sensitivity and Specificity of Tests for CD

Serology

Out of 3,982 citations identified by the search strategy for the Celiac 1 objective, 907 met level 2 screening criteria. Of these, 204 diagnostic test studies of one or more of the serological markers of interest (AGA, EMA, tTG) were identified. Sixty studies fulfilled the level 3 inclusion criteria (Appendix F; Evidence Table 1, Appendix I).[26-85] The most common reasons for failing level 3 inclusions were AGA studies conducted before 1990, studies utilizing an improper or an unbiopsied control group, or studies that did not give any description of the biopsy criteria defining CD. Five pairs of duplicate publications were identified.[27,28,45,46,58,65,73,74,84,86] Out of each duplicate pair, the study with the most complete data was abstracted,[27,45,46,58,74] bringing the total of included unique studies to 55. The majority of these studies assessed more than one serological marker, and some studied more than one age group. Of the included articles, 20 were conducted in or included an adult population, 33 were conducted in a population of children, and eight in a mixed population of adults and children of varying proportions. The statements in this section that relate to mixed studies or studies in children and adults refer to these eight studies, and not to a sample that we pooled from different studies.

To minimize clinical and statistical heterogeneity, the included articles of a particular antibody test were divided into groups by age of the included population (adults, children, mixed), the study design (case control, or relevant clinical population/cohort), by antibody type (IgA or IgG), and by test methodology (e.g., monkey esophagus [ME] or human umbilical cord [HUC]). Within these groups, further differences in study population, country of origin, and biopsy definitions (especially whether or not mild grades without villous atrophy were included) were assessed systematically. Studies that reported using the ESPGAN criteria for the diagnosis of CD were categorized as including patients with some degree of villous atrophy. Other potential causes of heterogeneity such as the cut-offs used to define a positive test were assessed.

Two articles were identified that assessed the diagnostic value of various antibodies in children[64] and in mixed-age populations[40] with IgA deficiency. As well, one study enrolled biopsy-proven CD patients who were known to be EMA negative.[66] These studies were considered separately from the others. Studies of using antibodies in combination were also assessed separately.

Pooled statistical estimates (with 95% CIs) are provided for studies without clinical and statistical heterogeneity, and summary ROC curves for the studied antibodies are provided in Appendix E. Sensitivity analyses by study design did not show a significant difference except for the analysis of IgA-tTG-guinea pig (GP) in adults. Therefore, apart from studies of IgA-tTG-GP in adults, pooled estimates, when available, included data from both study designs.

Note: Appendixes and Evidence Tables are provided electronically at http://www.ahrq.gov/clinic/tp/celiactp.htm

AGA. The diagnostic characteristics of IgA were assessed in 35 studies and the diagnostic characteristics of IgG-AGA were assessed in 30 studies. Of the 35 IgA-AGA studies, 11 were conducted in an adult population,[30,33,45,50,54,61-63,71,77,80] 21 in a population of children,[26,27,29,31,34,36,38,42,43,50,52,56,59,60,64,67,68,83,85,87,88] and five in a mixed population.[27,37,40,74,75]

Of the 30 IgG-AGA studies, seven were conducted in an adult population,[30,33,54,62,63,71,80] 19 were conducted in population of children,[26,27,29,31,34,36,38,42,43,50,52,58,59,64,66,68,69,83,85] and five in a mixed population.[27,37,40,74,75] Some studies provided data for more than one age group.

Some studies only provided summary statistics without the raw two-by-two table results,[33,34,54,58,59,69] however, the raw data was calculated from the presented sensitivity and specificity, and from the group sizes.

One study[66] was conducted in CD patients who were known to be IgA-EMA negative, and was not included in the main analysis. In this study of children, the sensitivity for IgA-AGA was 22% and the sensitivity for IgG-AGA was 33%, whereas, the specificity for IgA-AGA was 67% and the specificity for IgG-AGA was 58%; these values are considerably lower than those reported in other studies. Another two studies were conducted in patients with IgA deficiency.[40,64] The first demonstrated a sensitivity of 0% using IgA-AGA, but a sensitivity and specificity of 100% using IgG-AGA,[40] whereas the second showed a sensitivity of 0% with IgA-AGA, but a sensitivity of 100% and a specificity of 80% using IgG-AGA.

Despite clinical subdivision of the identified studies, significant heterogeneity was identified for each of the pooled AGA subgroup results (Tables 2 to 7). Heterogeneity can be visualized graphically in the ROC curves (Figures 2 to 4) and suggests that the heterogeneity is in part related to a serological test cut-off threshold effect. As well, two studies included CD patients with less than a Marsh IIIa grade;[37,45] these studies had lower than average sensitivities (61% and 67% for IgA-AGA) then that reported in other studies. The remaining heterogeneity likely represents a combination of the effects of different test kits, inter-lab variability, and differences in the study groups. For example, within the child population, two of the outlier studies were conducted in Turkey,[26,85] although apparently using standard methodology. Therefore, overall pooled estimates do not represent true summary statistics in these situations.

IgA-AGA. Despite the apparent heterogeneity, one can make some broad statements regarding the diagnostic value of AGA antibodies. IgA-AGA appears to offer fair to good performance in children (Table 2; Figure 2).[26,27,29,31,34,36,38,42,43,50,52,58,59,64,66,68,69,83,85] Ten of the 19 studies demonstrated a sensitivity of IgA-AGA of greater than 80%, and six of the studies demonstrated a sensitivity of greater than 90%. However, nine studies demonstrated sensitivities of less than 80%. The specificity was greater than 80% in 15 of the 19 studies, and greater than 90% in 11 studies. Only four studies showed a specificity of less than 80%.

Ten studies assessed IgA-AGA in adults (Table 3; Figure 3).[30,33,54,62,63,71,80] Five of the ten studies demonstrated sensitivities greater than 80%, and three of the studies demonstrated sensitivities of greater than 90%. However, four studies demonstrated sensitivities of less than 65%. The specificity was greater than 80% in eight studies and greater than 90% in three. Five studies had specificities between 80% and 90%, and only two studies had specificities less than 80%.

Among the studies that assessed IgA-AGA in a mixed population of adults and children,[27,37,74,75] two demonstrated poor sensitivities of less than 70% but with specificities between 90% and 92%, one demonstrated a sensitivity of 85% and a specificity of 85%, and the last demonstrated a sensitivity of 91% and a specificity of 98% (Table 4; Figure 4).

Table 2: Included studies for IgA-AGA in children

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev
Picarelli, 2000; Italy	Case-control	ESPGAN	22.2*	66.7*	50*	36.3*	0.60*
Gaetano, 1997; Italy	Case-control	ESPGAN	92	68	85.2	80.9	0.67
Carroccio, 1993; Italy	Case-control	Biopsies confirmed at diagnosis, on GFD, and rechallenge (severity grade - not reported)	68	91.7	86.1	79.7	0.43
Hansson, 2000; Sweden	Case-control	ESPGAN	95.5	73.9	77.8	94.4	0.49
Berger, 1996; Switzerland	Case-control	ESPGAN revised with complete villous atrophy	76	67	74	59	0.55
Lerner, 1994; USA, Israel	Case-control	Criteria of Townley modified by Ingkaran	52	94	87	74	0.52
Bahia, 2001; Brazil	Relevant clinical population	Severe villous atrophy	95.5	95.6	91.3	97.9	0.31
Russo, 1999; Canada	Relevant clinical population	ESPGAN	83.3	84.5	64.5	93.8	0.25
Bode, 1993; Denmark	Relevant clinical population	ESPGAN	64	99	90	97	0.07
Poddar, 2002; India	Relevant clinical population	ESPGAN (villous atrophy and unequivocal response to GFD)	94	91.5	92	93.5	0.52
Ascher, 1996; Sweden	Relevant clinical population	ESPGAN	100	94.4	95.7	100	0.55
Lindberg, 1985; Sweden	Relevant clinical population	ESPGAN; Alexander grading	88	88			0.31
Altuntas, 1998; Turkey	Relevant clinical population	Subtotal or total villous atrophy, crypt hyperplasia, increased IEL	23	90	75	48	0.55
Artan, 1998; Turkey	Relevant clinical population	ESPGAN ;	58	51	42.4	66.7	0.38
Rich, 1990; USA	Relevant clinical population	Not recorded - state "severe" lesion	53	93	72.7	85.7	0.25
Gonczi, 1991; Australia	Relevant clinical population (184 children with suspected CD)	ESPGAN no details on biopsy findings	95	92.4	76	98.6	0.20
Wolters, 2002; Netherlands	Relevant clinical population (identified retrospectively)	Subtotal villous atrophy with crypt hyperplasia	83	86	81	81	0.51
Lindquist, 1993; Sweden	Relevant clinical population (suspected celiac)	ESPGAN; subtotal or partial villous atrophy	86.5	92.7	93.7	85	0.55
Chirdo, 1999; Argentina	Relevant clinical trial	Total or subtotal villous atrophy	75	87.1	84	80	0.47
Chartrand, 1997; Canada	Relevant clinical population	ESPGAN - with flat mucosal biopsy	80	92	67	96	0.17
Meini, 1996; Italy	Relevant clinical population	Partial villous atrophy or total villous atrophy	0	100	0	91.7	0.08
*30 IgA-EMA-negative patients suspected of CD; 9 of 18 CD patients IgA deficient							

Table 3: Included studies for IgA-AGA in adults

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev (%)
Sategana-Guidetti, 1995; Italy	Case-control	Roy-Choudhury criteria; partial or total villous atrophy	55	100	100	55.9	35.0
Dahele, 2001; Scotland	Case-control	Included 6 with IEL, rest partial villous atrophy or greater	61	86	88.5	42.7	43.6
Bode, 1994; Denmark	Relevant clinical population	Crypt hyperplasia, villous atrophy and increase inflammatory cells	46	98	75	92	25.7
Kaukinen, 2000; Finland	Relevant clinical population	Villous height to crypt ratio <2.0; IEL and HLA also tested	83	45	75	92	57.0
Maki, 1991; Finland	Relevant clinical population	Severe pathology with crypt hyperplasia to total villous atrophy; mild changes considered normal	30.8	87.2	22.2	91.3	14.8
McMillan, 1991; Ireland	Relevant clinical population	Revised ESPGAN	100	100	100	100	31.5
Bardella, 2001; Italy	Relevant clinical population	Marsh; no grade reported	95	89	76	98	33.3
Gonczi, 1991; Australia	Relevant clinical population (184 children with suspected CD)	ESPGAN no details on biopsy findings	92	88.2	85.2	93.8	45.8
Valdimarsson, 1996; Sweden	Relevant clinical population+ a few dypeptic controls	Alexander's classification; partial or subtotal villous atrophy	79	70	28	96	36.8
Vogelsang, 1995; Austria	Relevant study population	Modified ESPGAN; flat mucosa; crypt hyperplasia raised IELs	81.6	83	81.6	83	48.0

Table 4: Included studies for IgA-AGA in studies including both children and adults

Author, year; country	Study type	Biopsy criteria	Notes	Sens	Spec	PPV	NPV	Prev
Cataldo, 2000; Italy	Case-control	Original & revised criteria?	20 IgA-deficient CD vs healthy IgA-deficient non-CD	0	100	0	33.3	0.7
Sulkanen, 1998; Finland	Case-control	ESPGAN		84.5	81.6	75.2	89	0.4
Ascher, 1996; Sweden	Relevant clinical population	ESPGAN		90.9	98.5	98	92.7	0.5
Carroccio, 2002; Italy	Relevant clinical population	Marsh, broken down by criteria; CD was diagnosed as enlarged crypts and/or villous atrophy-with normalization on GFD		67	90	86	75	0.5
Tesei, 2003; Argentina	Relevant clinical population	Marsh II to IV - with confirmation		64	92	92	64	0.6

Figure 2: IgA-AGA in children with CD

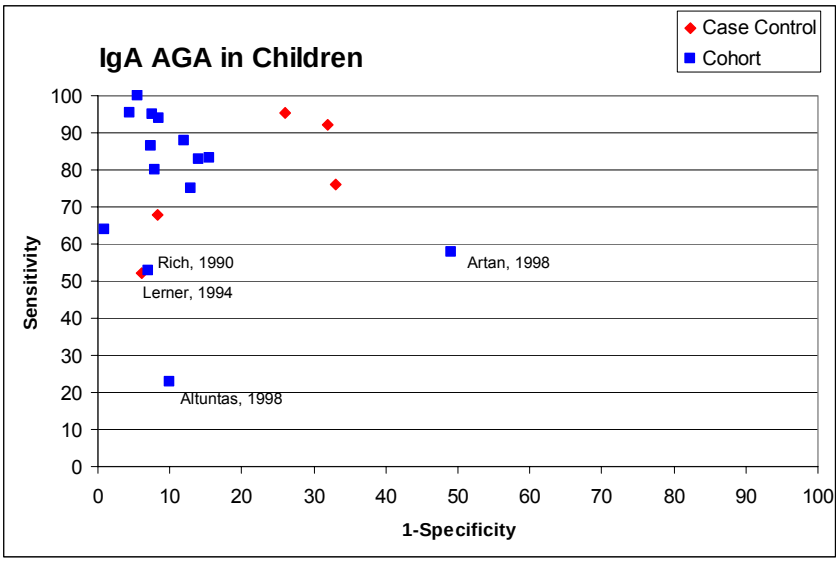


Figure 3: IgA-AGA in adults with CD

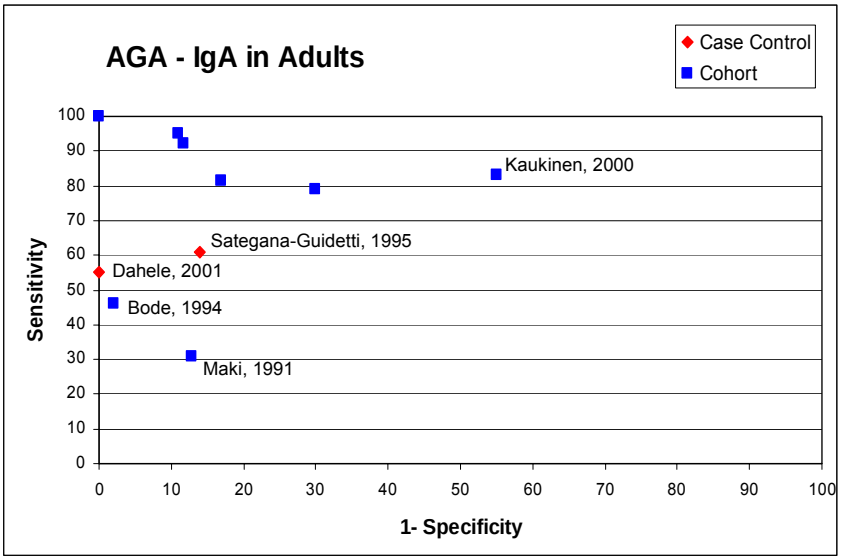
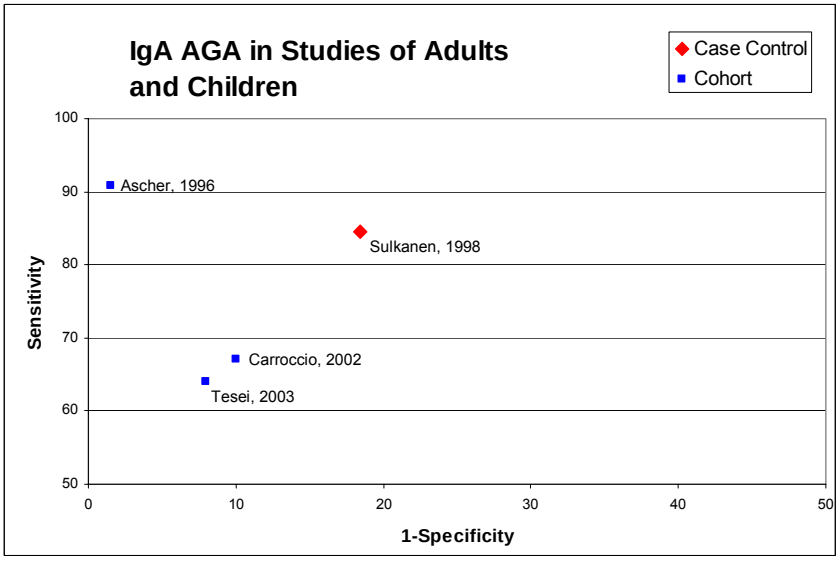


Figure 4: IgA-AGA in adults and children with CD



IgG-AGA. The seven studies of IgG-AGA in adults demonstrated considerably greater heterogeneity.[30,33,54,62,63,71,80] The sensitivity ranged from 17% to 100%, with little study grouping. However, there was less variation in the reported specificities. Five of the seven studies demonstrated specificities greater than 80%, whereas, the remaining two studies had specificities of greater than 70%. (Table 5; Figure 5)

In contrast, among the 17 analyzed studies (non-IgA deficient) of IgG-AGA conducted in children,[26,27,29,31,34,36,38,42,43,50,52,58,59,68,69,83,85] there seemed to be greater variability in the specificity than in the sensitivity (Table 6; Figure 6). Fifteen of the 17 studies demonstrated sensitivities that were greater than 80%, and six demonstrated sensitivities greater than 90%. Only two studies showed a sensitivity of less than 80%. In contrast, with regards to specificity, two groupings of studies become apparent. The first group consists of 11 studies, all of which had specificities greater than 79%, and except for one study, had sensitivities that were greater than 80%. In contrast, the second group of six studies all had specificities below 70%, and with the exception of one study, had sensitivities greater than 80%. (Tables and figures)

Four studies looked at IgG-AGA in a non-IgA-deficient mixed population of adults and children.[27,37,74,75] Two of these demonstrated sensitivities greater than 80%, one showed a sensitivity of 84%, whereas the second had a sensitivity of 96%. However, only the first study had specificity greater than 80%. In total, three of the four studies had specificities less than 80% (Table 7; Figure 7).

Table 5: Included studies for IgG-AGA in adults

Author, year, country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev (%)
Sategana-Guidetti, 1995; Italy	Case-control	Roy-Choudhury criteria; partial or total villous atrophy	78	80.7	87.6	67.6	56.7
Bode, 1994; Denmark	Relevant clinical population	Crypt hyperplasia, villous atrophy and increase inflammatory cells	62	97	73	94	34.8
Kaukinen, 2000; Finland	Relevant clinical population	Villous height to crypt ration <2.0; IEL and HLA also tested	17	86	14	93.5	15.1
Maki, 1991; Finland	Relevant clinical population	Severe pathology with crypt hyperplasia to total villous atrophy; mild changes considered normal	46.2	89	33.3	93.3	14.8
McMillan, 1991; Ireland	Relevant clinical population	Revised ESPGAN	57	85	64	81	28.1
Gonczi, 1991; Australia	Relevant clinical population (184 children with suspected CD)	ESPGAN no details on biopsy findings	100	69.7	69.4	100	61.0
Vogelsang, 1995; Austria	Relevant study population	Modified ESPGAN; flat mucosa; crypt hyperplasia raised IELs	73.5	73.6	72	75	49.0

Table 6: Included studies for IgG-AGA in children

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev
Picarelli, 2000; Italy	Case-control	ESPGAN	33.3	58.3	54.5	36.8	0.60
Gaetano, 1997; Italy	Case-control	ESPGAN	100	36	75.7	100	0.67
Carroccio, 1993; Italy	Case-control	Biopsies confirmed at diagnosis, on GFD, and rechallenge (severity grade – not recorded)	88.9	46.7	55.6	84.8	0.43
Hansson, 2000; Sweden	Case-control	ESPGAN	81.8	82.6	81.8	82.6	0.49
Berger, 1996; Switzerland	Case-control	ESPGAN revised with complete villous atrophy	69	59	68	53	0.55
Lerner, 1994; U.S.A, Israel	Case-control	Criteria of Townley modified by Ingkaran	88	92	88	92	0.52
Bahia, 2001; Brazil	Relevant clinical population	Severe villous atrophy	90.9	97.8	95.2	95.7	0.32
Russo, 1999; Canada	Relevant clinical population	ESPGAN	83.3	85.9	66.7	93.8	0.25
Bode, 1993; Denmark	Relevant clinical population	ESPGAN	71	99	100	98	0.07
Ascher, 1996; Sweden	Relevant clinical population	ESPGAN	100	66.7	75.6	100	0.55
Lindberg, 1985; Sweden	Relevant clinical population	ESPGAN; Alexander or Perea et al.	93	89	93.1	88.6	0.31
Altuntas, 1998; Turkey	Relevant clinical population	Subtotal or total villous atrophy, crypt hyperplasia, increased IEL	100	0	55	0	0.55
Artan, 1998; Turkey	Relevant clinical population	ESPGAN	83	59	55.6	85.2	0.38
Rich, 1990; USA	Relevant clinical population	Not reported - state "severe" lesion	100	58	44	100	0.25
Gonczi, 1991; Australia	Relevant clinical population (184 children with suspected CD)	ESPGAN no details on biopsy findings	100	92.4	76.9	100	0.20
Wolters, 2002; Netherlands	Relevant clinical population (identified retrospectively)	Subtotal villous atrophy with crypt hyperplasia	83	80	86	82	0.51
Chirdo, 1999; Argentina	Relevant clinical trial	Total or subtotal villous atrophy	85.7	80.6	80	86	0.47
Chartrand, 1997; Canada	Relevant clinical population	ESPGAN - with flat mucosal biopsy	83	79	45	96	0.17
Meini, 1996; Italy	Relevant clinical population	Partial villous atrophy or total villous atrophy	100	80	31.2	100	0.08

Table 7: Included studies for IgG-AGA in studies including both children and adults

Author, year; country	Study type	Biopsy criteria	Notes	Sens	Spec	PPV	NPV	Prev
Cataldo, 2000; Italy	Case-control	Original and revised criteria?	20 IgA-deficient CD vs healthy IgA-deficient non-CD	100	100	100	100	0.7
Sulkanen, 1998; Finland	Case-control	ESPGAN		69	73.4	63	78.3	0.4
Ascher, 1996; Sweden	Relevant clinical population	ESPGAN		96.4	69.2	72.6	95.7	0.5
Carroccio, 2002; Italy	Relevant clinical population	Marsh-broke down by criteria; CD was diagnosed as enlarged crypts and/or villous atrophy - with normalization on GFD		76	75	73.4	77.3	0.5
Tesei, 2003; Argentina	Relevant clinical population	Marsh II to IV - with confirmation		84	86	89	79	0.6

Figure 5: IgG-AGA in adults with CD

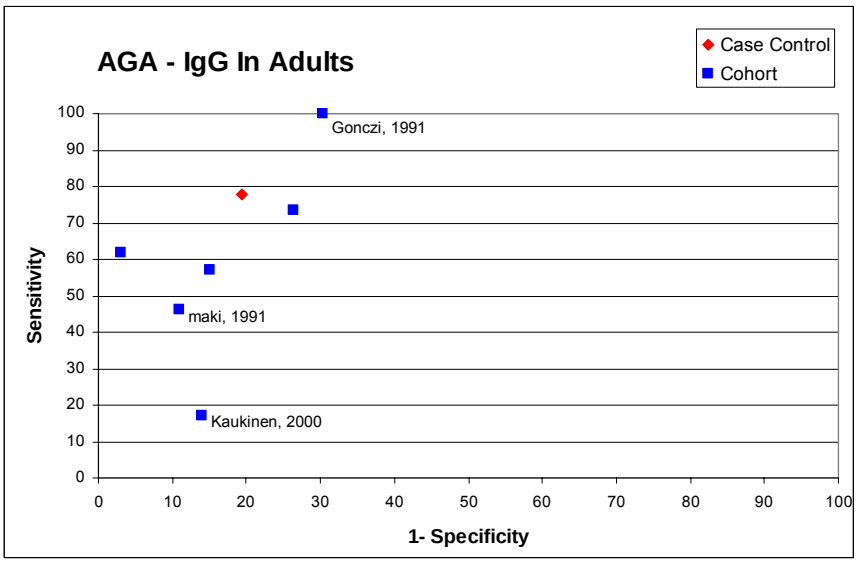


Figure 6: IgG-AGA in children with CD

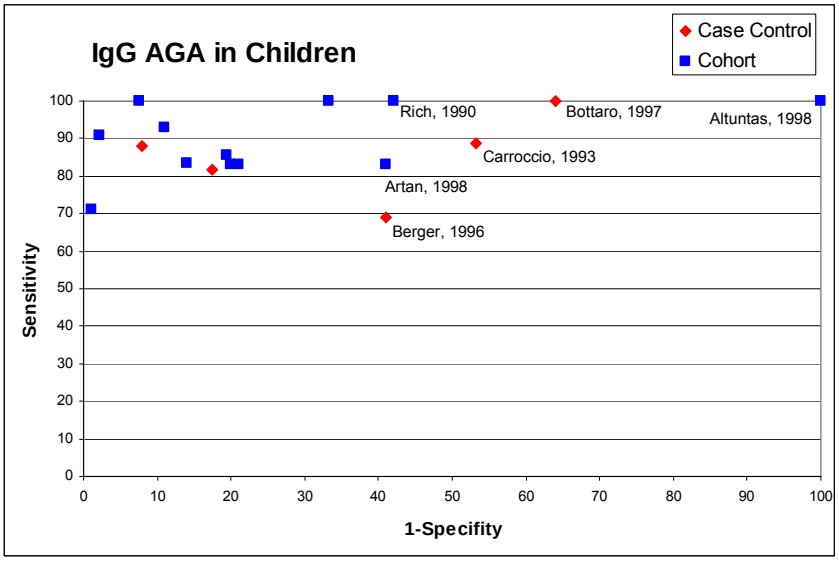
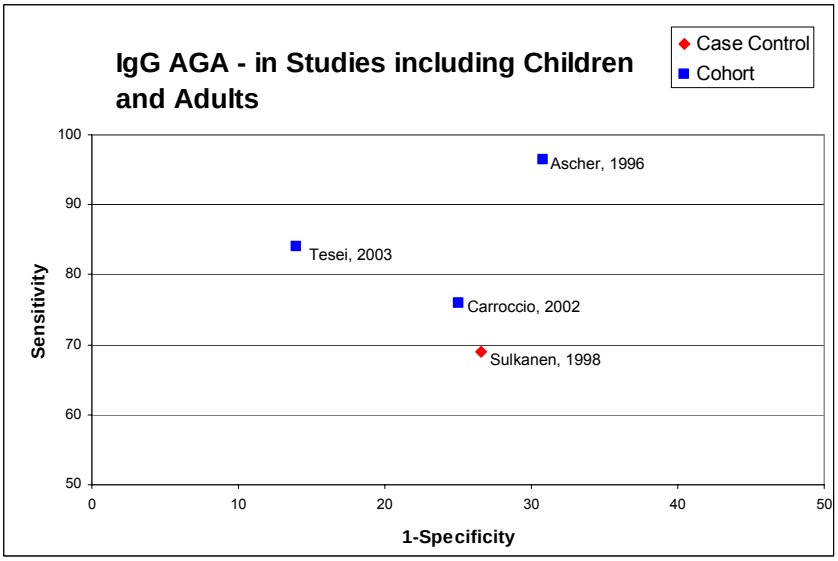


Figure 7: IgG-AGA in children and adults with CD



EMA

EMA—ME. The diagnostic characteristics of IgA-EMA-ME were assessed in 35 studies, and the diagnostic characteristics of IgG-EMA-ME were assessed in three studies. Of these included studies, 11 IgA-EMA-ME studies were conducted in adults,[30,32,39,51,57,63,71,77,78,80,81] 17 in children,[27,35,36,38,41,44,46,51,52,55,56,58,60,69,79,82,83] and five in a mixed population.[27,37,40,47,75] Some studies provided data for more than one age group. One study in children provided data on two different populations (including different control groups).[55] IgG-EMA-ME was assesed in one adult population,[63] one child population,[66] but not in any of the mixed-population studies.

One study was conducted in a population of known CD patients who had previously tested negative for EMA. In this study, the sensitivity and specificity of IgG EMA-ME were both 100%;[66] the performance of IgA-EMA was not reported. Another study that included CD patients with less than a Marsh IIIa grade,[37] demonstrated a sensitivity of 88%. Some studies only provided summary statistics without the raw two-by-two table results,[46,58,69] however, the raw data was abstracted based on the reported sensitivity and specificity, and the group sizes.

IgA-EMA-ME. Among the 11 studies of IgA-EMA-ME conducted in adults,[30,32,39,51,57,63,71,77,78,80,81] the specificity of the test was 100% in all except one, which showed a specificity of 97.2% (Table 8; Figure 8). The sensitivity of the test showed some slight variation among the studies. One outlier study demonstrated a sensitivity of only 74%;[77] however, the authors found that in the remaining five of 19 CD patients who tested negative for EMA, three were IgA deficient. If these patients were excluded, the sensitivity rose to 88%. The authors also go on to say that they seem to have a high proportion of IgA-deficient subjects in their referral base. The remaining ten studies showed sensitivities of 89% or greater. In fact, five studies showed a sensitivity of 100%, one a sensitivity of 99%, and another a sensitivity of 97%. In all, eight out of the 11 showed a sensitivity of 95% or greater, matching the very high specificity of this test. There was no statistical heterogeneity for this analysis. The pooled estimates for the sensitivity and specificity along with their 95% CI values were 97% (95% CI: 95.7-98.5) and 99.6% (95% CI: 98.8-99.9), respectively.

Among the 18 studies that assessed IgA-EMA-ME in children,[27,35,36,38,41,44,46,51,52,55,56,58,60,69,79,82,83] all but one outlier[69] were grouped together, and the sensitivities and specificities were both greater than 89% (Table 9; Figure 9). The outlier study demonstrated a sensitivity of only 74%, and also demonstrated low sensitivity for IgA-EMA-HU (see below).[69] The authors comment on the difficulties of interpretating immunofluorescence data as a likely explanation. Ten studies showed sensitivities greater than 95%, and except for one study with a sensitivity of 89%, the remaining seven studies had sensitivities between 90% and 95%. All these studies demonstrated specificities of 89% or greater, 16 had specificities greater than 90%, and 14 had specificities greater than 96%. There was no evidence of statistical heterogeneity in this analysis. The pooled sensitivity and specificity was 96.1% (95% CI: 94.4-97.3) and 97.4% (95% CI: 96.3-98.2), respectively.

Among the four studies in a mixed-age population that assessed IgA-EMA-ME,[27,37,47,75] all showed specificities of greater than 98% (Table 10; Figure 10). However, these studies showed some variation in the reported sensitivities. One study reported a very low sensitivity of 75%.[47] Two other studies showed a sensitivity of 86% and 88%, respectively, whereas the last showed a sensitivity of 98%.

Table 8: Included studies for IgA-EMA-ME in adults

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev (%)
Hallstrom, 1989; Finland	Case-control	Flat mucosa	90.6	100	100	88.9	51.8
Biagi, 2001; Italy	Case-control	Partial villous atrophy or greater	94.6	100	100	94.5	49.1
Ladinser, 1994; Italy	Case-control	Revised ESPGAN	100	100.0	100	100	21.1
Sategana-Guidetti, 1995; Italy	Case-control	Roy-Choudhury criteria; partial or total villous atrophy	100	100	100	100	63.7
Valentini, 1994; Italy	Case-control	Partial villous atrophy or greater	99	100	100	96.7	76.2
Volta, 1995; Italy	Case-control	Roy-Choudhury criteria	95	100	100	97.1	35.6
Carroccio, 2002; Italy	Relevant clinical population	Ferguson and Murray; partial or total villous atrophy	100	100	100	100	11.6
McMillan, 1991; Ireland	Relevant clinical population	Revised ESPGAN	89.2	100	100	95.3	28.1
Bardella, 2001; Italy	Relevant clinical population	Marsh	100	97.2	93	100	28.7
Valdimarsson, 1996; Sweden	Relevant clinical population+ a few dypeptic controls	Alexander's classification; partial or subtotal villous atrophy	74	100	100	96	9.7
Vogelsang, 1995; Austria	Relevant study population	Modified ESPGAN; flat mucosa; crypt hyperplasia raised IELs	100	100	100	100	48.0

Table 9: Included studies for IgA-EMA-ME in children

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev
Chirdo, 2000; Argentina	Case-control	ESPGAN	92.4	100	100	85.2	0.7
Kolho, 1997; Finland	Case-control	Revised ESPGAN	95	100	100	97	0.3
Kolho, 1997; Finland	Case-control	Revised ESPGAN	100	100	100	100	0.5
Whelan, 1996; Ireland	Case-control	Subtotal villous atrophy	100	100	100	100	0.4
Bonamico, 2001; Italy	Case-control	ESPGAN	95.1	98.2	90	44.3	0.5
Gaetano, 1997; Italy	Case-control	ESPGAN	96	96	97.9	92.3	0.7
Carroccio, 1993; Italy	Case-control	Biopsies confirmed at diagnosis, on GFD, and rechallenge (severity grade - not reported)	100	96.7	95.7	100	0.4
Di Leo, 2003; Italy	Case-control	ESPGAN	100	96.5	93.5	100	0.4
Vitoria, 2001; Italy	Case-control	Subtotal villous atrophy	100	100	100	100	0.6
Hansson, 2000; Sweden	Case-control	ESPGAN	95.5	100	100	95.8	0.5
Lerner, 1994; USA, Israel	Case-control	Criteria of Townley modified by Ingkaran	97	98	97	98	0.5
Hallstrom, 1989; Finland	Case-control	Flat mucosa	100	100	100	100	0.4
Chan, 2001; Canada	Relevant clinical population	Villous atrophy, crypt hyperplasia, increased lymphocytes	89	97	80	98	0.1
Russo, 1999; Canada	Relevant clinical population	ESPGAN	75	88.7	69.2	91.3	0.3
Ascher, 1996; Sweden	Relevant clinical population	ESPGAN	95.4	100	100	94.7	0.6
Wolters, 2002; Netherlands	Relevant clinical population (identified retrospectively)	Subtotal villous atrophy with crypt hyperplasia	92	90	90.5	92	0.5
Lindquist,1993; Sweden	Relevant clinical population (suspected CD)	ESPGAN; subtotal or partial villous atrophy	98.1	92.7	94.4	97.5	0.6
Kumar, 1989; USA, Israel	Relevant clinical population and control cases	ESPGAN + Townley	96.0	89.0	87.0	96.7	0.2

Table 10: Included studies for IgA-EMA-ME in studies including both children and adults

Author, year; country	Study type	Biopsy criteria	Notes	Sens	Spec	PPV	NPV	Prev
Cataldo, 2000; Italy	Case-control	Original & revised criteria?	20 IgA-deficient CD vs healthy IgA-deficient non-CD	0	100	0	33.3	0.7
Dickey, 2001; Northern Ireland	Case-control	Villous atrophy		75.3	98.3	98.2	76	0.6
Ascher, 1996; Sweden	Relevant clinical population	ESPGAN		98.2	100	100	98.5	0.5
Carroccio 2002; Italy	Relevant clinical population	Marsh - broke down by criteria; CD was diagnosed as enlarged crypts and/or villous atrophy - with normalization on a GFD		88	99	98.7	90	0.5
Tesei, 2003; Argentina	Relevant clinical population	Marsh II to IV - with confirmation		86	100	100	83	0.6

Figure 8: IgA-EMA-ME in adults with CD

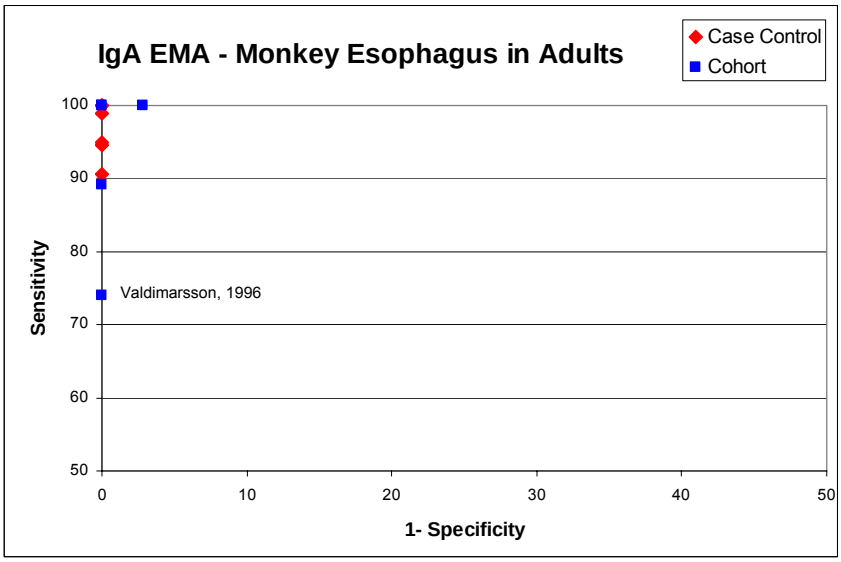


Figure 9: IgA-EMA-ME in children with CD

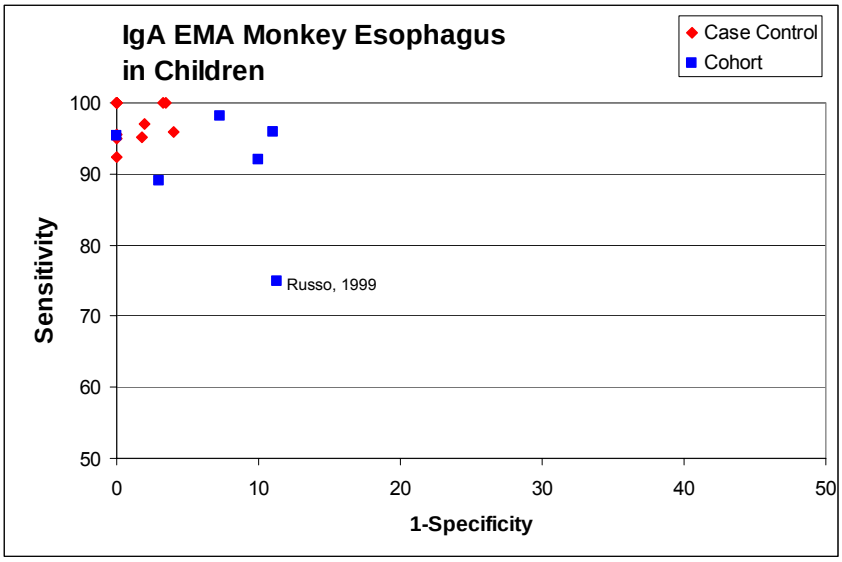
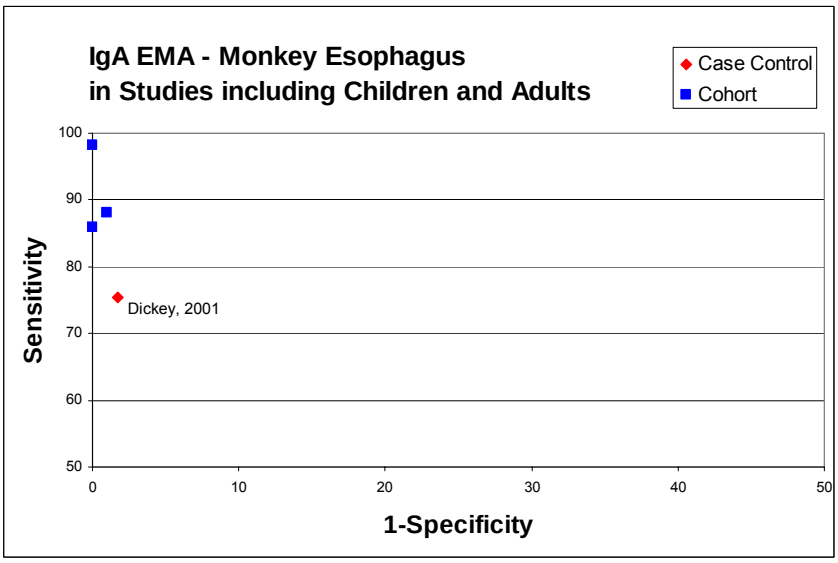


Figure 10: IgA-EMA-ME in adults and children with CD



IgG-EMA-ME. Only two studies meeting our inclusion criteria assessed IgG-EMA-ME, one in adults (Table 11),[63] and one in children (Table 12).[66] In the single adult study,[63] the sensitivity of the test was found to be 39%, whereas, the specificity was 98%. In a case-control study design, Picarelli et al. studied 30 IgA-EMA-negative children suspected of having CD.[66] Of these 30 children, 18 were subsequently found to have CD by duodenal biopsy and nine of the 18 were

found to be IgA deficient. In this highly selected population, the reported sensitivity and specificity of IgG-EMA-ME were both 100%.

Table 11: Included studies for IgG-EMA-ME in adults

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev (%)
McMillan, 1991; Ireland	Relevant clinical population	Revised ESPGAN	39	98.3	92	78	13.5

Table 12: Included studies for IgG-EMA-ME in children

Author, year; country	Study type	Biopsy criteria	Notes	Sens	Spec	PPV	NPV	Prev
Picarelli, 2000; Italy	Case-control	ESPGAN	30 IgA-EMA neg. pts suspected of CD; 9/18 CD patients IgA deficient	100	100	100	100	0.1

EMA—HU. IgA-EMA-HU was assessed in 13 studies. Six of these studies were conducted in adults,[45,49,54,57,61,70,89] five in children,[36,53,55,69,70] and two in a mixed population.[72,74] One study provided summary statistics without the raw two-by-two table results,[69] however the raw data was calculated from the reported sensitivity and specificity and the group numbers. One study provided data on two different populations (including different control groups).[55]

IgG-EMA-HU was not assessed in any of the studies meeting our inclusion criteria.

Two studies included CD patients (both adult and children) with less than a Marsh IIIa grade, and reported IgA-EMA-HU sensitivities of 87% and 100%.[45]
[70]

IgA-EMA-HU. Six studies in adults assessed IgA-EMA-HU (Table 13; Figure 11).[45,49,54,57,61,70,89] In all six, the specificity was reported to be 100%. There was, however, variability in the reported sensitivities, which ranged from 87% to 100%. Three studies demonstrated sensitivities between 87% and 89%, two between 90% and 95% and one showing a sensitivity of 100%. There was no observed statistical heterogeneity for this analysis. The pooled sensitivity and specificity was found to be 90.2% (95% CI: 85.9-93.4) and 100% (95% CI: 99.1-100), respectively.

Five studies with six separate child populations assessed IgA-EMA-HU (Table 14; Figure 12).[36,53,55,69,70] Four of the six studies were grouped together and revealed sensitivities between 94% and 100%, and specificities of 100%. Of the two outliers,[90] one showed a sensitivity of 100% and a specificity of 77%. The other study,[69] was an outlier in other analyses, and demonstrated a sensitivity of 46% and a specificity of 96%. The authors comment on difficulties of interpretation of the immunofluorescence as a likely explanation. After accounting for this study, there was no statistical heterogeneity documented for sensitivity. The pooled sensitivity for this analysis was 96.9% (95% CI: 93.5-98.6). A pooled specificity for this analysis was not calculated, but is likely close to 100% given that four of the five grouped studies demonstrated a specificity of 100%.

Two studies assessed IgA-EMA-HU in a mixed-age population (Table 15; Figure 13).[72,74] In both these studies, the specificity was 100% (95% CI: 97.5-100) and the sensitivity 93% (95% CI: 88.1-95.4).

Table 13: Included studies for IgA-EMA-HU in adults

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev (%)
Gillbert, 2000; Canada	Case-control	Mild, moderate, severe villous atrophy	100	100	100	100	33.3
Ladinser, 1994; Italy	Case-control	Revised ESPGAN	90	100	100	98	18.9
Salmaso, 2001; Italy	Case-control	Grades I-IV Marsh with response to a GFD	87	100	100	95.1	24.7
Volta, 1995; Italy	Case-control	Roy-Choudhury criteria	95	100	100	97.1	35.6
Dahele, 2001; Scotland	Case-control	Included 6 with IEL, rest partial villous atrophy or greater	87	100	100	81.3	55.3
Kaukinen, 2000; Finland	Relevant clinical population	Villous height to crypt ration <2.0; IEL and HLA also tested	88.9	100	100	98.9	7.6

Table 14: Included studies for IgA-EMA-HU in children

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev
Kolho, 1997; Finland	Case-control	Revised ESPGAN	95	100	100	97	0.3
Kolho, 1997; Finland	Case-control	Revised ESPGAN	100	100	100	100	0.5
Gaetano, 1997; Italy	Case-control	ESPGAN	94	100	100	89.2	0.7
Salmaso, 2001; Italy	Case-control	Grades I-IV Marsh with response to GFD	100	100	100	100	0.6
Russo, 1999; Canada	Relevant clinical population	ESPGAN	45.8	95.8	78.6	84	0.3
Iltanen, 1999 Finland	Relevant clinical population	ESPGAN - CD confirmed at follow-up	100	77.1	60.1	100	0.3

Table 15: Included studies for IgA-EMA-HU in studies including both children and adults

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev
Sblaterro, 2000; Italy	Case-control	ESPGAN	93	100	100	80	0.8
Sulkanen, 1998; Finland	Case-control	ESPGAN	92.6	99.5	99.2	94.9	0.4

Figure 11: IgA-EMA-HU in adults with CD

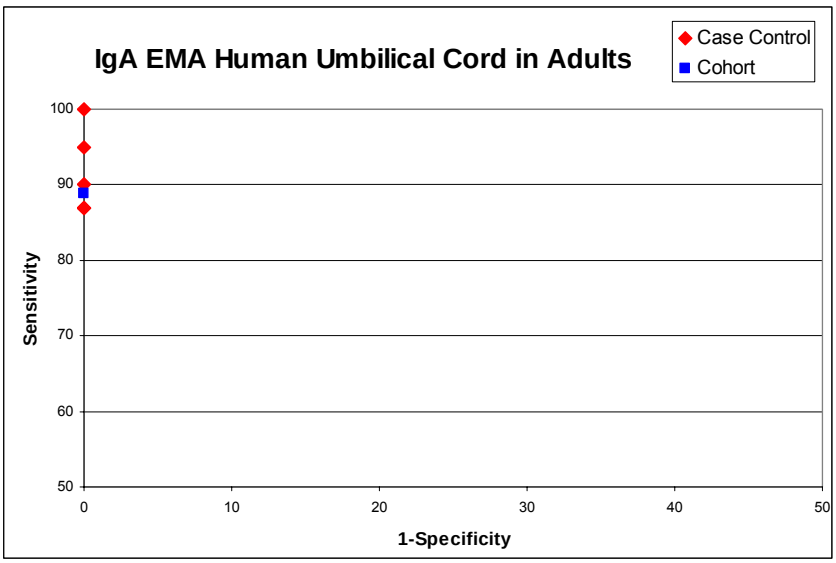


Figure 12: IgA-EMA-HU in children with CD

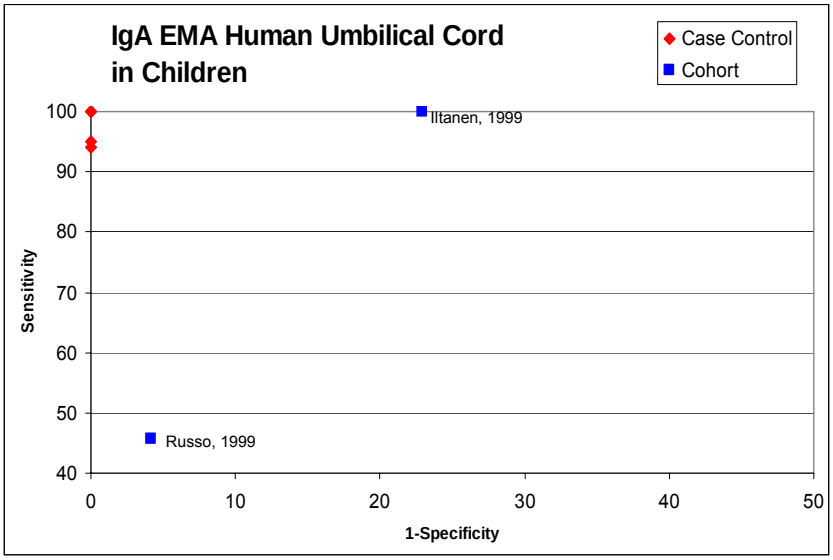
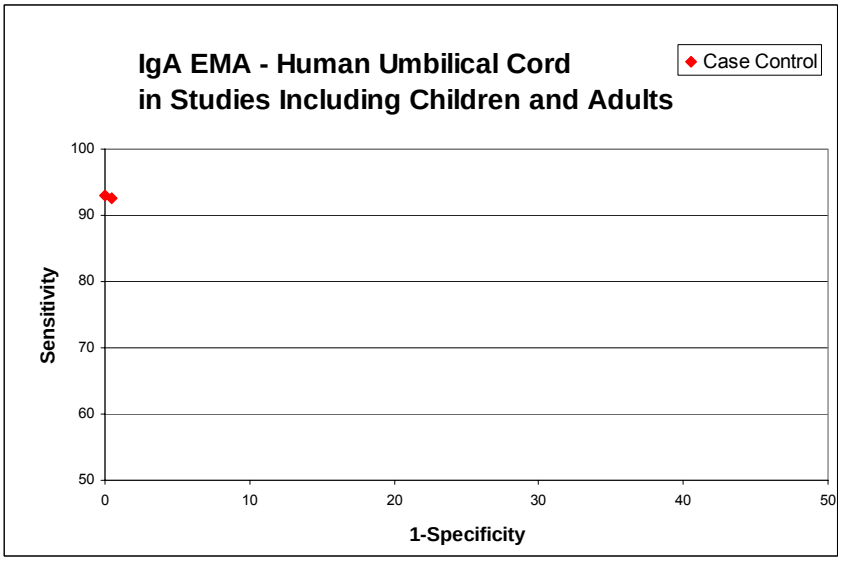


Figure 13: IgA-EMA-HU in adults and children with CD



tTG antibodies

tTG—GP liver. The diagnostic characteristics of IgA-tTG-GP were assessed by ELISA in nine studies**,** and the diagnostic characteristics IgG-tTG-GP assessed by ELISA in three studies. Of the IgA-tTG–GP studies, five were conducted in adults,[30,32,39,45,70] five in children,[35,41,52,70,83] and four in a mixed population.[47,72,74,76] One study provided separate data for more than one age group.[70]

Of the IgG-tTG-GP studies that met the inclusion criteria, none were in adults or children, although two studies were in a mixed population.[72,76]

Two studies included CD patients with less than a Marsh IIIa grade.[45,70] These studies demonstrated sensitivities of 81% and 95% for IgA-tTG-GP.

IgA-tTG-GP. In the analysis of IgA-tTG-GP in adults, five studies grouped themselves by study design.[30,32,39,45,70] The two cohort studies (relevant clinical population)[30,39] both showed sensitivities of 100%, and specificities of 92% and 98%, respectively. On the other hand, the three case-control studies[32,45,70] demonstrated high specificities (97% to 98%), but sensitivities of only 81% to 88% (Table 16; Figure 14). This analysis did not show statistical heterogeneity, but the differences by study design were striking, so a pooled estimate for sensitivity was not performed. The pooled specificity was 95.3% (95% CI: 92.5-98.1).

The analysis of IgA-tTG-GP in children showed very little variability in either the sensitivity, or specificity (Table 17; Figure 15). Among these five studies,[35,41,52,70,83] the sensitivities ranged from 89% to 96%. The specificities were all greater than 92%, with three studies showing specificities greater than 96%,[41,52,83] and two studies having a sensitivity of 100%.[35,70] The pooled estimates of the sensitivity and specificity were 93.1% (95% CI: 88.8-95.9) and 96.3% (95% CI: 93.1-98.0), respectively (Table 17; Figure 15)

Among the studies of mixed-age groups,[47,72,74,76] there was one outlier study with a sensitivity of only 84% but a specificity of 100% (Table 18).[72] The specificities of the remaining studies were all greater than 94% (Table 18; Figure 16), and the sensitivities were between 92% and 95%. Heterogeneity was detected in the estimates of sensitivity, but not for specificity. The pooled specificity was 95.4% (95% CI: 92.7-97.2).

Table 16: Included studies for IgA-tTG-GP in adults

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev (%)
Biagi, 2001; Italy	Case-control	Partial villous atrophy or greater	87.5	98.1	98	87.1	46.3
Salmaso, 2001; Italy	Case-control	Grades I-IV Marsh with response to a GFD	87	97	90.9	94.9	27.2
Dahele, 2001; Scotland	Case-control	Included 6 with IEL, rest partial villous atrophy or greater	81	97	97.9	74.1	52.5
Carroccio, 2002; Italy	Relevant clinical population	Ferguson and Murray; partial or total villous atrophy	100	92	60	100	18.8
Bardella, 2001; Italy	Relevant clinical population	Marsh	100	98.2	83.3	100	10.0

Table 17: Included studies for IgA-tTG-GP in children

Author, country; year	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev
Bonamico, 2001; Italy	Case-control	ESPGAN	90.3	100	100	30.3	0.5
Salmaso, 2001; Italy	Case-control	Grades I-IV Marsh with response to a GFD	95	100	100	94.1	0.6
Hansson, 2000; Sweden	Case-control	ESPGAN	90.9	95.7	95.2	91.7	0.5
Chan, 2001; Canada	Relevant clinical population	Villous atrophy, crypt hyperplasia, increase lymphocytes	89	94	67	98	0.1
Wolters, 2002; Netherlands	Relevant clinical population (identified retrospectively)	Subtotal villous atrophy with crypt hyperplasia	96	92	92.6	95.7	0.5

Table 18: Included studies for IgA-tTG-GP in studies including both adults and children

Author, year, country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev
Dickey, 2001; Northern Ireland	Case-control	Villous atrophy	93.2	96.6	97.1	91.8	0.6
Sblaterro, 2000; Italy	Case-control	ESPGAN	84	100	100	62.5	0.8
Sulkanen, 1998; Finland	Case-control	ESPGAN	95	93.7	90.8	96.5	0.4
Troncone, 1999; Italy	Relevant clinical population	ESPGAN	91.7	98	98	94	0.4

Figure 14: IgA-tTG-GP in adults with CD

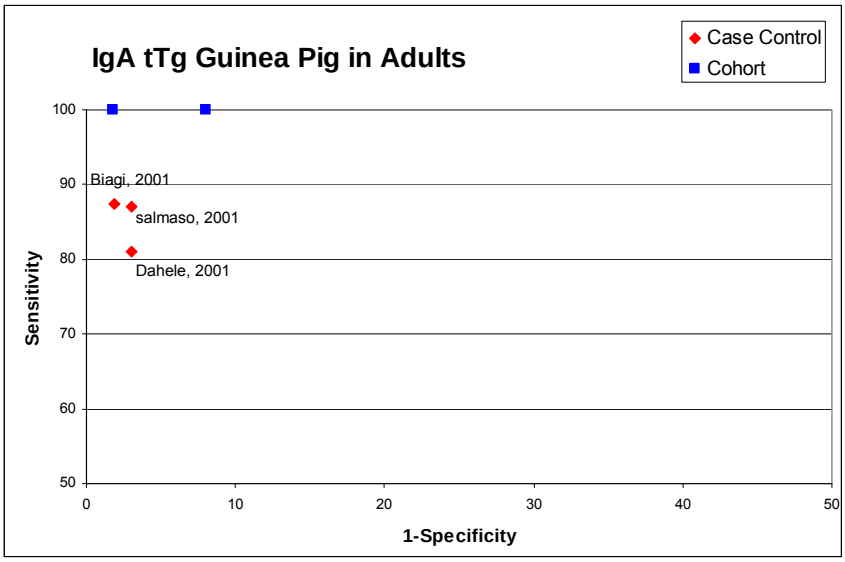


Figure 15: IgA-tTG-GP in children with CD

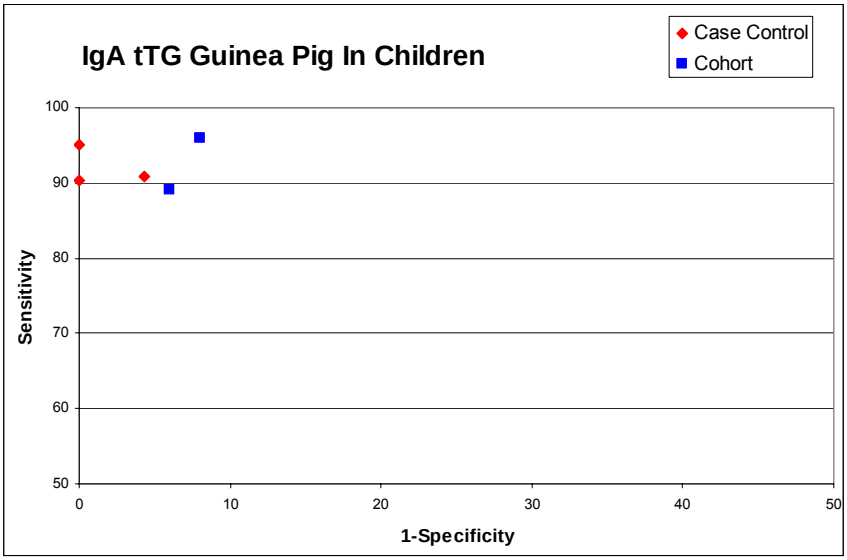
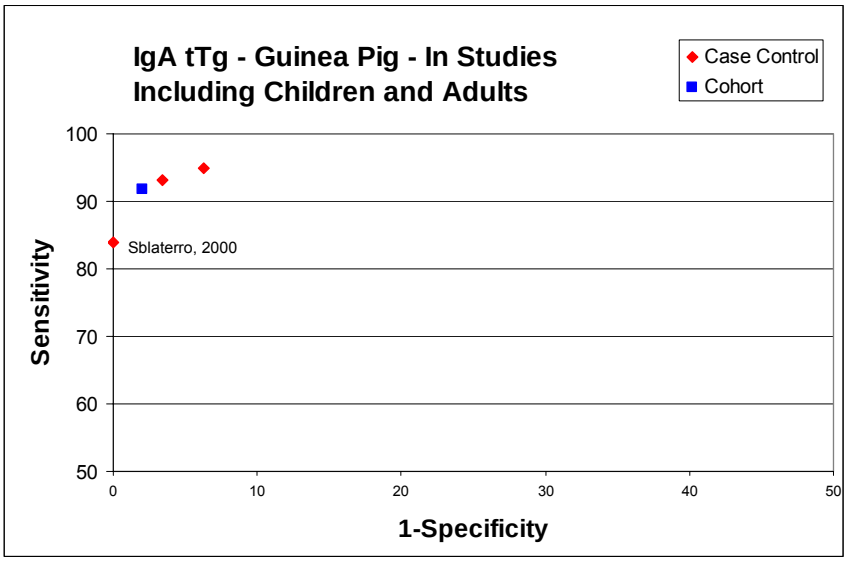


Figure 16: IgA-tTG-GP in adults and children with CD

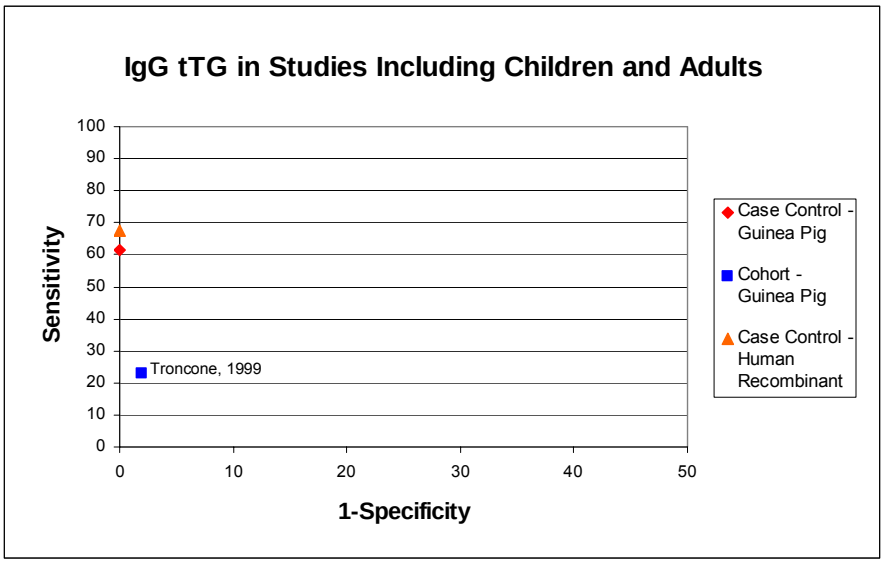


IgG-tTG-GP. Two studies in a mixed-age population assessed IgG-tTG- GP (Table 19; Figure 17).[72,76] The specificities in both studies were greater than 98%, but the sensitivities were 23% and 62%, respectively.

Table 19: Included studies for IgG-tTG-GP in studies including both children and adults

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev
Sblaterro, 2000; Italy	Case-control	ESPGAN	61.5	100	100	44.4	0.8
Troncone, 1999; Italy	Relevant clinical population	ESPGAN	23	98	92	63	0.4

Figure 17: IgG-tTG-GP in adults with CD



tTG – human recombinant (HR)

IgG-tTG-HR. The diagnostic characteristics of IgA-tTG-HR were assessed by ELISA in ten studies, and the diagnostic characteristics IgG-tTG-HR were assessed by ELISA in two studies. Of the IgA-tTG-HR studies, three were conducted in adults,[39,49,54] three in children,[52,79,83] and three in a mixed population.[40,72,75]

Of the IgG-tTG-HR studies, two were conducted in a mixed population (Table 20),[40,72] but none were conducted in adults or children. One study was conducted in IgA-deficient patients and is described below.[40]

Two studies included CD patients with less than a Marsh IIIa grade.[45,70] These studies demonstrated sensitivities of 81% and 95% for IgA-tTG-GP.

One study was conducted in a mixed-age population of patients with known IgA deficiency.[40] In this study, the sensitivity of IgA-tTG–HR was 0%, wheras, the sensitivities and specificities of IgG-tTG-HR were 100% and 80%, respectively.

Table 20: Included studies for IgG-tTG-HR in studies including both children and adults

Author, year; country	Study type	Biopsy criteria	Notes	Sens	Spec	PPV	NPV	Prev
Cataldo, 2000; Italy	Case-control	Original & revised criteria?	20 IgA-deficient CD vs healthy IgA-deficient non-CD	100	80	90.1	100	0.7
Sblaterro, 2000; Italy	Case-control	ESPGAN		67.6	100	100	48.7	0.8

IgA-tTG-HR. Three studies assessed IgA-tTG-HR in an adult population (Table 21; Figure 18).[39,49,54] There was very little variability in the reported values for the sensitivities and specificities. The sensitivities were 100% in two studies, and 95% in the other. The specificities were 100% in two studies, and 97% in another. The pooled estimates of the sensitivity and specificity were 98.1% (95% CI: 90.1%-99.7%) and 98.0% (95% CI: 95.8-99.1), respectively.

Among the three studies in children (Table 22; Figure 19),[52,79,83] the sensitivities were 96% in two studies and 95% in one. The specificities were 100% in two studies, and 96% in one. The pooled estimates of the sensitivity and specificity were 95.7% (95% CI: 90.3-98.1) and 99.0% (95% CI: 94.6-99.8), respectively.

Only two studies assessed the IgA-tTG-HR in a mixed-age population without IgA deficiency (Table 23; Figure 20).[72,75] The sensitivities and specificities were 92% and 100%, respectively, for the first study, and 91% and 96%, respectively, for the second. The pooled estimates of the sensitivity and specificity were 90.2% (95% CI: 86.4-93.0) and 95.4% (95% CI: 91.5- 97.6), respectively.

Overall, these studies demonstrated a specificity of close to 100% and sensitivity in the range of 90% to 96%.

Table 21: Included studies for IgA-tTG-HR in adults

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev (%)
Carroccio, 2002; Italy	Relevant clinical population	Ferguson and Murray; partial or total villous atrophy	100	97	80	100	14.5
Gillbert, 2000; Italy	Case-control	Mild, moderate, severe villous atrophy	95.2	100	95.2	100	31.7
Kaukinen, 2000; Finland	Relevant clinical population	Villous height to crypt ration <2.0; IEL and HLA also tested	100	100	100	100	8.7

Table 22: Included studies for IgA-tTG-HR in children

Author, year; country	Study type	Biopsy criteria	Sens	Spec	PPV	NPV	Prev
Vitoria, 2001; Italy	Case-control	Subtotal villous atrophy	95	100	100	93	0.6
Hansson, 2000; Sweden	Case-control	ESPGAN	95.5	95.7	95.5	95.7	0.5
Wolters, 2002; Netherlands	Relevant clinical population (identified retrospectively)	Subtotal villous atrophy with crypt hyperplasia	96	100	100	96	0.5

Table 23: Included studies for IgA-tTG-HR in studies including both children and adults

Author, year; country	Study type	Biopsy criteria	Notes	Sens	Spec	PPV	NPV	Prev
Cataldo, 2000; Italy	Case-control	Original & revised criteria?	20 IgA deficient CD vs healthy IgA-deficient non-CD	0	100	0	33.3	0.7
Sblaterro, 2000; Italy	Case-control	ESPGAN		91.5	100	100	76.9	0.8
Tesei, 2003; Argentina	Relevant clinical population	Marsh II to IV - with confirmation		91	96	97	87	0.6

Figure 18: IgA-tTG-HR in adults with CD

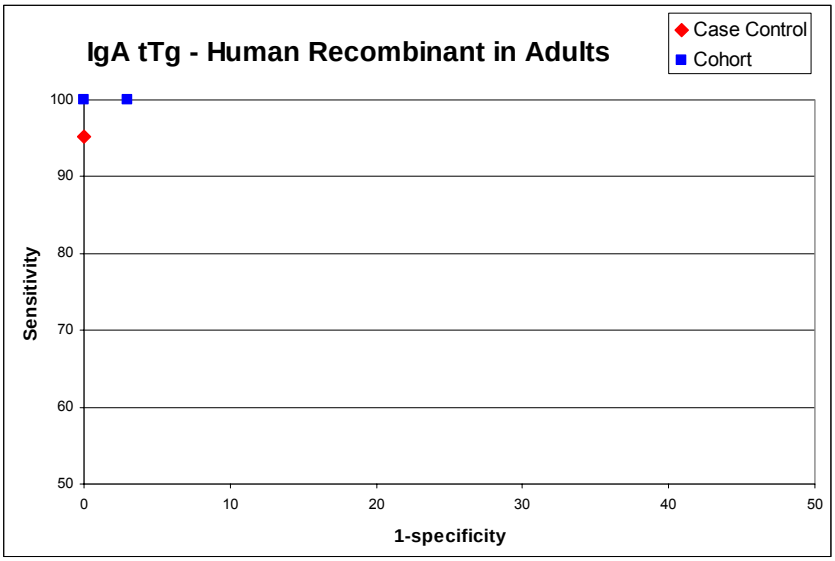


Figure 19: IgA-tTG-HR in children with CD

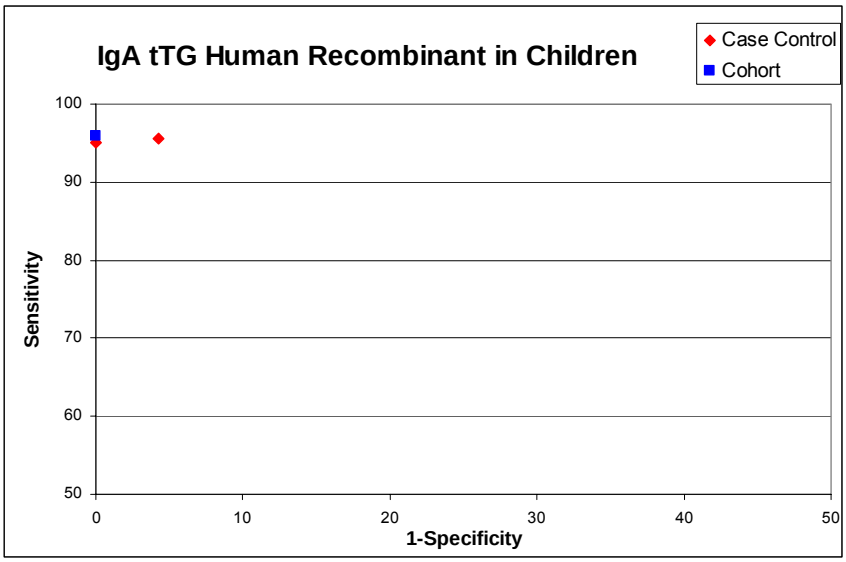
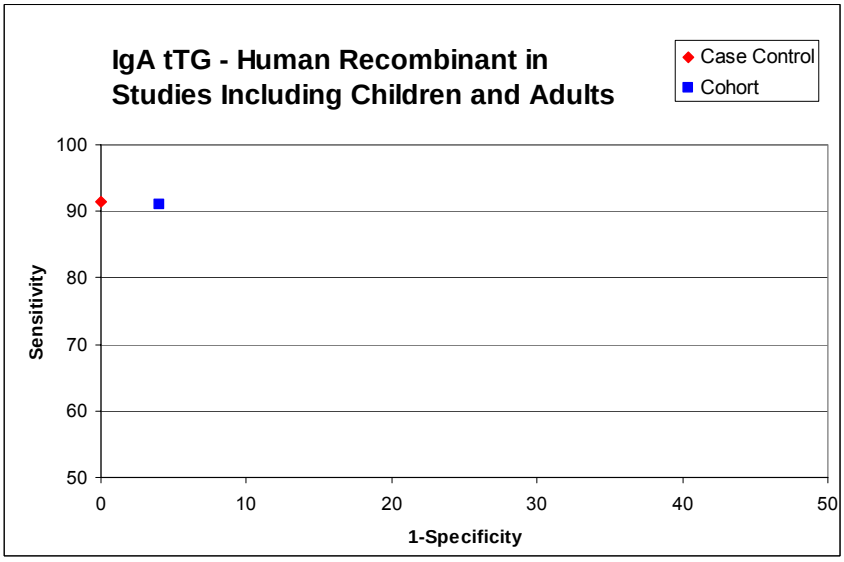


Figure 20: IgA-tTG-HR in adults and children with CD



IgG-tTG-HR, IgA deficient. Only one study of IgG-tTG-HR, conducted in an IgA-deficient population, was identified.[72] In this study, the sensitivity and specificity of IgG-tTG-HR was 68% and 100%, respectively.

Mixed-antibody combinations. Several studies were identified that tested different antibodies in combination. Six studies in children assessed the use of IgA- and IgG-AGA (Table 24).[34,42,48,50,59,85] When either of these tests were positive, the resulting sensitivities ranged from 83% to 100%, and the specificities ranged from 71% to 99%. One study, that apparently used similar methodologies, had the lowest sensitivity (83%) and specificity (36%) of the group.[85] When the same authors tested the antibodies under the requirement of both tests being concordant, the sensitivity fell, as would be expected, to 50%, and the specificity rose to 67%.[85] Three adult studies were identified that used IgA- and IgG-AGA in an either/or protocol (Table 24).[33,50,78] As was observed in the studies of children, significant between-study differences existed, making pooled estimates inappropriate. Nonetheless, in these studies the sensitivity ranged from 77% to 100%, while the specificity ranged from 90% to 97%.

One study in a mixed-age population assessed the use of a combination of IgA- and IgG-tTG-HR antibodies (Table 25).[72] In this study, the sensitivity when either test was positive was 98.5%, while the specificity remained high at 100%. Another study in children assessed the combination of IgA-AGA and IgA-EMA-HU when either test was positive, and found a sensitivity of 100% and a specificity of 73% (Table 26).[69] This same study assessed the same antibodies under the situation where both tests needed to be concordant. In this circumstance, the sensitivity remained 100% and the specificity rose to 93%.

In general, combining tests when either test is positive tended to improve sensitivity at the cost of specificity, while a requirement for the tests to be concordant tended to improve specificity.

Table 24: Included studies for combination IgA and IgG AGA, when either test is positive

Author, year; country	Study type	Biopsy criteria	Notes	Sens	Spec	PPV	NPV	Prev
Valentini, 1994; Italy	Case-control	Partial villous atrophy or greater	Adults	92	90	96.8	77.1	0.76
Bode, 1994; Denmark	Relevant clinical population	Crypt hyperplasia, villous atrophy and increase inflammatory cells	Adults	77	95	71	97	0.41
Gonczi, 1991; Australia	Relevant clinical population (184 children with suspected celiac)	ESPGAN no details on biopsy findings	Adults	100	97.1	96.2	100	0.44
Bode, 1993; Denmark	Relevant clinical population	ESPGAN	Children	86	99	92	99	0.1
Falth-Magnusson, 1994; Sweden	Relevant clinical population	ESPGAN + Alexander grading IV, grade III to IV challenge	Children	88.5	93.7	88.8	93.5	0.4
Lindberg, 1985; Sweden	Relevant clinical population	ESPGAN, Alexander grading	Children	97	83	41.8	98.2	0.3
Artan, 1998; Turkey	Relevant clinical population	ESPGAN	Children: IgA AGA or IgG AGA	83	36	44	77.8	0.3
Gonczi, 1991; Australia	Relevant clinical population (184 children with suspected CD)	ESPGAN no details on biopsy findings	Children	100	98.7	95.2	98.7	0.2
Chartrand, 1997; Canada	Relevant clinical population	ESPGAN – with flat mucosal biopsy	Children	93	71	43	98	0.2

Table 25: Included studies for combination IgA and IgG tTG-HR, when either test is positive

Author, year; country	Study type	Biopsy criteria	Notes	Sens	Spec	PPV	NPV	Prev
Sblaterro, 2000; Italy	Case-control	ESPGAN	Adults and children	98.5	100	100	95.2	0.8

Table 26: Included studies for combination IgA-AGA and IgG-EMA-HU, when either test is positive

Author, year; country	Study type	Biopsy criteria	Notes	Sens	Spec	PPV	NPV	Prev
Russo, 1999; Canada	Relevant clinical population	ESPGAN	Children	100	73	57	82	0.3

Prevalence of CD and the positive predictive value (PPV) and negative predictive value (NPV) of serology. The prevalence of CD in the tested populations is presented in Tables 2 to 26 for the individual studies, and in Table 27 for the pooled estimate for the analysis groups.

The minimum prevalence of CD in individual study populations was greater than 25% in most of the studied analysis groups (i.e., IgA-AGA, IgG-AGA, etc), except for ten analysis groups where the minimum prevalence was between 9% and 12%. In all the analysis groups, the maximum prevalence ranged from 30% to as high as 70%. The pooled prevalence for the analysis groups was predominantly between 30% and 45%.

In assessing the IgA-EMA and IgA-tTG analysis groups, the pooled prevalence ranged from 33% to 46% except for the analysis of IgA-tTG-HR in adults, which showed a pooled prevalence of 16%. Figure 21 is a plot of the individual study prevalence versus the study's PPV, and suggests that below a CD prevalence of about 35% to 40%, the PPV of these IgA-based tests tends to drop from about 90% to 100%, to about 80% or less. As expected, Figure 22 demonstrates the reverse relationship, with the NPV being between 95% and 100% up to a CD prevalence of about 45%, and then dropping off.

Table 27: Weighted pooled estimates with 95% CIs and heterogeneity identified

Analysis	Sens	L 95% CI:	U 95% CI:	Spec	L 95% CI:	U 95% CI:	Prev	L 95% CI:	U 95% CI:	PPV	L 95% CI:	U 95% CI:	NPV	L 95% CI:	U 95% CI:
IgA-AGA–ADULT	H	H	H	H	H	H	0.358	0.332	0.385	H	H	H	H	H	H
IgG-AGA–ADULT	H	H	H	H	H	H	0.367	0.335	0.401	H	H	H	H	H	H
IgA-EMA-ME–ADULT	0.974	0.957	0.985	0.996	0.988	0.999	0.398	0.371	0.425	0.974	0.957	0.985	0.996	0.988	0.999
IgG-EMA-ME–ADULT (one study)	0.393	0.236	0.576	0.984	0.913	0.997	0.135	0.079	0.221	0.393	0.236	0.576	0.984	0.913	0.997
IgA-EMA-HU–ADULT	0.902	0.859	0.934	1.000	0.991	1.000	0.331	0.297	0.368	0.902	0.859	0.934	1.000	0.991	1.000
IgA-tTG-GP—ADULT	0.859	0.808	0.898	0.953	0.930	0.969	0.312	0.279	0.348	0.859	0.808	0.898	0.953	0.930	0.969
IgA-tTG-HR–ADULT	0.981	0.901	0.997	0.981	0.958	0.991	0.160	0.126	0.202	0.981	0.901	0.997	0.981	0.958	0.991
IgA-AGA–CHILD	H	H	H	H	H	H	0.363	0.341	0.385	H	H	H	H	H	H
IgG-AGA–CHILD	H	H	H	H	H	H	0.437	0.413	0.462	H	H	H	H	H	H
IgA-EMA-ME–CHILD	0.961	0.945	0.973	0.974	0.963	0.982	0.400	0.378	0.423	0.961	0.945	0.973	0.974	0.963	0.982
IgA-EMA-HU–CHILD	0.969	0.935	0.986	H	H	H	0.447	0.402	0.493	0.969	0.935	0.986	0.949	0.915	0.970
IgA-tTG-GP–CHILD	0.931	0.888	0.959	0.963	0.931	0.980	0.446	0.401	0.493	0.931	0.888	0.959	0.963	0.931	0.980
IgA-tTG-HR–CHILD	0.957	0.903	0.981	0.990	0.946	0.998	0.519	0.452	0.584	0.957	0.903	0.981	0.990	0.946	0.998
IgA-AGA–MIXED	H	H	H	H	H	H	0.415	0.386	0.444	H	H	H	H	H	H

H = significant heterogeneity by Pearson's Chi square

Table 27 (cont'd): Weighted pooled estimates with 95% CIs and heterogeneity identified

Analysis	Sens	L 95% CI:	U 95% CI:	Spec	L 95% CI:	U 95% CI:	Prev	L 95% CI:	U 95% CI:	PPV	L 95% CI:	U 95% CI:	NPV	L 95% CI:	U 95% CI:
IgG-AGA–MIXED	H	H	H	H	H	H	0.510	0.480	0.540	H	H	H	H	H	H
IgA-EMA-ME–MIXED	H	H	H	0.995	0.982	0.999	0.467	0.434	0.500	0.859	0.825	0.888	0.995	0.982	0.999
IgA-EMA-HU–MIXED	0.925	0.881	0.954	0.996	0.975	0.999	0.437	0.391	0.484	0.925	0.881	0.954	0.996	0.975	0.999
IgA-tTG-GP–MIXED	H	H	H	0.954	0.927	0.972	0.463	0.425	0.501	0.913	0.877	0.939	0.954	0.927	0.972
IgG-tTG-GP–MIXED	0.451	0.363	0.543	0.988	0.935	0.998	0.265	0.208	0.331	0.451	0.363	0.543	0.988	0.935	0.998
IgA-tTG-HR–MIXED	0.902	0.864	0.930	0.954	0.915	0.976	0.573	0.530	0.616	0.902	0.864	0.930	0.954	0.915	0.976
IgG-tTG-HR–MIXED (one study)	0.677	0.556	0.778	1.000	0.839	1.000	0.518	0.413	0.621	0.677	0.556	0.778	1.000	0.839	1.000

H = significant heterogeneity by Pearson's Chi square
Note: see Appendix G for raw pooled data by antibody test

Figure 21: PPV and prevalence from individual studies

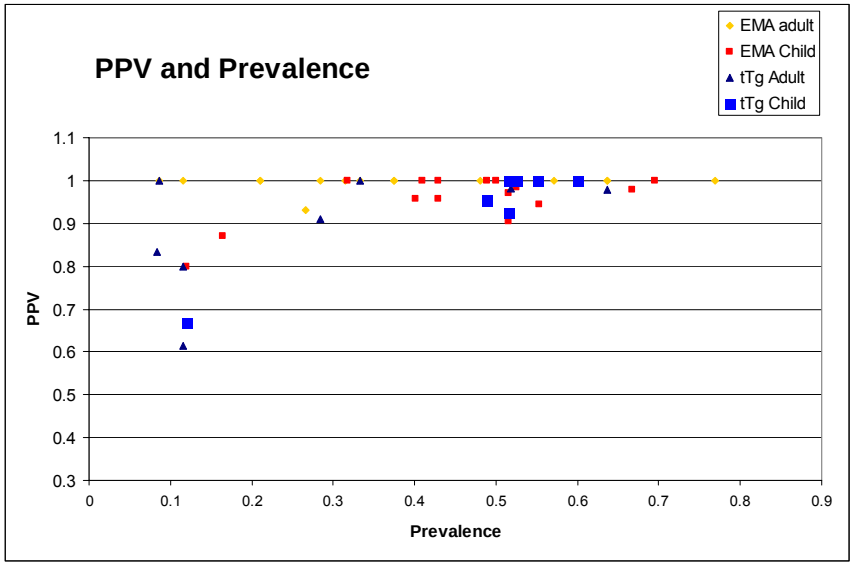


Figure 22: NPV and prevalence from individual studies

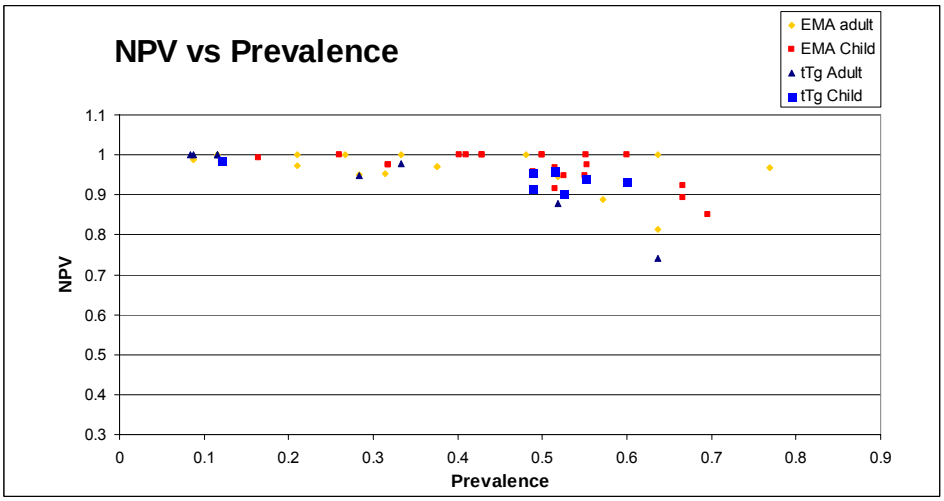
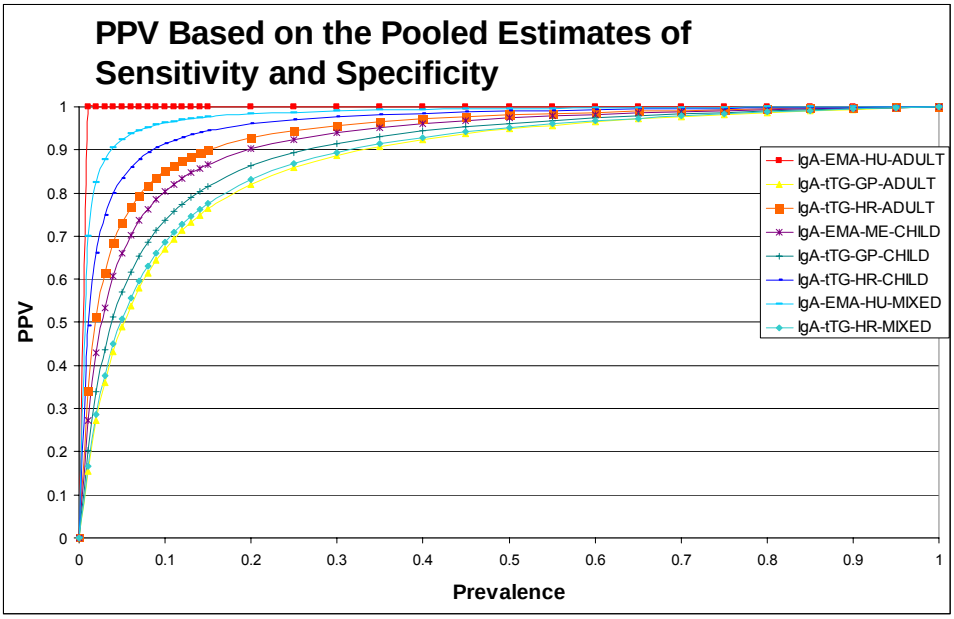


Figure 23: PPV based on the pooled estimates of sensitivity and specificity



HLA DQ2/DQ8

We identified 99 potentially relevant HLA articles that appeared to address HLA DQ2/DQ8 in a CD population (Appendix F).[8-11,15,53,54,62,91-100,100-177] These studies were not designed to determine the diagnostic utility of DQ2 or DQ8 per se.

Of the identified studies, 54 allowed estimation of the prevalence, sensitivity or specificity of HLA DQ2/DQ8 in the studied population.[8-10,15,53,54,93,100,109,120,134-177] In one study, DQ2 data could not be reliably extracted.[169] The authors of one study[9] explicitly stated that the patients used were the same as in two of their other publications.[8,93] In two other publications by the same authors,[9,10] the patients appear to be different and the authors do not indicate that they used patients from a previous study. However, the possibility that these two studies[9,10] share a subset of patients cannot be excluded. Another two studies addressing different topics but with extractable HLA data, also appeared to have used the same patients.[53,136] In cases of duplicate publications, the studies with the greater number of patients were used.[9,136]

The study designs and strictness of CD diagnosis in these articles varied, as did the inclusion of a control group. Most of the CD cases were diagnosed based on the ESPGAN criteria, although in some studies CD was diagnosed based on serology and then in most cases later confirmed by biopsy.[15,109,120,160,161,164,168,170,172,177] Nine of the studies were classified as cross-sectional studies,[169-177], 32 were case-control studies,[8-10,53,100,120,134-159] and 12 were mixed cross-sectional/case-control studies or could be considered as diagnostic cohort studies.[15,54,109,160-168] Four of the mixed design studies[109,164,166,178] used screen-negative patients as the control group, whereas the rest used a control group that was separate from the screened population. The study populations were also variable. The case-control studies used known CD cases compared with variously defined CD negative controls.

Seven studies used relatives of CD patients,[158,161,164,166,169,172,177] four used a population with Down Syndrome,[109,134,160,170] two used a population with type I diabetes,[165,173] and one used a mixed group of patients with CD including some with Down's and others with diabetes.[173] The mixed-design/cohort studies used patients suspected of CD on clinical grounds or subjects who belonged to a high-risk group, such as type 1 diabetics or first-degree relatives of patients with CD. The remaining articles used a screened healthy population or another specific group.

The articles with extractable data stated the frequency of HLA DQ2, and to a lesser extent the frequency of HLA DQ8, in their CD group. The cross-sectional studies did not include a control group. Only the frequency as a surrogate of sensitivity was available. None of the case-control or mixed-design studies calculated the sensitivity or specificity of HLA DQ2 or DQ8. However, these studies allowed us to derive estimates of these statistics from their results or tables. The considerable degree of clinical and methodological heterogeneity between the identified studies did no allow for statistical pooling of the results.

Two studies fulfilled our inclusion requirement of both cases and control groups undergoing intestinal biopsy (Evidence Table 2, Appendix I; Table 28).[136,152] The remaining studies had various control group types: unbiopsied, healthy controls, disease controls, or serology-negative controls. **These studies provide useful information and are presented at the end of the HLA results section for reference**.

The study by Iltanen et al.,[136] was conducted in a group of Finnish children to assess the density of gamma delta positive intraepithelial lymphocytes (γδ+ IELs) in: patients with CD by biopsy (ESPGAN); patients with suspected CD where the diagnosis was excluded by biopsy; and, in a group of biopsy-negative patients who underwent endoscopy for dyspepsia. The biopsy

aspect of this study is presented in its respective section. In this study, HLA DQ2 was found in 19 of 21 (90.5%) of patients with CD as apposed to 29 out of 67 (29.9%) of the control patients. Elevated γδ+ IEL density was significantly associated with DQ2 positivity. The calculated diagnostic measures for this study are presented in Table 28. In this population, DQ2 demonstrated a high sensitivity of 90.5% but a relatively modest specificity of only 70%, which is understandable given that the control population had a fairly high frequency of DQ2 positivity. The prevalence of CD in the study population was 1:4.2 (or 24%). The PPV was 49% and the NPV was 96%, suggesting that a negative DQ2 test result provides the greatest diagnostic information.

Sacchetti et al.[152] studied a group of Italian children suspected of having CD. Patients fulfilling the ESPGAN criteria were classified as having CD (n = 48 of 80), whereas, the remainder (n=32) were considered disease controls. The authors also used a second retrospectively defined group of known CD patients by ESPGAN criteria (n = 74), and a second group control of 180 unbiopsied healthy subjects. HLA DQ2 was determined in the CD group as a whole and in the two control groups, with the results presented in Table 28. In this study, the sensitivity of HLA DQ2 was 88.9% and the specificity was 81% for the comparison with the biopsied controls; the sensitivity of HLA DQ2 was 88.9% and the specificity was 73% for the comparison with the unbiopsied controls. Interestingly, in this study only 18.8% of the biopsy-negative controls were positive for HLA DQ2, whereas, 26.7% of the unbiopsied controls were HLA DQ2 positive. This difference accounts for the higher specificity seen for HLA DQ2 in the comparison with the biopsy-negative control group as compared with the comparison with the healthy controls. The prevalence of CD in the studied population was also quite high in both portions of this study (79% for comparison with biopsied controls and 51% for the comparison with unbiopsied controls). As such the PPV and the NPV of HLA DQ2 in this study were 95% and 62%, respectively. The difference in prevalence between this and the Iltanen study accounts for the differences seen in the PPVs and NPVs.

Table 28: HLA studies with biopsied cases and controls

Author, year; country	Prev of CD	DQ2 in CD	DQ2 in controls	Sensitivity	Specificity	PPV	NPV	CD population
Iltanen, 1999; Finland	0.24	90.48	29.85	90%	70%	49%	96%	Known CD versus biopsied controls
Sacchetti, 1998; Italy	0.79	86.89	18.75	87%	81%	95%	62%	Known CD versus biopsied controls
	0.51	86.89	26.72	87%	73%	77%	84%	Versus unbiopsied healthy controls

HLA all study data. The following section presents the data of the HLA studies that failed to be included on the basis that the control groups were not assessed with the gold standard test for CD (biopsy). These studies collectively provide useful information on the diagnostic value of HLA testing, but have to be interpreted with caution.

The prevalence of DQ2 and DQ8 in these studies is presented in Table 29, while the results of the diagnostic value of HLA DQ2 and HLA DQ8 are presented in Tables 30 and 31. Unfortunately, none of these studies were actual studies of the diagnostic value of HLA DQ2 or HLA DQ8 for the diagnosis or screening of CD. However, as presented in the Tables, the crude data was abstracted and the diagnostic characteristics were calculated. Significant clinical and statistical heterogeneity existed between these studies, making arithmetic pooling of the studies unjustified. Figure 24 and Figure 25 represent the plotting of each study's sensitivity (true positives) versus 1-specificity (false positives) to create a ROC presentation. The value of these figures lies in the global picture they represent regarding the results of each of the studies. Figure 24 demonstrates that the vast majority of the studies cluster together in a region where the sensitivity of HLA DQ2 is greater than 80%, with most studies lying above the 90% sensitivity mark. In contrast, these same studies have specificities in the range of 55% to 80%. Outlier studies are identified by author name. The best sensitivities and specificities were seen in two studies. The first, by Kaur et al.,[163] was a study from India where only 4.6% of the control population was positive for HLA DQ2. The second study, by Tighe et al.,[149] was conducted in a group of patients with CD and ethnically-matched control subjects from Rome, Italy. The prevalence of CD was quite high in the studied group (51%), and the frequency of HLA DQ2 in the control population of 12.2% was much lower than that observed in other Italian studies.

The remaining outlier studies were divided into a low-sensitivity/high-specificity group (Group 1), and a high-sensitivity/low-specificity group (Group 2). In the first case, all the studies were conducted in a non-Western European population. In particular, the worst performance of HLA DQ2 occurred in a study from Chile,[137] where the frequency of HLA DQ2 was very low in both the patients with CD and the control subjects. It is important to note, however, that not all non-Western populations deviated from the main cluster of studies. For example, Catassi et al.[120] found that 91% of Saharawi Arabs (Algeria) with CD carried HLA DQ2 compared with 38.9% of Saharawi controls. These values are similar to those seen in most Western populations. The second group all showed relatively poor specificity, although the sensitivity was preserved. As would be expected, the control groups of these studies were at high risk of having CD (relatives of CD[158,164,166]), or were a population with a known higher frequency of HLA DQ2 (individuals with diabetes[161,165]). As such, the high frequency of HLA DQ2 in these control populations makes the specificity of HLA DQ2 rather poor.

The frequency of HLA DQ8 in Western European populations with CD varies from approximately 2.7% to 6% (Table 29). The frequency is slightly higher in studies from Italy, the UK, and France (5.6% to 8% of CD patients). The frequency of HLA DQ8 in a subset of patients who had HLA testing in a large American serology screening study for CD was 22%,[15] which is quite a bit higher than that reported in the European studies.

A small group of studies allowed the estimation of the sensitivity and specificity of having either HLA DQ2 or DQ8. The results of these studies are presented in Table 33 and Figure 25. As can be seen in the figure, these studies confer a wide variation. Clearly, the sensitivity of using this strategy is quite high and is likely close to 100% in Western populations. The study by Balas et al.,[155] likely represents the closest to the truth, as this was a typical case-control design in patients with know CD compared with healthy controls. The Fasano et al. study[15]

represents the largest study and gives similar results to those obtained by Balas et al., however, the higher frequency of HLA DQ8 in their control group compared with other studies is of concern. Once again, the remaining studies can be grouped into high-specificity/low-sensitivity (Group 1) and high-sensitivity/low-specificity (Group 2). As was the case for HLA DQ2, Group 1 consists of two studies of non-Western populations, whereas, Group 2 represents studies with first-degree relatives and a study that used patients with diabetes as their control group.

Table 29: Prevalence/frequency of HLA DQ2 and HLA DQ8 in prevalence and mixed-design studies, and in case-control studies with HLA DQ8 data

Author	Year	Country	# of CD	% DQ2	% DQ8	% DQ2/8	Population with CD
Lewis	2000	USA	101	90.10	n/a	n/a	Confirmed cases among CD relatives
Book	2001	USA	8	87.50	12.50	100	Down Syndrome
Book	2003	USA	34	n/a	n/a	97.06	Affected 1st-degree relatives of CD sib. pairs
Csizmadia	2000	Netherlands	10	100	20	n/a	Down Syndrome
Fasano	2003	USA	98	83.67	22.45	100	Screened large population only subset tested for HLA
Iltamen	1999	Finland	5	100	n/a	n/a	Sjogren's syndrome
Kaukinen	2000	Finland	6	100	n/a	n/a	Known CD
Maki	2003	Finland	56	85.71	n/a	n/a	Screen of school-age children
Mustalahti	2002	Finland	29	100	n/a	n/a	Relatives of CD or DH
Catassi	2001	Algeria	79	91.3	n/a	95.6	Saharawi Arabs
Lui	2002	Finland	260	96.92	2.69	99.62	Family members of celiacs
Polvi	1996	Finland	45	100	n/a	n/a	Known CD
Ploski / Sollid	1996	Sweden	135	91.85	4.44	96.30	Known CD
Popat	2002	Sweden	62	93.55	n/a	n/a	Known CD
Larizza	2001	Italy	7	100	n/a	n/a	Children with autoimmune thyroid disease, EMA+biopsy
Failla	1996	Italy	7	14.29	n/a	n/a	Down Syndrome (only 7 CD cases)
Farre	1999	Spain	60	93.33	n/a	n/a	1st-degree relatives of celiacs
Balas	1997	Spain	212	94.81	4.25	99.06	Known CD
Zubillaga	2002	Spain	135	92.59	3.70	96.0 (calc)	Mostly CDs, some CD in subjects with Down Syndrome and subjects with diabetes
Karell	2003	France	92	86.96	6.52	93.48	Known CD
		Italy	302	93.71	5.63	89.40	
		Finland	100	91	5.00	96.00	
		Norway/ Sweden	326	91.41	5.21	96.63	
		Uk	188	87.77	7.98	95.74	
		Total	1008	93.71	5.95	93.95	
Kaur	2002	India	35	97.14	n/a	n/a	Known CD
Neuhausen	2002	Israel	23	82.61	56.52	100	Bedouin Arabs
Tuysuz	2001	Turkey	55	83.64	16.36	90.91	Children with known CD
Bouguerra	1996	Tunisia	94	84.04	n/a	n/a	Known CD
Sumnik	2000	Czech	15	80	66.67	100	Diabetics
Perez-Bravo	1999	Chile	62	11.29	25.81	37.10	Chileans

DH = dermatitis herpetiformis

Table 30: Sensitivity/specificity (calculated) for HLA DQ2 in case-control studies

Author, year; country	Prev of CD	% DQ2 in CD	% DQ2 in Controls	Sens	Spec	PPV	NPV	CD population
Fine, 2000; USA	0.06	88 (22/25)	31.24 (134/429)	0.88	0.69	0.14	0.99	Known CD
Howell, 1995; UK	0.38	91.21 (83/91)	23.18 (35/151)	0.91	0.77	0.7	0.94	Known CD
Michalski, 1995; Ireland	0.62	96.67 (87/90)	39.29 (22/56)	0.97	0.61	0.8	0.92	Known CD
Colonna, 1990; Italy	0.36	94.59 (140/148)	40.82 (109/267)	0.95	0.59	0.56	0.95	Known CD
Catassi, 2001; Algeria	0.37	91.1 (72/79)	38.9 (53/136)	0.91	0.61	0.58	0.92	Saharawi Arabs
Congia, 1991; Italy	0.2	96 (24/25)	34 (34/100)	0.96	0.66	0.41	0.99	Known CD
Ferrante, 1992; Italy	0.48	88 (44/50)	16.36 (9/55)	0.88	0.84	0.83	0.88	Known CD
Mazzilli, 1992; Italy	0.5	92 (46/50)	18 (9/50)	0.92	0.82	0.84	0.91	Known CD
Tighe, 1992; Italy	0.49	70.59 (39/43)	8.33 (5/41)	0.91	0.88	0.89	0.9	Known CD
Castro, 1993; Italy	0.38	80 (4/5)	37.5 (3/8)	0.8	0.63	0.57	0.83	Down Syndrome
Lio, 1997; Italy	0.45	100 (18/18)	63.64 (14/22)	1	0.36	0.56	1	Known CD
Sacchetti, 1998; Italy	0.79	86.89 (106/122)	18.75 (6/32)	0.87	0.81	0.95	0.62	Known CD and biopsied controls
Sacchetti, 1998; Italy	0.51	86.89 (106/122)	26.72 (31/116)	0.87	0.73	0.77	0.84	Healthy controls
Iltamen, 1999; Finland	0.24	90.48 (19/21)	29.85 (20/67)	0.9	0.7	0.49	0.96	Known CD
Ploski/Sollid, 1993; Sweden	0.34	94.68 (89/94)	25.97 (47/181)	0.95	0.74	0.65	0.96	Known CD
Pattersson, 1933; Sweden	0.4	92.31 (60/65)	43.75 (42/96)	0.92	0.56	0.59	0.92	Known CD
Ploski/Sollid, 1996; Sweden	0.43	91.85 (124/135)	22.35 (40/179)	0.92	0.78	0.76	0.93	CD vs blood donors
Fernandez-Arquero, 1995; Spain	0.36	92 (92/100)	25.56 (46/180)	0.92	0.74	0.67	0.94	Known CD
Arranz, 1997; Spain	0.5	92 (46/50)	24 (12/50)	0.92	0.76	0.79	0.9	Known CD
Balas, 1997; Spain	0.22	94.81 (201/212)	29.25 (217/742)	0.95	0.71	0.48	0.98	Known CD

Table 30 (cont'd): Sensitivity/specificity (calculated) for HLA DQ2 in case-control studies

Author, year; country	Prev of CD	% DQ2 in CD	% DQ2 in Controls	Sens	Spec	PPV	NPV	CD population
Ruiz Del Prado, 2001; Spain	0.04	94.74 (36/38)	39.22 (351/895)	0.95	0.61	0.09	1	Known CD
Dijilali-Saiah, 1994; France	0.27	88.75 (71/80)	21.13 (45/213)	0.89	0.79	0.61	0.95	Known CD
Dijilali-Saiah, 1998; France	0.44	83.17 (84/101)	20 (26/130)	0.83	0.8	0.76	0.86	Known CD
Tighe, 1993; Israel	0.51	90.7 (24/34)	12.2 (3/36)	0.71	0.92	0.89	0.77	Ashkenazi Jews, known CD
Arnason, 1994; Iceland	0.13	84 (21/25)	36.36 (60/165)	0.84	0.64	0.26	0.96	Known CD
Boy, 1994; Sardinia	0.5	96 (48/50)	32 (16/50)	0.96	0.68	0.75	0.94	Known CD
Congia, 1994; Sardinia	0.42	90.77 (59/65)	39.33 (35/89)	0.91	0.61	0.63	0.9	Known CD
Erkan, 1999; Turkey	0.5	40 (12/30)	6.67 (2/30)	0.4	0.93	0.86	0.61	Known CD
Tumer, 2000; Turkey	0.3	51.52 (17/33)	25.97 (20/77)	0.52	0.74	0.46	0.78	Turkish, known CD
Tuysuz, 2001; Turkey	0.52	83.64 (46/55)	24 (12/50)	0.84	0.76	0.79	0.81	Turkish, known CD
Perez-Bravo, 1999; Chile	0.33	11.29 (7/62)	2.42 (3/124)	0.11	0.98	0.7	0.69	Chilean

Table 31: Sensitivity/specificity (calculated) for HLA DQ2 in mixed-design studies

Author, year; country	Prev of CD	% DQ2 in CD	% DQ2 in controls	Sens	Spec	PPV	NPV	CD population
Book, 2001; USA	0.09	87.50 (7/8)	15.58 (12/77)	0.88	0.84	0.37	0.98	Down Syndrome
Csizmadia, 2000; Netherlands	0.11	100 (10/10)	28 (25/90)	1.00	0.72	0.29	1.00	Down Syndrome
Fasano, 2003; USA	0.52	83.67 (82/98)	42.39 (39/92)	0.84	0.58	0.68	0.77	9019 at risk, 4126 not at risk
Larizza, 2001; Italy	0.08	100 (7/7)	34.62 (27/78)	1	0.65	0.21	1	Children with autoimmune thyroid disease, EMA+biopsy
Polvi, 1996; Finland	0.58	100 (45/45)	28.13 (9/32)	1	0.72	0.83	1	CD vs various controls
Iltamen, 1999; Finland	0.15	100 (5/5)	n/a	1	n/a	n/a	n/a	Sjogren's syndrome
Kaukinen, 2000; Finland	0.17	100 (6/6)	n/a	1	n/a	n/a	n/a	CD vs disease controls
Lui, 2002; Finland	0.52	96.92 (252/260)	57.38 (136/237)	0.97	0.43	0.65	0.93	Family members of celiacs (controls=unaffected family members)
Farre, 1999; Spain	0.55	93.33 (56/60)	18 (9/50)	0.93	0.82	0.86	0.91	CD vs healthy controls
	0.26	93.33(56/60)	63.91(108/169)	0.93	0.36	0.34	0.94	CD vs relatives of CD
Sumnik, 2000; Czech	0.07	80 (12/15)	49.46 (92/186)	0.8	0.51	0.12	0.97	Diabetes (control=EMA neg.)
Kaur, 2002; India	0.11	97.14 (34/35)	4.64 (13/280)	0.97	0.95	0.72	1	CD vs healthy controls
Neuhausen, 2002; Israel	0.31	82.61 (19/23)	61.54 (32/52)	0.83	0.38	0.37	0.83	Bedouin Arabs (some cases and controls not biopsied)

Table 32: Sensitivity/specificity (calculated) for HLA DQ8

Author, year; country	Prev of CD	DQ8 in CD	DQ8 in controls	Sens	Spec	PPV	NPV	CD population
Csizmadia, 2000; Netherlands	0.11	20 (2/10)	20 (18/90)	0.20	0.80	0.10	0.90	Down Syndrome
Fasano, 2003; USA	0.52	22.45 (22/98)	20.65 (19/92)	0.22	0.79	0.54	0.49	Screened at-risk and not-at-risk populations
Lui, 2002; Finland	0.52	2.69 (7/260)	10.55 (25/237)	0.03	0.89	0.22	0.46	Family members of CD patients (controls=unaffected family members)
Ploski/Sollid 1996; Sweden	0.43	4.44 (6/135)	25.14 (45/179)	0.04	0.75	0.12	0.51	Known CD
Balas, 1997; Spain	0.22	4.25 (9/212)	16.85 (125/742)	0.04	0.83	0.07	0.75	Known CD
Sumnik, 2000; Czech	0.07	66.67 (10/15)	65.59 (122/186)	0.67	0.34	0.08	0.93	Diabetes
Neuhausen, 2002; Israel	0.31	56.52 (13/23)	25 (13/52)	0.57	0.75	0.5	0.8	Bedouin Arabs
Tuysuz, 2001; Turkey	0.52	16.36 (9/55)	8 (4/50)	0.16	0.92	0.69	0.5	Turkish known CD
Perez-Bravo, 1999; Chile	0.33	25.81 (16/62)	12.9 (16/124)	0.26	0.87	0.5	0.7	Chileans

Table 33: Sensitivity/specificity (calculated) for HLA DQ2 or DQ8

Author; year; country	Prev of CD	DQ2 or DQ8 in CD	DQ2 or DQ8 in controls	Sens	Spec	PPV	NPV	Notes
Fasano, 2003; USA	0.52	100 (98/98)	59.78 (55/92)	1	0.4	0.64	1	Screened at-risk and not-at-risk populations
Catassi, 2001; Algeria	0.37	96.2 (76/79)	41.9 (57/136)	0.96	0.58	0.57	0.96	Saharawi Arabs
Lui, 2002; Finland	0.52	99.62 (259/260)	67.93 (161/237)	1	0.32	0.62	0.99	Family members of CD (controls=unaffected family members)
Balas, 1997; Spain	0.22	99.06 (210/212)	46.09 (342/742)	0.99	0.54	0.38	1	Known CD
Sumnik, 2000; Czech	0.07	100 (15/15)	87.63 (163/186)	1	0.12	0.08	1	Diabetes
Tuysuz, 2001; Turkey	0.52	90.91 (50/55)	32 (16/50)	0.91	0.68	0.76	0.87	Turkish Known CD
Neuhausen, 2002; Israel	0.31	100 (23/23)	86.54 (45/52)	1	0.13	0.34	1	Bedouin Arabs
Perez-Bravo, 1999; Chile	0.33	37.1 (23/62)	15.32 (19/124)	0.37	0.85	0.55	0.73	Chileans

Figure 24: HLA DQ2

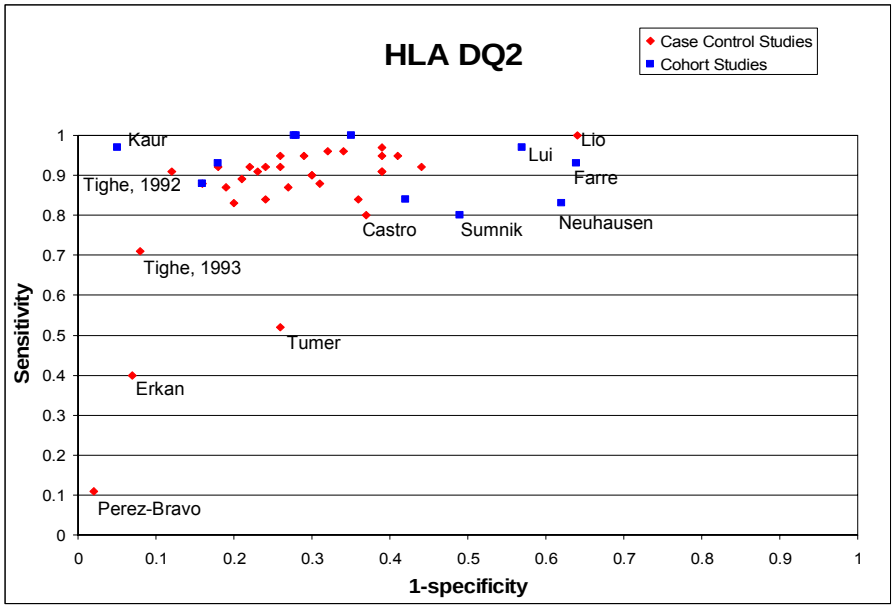
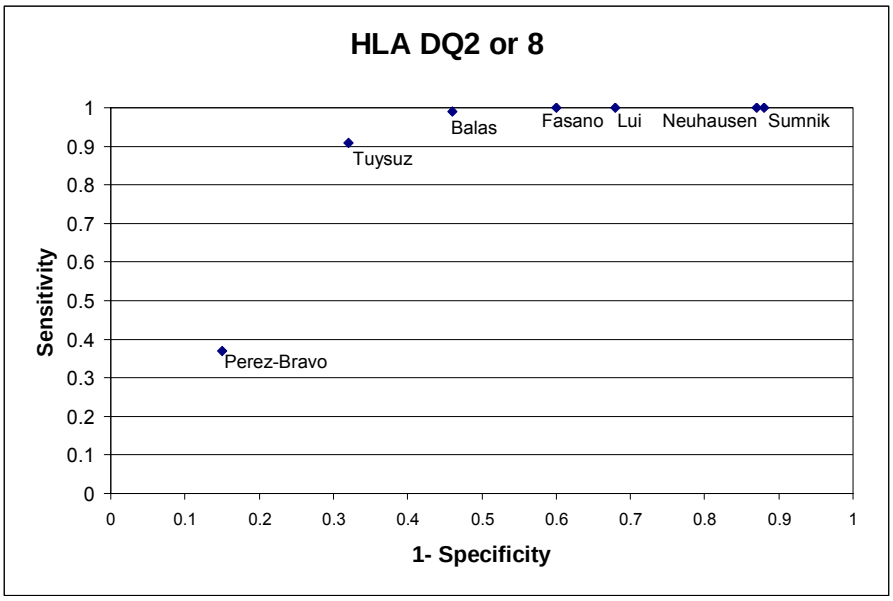


Figure 25: HLA DQ2 and DQ8



Biopsy

Using epidemiologically appropriate eligibility criteria, our comprehensive literature search did not identify any studies that specifically addressed the question of the sensitivity or specificity of biopsy for the diagnosis of CD.

However we sought to obtain indirect evidence regarding the diagnostic performance of biopsy as a test for CD. Some data was available from those studies identified for other review objectives, such as the cross-sectional screening studies, the HLA DQ2/8 studies, and studies of IELs. We also sought studies of follow-up of biopsy negative patients suspected of CD, and studies of silent and latent CD. The findings from these studies are presented in the Discussion and in Appendix H.

Quality Assessment

Overall, the quality of the diagnostic studies assessed in the Celiac 1 objective was quite good (Appendix J, Table 1). However, 59% of the studies reported using a selected patient population that may not be representative of a clinically relevant population. This is likely related to study design. Only 11% of the studies reported on whether the reference test was reported without knowledge of the index test. We felt that this was not a major threat to the validity of the studies.

Celiac 2: Incidence and Prevalence of CD

The literature search yielded 2,116 references (Appendix F). A first-level screen of the titles, abstracts and keywords, for articles that related to the incidence or prevalence of CD, excluded 1,506 references. Full-text versions of each of the 610 retained references were obtained and used for a second-level screen for articles, with a focus on the incidence and/or prevalence of CD. Review articles were also identified and kept for reference (n = 71). Three hundred and forty-eight out of the 610 references were excluded. The remaining 262 references were screened at a third level (Appendix F). Studies were included if they reported the prevalence and/or incidence of CD in the following groups: (1) general populations from North America or Western Europe; (2) first-degree relatives of patients with CD; (3) patients with type 1 diabetes; (4) patients being investigated for anemia; (5) patients with osteoporosis or osteopenia; (6) patients with suspected CD on the basis of their clinical presentations. We did not use any geographic restriction for the studies of populations at risk (first-degree relatives and type 1 diabetics) or of associated clinical presentations (suspected CD, anemia, or metabolic bone disease). Studies of prevalence or incidence that used AGA tests conducted prior to 1990 were excluded after discussion with the AHRQ because of potential problems with the reliability of older AGA assays. Reports which were not sufficiently explicit for data extraction also had to be excluded.[179-181]

We defined incidence studies as those studies that reported the total number of new cases of CD for a given territory and period, over a unit of population density. Therefore, studies of incidence where there was no population denominator were excluded. When multiple studies of incidence of CD were available for a similar country or geographic area, the most recent and/or most encompassing was selected. In general, we excluded the studies whose observation periods pertained exclusively to a period prior to 1990.

A total of 133 publications were selected. Of these, 14 publications were identified as duplicates on the basis that the same study population was reported on elsewhere, or as part of a larger cohort.[122,182-194] The remaining 119 original studies on prevalence and/or incidence of CD in the populations of interest were included and their data abstracted. Of these included studies, 42 assessed the prevalence and/or incidence of CD in a general population. Twelve of the 42 reported on the incidence of CD,[128,195-205] and 30 reported on the prevalence, either in the US (three studies[206-208]), Scandinavia (11 studies[209-219]), Italy and San Marino (seven studies[126,220-225]), UK (four studies[226-229]), or other countries (Spain[230], the Netherlands,[231,232] Switzerland,[233] and Germany[234]).

Studies of the prevalence of CD in populations at risk were divided as follows: 18 studies of the first-degree relatives of CD patients,[129,167,206,235-249] and 34 studies in patients with type 1 diabetes.[234,250-282]

Studies of the prevalence of CD in patients with associated clinical presentations were divided as follows: 12 studies in anemia and/or iron deficiency,[283-294] four studies in metabolic bone disease,[295-298] and 13 studies of patients with suspected CD on the basis of their clinical presentation.[206,238,299-309] The clinical manifestations that were included in the "suspected CD category" were: chronic diarrhea, weight loss, malabsorption or abdominal pain in adults and failure to thrive, short stature, malabsorption, chronic diarrhea, and abdominal pain in children. Four studies included groups at multiple-risk levels.[206,234,238,272]

Incidence of CD in the General Population

The incidence of CD in North America and Western Europe was derived from studies from the following countries: US,[128,205] England,[201] Italy,[202], Sicily,[203] Spain,[204] Netherlands,[200] Sweden,[195] Denmark,[196,197] and Finland (Evidence Table 3, Appendix I; Table 34).[198,199] In the report, crude incidence is defined as the number of new cases per 100,000 population-at-risk per year and cumulative incidence as the number of new cases per 1,000 live births; cumulative incidence is age-specific and its denominator reflects the total number of individuals from the same year of birth (i.e., birth cohort).

Table 34: Included studies of incidence of CD in the general population

Study	Country, period	Group at risk	Period related to results	Incidence	
				Crude incidence (# cases/100,000 patient year)	**Cumulative incidence** (# cases/1,000 births)
Ivarsson, 2003 Duplicate Ivarsson, 2000[193]	Sweden, 1973-97	Children	1997 (0-2 y)	51 (95% CI: 36-70)	Age 2 (1995): 1.7 (95% CI: 1.3-2.1)
			1996 (2-5 y)	33 (95% CI: 24-44)	
			1996 (5-15 y)	10 (95% CI: 7-13)	
Weile, 1993 Duplicate Weile, 1993[196]	Denmark, 1960-88	Children	1960-88		Age 5 (1988): 0.118
Maki, 1990 Duplicate ref[194]	Finland, 1960-84	Children	1974-83	3.46 (95% CI: n/r)	
Hawkes, 2000	England, 1981-95	Children	1991-95	2.15 (95% CI: n/r)	
Magazzu, 1994	Sicily 1975-89	Children	1989 birth cohort		Age 5 (1989): 1.16 95% CI: 0.92-1.42
Lopez-Rodriguez, 2003	Spain, 1981-99	Children 0-14 y	1981-90	6.87 (95% CI: 5.26-8.83)	
			1991-99	16.04 (95% CI: 12.99-19.59)	
		Children 0-4 y	1991-99	42.04 (95% CI: n/r)	
Hoffenberg, 2003	US (Denver, Colorado), 1993-99	Children	1993-99		Age 5 (1999): 9 (95% CI: 4-20)
Jansen, 1993	Netherlands 1990-92	All ages	1991-92	1.0 (95% CI: n/r)	
Corrao, 1995	Italy 1990-91	All ages	1990-91	2.13 (95% CI: n/r)	Age 5 (1991): 0.81
Talley, 1994	US 1960-90 Olmstead County	All ages	1960-90	1.2 (95% CI: 0.7-1.6)	
			1980-90	1.7 (95% CI: n/r)	
Bodé, 1996	Denmark, 1976-91	Adults	1976-91	1.27 (95% CI: n/r)	
Collin, 1997	Finland, 1975-94	Adults	1990-94	17.2 (95% CI: n/r)	
Hawkes, 2000	England, 1981-95	Adults	1991-95	3.08 (95% CI: n/r)	

Incidence in children: The crude incidence of CD in children age 0 to 15 years varied from 2.15 to 51 cases per 100,000 patient years.[193-195,198,201,204] When reported, the relative risk (RR) of CD was greatest for the 0- to 2-year age group, as well as for women, and varied from 32.26 to

42.4[193,195,204] and from 1.9 to 3.34,[128,193,195] respectively. The cumulative incidence at age 5, when reported, varied between 0.089 and 9 cases per 1,000 live births.[128,196,202,203] (see Table 34).

The incidence of CD has been most studied in the Scandinavian countries, particularly Sweden,[193,195,310-313] Denmark,[196,197,313,314] and Finland,[194,198,199] where important disparities have been observed over time and between countries. Reports from these countries have the advantage of being derived from comprehensive prospective databases and from populations which are genetically fairly stable, shedding light on potential environmental causal exposures,[195,196] or on variations in practice patterns.

In Scandinavia, the highest incidences of CD in children were found in Sweden for the 0- to 2-year age group from 1987 to 1997, where an average of 198 new cases per 100,000 patient years (95% CI: 186-210) were observed.[193,195] This peak in incidence was followed by a rapid decline, observed during 1995-97, where incidences dropped to an average of 51/100,000 patient years (95% CI: 36-70). In contrast, the incidence of CD in children aged 2 to 4.9 years and 5 to 15 years was only slightly increased over the 1973-97 period, with a peak in 1996 of 33 cases (95% CI: 24-44) per 100,000 patient years and 10 cases (95% CI: 7-13) per 100,000 patient years for these respective age groups. A cohort effect was noted in that the cumulative incidences at 2 years of age for the children belonging to birth cohorts from 1984 to 1994 were on the gradual rise (up to 4.4 cases/1,000 births [95% CI: 3.8-5.0] for the 1993 cohort), while a progressive decline was observed for birth cohorts from 1994 to 1996 (down to 1.7 cases [95% CI: 1.3-2.1] per 1,000 births for the 1995 cohort). Most of these cases were symptomatic, so that these observations are unlikely to be due to changes in screening practices. Interestingly, these changes mirrored changes in the composition of infant formulas, with the highest values of a wheat/rye/barley exposure index during the years 1982-1994.

In contrast, the incidence of CD in Denmark, a neighbouring country, has been significantly lower and very stable from 1960 to 1988,[196] with an average incidence of 0.089/1,000 live births for that period.[313] A comparison of dietary exposures between Swedish and Danish children diagnosed with CD between 1972 and 1989 showed that by the age of 8 months, the Swedish diet contained more than 40 times more gliadin than the Danish diet.[313] In Finland, incidences have also been fairly stable, and have in fact decreased among infants but increased among older children.[198] However, these observations date back to 1984 and can therefore not be compared with the Swedish epidemics.

Spain has also seen an increased incidence of CD over the past 25 years, from 6.87 (95% CI: 5.26-8.83) cases/100,000/year in 1981-90 to 16.04 cases/100,000/year (95% CI: 12.99-19.59) in 1991-99,[204] an observation that was correlated with an increased proportion of silent or atypical presentations at diagnosis (i.e., inferring a role for changes in clinical practice). The age at diagnosis also correlated positively with the age at which gluten was introduced in the diet.

The role of dietary exposure during infancy is also highlighted in studies from the UK, where recommendations on infant feeding, promoting breastfeeding and later introduction of starches, were published in 1974. Subsequent to these recommendations, there was a fall in the incidence of childhood CD;[315,316] however, this data is not presented in detail because we focused on reports from the past 15 years.

As opposed to the incidences derived from reported cases, the incidence observed from a prospective screening protocol are not subject to variations related to practice patterns and are obviously more comprehensive and accurate. Hoffenberg et al., from the US, conducted the only prospective CD screening study available to date.[128] Between December 1993 and September 1999, a total of 22,346 newborns in Denver, Colorado were screened for HLA genotypes

associated with CD and type 1 diabetes. A representative sample of at risk HLA DRB1*03 positive infants were prospectively followed (n=987), for as long as the first seven years of life. Serological screening was performed at nine, 15 and 24 months of age, then yearly. Small bowel biopsies were recommended if the serology (tTG in most cases) was positive on two separate occasions, or in the presence of clinical suspicion. Between 1993 and 1999, 19 children were found to have evidence of CD, ten children had biopsy-confirmed CD, whereas, nine children had a positive tTG result at least twice. The mean age at presentation of evidence of CD was 4.6 years (range 2.6-6.5). Compared with HLA-DR3-negative children, the RR for evidence of CD was 5.6 (1.5-21, p=0.009) and 9.1 (1.7-48, p=0.003), for those expressing one and two HLA-DR3 alleles, respectively. The RR of CD in females was 3.34 (1-10.9, p=0.048) times that of males. Cognisant of the prevalence of HLA-DR mono- and heterozygotes among the same birth cohort, the authors calculated that by the age of 5, the estimated cumulative incidence of CD in the general population (defined as either biopsy-proven CD or persistently elevated tTG) was 9/1000 births (95% CI: 4-20), or 1:104 (1:49 to 1:221). This remarkably high cumulative incidence (i.e., twice that of the highest value among Swedish children at 4 years of age – 5.0 [95% CI: 4.4-5.7][193]) has to be interpreted in light of the fact that only ten out of the 19 cases had been biopsied; the remaining nine cases were diagnosed on the basis of a persistently elevated tTG titre, the PPV of which the same authors reported to be only 70% to 83%.[317] However, as mentioned above, these results are derived from an actual prospective and systematic screening intervention for CD, where asymptomatic cases would be detected. In all likelihood, there is therefore an important proportion of CD cases who remain undiagnosed during early childhood.

Incidence in adults: The crude incidence of CD in adults varied from lows of 1.27 in Denmark[197] and 3.08 in England, [201] to a high of 17.2 cases per 100,000 patient years in Finland,[199] where specific efforts had been untaken to encourage screening for CD (see Table 34).

As has been observed for children, the incidence of CD in adults seems to have increased over the past 20 years.[199,201] This is largely explained by a change in practice patterns: physicians are more aware of the condition, its atypical manifestations and associated condition, while at the same time, serological testing has become widely available. There are therefore more diagnoses made on the basis of case-finding. This is reflected by the fact that the proportion of patients being diagnosed with CD in the absence of symptoms, or as a result of serological testing, has also increased.[199,201,318-320]

In Finland over the period 1975-94, Collin et al.[199] have observed a ten-fold rise in the incidence of CD. The authors attributed this to the use of serologic screening (physicians were actively told to screen patients with type I insulin-dependent diabetes (IDDM), autoimmune thyroid disease, connective tissue diseases, women with infertility, patients with neurologic symptoms and first-degree relatives of CD patients), the routine performance of intestinal biopsies on all patients undergoing gastroscopy, and to the opening of open-access endoscopy clinics, creating the ability of all general practitioners to refer patients for gastroscopy.

In Italy, a gradual increase in the number of annual new CD diagnoses was observed between 1968 and 1992;[318,320] this increase correlated with an increased proportion of patients with subclinical presentations being identified.[318,320] Interestingly, despite the changing clinical presentation, there was no statistical difference between the histological grades at diagnosis.[320]

The incidence of CD in individuals of all ages varies from 1.0 in the Netherlands[200] to 2.13 in Italy.[202] In Italy, the RR of CD in adults ranged from 0.11 in the >60 year group to 0.33 in the 16-39 year group, compared with children.[202] The RR of CD for females was 1.90 (95% CI: 1.48-2.45).[202]

In the US, the 30-year incidence (1960-90) for Olmstead County was 1.2 (95% CI: 0.7-1.6), and the incidence for 1980-90 was slightly higher at 1.7 (95% CI: not reported).[205] This observation contrasts with the cumulative incidence of 9/1000 by age 5 reported by Hoffenberg from Denver, Colorado;[128] clearly, further knowledge of the epidemiology of CD in the US is required.

The point prevalence of CD can be calculated from registers of CD cases and the size of the population at risk; we found reports of such an observation in three of the included incidence studies.[199,199,205] The point prevalence of CD was 21.8/100,000 in Olmstead County in 1991,[205] 2.7/100,000 (95% CI: 11.0-14.5) in the Netherlands in 1992,[200] and 204/100,000 (95% CI: 181-231) in Finland in 1994.[199] Of note, the later prevalence from Finland was observed in a community where intense efforts had been carried to screen the population at risk for CD.

Prevalence of CD in the General Population—Different Geographic and Racial/Ethnic Populations

Thirty-seven studies reported on the prevalence of CD in a general population (Evidence Table 4, Appendix I; Table 35). Three of these were conducted in the US,[206-208] 16 in the Scandinavian countries,[184-187,209-219,232]) eight in Italy,[126,182,183,220,221,223-225] five in the UK,[188,226-229] and five in other countries (Spain,[230] Republic of San Marino,[222] the Netherlands,[231] Switzerland[233] and Germany[234]). Several pairs of duplicate publications were identified including two triplets,[182-188,211,213,218,220,321] which brought the total number of included unique articles down to 30. The articles with the most complete data were used for the report.[126,206-234] Only seven studies were conducted in a child population,[206,209,215,220,221,223,232,233] but one large American study included separate data for both adults and children.[206] All the included studies were conducted between 1992 and 2003. A summary of the included study characteristics is presented in Table 35. A breakdown of the included studies by screening test and age group is provided in Table 36.

The prevalence of CD by serology in the general unselected populations of North America and Western Europe, ranged widely from 152 per 100,000 (0.152% or 1:658) to 2,670 per 100,000 (2.67% or 1:37). The prevalence by biopsy ranged from 152 per 100,000 (0.152% or 1:658) to 1,870 per 100,000 (1.87% or 1:53). In four of the studies, a large proportion of the serology-positive subjects did not undergo biopsy.[206,216,224,232]

Among the included studies, there was no clear pattern relating prevalence to study age group, or in a consistent way to country, with large numbers of studies clustering around a prevalence range of 0.0025 to 0.014 by serology and 0.0025 to 0.010 by biopsy (Table 35; Figure 26, 27). In fact for prevalence by serology, the 50th, 75th, and 80th percentiles occurred at a prevalence of 0.00637 (0.64%), 0.0117 (1.2%), and 0.0125 (1.3%), respectively, while by biopsy the 80th percentile was at a prevalence of 0.0074 (0.74%) (Table 37; Figure 26, 27). Categorizing the studies by screening test and age group reduced the variability somewhat, but significant between study variation persisted. There were not enough studies to divide an analysis by screening test, age group, and country, simultaneously.

Among the studies conducted in the US,[206-208] the prevalence ranged from 0.00312 (0.312% or 1:320—only child population in this group) to 0.00949 (0.949% or 1:105). The largest of these, by Fasano et al.,[322] found a prevalence of CD in "not at risk" populations to be 0.95% in adults, 0.31% in children, and 0.75% overall (0.0075 or 1:133). This study included a predominately Caucasian population, although other ethnic groups were included (94% white; 3% black; 1.5% hispanic; 1% asian; 0.5% other). Not et al.[208] found the prevalence by EMA confirmation of initial AGA testing to be 0.004 (0.4% or 1:250) in another predominately Caucasian population that also included other ethnic backgrounds (Caucasian [87%], African-American [11.5%], and Asian [1.5%]). Finally, Green et al.[207] found a prevalence of 0.005 (0.5% or 1:200) in 1,749 patients undergoing upper endoscopy. The reason for the initial endoscopy in this study was not clearly described, and only those patients with endoscopic features suggestive of CD were biopsied, which may have underestimated the true prevalence of CD. The prevalence of CD among the six Italian studies was similar to that seen in the American studies, showing a range from 0.2% to 0.86%.[126,221,223-225,230] The prevalence of CD in other countries is presented in Table 35.

Only four studies demonstrated a prevalence of CD of greater than 0.015 (1.5%) (UK,[323] Sweden,[209,219] Germany[234]), and an additional six showed a prevalence of between 0.010 (1.0%) and 0.015 (1.5%) (UK,[228] Sweden,[216] Netherlands,[232] Ireland,[229] Finland[214,215]). These studies would suggest a potentially higher prevalence of CD in these countries, though it should be kept in mind that other studies from these same countries showed a prevalence of less than 1.0%, including four studies from Sweden[211,213,216,217] (Figure 28). Only three of the eight studies conducted in a child population demonstrated a prevalence of CD of greater than 1.0% (Finland,[215] Sweden,[209] Netherlands[232]).

Among the 30 included studies, there was a considerable amount of variation in the point estimates for the prevalence of CD both by serology and by biopsy due to differences in serological test strategies, biopsy definitions and patient sampling, making pooled estimates unreliable. To further explore the potential sources of variability in the observed prevalence of CD, we plotted the studies' prevalence versus its sample size (Figure 29). This scatter diagram visually illustrates the distribution of the prevalence of CD among the included studies. The study with the highest reported prevalence of CD (2.67%), was also the one with the smallest sample size of 150 healthy patients, and also included several other at-risk groups, which were the primary focus of that study.[234] Overall, studies with the smallest sample sizes tended to produce both the highest and lowest prevalence of CD. Using an arbitrary cut-off of 1,600 patients to divide "small" and "large" sample size studies, the prevalence by serology ranged fairly evenly from 0.17% to 2.67% for the 13 small studies, while 12 of the 18 large studies were located within a range of 0.5% to 1.26% (one study did not provide prevalence by serology).

Table 35: Prevalence of CD by country

Author, year	Country	Age group	Test	Total patients	Prevalence by serology	Prevalence by biopsy	Notes
Fasano, 2003	USA	Adults	EMA - ME; all positive EMA tested with tTG-HU	2,845	0.00949		116/350 biopsied
Green, 2000	USA	Adults	EGD/biopsy	1,749		0.00515	Not all sytematically biopsied; only those with suggestive endoscopic features
Not, 1998	USA	Adults	IgG- and IgA-AGA - ELISA; confirmed with IgA-EMA ME or HU	2,000	0.00400		
Fasano, 2003	USA	Children		1,281	0.00312		
Johnston, 1998	UK	Adults	IgA-AGA, IgA-EMA	1,823	0.00823		
Sanders, 2003	UK	Adults	IgG- and IgA - ELISA; EMA-ME	1,200	0.01917	0.01000	22/23 biopsied
West, 2003	U.K.	Adults	IgA EMA-ME, IgA-tTGA	7,527	0.01156		
Rutz, 2002	Switzerland	Children	IgA-EMA-ME, IgA-tTG, IgG-AGA and IgA-AGA	1,450	0.00759	0.00690	10/11 biopsied
Borch, 2001	Sweden	Adults	Biopsy, IgA- and IgG-AGA; IgA-EMA-ME	482	0.01452	0.01867	
Grodzinsky, 1996	Sweden	Adults	IgA-AGA; IgA-EMA	1,866	0.00589	0.00375	Prevalence by IgA-EMA not reported
Ivarsson, 1999	Sweden	Adults	IgA- and IgG-AGA - ELISA, cut-off not recorded; IgA-EMA - ME; serum IgA level	1,894	0.00475	0.00475	
Sjoberg, 1994	Sweden	Adults	IgG- and IgA-AGA	1,537	0.01431	0.00065	13/22 biopsied
Sjoberg, 1999	Sweden	Adults	IgA-AGA, IgA confirmed with EMA-ME	1970	0.00152	0.00152	

EGD=esophagogastroduodenoscopy; IF=immunofluorescence; prevalence expressed as proportion (multiply by 100 for percent, or 100,000 for per 100,000 value)

Table 35 (cont'd): Prevalence of CD by country

Author, year	Country	Age group	Test	Total patients	Prevalence by serology	Prevalence by biopsy	Notes
Carlsson, 2001	Sweden	Children	AGA, EMA, biopsy using Watson capsule	690	0.01884	0.01594	
Riestra, 2000	Spain	Adults	IgG/IgA-AGA, IgA-EMA; the study was conducted as a 1) two-step protocol (determination of IgA/IgG-AGA, if positive measuring IgA-EMA); and a 2) one-step protocol (measuring IgA-EMA)	1,170	0.00171	0.00256	1 CD picked up when AGA and EMA was neg.
Corazza, 1997	Republic of San Marino	Adults	IgA-EMA; biopsy	559	0.00179	0.00179	
Hovdenak, 1999	Norway	Adults	IgA- and IgG-AGA; IgA-EMA	2,069	0.00387	0.00338	
Rostami, 1999	Netherlands	Adults	IgA-EMA	1,000	0.00300	0.00300	
Csizmadia, 1999	Netherlands	Children	IgA-EMA	6,127	0.01224	0.00506	57/75 biopsied
Pittschieler, 1996	Italy	Adults	IgA- and IgG-AGA; IgA-EMA; biopsy	4,615	0.00195	0.00195	38 of 140 biopsied
Trevisiol, 1999	Italy	Adults	IgA-EMA; biopsy	4,000	0.00250	0.00250	
Volta, 2001	Italy	Adults (mostly)	IgA-EMA-HU; biopsy	3,483	0.00574	0.00488	Prevalence of 0.57% (20/3483) if included 3 patients with normal villous but with increased IELs
Catassi, 2000	Italy	Children	IgG-AGA (7 AU); IgA-AGA (15 AU); IgA-EMA indirect IF (1:5 dilution); biopsy	2,096	0.00859		
Catassi, 1996	Italy	Children	IgA- or IgG-AGA; confirmed with EMA and biopsy	17,201	0.00645	0.00477	

EGD=esophagogastroduodenoscopy; IF=immunofluorescence; prevalence expressed as proportion (multiply by 100 for percent, or 100,000 for per 100,000 value)

Table 35 (cont'd): Prevalence of CD by country

Author, year	Country	Age group	Test	Total patients	Prevalence by serology	Prevalence by biopsy	Notes
Di Pietralata, 1992	Italy	Children	IgA-AGA; biopsy	3,022	0.00629	0.00596	
Dickey, 1992	Ireland	Adults	IgA AGA	443	0.01129		
Jager, 2001	Germany	Mixed - mostly adults	IgA-AGA, IgG-AGA, IgA-tTG -	150	0.02667		Mixed group of at-risk populations, healthy group used
Kolho, 1998	Finland	Adults	EMA -HU	1,070	0.01028	0.00748	
Maki, 2004	Finland	Children	IgA and IgG tTG; IgA and IgG EMA - IF; total serum IgA; HLA DR, DQ2 and DQ8	3,654	0.01259	0.00739	
Collin, 2002	Finland	Mixed - mostly adults	Biopsy	2,974	0.00605		
Weile, 2001	Denmark and Sweden	Adults	Serum IgA: IgG-AGA; IgA-AGA, cut-off >40 units; EMA; in cases of IgA <0.07g/L, IgG-AGA was analyzed	1,573	0.00254		

EGD=esophagogastroduodenoscopy; IF=immunofluorescence; prevalence expressed as proportion (multiply by 100 for percent, or 100,000 for per 100,000 value)

Table 36: Prevalence of CD by serological screening test

Screening test	Age group	Number of studies	Total patients	Prevalence range
Primary biopsy	Adults	2[207,210]	4,723	0.00515 - 0.00605
IgA AGA	Overall	2[223,229]	3,465	0.00629 - 0.01129
	Adults	1[229]	443	0.01129
	Children	1[223]	3,022	0.00629
IgA / IgG AGA	Adults	1[216]	1,537	0.01431
IgA AGA - IGA EMA	Overall	6[208,209,211,217,219,226]	8,831	0.00152 - 0.01884
	Adults	5[208,209,211,217,219]	6,999	0.00152 - 0.01884
	Children	1[321]	1,823	0.00823
IgA/IgG AGA – IgA EMA	Overall	7[212,213,218,220,221,224,227]	30,648	0.00195 - 0.01917
	Adults	5[212,213,218,224,227]	11,351	0.00195 - 0.01917
	Children (Italy)	2[220,221]	19,297	0.00645 - 0.00859
IgA/IgG AGA – IgA tTG	Mostly adults (Germany)	1[234]	150	0.02667
IgA EMA	Overall	7[126,214,222,225,230-232]	17,409	0.00171 – 0.01224
	Adults	7[126,214,222,225,230,231]		0.00171 – 0.01028
	Children (Netherlands)	1[232]	6,127	0.01224
IgA EMA – IgG tTG	Overall	4[206,215,228,233]	16,757	0.00312 - 0.01259
	Adults (USA, UK)	2[206,228]	10,372	0.00949 - 0.01156
	Children	3 (includes Fasano Child Group)[206,215,233]	6,385	0.00312 - 0.01259
Note: Country of study was indicated when possible; prevalence expressed as proportion (multiply by 100 for percent, or 100,000 for per 100,000 value)				

Table 37: Prevalence of CD by statistical percentiles

Percentiles	Serology	Biopsy
5	.0016255	.0007378
10	.0018050	.0015761
25	.0030919	.0025321
50	.0063702	.0047672
60	.0084439	.0050768
75	.0117290	.0071429
80	.0125193	.0074416
90	.0184088	.0147536
95	.0225417	.0183992
100	.0266667	.0186722
Minimum	.00152	.00065
Maximum	.02667	.01867

Prevalence expressed as proportion (multiply by 100 for percent, or 100,000 for per 100,000 value)

Figure 26: Frequency distribution of prevalence of CD by serology among included studies

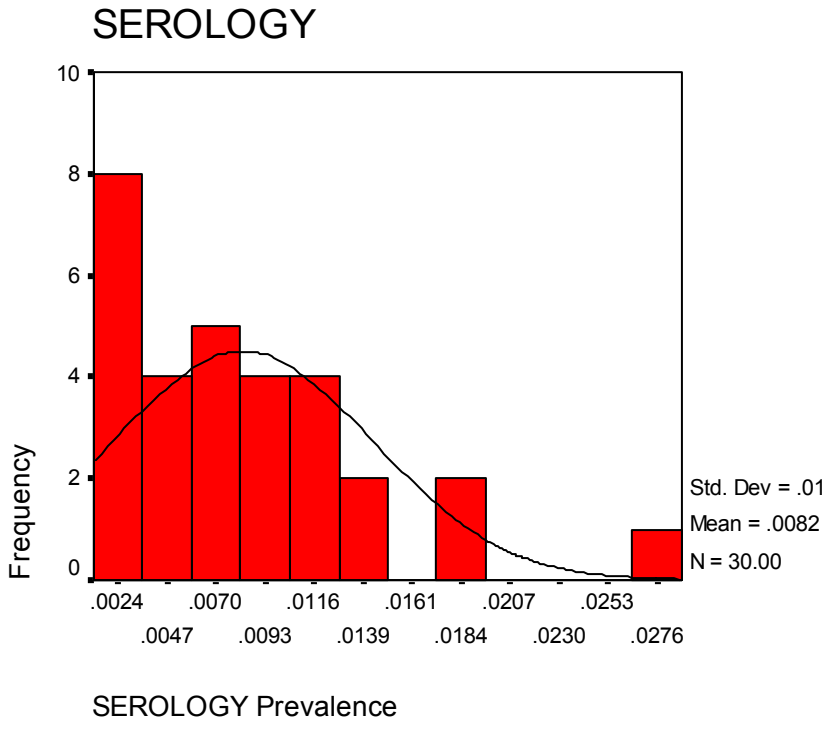


Figure 27: Frequency distribution of prevalence of CD by biopsy among included studies

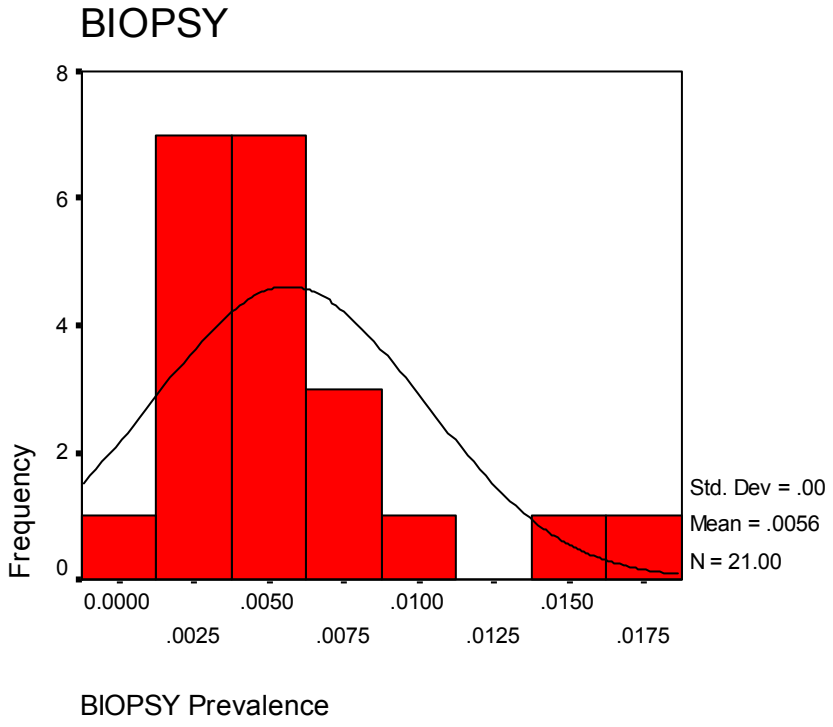


Figure 28: Prevalence of CD by country

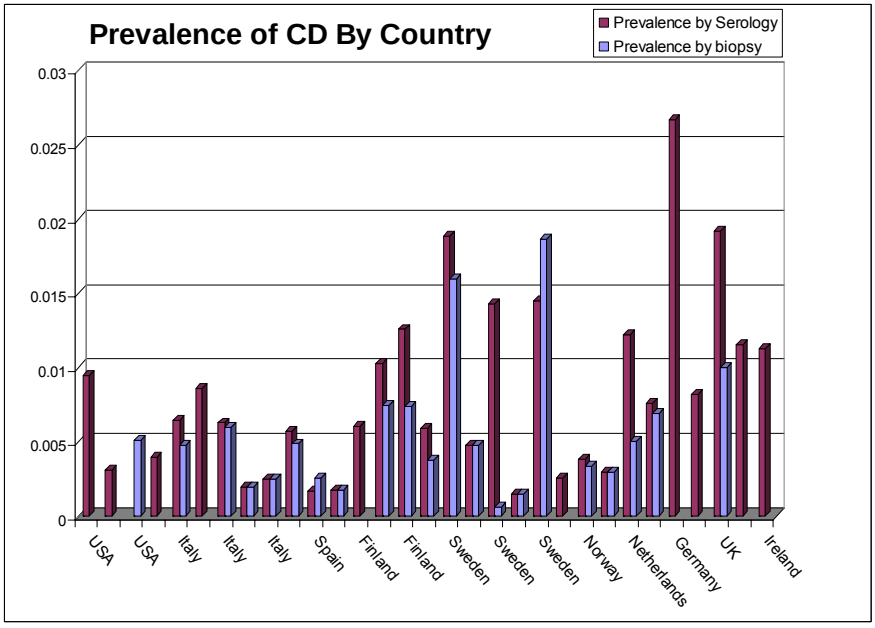
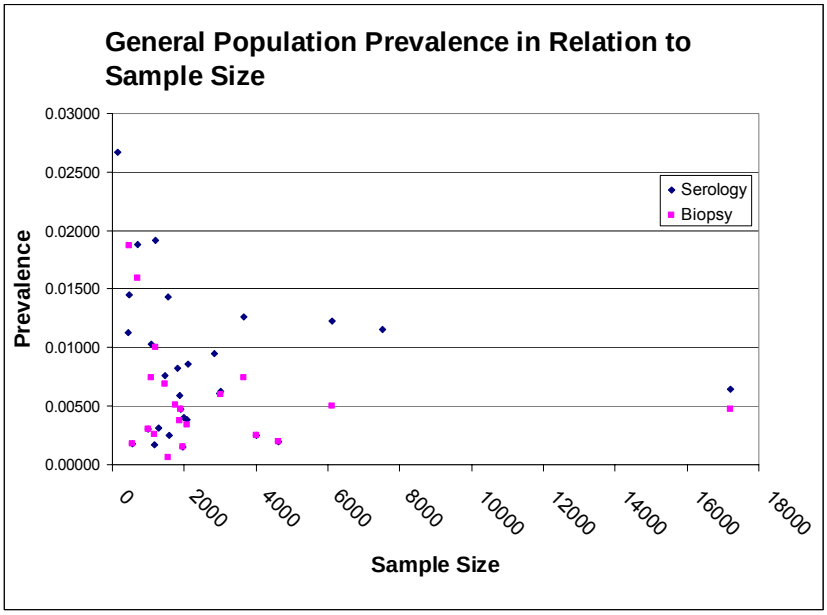


Figure 29: General population prevalence in relation to sample size



Prevalence of CD in Patients with Suspected CD

Adults: The prevalence of CD in adults suspected of the diagnosis was reported in four studies (Evidence Table 5, Appendix I; Table 38); three from Italy,[300,301,303] and one from the US.[206] The following reasons for suspecting a diagnosis of CD were documented: anemia, persistent iron deficiency, bowel disturbances, chronic intermittent diarrhea, abdominal pain, constipation, dyspepsia, severe malabsorption, tiredness and weight loss, mineral metabolism deficiencies, osteoporosis, arthralgias, arthritis, dermatitis, hypertransaminasemia, type I diabetes mellitus, infertility, and gluten intolerance in childhood not further investigated.

All three Italian studies were from referral centers, and intestinal biopsies were performed on all suspected cases, which cumulated to 347. The prevalence of CD was very high in these series, i.e., 43%,[300] 50%,[301], and 12%.[303]

In a large study of prevalence of CD in at-risk and not-at-risk individuals in the US, a total of 1,910 adults with CD-associated symptoms or disorders underwent serological testing with EMA. Fifteen of the 28 EMA-positive subjects (53.6%) consented to a biopsy, which was confirmatory in all cases.[206] The source of these patients and their mode of recruitment/referral were not reported. Based on the EMA result, the prevalence of CD in these adults with suspected CD was 1.5%.

Children: The prevalence of CD in children suspected of the diagnosis was reported in nine studies (Table 38); three from Canada,[304,305,307] two from the US,[206,324] and one each from Denmark,[302] England,[308] Italy,[309] and New Zealand.[306] The following reasons for suspecting a diagnosis of CD were documented: abdominal pain,[238,304,305,307,309] diarrhea,[238,304,305,308] failure to thrive/short stature,[206,238,304-306,309] weight loss,[305] vomiting,[304,305] abdominal distension,[304,305]

chronic GI symptoms,[306] inflammatory bowel disease,[304] family history of CD,[238,304,306,309] type I diabetes mellitus,[206,238,306] iron deficiency anemia (IDA),[309] thyroid disease,[304] trisomy 21,[206,304,309] as well as enamel hypoplasia, recurrent aphtous stomatitis, autoimmune diseases, IgA deficiency, and occult hypertransaminasemia.[309]

Five of the eight studies came out of referral centers where all suspected cases (cumulating to 978) were biopsied.[302,304-306,308] The prevalence of CD in these children ranged from 4.6%[306] to 17%.[305]

In a case-finding study among 26 family pediatricians in Italy, 240 children were screened with EMA based on the presence of risk factors, and 18 diagnoses of biopsy-proven CD were made, resulting in a prevalence of 7.5%.[309]

Three studies, two American[206,238] and one Canadian,[307] reported the prevalence of CD in children with related symptoms or conditions based on EMA testing. The cumulative number of children was 2,426, and the prevalence ranged from 1.1% in the Canadian study of children with chronic abdominal pain,[307] to 4.0% in the large American study of CD prevalence in at-risk and not-at-risk populations.[206]

All ages: Hin et al., performed a case-finding study through nine primary care clinics of central England that served a total population of 70,000 (Table 38).[299] A thousand patients were enrolled for serological screening, satisfying the following entry criteria: irritable bowel syndrome, anemia, family history of CD, malabsorption symptoms, diarrhea, fatigue, thyroid disease, diabetes mellitus, weight loss, short stature, failure to thrive, epilepsy, infertility, arthralgia, or eczema. The mean age of the screened subjects was 42.8 years; 5.3% were aged under 10, and 3.1% were aged 80 to 90 years. Thirty patients were EMA-positive, all of whom were confirmed by biopsy to have some enteropathy (90% had subtotal or total villous atrophy), and only one out of 30 patients had only IELs in the absence of villous atrohpy. The mean age of the 30 cases with CD was 42.8 years, and there was only one child diagnosed with CD. The prevalence of CD was 3.0%.

Table 38: Included studies for prevalence of CD in patients with suspected CD

Study, year; country	Clinical setting	Age group	Dx criteria	N tested	Prevalence (%)
Bardella, 1991; Italy	Referral center	Adults	Biopsy	60	43.3
Bardella, 2001; Italy	Referral center	Adults	Biopsy	80	50.0
Carrocio, 2002; Italy	Referral center	Adults	Biopsy	207	11.6
Fasano, 2003; USA	Not reported	Adults	EMA	1,910	1.5
Bode, 1993; Denmark	Referral center	Children	Biopsy	191	7.3
Day, 2000; New Zealand	Referral center	Children	Biopsy	153	4.6
Thomas, 1992; England	Referral center	Children	Biopsy	381	7.9
Chan, 2001; Canada	Referral center	Children	Biopsy	77	13.0
Chartrand, 1997; Canada	Referral center	Children	Biopsy	176	17.0
Ventura, 2001; Italy	Community pediatricians	Children	Biopsy	240	7.5
Fitzpatrick, 2001; Canada	Community pediatricians	Children	EMA	92	1.1
Fasano, 2003; USA	Not reported	Children	EMA	1,326	4.0
Hill, 2000; USA	Referral center	Children	EMA	1,008	2.5
Hin, 1999; England	Community practice	All ages	Biopsy	1,000	3.0

Prevalence of CD in with Type I Diabetes

The literature search identified 36 studies that assessed the prevalence of CD in patients with type I diabetes (insulin-dependent diabetes mellitus [IDDM]).[191,192,234,250-282] Two sets of duplicate publications were identified.[191,192,277,282] The publications with the most complete data sets were used.[277,282] Of the 34 unique studies (Evidence Table 6, Appendix I; Table 39), seven were conducted in an adult population,[257,263,266,270,273,277,279] 21 in a child population,[250-252,254-256,260-262,264,265,267,271,272,274-276,278,280-282] and six were conducted in a mixed population of adults and children.[234,253,258,259,268,269]

All the included studies initially screened the study population with one or more antibodies. Three studies did not confirm positive serology with biopsy,[265-267] whereas in nine studies confirmatory biopsies were performed in less than 75% of the screened-positive patients.[253,259,264,269,272,274,277-279] These studies were not included in the pooled estimates of the prevalence of CD by biopsy. All the studies that reported biopsy criteria used partial villous atrophy or greater to define CD.

For all the included studies, the minimum prevalence of CD in IDDM by serology was 1% and the maximum was 12%. By biopsy, the minimum and maximum prevalence was 1% and 11%, respectively. Within a given study, the prevalence by serology was almost uniformly greater than the prevalence by biopsy, as would be expected. Table 39 (individual studies) and Table 40 (pooled summaries) list the study details, the individual study estimates of CD prevalence and the pooled estimates of prevalence when appropriate.

The prevalence of CD in adults was assessed in seven studies.[257,263,266,270,273,277,279] Six of these studies used IgA EMA as the screening test,[257,263,266,270,273,279] whereas the largest study used IgA- and IgG-AGA, followed by EMA for confirmation.[277] In this last study, EMA confirmation was positive in 22 of the initially screened sample of 848 patients (2.6%), but biopsy confirmation was only performed in 14 of these patients, making the estimate of 0.83% prevalence by biopsy unreliable. The second largest study (n=509) did not confirm the EMA-positive patients with biopsy, and demonstrated the lowest prevalence of CD by EMA (1.4%) of all of the studies.[266] In another study of 185 patients,[279] the prevalence of CD by EMA was 4.9%, but only five of nine screen-positive patients were biopsied, making the prevalence of 2.2% (4/185) by biopsy a likely underestimation since four of the five biopsied EMA-positive patients were diagnosed with CD. A small study of 62 patients used biopsy as the screening test and found the prevalence of CD to be 11.3%, which is the highest prevalence of the group.[263] The remaining studies had uniform biopsy confirmation.[257,270,273] In these studies the prevalence of CD by EMA ranged from 3.1% to 7.9%, and the prevalence of CD by biopsy ranged from 2.6% to 6.4%.

Twenty-one studies assessed the prevalence of CD in children with IDDM.[250-252,254-256,260-262,264,265,267,271,272,274-276,278,280-282] Six of these studies used IgA-AGA or -AGA in combination with either IgG-AGA or other antibody tests.[254,256,267,271,274,276] The largest study tested 776 children with AGA and ARA (reticulin antibodies), and found a prevalence of CD by serology of 9.8%.[274] However, only 35 of 76 serology-positive patients were biopsied, making the reported prevalence by biopsy of 2.5% a likely underestimation. A single study of 459 patients that used IgA-AGA as the screening test found the prevalence of CD by serology to be 4.1%, and the prevalence of CD by uniform biopsy confirmation to be 4.6%.[276] The second largest study (n=498) used a combination of IgA- and IgG-AGA, and found a prevalence of CD by serology of 6.0% and a prevalence of CD by biopsy of 3.2%.[254] Two other studies that used IgA and IgG-AGA[271] or paired IgA-AGA measurements,[256] found a very similar prevalence by serology of 10.7% and 8.5%, respectively, and a prevalence by biopsy of 3.95% and 2.8%, respectively. The last study in this group did not perform biopsy confirmation of the IgA- and IgG-AGA derived prevalence of 3.76%.[267]

Seven studies used IgA-EMA to screen for CD in children with IDDM.[251,252,255,260,272,275,281] One Hungarian study of 205 children demonstrated a relatively high prevalence by serology and biopsy of 11.7% and 8.3%, respectively,[252] whereas an Austrian study of 403 children demonstrated a relatively low prevalence by serology and biopsy of 3.0% and 1.5%, respectively.[275] A study by Rossi et al.[272] from the US demonstrated a prevalence of CD of 4.7%. The remaining studies demonstrated fairly consistent results, with the prevalence of CD by serology ranging from 5.5% to 7.8%, and the prevalence by biopsy ranging from 3.3% to 6.5%.[251,255,260,281]

Three studies used IgA-tTG either alone[264] or in combination with IgG-tTG.[265,278] IgA-tTG was used alone in a study of 503 children which demonstrated a prevalence by serology of 4.4%. Ten of the 23 serology-positive patients did not undergo biopsy confirmation, making the

reported prevalence of 1.7% a likely underestimation. Of the two studies that used IgA- and IgG-tTG, the first did not perform biopsy confirmation and reported a prevalence of CD by serology of 8.4%,[265] whereas, the other found a prevalence of CD by serology of 6.3%, and by biopsy of 2.9%, although only eight of 13 serology-positive patients underwent biopsy.[278]

Five studies used a combination of IgA-EMA and one or more other antibodies, to assess the prevalence of CD in children with IDDM.[250,261,262,280,282] In three studies, EMA was combined with AGA,[250,261,280] in one it was combined with tTG,[262] and in the one it was combined with AGA and tTG.[282] In one study, only the confirmed biopsy prevalence of 8.3% was reported.[280] Overall, this group reported prevalences by serology ranging from 5.0% to 9.6%, and by biopsy ranging from 3.7% to 8.6%.

The remaining six studies assessed the prevalence of CD in a mixed-age population of patients with IDDM.[234,253,258,259,268,269] One study of 1,785 patients found the prevalence of CD by IgA AGA to be 4.1%. In this study, only 49 of 73 screen-positive patients underwent biopsy confirmation, making the reported prevalence by biopsy of 0.73% an underestimation.[269] Another large study of 1,114 patients used IgA and IgG AGA as an initial screen of screen-positive patients, and then performed a second level screen with IgA EMA before moving on to biopsy.[259] The EMA confirmed prevalence of CD was 4.9%, whereas, the reported biopsy confirmed prevalence was a relatively high 5.7%. In this study, 78 of 121 initial AGA-positive patients underwent biopsy, suggesting that most of the EMA-positive patients were biopsied.

Among the two studies that used IgA EMA as the screening test in a mixed-age population, the prevalence of CD by serology was 2.3%[258] and 5.7%.[268] It was unclear in the first study how the final confirmed prevalence of CD of 0.75% was arrived at,[258] whereas, in the other study the uniformly confirmed biopsy prevalence was 5.7%.[268]

The final two studies assessed the prevalence of CD in a mixed-age population of diabetics using IgA-tTG.[234,253] The prevalence of CD by serology was fairly high in both these studies: 9.6%[234] and 11.5%.[253] The first study did not perform biopsy confirmation, whereas, in the last study only 20 of 98 screen-positive patients were biopsied, making the reported prevalence of CD by biopsy of 1.8% a likely underestimation.

Clinical heterogeneity existed for some subgroups of this analysis making an overall pooled estimate of the prevalence of CD in children and adults with IDDM not entirely possible. However, a summary table (Table 40) is provided which presents the data grouped by age group and screening test, and Figure 30 presents the prevalence of CD in diabetes by study size. For similar studies a weighted pooled prevalence is provided, and individual study data with annotation is presented for studies that could not be pooled.

Table 39: Included studies of prevalence of CD in type I diabetes

Author, year; country	Total patients	Age group	Screening test(s)	First serology	Confirmatory serology	Biopsy proven	Biopsy criteria & description	Prevalence by serology	Prevalence by biopsy	Notes
Li Voon Chong, 2002; UK	509	Adults	EMA	7	None	n/a	None done	0.0138	n/a	
Talal, 1997; USA	185	Adults	EMA	9	None	4	ESPGAN	0.0486	0.0216	Only 5/9 biopsied
Rossi, 1993	211	Children, some adults	EMA	10	None	3	ESPGAN	0.0474	0.0142	Only 3/10 biopsied
Kaukinen, 1999; Finland	62	Adults	EMA		None	7	ESPGAN	0.0000	0.1129	
Sjoberg, 1998; Germany	848	Adults	AGA - IgG or IgA; EMA	258	22	7	Marsh	0.0259	0.0083	Only 14/22 biopsied
Sategna-Guidetti, 1994; Italy	383	Adults	EMA	12	None	10	Roy-Choudhury	0.0313	0.0261	10/12 biopsied
Rensch, 1996; USA	47	Adults	EMA	3	None	3	Loss of villous architecture, crypt hyperplasia, and increased IELs	0.0638	0.0638	
Frazer-Reynolds, 1998; Canada	263	Children	EMA	17	None	12	Carey capsule; Marsh criteria;	0.0646	0.0456	17/19 biopsied
Gillett, 2001; Canada	233	Children	EMA or AGA	19	None	14	Not reported	0.0815	0.0601	18/19 biopsied
Hansen, 2001; Denmark	104	Children	EMA or tTG	10	None	9	Partial or total villous atrophy, crypt hyperplasia and IEL infiltration	0.0962	0.0865	9/10 biopsied
Saukkonen, 1996; Finland	776	Children	AGA or ARA	76	None	19	Not reported	0.0979	0.0245	Only 35/76 biopsied
Spiekerkoetter, 2002; Germany	205	Children	tTG IgA or IgG	13	None	6	Marsh	0.0634	0.0293	Only 8/13 biopsied

Table 39 (cont'd): Included studies of prevalence of CD in type I diabetes

Author, year; country	Total patients	Age group	Screening test(s)	First serology	Confirmatory serology	Biopsy proven	Biopsy criteria & description	Prevalence by serology	Prevalence by biopsy	Notes
Arato, 2003; Hungary	205	Children	EMA	24	None	17	n/r	0.1171	0.0829	
Barera,1991; Italy	498	Children	AGA IgA then if neg IgG AGA	30	None	16	Subtotal villous atrophy	0.0602	0.0321	22/30 biopsied
Barera, 2002; Italy	273	Children	EMA, second EMA	15	10	9	Marsh; type II or III lesion	0.0549	0.0330	
Valerio, 2002; Italy	383	Children	EMA or IgG AGA	n/r	None	32	ESPGAN	n/r	0.0836	
Carelo, 1996; Spain	141	Children	IgA AGA if positive on two occaions	12	None	4	Subtotal villous atrophy	0.0851	0.0284	
Roldan, 1998; Spain	177	Children	IgA, IgG AGA, (and known cases, and some tested with EMA)	19	None	7	ESPGAN	0.1073	0.0395	Mixed group diagnosed by different means
Juan, 1998; Spain	93	Children	EMA	7	None	6	ESPGAN	0.0753	0.0645	
Sigurs, 1993; Sweden	459	Children	AGA	19	None	21	Watson Capsule	0.0414	0.0458	18/19 biopsied included known CD
Agardh, 2001; Sweden	162	Children	AGA, EMA, or tTG IgG or IgA	8	8	6	As described by Carlsson et al. 1999, Pediatrics 103:1248	0.0494	0.0370	Only 6 of 8 biopsied
Acerini, 1998; UK	167	Children	EMA or AGA	11	None	8	ESPGAN	0.0659	0.0479	9/11 biopsied
De Block, 2001; Belgium	399	Mixed	EMA	9	None	3	No biopsy performed	0.0226	0.0075	Unclear how the 3 cases confirmed

Table 39 (cont'd): Included studies of prevalence of CD in type I diabetes

Author, year; country	Total patients	Age group	Screening test(s)	First serology	Confirmatory serology	Biopsy proven	Biopsy criteria & description	Prevalence by serology	Prevalence by biopsy	Notes
Jager, 2001	197	Mixed	tTG	19	None		n/r	0.0964		
De Vitis, 1996; Italy	1114	Mixed	IgA, IgG then IgA EMA	121	55.00	63	Marsh - "villous atrophy"	0.1086	0.0566	78/121 biopsied
Not, 2001; Italy	491	Mixed	EMA	28	None	28	Intestinal biopsy; Marsh's modified classification	0.0570	0.0570	
Bao, 1999; USA	847	Mixed	tTG	98	None	15	n/r	0.1157	0.0177	Only 20/98 biopsied
Kordonouri, 2000; Germany	520	Mixed - mostly children	tTG	23	None	9	Marsh criteria	0.0442	0.0173	10/23 not biopsied
Aktay, 2001; USA	218	Mixed - mostly children	EMA	17	None	10	Partial or total villous atrophy, inflammation in lamina propria with increased IELs, and hyperplasia of crypts; classified as partial or total villous atrophy	0.0780	0.0459	14/17 biopsied
Cronin, 1997; Ireland	101	Mixed - mostly adults	EMA	8	None	5	n/r	0.0792	0.0495	
Schober, 2000; Austria	403	Mixed - mostly children	EMA	12	None	6	Modified Marsh and Crowe; Watson-type capsule	0.0298	0.0149	11/12 biopsied
Lampasona, 1999; Italy	287	Mixed - mostly children	tTG IgA or IgG	24	None	n/a	No biopsy	0.0836	n/a	

Table 39 (cont'd): Included studies of prevalence of CD in type I diabetes

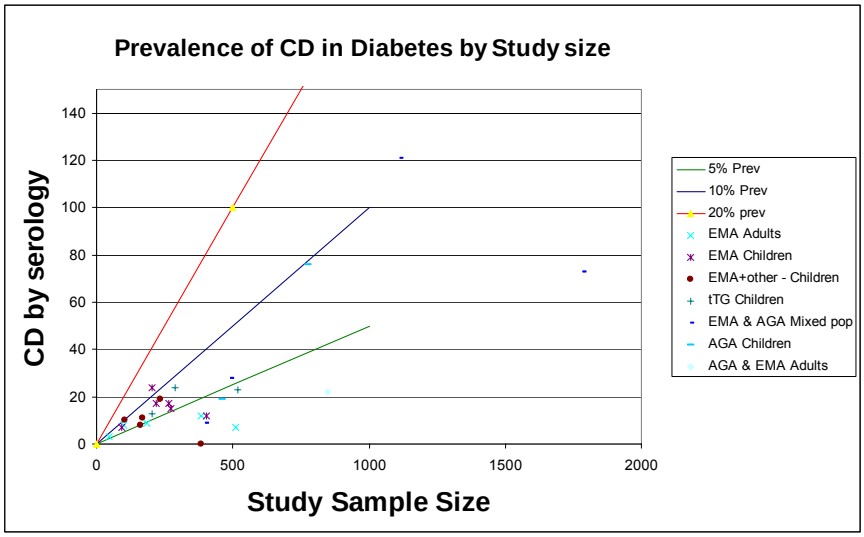
Author, year; country	Total patients	Age group	Screening test(s)	First serology	Confirmatory serology	Biopsy proven	Biopsy criteria & description	Prevalence by serology	Prevalence by biopsy	Notes
Lorini, 1996; Italy	133	Mixed - mostly children	AGA IgA or IgG	5	None	n/a	No biopsy	0.0376	n/a	
Page, 1994; Mixed	1785	n/a	AGA	73	None	13	n/a	0.0409	0.0073	Only 49/73 biopsied

Table 40: Summary of prevalence of CD in type I diabetes by age groups and screening test

Number of studies	Total patients	Age group	Screening test(s)	Prevalence by serology	Prevalence by biopsy
1[277]	848	Adults	AGA - IgG or IgA; then EMA	0.0259	0.0083*
1[266]	509	Adults	EMA	0.0138	n/a
1[279]	185	Adults	EMA	0.0486	0.0216*
1[263]	62	Adults	EMA	n/a	0.1129
3[257,270,273]	531	Adults	EMA	0.0433	0.0339
1[274]	776	Children	AGA or ARA	0.0979	0.0245*
1[276]	459	Children	AGA	0.0414	0.0458
4[254,256,267,271]	949	Children	AGA – various combinations	0.0695	0.0331
1[252]	205	Children	EMA	0.1171	0.0829
1[275]	403	Children	EMA	0.0298	0.0149
5[251,255,260,272,281]	1058	Children	EMA	0.0624	0.0437
4[251,255,260,281]	847	Children	EMA	0.0661	0.0437
5[250,261,262,280,282]	1049	Children	EMA - combinations	0.0721	0.0658
1[265]	287	Children	tTG IgA with IgG	0.0836	n/a
1[278]	205	Children	tTG IgA with IgG	0.0634	0.0293*
1[264]	520	Children	tTG	0.0442	0.0173*
1[269]	1785	Mixed	AGA	0.0409	0.0073*
1[259]	1114	Mixed	IgA, IgG-AGA then IgA-EMA	0.0494	0.0566*
1[268]	491	Mixed	EMA	0.0570	0.0570
1[258]	399	Mixed	EMA	0.0226	0.0075†
1[234]	197	Mixed	tTG	0.0964	n/a
1[253]	847	Mixed	tTG	0.1157	0.0177*

*large proportion of serology-positive patients not biopsied,[253,259,264,269,272,274,277-279] these were not included in the pooled analysis of prevalence by biopsy
**no description of how diagnosis made – result not pooled

Figure 30: Prevalence of CD in diabetes by study size



Prevalence of CD in Relatives of Patients with CD

There were 18 studies on the risk of CD in first-degree relatives of patients with biopsy-proven CD,[129,167,206,235-249], four of which also provided data on the risk of CD in second-degree relatives (Evidence Table 7, Appendix I; Table 41).[206,235,238,239]

First-degree relatives: First-degree relatives were directly evaluated with small bowel biopsy in five studies; three were performed in England in the 1970's,[242,243,245] and two in Finland during the 1990's.[129,167] The biopsy criteria for a diagnosis of CD was not reported in one study,[243] and implied at least some degree of villous atrophy in the other four.[129,167,242,325] The percent of all at-risk family members that were studied varied from 34%[245] to 100%.[243] The study size varied between 29[242] and 182,[245] and the cumulative number of patients tested was 494. The prevalence of CD among first-degree relatives undergoing intestinal biopsy varied from 5.5%[243] to 22.5%;[245] the pooled prevalence was 16%.

Serological screening of the first-degree relatives of patients with biopsy-proven CD was performed in 12 studies.[206,235-237,239-241,244,246-249] In seven of those studies, intestinal biopsy was performed on at least 80% of the subjects who tested positive serologically, i.e., in 84 % of subjects in one study,[237] and in 100% of subjects in the other six studies.[236,239,244,247-249] Serological screening was performed with AGA alone in one study,[236] whereas, the other six studies used EMA, either alone[239] or in combination.[237,244,247-249] Six studies used criteria implying some degree of villous atrophy,[236,237,239,244,247,248] whereas, one study included cases with Marsh I changes.[249] The study size varied from 92[248] to 943[239] subjects, for a cumulative number of 2,607 subjects. For the studies that required some degree of villous atrophy for diagnosis, the prevalence varied from 4%[236] to 12%,[248] and the mean prevalence was 7.6%.

However, when Marsh I lesions were also considered diagnostic, the prevalence of CD among first-degree relatives was reported at 44.1%.[249]

In five other studies of first-degree relatives,[206,235,240,241,246] confirmatory biopsy was not routinely performed (available in 9%[246] to 58%[241] of the cases), and the reported prevalence of CD was based on the serology results. EMA was used for serological screening in all of these studies, either alone,[206,240] or in combination with AGA [241,246] or tTG.[235]

Two of these studies were performed in families where at least two index cases prevailed and are, therefore, reviewed separately.[235,241] Ninety percent of the at-risk populations from these two studies were tested, which represents a cumulative number of 629 subjects. The prevalence of CD among these first-degree relatives from families where there are at least two index cases of known CD or dermatitis herpetiformis (DH) was 9.4%[241] and 17.2%.[235]

The study size of the other three studies varied from 115[240] to 4,508,[206] and the cumulative number of first-degree relatives tested was 5,265. The prevalence of CD among these serology-tested first-degree relatives varied between 2.8%[246] and 4.5%[206] (mean prevalence 4.3%).

Other relatives: One study from the US[238] reported an EMA-based prevalence of 4.7% in 192 first- and second-degree relatives; the prevalence from each of the groups of relatives was not reported separately.

An American study by Book et al.[235] studied the prevalence of CD in second-degree relatives and first cousins of CD sibling pairs (i.e., families with two affected index cases). Eighty-two second-degree relatives and 47 first cousins were tested with EMA and tTG, and the diagnosis was biopsy confirmed in 40% of the cases. The serology-based prevalence was 19.5% in second-degree relatives and 17.0% in first cousins.

Two other studies, one large (n=1,275) American study of prevalence of CD in at-risk and not-at-risk subjects,[206] and one Hungarian study,[239] provided data on the prevalence of CD in second-degree relatives. The EMA-based prevalence of CD in those groups was 2.6% and 5.5%, respectively (mean prevalence 2.7% on a cumulative number of 1,329 second-degree relatives).

Table 41: Prevalence of CD in relatives of CD patients

Study, year; country	Relative Type	Index case	Screening	Dx criteria	N tested	Prevalence (%)
Polvi, 1996; Finland	1st degree	CD in family	Biopsy	ESPGAN	90	20
Holm, 1993; Finland	1st degree	CD in family	Biopsy	Some VA	121	10.7
Robinson, 1971; England	1st degree	CD child in family	Biopsy	Some VA	29	10.3
Rolles, 1974; England	1st degree	CD child in family	Biopsy	Not reported	72	5.6
Stokes, 1976; England	1st degree	CD in family	Biopsy	Some VA	182	22.5
Tursi, 2003; Italy	1st degree	CD in family	Biopsy	Marsh I-IV	111	44.1
Corazza, 1992; Italy	1st degree	CD adult in family	AGA	Some VA	328	4.0
Pittschieler, 2003; Italy	1st degree	CD in family	EMA, TTG	Some VA	92	12.0
Rostami, 2000; Netherlands	1st degree	CD in family	AGA, EMA, Hx	ESPGAN	338	10.9
Hogberg, 2003; Sweden	1st degree	CD in family	AGA, EMA, TTG	Some VA	120	8.3
Korponay-Szabo, 1998; Hungary	1st degree	CD in family	EMA	Some VA	943	9.1
Farre, 1999; Spain	1st degree	CD in family	AGA, EMA	Some VA	675	5.6
Kotze, 2001; Brazil	1st degree	CD in family	EMA	+ve serology*	115	3.5
Fasano, 2003; US	1st degree	CD in family	EMA	+ve serology	4,508	4.5
Vitoria, 1994; Spain	1st degree	CD in family	AGA, EMA	+ve serology	642	2.8
Mustalahti, 2002; Finland	1st degree	>1 DH or CD sib	AGA, EMA	+ve serology	466	9.4
Book, 2003; US	1st degree	CD sib pairs	EMA, TTG	+ve serology	163	17.2
Hill, 2000; US	1st & 2nd degree	CD in family	EMA	+ve serology	192	4.7
Fasano, 2003; US	2nd degree	CD in family	EMA	+ve serology	1,275	2.6
Korponay-Szabo, 1998; Hungary	2nd degree	CD in family	EMA	+ve serology	54	5.6
Book, 2003; US	2nd degree	CD sib pairs	EMA, TTG	+ve serology	82	19.5
Book, 2003; US	1st cousins	CD sib pairs	EMA, TTG	+ve serology	47	17.0

*EMA titre $\geq$ = 1/5
VA = villous atrophy; DH = dermatitis herpetiformis

Prevalence of CD in Patients with IDA

Twelve studies were identified that allowed for the extraction of the prevalence of CD among patients who were evaluated for anemia (Evidence Table 8, Appendix I; Table 42).[283-294] In all of these, IDA was the primary focus of the study or made up the cause of anemia in the majority of the study patients. Tables 42 and 43 summarize the characteristics of the included studies.

Three studies assessed the prevalence of CD in IDA patients with GI symptoms.[283,288,290] The prevalence of CD in these studies ranged from 10.3% to 15% of the studied group. One small study assessed the prevalence of CD in a group of patients who had IDA but no identified GI source.[286] In this study, the prevalence of CD by AGA and confirmed by EMA was 30%.

In another study, the authors assessed the prevalence of CD in pre-menopausal women with IDA.[293] The overall prevalence of CD in this population was found to be 12.9% by tTG, and 8.5% after biopsy confirmation. CD was found in 1 of 22 (4.5%) of women with heavy periods, and 4 of 18 (22%) of women with normal menstrual flow.

Four studies assessed the prevalence of CD in asymptomatic IDA patients by serology.[285,287,291,292] Two of these used EMA screening,[291,292] whereas the other two initially screened with AGA and then confirmed with EMA.[285,287] The prevalence of CD in this group ranged from 2.3% to 5.0%. Another three studies assessed the prevalence of CD by biopsy in asymptomatic IDA patients, finding it to be between 2.9% and 6%.[284,289,294]

Table 42: Included studies of CD in adult patients with anemia

Author, year; country	No. of pts	Age group	Population	Anemia type	Screening test	First serology	Confirmatory serology	Biospsy proven	Biopsy criteria	Prevalence by serology	Prevalence by biopsy
Akerman, 1996; Israel	93	Adult - some teens	Out-patients with IDA (50% symptomatic)	IDA	EGD/ biopsy			13	Subtotal or greater villous atrophy	n/a	0.139785
Annibale, 2001; Italy	71	Adults	Asymptomatic	IDA	EGD/ biopsy			4	Marsh	n/a	0.056338
Corazza, 1995; Italy	200	Adults	Referred to hematology	IDA	IgA/IgG-AGA then EMA then biopsy	16	10	10	Not mentioned	0.05	0.05
Dickey, 1997; UK	10	Adults	Asymptomatic, previously investigated no gross GI cause found	IDA	IgA AGA then EMA	4	3		Endoscopic biopsy; criteria n/r; finding of villous atrophy and IELs in duodenal biopsy	0.3	n/a
Howard, 2002; UK	258	Adults	IDA identified through lab	IDA, folate	IgA/IgG-AGA and EMA then biopsy	28		12	Not applicabe	0.10852713	0.046512*
Kepczyk, 1995; USA	39	Adults	Mostly symptomatic out-patients with IDA	IDA	EGD/ biopsy			4	Villous atrophy, crypt hyperplasia, inflammatory infiltrate	n/a	0.102564
McIntyre, 1993; UK	50	Adults	Out-patients with IDA	IDA	EGD/ biopsy			3	Not reported	n/a	0.06

*24/28 biopsied

Table 42 (cont'd): Included studies of CD in adult patients with anemia

Author, year; country	No. of pts	Age group	Population	Anemia type	Screening test	First serology	Confirmatory serology	Biospsy proven	Biopsy criteria	Prevalence by serology	Prevalence by biopsy
Oxentenko, 2002; USA	113	Adults	Undergoing EGD for IDA	IDA	EGD/ biopsy			17	CD was defined as total or partial villous atrophy with IELs	Not applicable	0.150442
Ransford, 2002; UK	484	Adults	Referred to hematology	IDA	EMA then EGD/ biopsy	17		11	Revised ESPGAN; duodenal histologic changes were graded according to Marsh I-III	0.03512397	0.022727[†]
Unsworth, 2000; UK	483	Adults	Blood donors	Anemia unspecified	IgA-EMA then biopsy	32		22	n/r	0.06625259	0.045549[‡]
Annibale, 2003; Italy	59	Adult	Pre-menopausal women with IDA	IDA	IgA tTG then biopsy	7		5	Marsh	0.11864407	0.084746[**]
Van Mook, 2001; The Netherlands	35	Adult	Asymptomatic	IDA	EGD / biopsy			1	Marsh I	Not applicable	0.028571

[†]5 Marsh I identified by CD3
[‡]25/32 biopsied
[**]5/7 biopsied; 30 had heavy periods; CD in 1/22 with heavy periods, and 4/18 with normal periods

Table 43: Summary of prevalence of CD in adult patients with anemia by population and screening test

No. of studies	Total patients	Population	Screening test(s)	Prevalence by serology	Prevalence by biopsy
3[283,288,290]	245	Symptomatic IDA	Biopsy	n/a	0.139
1[286]	10	Asymptomatic, previously no gross GI cause found investigated	IgA-AGA then EMA	0.3	n/a
1[293]	59	Pre-menopausal women with IDA	IgA-tTG then Biopsy	0.119	0.085
4[285,287,291,292]	1,425	Asymptomatic serology screened	IgA-EMA, or-AGA followed by EMA; all biopsy confirmed	0.061	0.039
3[284,289,294]	156	Asymptomatic biopsy screened	Biopsy	n/a	0.051

Prevalence of CD in Patients with Low Bone Mineral Density (BMD)

Four articles were identified that assessed the prevalence of CD in patients with low BMD (Evidence Table 9, Appendix I).[295-297,326] The study characteristics and definitions used to define low BMD, osteopenia, and CD are presented in Table 44. Three of these studies determined BMD using dual energy X-ray absorptiometry (DXA), and defined osteoporosis as a BMD less than 2.5 standard deviations from the peak bone mass of sex-matched control,[295,297,326] whereas, the other used single photon absorptiometry (SPA).[296] One study included patients with non-traumatic fractures,[295] whereas, in the others, idiopathic osteoporosis was sufficient for inclusion. All four studies used serology screening with biopsy confirmation of screen-positive patients. Three studies relied on AGA testing as the initial screen[295,296,326] followed by biopsy,[296] or further confirmatory serology testing with EMA[295] or tTG[326] prior to biopsy. The final study screened with EMA-ME, with positive screens moving on to biopsy.[297] Two studies defined the biopsy criteria for CD and used a fairly standard but rigid requirement of subtotal or greater villous atrophy.[295,297]

In the studies that used this test as the initial screen, AGA was positive in 6% to 21% of the patients with osteoporosis. However, in these studies CD was confirmed by biopsy in only 0.9% to 3% of patients.[295,296,326] The study that used EMA-ME as a screening test identified potential CD cases in 7.3% of patients, but none of these met the authors' biopsy criteria for CD.[297]

Table 44: Prevalence of CD in patients with low BMD

Author, year; country	Population	BMD definition	Test	Prevalence
Lindh, 1992, Sweden	92 consecutive patients with idiopathic osteoporosis screened for CD; 91% F (mean age 66+-12 Y); and 9 M (mean age 50+-12 Y)	Bone mineral content by photon absorptiometry (SPA) of non-dominant forearm; criteria n/r	IgA-AGA ELISA; cut-off was 2 SD above the mean of blood donors; confirmatory biopsy in 6 - criteria n/r	11/92 (12.0%) AGA +ve.; 3% (3/92) biopsy confirmed Mean proximal SPA 0.97 g/cm^2 Mean distal SPA 0.67 g/cm^2
Gonzalez, 2002; Argentina	127 postmenopausal women with osteoporosis; age (Y): mean 68, range 50-82; 747 controls; age (Y): mean 29, range 16-79	History of non-traumatic fractures and lumbar spine and/or femoral neck BMD below T-score -2.5 DXA	IgA and IgG-AGA ELISA; cut-off levels: for IgA - 15 AU/mL; for IgG - 20 AU/mL; positives confirmed with IgA-EMA-ME positive at 1:5 dilution; positives confirmed with biopsy in EMA positives; showing villous atrophy, crypt hyperplasia and IEL >30%	1/127, or 7.9 x 1000 (95% CI: 0.2-43.1); test positivity: AGA found in 8 of 127 (6.3%) pts on level 1; 1 of these 8 pts was EMA positive on the 2nd level and eligible for biopsy which established a diagnosis of CD in 1 (0.9%)
Mather, 2001; Canada	Idiopathic low BMD; mean age 57 Y; range 18-86 Y; 81.3% (78) F; 18.7% M (18) All osteopenic; 45/78 F and 13/18 M osteoporotic	DXA Osteopenia: BMD <1 SD of mean sex-matched peak BMD Osteoporosis: BMD <2.5 SD of mean sex-matched peak BMD	IgA- EMA-ME titers of ≥1:10; and biopsy confirmation based on subtotal or greater villous atrophy	7 (7.3%) of 96 pts were EMA +ve; all biopsies were negative based on subtotal or greater villous atrophy prevalence of 0%
Nuti, 2001; Italy	255 females with osteoporosis; mean age 66.6 Y range 36-65 Y	DXA BMD below T-score -2.5	IgA-AGA ELISA-cut-off level of 10 AU/mL-1; IgA-tTg cut-off >22 AU; confirmatory biopsy criteria n/r	53/255 (20.8%) +ve IgG-AGA; 24/53 +ve for tTG antibody (9.4%); intestinal biopsy in 10/24 resulted in 6 (2.4%) with confirmed CD
F=female; M-male; DXA=dual X-ray absorptiometry; Y=years; n/r=not recorded				

Quality Assessment

Using the cross-sectional checklist, the overall quality of reports of the included studies for the Celiac 2 objective, was marginal to fair (Appendix J, Table 2). For example, most of the studies did not report on whether the patients were consecutively enrolled, which could possibly lead to selection bias.

Celiac 3: Risk of Lymphoma in CD

Literature Search

Out of 379 references resulting from the literature search on CD and lymphoma, 150 were initially excluded because they did not directly address this topic (Appendix F). Of the 229 studies that were screened using full reports of the studies, 211 were excluded for the following reasons: review articles (n=73; 19.3% of level 2 articles); did not address the topic (n=33); assessed the risk of CD in lymphoma (n=28); were uncontrolled studies, including surveys (n=53); or, studied the basic mechanisms and the pathogenesis of lymphoma in CD (n=24).

The following eight exclusions were made from the 18 publications that reached level 3 (i.e., eligibility criteria): duplicate publications (n=7);[127,327-332] (for two of these reports,[328,332] patients originated from the same center [i.e., General Hospital, Birmingham]) and the reports were conducted during the same periods as other reports,[329,330,333] and we could not rule out that they were not similar series); data was not extractable (n=1).[334]

The nine controlled studies selected for data extraction were grouped as follows: eight cohort studies,[333,335-341] and one case-control study[342] (Evidence Table 10, Appendix I; Table 45). Mortality data from one controlled study in refractory CD is presented at the end of this section for reference.[343]

Table 45: Included studies for risk of lymphoma in CD

Study, year; country, period	Study type	Participants	Risk of lymphoma	Mortality	Other observations
Cottone, 1999; Sicily, 1980-97	Retrospective cohort	• 228 CD patients • 76% females • mean age at Dx 34.7 • 98% adult Dx • 100% on strict GFD	• Incidence NHL 3.1% • SIR NHL 3.75, p <0.01	SMR all causes 3.8 (1.9-6.7)	
Holmes, 1989; England, 1941-85	Prospective cohort	• 210 CD patients • 55% females • 51% on strict GFD	• Incidence NHL 4.3% • SIR NHL 42.7 (19.6-81.4)	SMR not reported	SIR NHL vs GFD compliance: • Strict GFD 44.4 • Gluten diet 100
Logan, 1989; Scotland, 1979-1986	Prospective cohort	• 653 CD patients • 60% females		Mortality from NHL 2.6% SMR from lymphoma 31 p<0.001 SMR all causes 1.9 (1.5-2.2)	SMR childhood Dx 1.4 (0.4-3.7) SMR adult dx 1.9 (1.5-2.3)
Askling, 2002; Sweden, 1964-94	Retrospective cohort	• 11,019 CD patients • 59% females • Mean age at Dx 17.4 (range 0->70)	• Incidence NHL 0.34% • SIR NHL 6.3 (4.2-125)	SMR from NHL 11.4 (7.8-16) SMR all causes 2 (1.8-2.1)	SIR NHL childhood Dx 1.9 (0.4-5.5) SIR NHL adult Dx 7.0 (5.0-9.5)
Collin, 1996; Finland, 1970-93	Prospective cohort	• 383 CD patients • 73% females • Mean age at Dx 41.8 (range 16-78) • 75% on strict GFD	• Incidence NHL 0.26% • SIR NHL 2.66 (0.07-14.8)		
Corrao, 2001; Italy, 1962-94	Prospective cohort	• 1,072 CD patients • 76% females • mean age at Dx 35.7 (range 18->50) • 59% on strict GFD		SMR from NHL: 69.3 (40.7-112.6) SMR all causes: 2.0 (1.5-2.7)	SMR age 18-29 at Dx: 2.5 (0.5-7.3) SMR age 30-49 at Dx: 2.4 (1.3-4.0) SMR age >50 at Dx: 1.9 (1.3-2.6) SMR strict GFD: 0.5 (0.2-1.1) SMR unlikely GFD: 6.0 (4.0-8.8)
Green, 2003; USA, 1981-2000	Prospective cohort	• 381 CD patients • 64% females • mean age at Dx 44 +/- 18	• Incidence NHL 1.3% • SIR NHL 6.2 (2.9-14)		
Selby, 1979; Australia, 1959-78	Retrospective cohort	• 93 CD patients • 67% females • mean age at Dx 40 (range 14-70)	• Incidence NHL 4.3% • SIR NHL 4.94, p<.0005		
Delco, 1999; USA, 1986-95	Case-control	• 458 CD patients • 4% females	• OR NHL 4.53 (2.01-10.23)		
Dx=diagnosis; SIR=standardized inidence ratio; NHL=non-Hodgkin's lymphoma; SMR=standardized mortality					

Measures of Risk

Eight out of nine studies were cohort studies, either prospective or retrospective. The standardized incidence ratio (SIR) was the most commonly reported measure of association; it was calculated as the incidence observed in the patient cohort divided by the expected incidence from the control population, along with a measure of precision (i.e., its 95% CI). The results were expressed either as SIRs of lymphoma or as the standardized mortality ratio (SMR) from lymphoma (SMR-NHL). The all-cause mortality was also reported in some studies.

It was not possible to pool these measures of risk, since SIRs by definition incorporate variables inherent to each population. The attributable risk (AR), was calculated whenever the incidence rates of NHL in CD patients and in the age-adjusted general population, were available.

Study Characteristics

There were eight cohort studies (five prospective [333,336,338-340] and three retrospective [335,337,341]) and one case-control study.[342] Two studies were from Italy,[335,339] two from the UK,[333,336] two from Scandinavia,[337,338] two from the US,[340,344] and one from Australia.[341] The observation periods varied from 7 years[336] to 44 years (1941-85;[333]), and the mean duration of patient follow-up varied from 6 years[335,339-341] to 18.6 years.[333] Patients were either selected from a national patient register,[336] from hospital discharge databases,[337,342] or represented all consecutive cases from a single[333,335,338,340,341] or multiple[339] institution(s). The cohort sizes varied from 93[341] to 11019;[337] 55% to 76% of patients with CD were female, except for the study by Delco et al.,[342] which used discharge diagnoses databases from the US Veterans Affairs hospitals (4% female CD patients). The mean age at diagnosis of CD was reported in six studies: in four studies, the diagnosis of CD was made almost exclusively in adulthood.[335,338,339,341] The mode of presentation was reported in four studies.[335,338,339,341] Adherence to a GFD was reported in five studies,[333,335,338,339,341] and could be used in the analysis in three of them.[333,338,339] Control data for the cohort studies was derived from local and national mortality data and cancer registers.

Types of Lymphomas

The total number of lymphomas diagnosed in each study and their histological type was not uniformly reported. Of the 84 lymphomas that were mentioned within these nine studies, 64 were referred to as "non-Hodgkin lymphoma (NHL)" not otherwise specified, one as "lymphoma," nine as " enteropathy-associated T-cell lymphoma (ETCL)," five as "B cell lymphoma", two as "large cell lymphoma," one each as a "T-cell other than ETCL," "lymphosarcoma" (currently classified as small cell lymphoma), and "histiocytic medullary reticulosis" (currently termed hairy-cell leukemia). Logan et al.[336] reported that they found "mostly lymphosarcomas (i.e., small-cell lymphomas) or reticulum-cell sarcomas (i.e., large-cell lymphomas) as well as two Hodgkin's lymphomas," whereas, the remaining authors systematically excluded Hodgkin's lymphomas from their respective analyses.

Incidence of Lymphoma and Related Mortality Data

The case definition of CD differed between the reports of institutional series and those derived from database analysis. The results will therefore be presented differently according to each of these two study designs.

Institutional series: By institutional studies, we mean reports on the evolution of cases consecutively diagnosed with CD and followed in one or several selected institution(s) over a specific period. Six out of the nine controlled studies were performed in that setting; in five out of six studies, the data originated from a single referral center.[333,335,338,340,341] The sixth study is the product of a collaborative effort between nine Italian centers.[339] In these studies, all cases were biopsy-proven CD.

Holmes et al., from Birmingham England, reported on a series of 210 biopsy-proven CD patients diagnosed and followed between 1941 and 1985.[333] This series was originally reported by Harris in 1967,[329] and reviewed in 1976[330] and in 1989[333] by Holmes. By this third publication, the authors had excluded all non biopsy-proven cases of CD, as well as the cases of cancer that arose either prior to or within 12 months of diagnosis of CD. The length of follow-up was of a minimum 13 years, 17.4 patient-years for men and 19.4 patient-years for women. Based on the original publication by Harris, we can assume that a large proportion of these patients (80% in Harris' series) were diagnosed with CD in adulthood. There were nine cases of NHL, compared with an expected 0.21, resulting in a SIR-NHL of 42.7 (95% CI: 19.6-81.4), which was the highest reported degree-of-risk for lymphoma among the controlled studies we identified.

Green et al.[340] prospectively followed 381 patients with biopsy-proven CD from New York City, most of whom were of European descent, and diagnosed between 1981 and 2000. The mean age at CD diagnosis was 44 +/- 18 years, and the duration of CD-related symptoms prior to diagnosis was 5 +/- 8 years. The mean follow-up was 6 +/- 11 years, for a total of 1,977 patient-years following the diagnosis of CD. There were a total of nine cases of NHL, occurring any time before or after the diagnosis of CD, leading to an attributable risk of NHL from CD of 120.2 cases per 100,000 patient years. The SIR-NHL, diagnosed at any time, was 9.1 (95% CI: 4.7-13), and the SIR-NHL for any lymphoma diagnosed at least one month after the diagnosis of CD was 6.2 (95% CI: 2.9-14).

Cottone et al.[335] reported on 228 patients with biopsy-proven CD and followed from 1980 to 1997, from a large referral center in Sicily. Ninety-eight percent of the patients had been diagnosed with CD during adulthood and the mean age at diagnosis was 34.7 years. The mean duration of follow-up was 6 years (range: 1 month to 17 years). No case of refractory CD was mentioned. There were seven cases of NHL, compared with an expected number of 1.824 (SIR-NHL of 3.75 ($p<0.01$)). The cumulative incidence of NHL was 3%, compared with an expected of 0.8%, leading to a risk difference or AR of 2.2%. The mean age at diagnosis of lymphoma was 59.4 years, and the mean time from the diagnosis of CD was 6.5 years. Lymphomas occurring prior to or within 6 months of CD diagnosis were excluded.

A large Italian multicenter study by Corrao et al.,[339] prospectively followed 1,072 patients with CD and spanned from 1962 to 1994, totaling 6,444 patient years. The mean follow-up was 6 years, and all patients were diagnosed with CD during adulthood (mean age at diagnosis of CD 35.7 years). The outcomes were strictly measured in terms of mortality data, i.e., mortality from NHL and from all causes. Events occurring at the time of CD diagnosis were included. There

were 16 instances of death from NHL. The SMR-NHL was 69.3 (95% CI: 40.7-112.6), whereas, the SMR of death from all cause (SMR-all cause) was 2.0 (95% CI: 1.5-2.7), showing that the risk of death from NHL in CD is disproportionately elevated.

Selby et al.[341] reported on a series of 93 patients with CD that were followed at a single institution in Australia between 1959 and 1978, for a mean duration of 6 years. Patients presented either during the teenage or adulthood, all were symptomatic at the time of diagnosis, and there were no refractory cases. There were four patients with NHL (simultaneous CD and lymphoma diagnosis included), compared with an expected of 0.081 (SIR-NHL 4.94, $p<0.0005$).

Collin et al.[338] reported on a prospective cohort of 383 patients with CD, diagnosed and followed at a single institution over the 1970-93 period, for a mean follow-up of 8.1 years (3,107 patient years in total). The mean age at diagnosis was advanced: 41.8 years, with a range of 16 to 78 years. Seventy-five percent of the patients adhered to a strict GFD and 82% of patients were symptomatic at the time of CD diagnosis. Simultaneous lymphoma and CD diagnoses were not excluded. There was a single case of lymphoma, compared with an expected 0.4 (SIR-NHL 2.66 [95% CI: 0.07-14.8]). As well, the 10- and 15-year survival of CD patients did not differ significantly from those of the general population.

Large database and register series: Logan et al.[336] reviewed the death certificates of CD patients belonging to a comprehensive register of CD patients that exists in Scotland since 1979, constituting a cohort of 653 CD patients gathered from 1979 to 1986. There were 17 deaths attributed to lymphoma, instead of an expected 0.55. Both Hodgkin and NHL were included, and so were those lymphomas occurring simultaneously to the diagnosis of CD. The SMR-lymphoma was 31 ($p<0.001$), which was disproportionately increased compared with the SMR-all causes, which was 1.9 (95% CI: 1.5-2.2).

Askling et al.[337] reported on the largest CD patient cohort (n=11,019), gathered from a comprehensive Swedish database of hospital discharge diagnoses over 1964 to 1994. It was not possible to ascertain how the diagnosis of CD was made or confirmed. The mean age at diagnosis of CD was 17.4 (range 0 to >70), and the mean follow-up was 9.8 years (range 0-32), for a total of 97,236 patient years. The ascertainment of outcome was achieved through the Swedish cancer register, as well as the register of causes of death. Lymphomas arising prior to or within 12 months of CD diagnosis were excluded, as for the incident lymphomas found at autopsy. There were 38 cases of NHL, and a SIR-NHL of 6.3 (95% CI: 4.2-125) was calculated. The SMR-NHL was 11.4 (95% CI: 7.8-16), which was disproportionately elevated compared with the SMR-all causes (2.0 [95% CI: 1.8-2.1]).

Delco et al.[342] used the database of discharge diagnoses from all US Veteran Affair hospitals to gather a total of 458 CD patients, hospitalized between 1986 and 1995. The concomitant diagnoses received by those patients were compared with those of five controls per CD patient, randomly selected from the same year's discharge database (total 2,692 controls). The mean age of the CD group was 63.8 +/- 12.4 years and the mean age of the control group was 59.7 +/- 14.8 years ($p<0.001$). Ninety-three percent of the patients with CD were white, compared with 74% of the control subjects ($p<0.0001$). The odds ratio (OR) of NHL (OR-NHL) in CD, was 4.53 (2.01-10.23).

Role of a GFD

The impact of GFD compliance was analyzed and reported in only two of the nine studies. Holmes et al.[333] reported a SIR of NHL in patients on a strict GFD (SIR 44.4), versus those who did not adhere to a GFD (SIR 100). Corrao et al.[339] observed that the mortality from all causes was lower in patients on a strict GFD, as opposed to those who were unlikely GFD-compliant (SMR 0.5 [95% CI: 0.2-1.1] and 6.0 [95% CI: 4.0-8.8], respectively). Although, in the study by Askling[337], compliance could not be directly ascertained, the SIR of lymphoma 1 to 4 years after diagnosis was 9.7 (95% CI:6.3-14), wheres, it dropped to 3.8 (95% CI: 2.2-6) five or more years after diagnosis, suggesting that the risk of lymphoma decreases over time on a GFD.

Risk of Lymphoma Versus Symptoms

The mode of presentation leading to the diagnosis of CD was not commonly reported. The reports from Italy[335,339] were unique in that they both detailed the circumstances by which the diagnosis of CD was diagnosed, portraying their cohorts as largely asymptomatic, since 45%[335] and 70%[339] of their patients had subclinical presentations, i.e., either mild symptoms, anemia, or were detected through screening. Conversely, it is reasonable to suggest that the studies that used hospital discharge diagnoses of CD as entry criteria would be largely made up of symptomatic CD patients. Unfortunately, it is not possible to compare the measured risk of lymphoma in the Italian studies to those of our other reports, because of the great disparities in populations, data collection and analyses amongst them.

The presence or absence of symptom at the time of CD diagnosis was not evaluated as a risk factor for lymphoma per se. Corrao et al.[339] did, however, analyze the impact of the mode of presentation on the mortality from all causes in CD. They observed that patients diagnosed with mild symptoms or by antibody screening did not show any relevant excess mortality, compared with the symptomatic group (SMR 1.2 [95% CI: 0.1-7.0] and 2.5 [95% CI: 1.8-3.4], respectively).[339]

Impact of the Age at Diagnosis of CD

Several studies analyzed the risk of lymphoma with respect to the age at diagnosis of CD. Patients who were diagnosed with CD during adulthood were either 1) asymptomatic during childhood or 2) symptomatic but eluded the diagnosis. For the later circumstance, authors have referred to "diagnostic delay" as a symptomatic period in the absence of diagnosis or treatment. The impact of the diagnostic delay was analyzed in two studies.[336,339] Corrao et al.[339] compared the mortality from all causes in patients who had suffered a diagnostic delay of more than 10 years, one to 10 years, or less than 1 year (no diagnostic delay), and found that the longer the untreated symptomatic period, the greater the mortality from all causes (SMR 3.8 [95% CI: 2.2-6.4], 2.6 [95% CI: 1.6-4.1], and 1.5 [95% CI: 0.9-2.3], respectively). Logan et al.,[336] on the other hand, reported opposite results: while the SMR-all causes was significantly greater than 1 for their entire cohort (1.9 [95% CI: 1.5-2.2]), for those CD patients diagnosed only in adult-life despite an obvious childhood illness typical of CD, all-cause mortality was similar to that of other CD patients diagnosed in adult life. A difference in methodology might explain this discrepancy, since the ascertainment of outcomes was derived from registers in Logan's study

and was probably not as accurate and reliable for outcomes such as the presence or absence of symptoms during childhood.

Logan et al.[336] also reported that the all-cause mortality was increased in the patients diagnosed as adults, but not those who were diagnosed as children (SMR 1.9 [95% CI: 1.5-2.3] and 1.4 [95% CI: 0.4-3.7], respectively).

The patients from Corrao's cohort were exclusively diagnosed with CD as adults. The SMR-all causes for patients diagnosed between 18 and 29 years was slightly less, and not significantly different from 1.0, compared with those who were diagnosed later on in life, i.e., 2.5 (95% CI: 0.5-7.3) for those diagnosed at age 30 years versus 2.4 (95% CI: 1.3-4.0) for those diagnosed at age 49 years and 1.9 (95% CI: 1.3-2.6) for those diagnosed at age >50 years.

Askling et al.[345] reported on 11,019 patients with CD, diagnosed at all ages, and found that the SIR-NHL was not significantly greater than one in CD patients who were diagnosed during childhood, in contrast with those who were diagnosed as adults (SIR-NHL 1.9 [95% CI: 0.4-5.5] for diagnoses made at ages 0 to 19 years compared with 7.7 [95% CI: 4.9-12] for those diagnosed between 20 and 59 years). Part of the increased risk in adults may be explained by the fact that in some of these cases the diagnosis of lymphoma can be made simultaneously or soon after that of CD. However, cases of lymphoma diagnosed within 12 months of CD diagnosis were excluded from Askling' study, so that the risk of lymphoma in adult CD diagnosis remains elevated independently of cases with simultaneous presentation.

Risk of Lymphoma in Refractory CD

We were unable to identify a single source of controlled data on the risk of lymphoma in refractory CD. There was one indirect source of controlled evidence on the mortality in CD. Nielsen et al.,[343] from Denmark, published the mortality data from 98 patients with CD diagnosed between 1964 and 1982, 24% of which were treated with prednisone because they did not respond to a GFD, i.e., probable refractory CD. The mortality in CD exceeded that of the general population (controlled for age and sex) by a factor of 3.4 ($p<0.025$); in GFD-responders, this factor was 2.2 ($p<0.025$), whereas it was 5.8 ($p<0.005$) in the non-responders. The causes of death were poorly documented, and therefore, will not be described here.

Quality Assessment

The overall quality of the included studies was good (Appendix J, Tables 3-5). For example the assessment of outcomes was complete in the included studies.

Celiac 4: Consequences of Testing for CD

Out of 1,199 citations that were identified by the search strategy for the Celiac 4 objective, 140 met the level 1 screening criteria (excluded 1059) (Appendix E). Of these, 126 met the level 2 screening criteria (excluded 14). At level 3, 35 articles satisfied the screening criteria (Evidence Table 11, Appendix H)[346-380] (excluded 72 articles at level 3). Eleven relevant articles were identified in other celiac objectives: five from Celiac 2;[381-385] four from Celiac 3;[331,335,336,343] and two from Celiac 5.[386,387]

The search strategy did not identify any studies that would allow us to address the specific benefits and harms of testing with different strategies for CD. The consequences such as false-positive results were dealt with in Celiac 1. We address the response to treatment in the sections that follow.

For the consequence of osteoporosis/fracture, an additional search was conducted with the search terms osteoporosis and CD, and five additional relevant studies were identified.[388-392]

The consequences that were included in this review were: 1) costs, 2) patients complying with treatment, 3) response to treatment in terms of symptoms, and 4) clinical outcomes such as reduced risk of complications–osteoporosis, mortality, anemia.

Given the recent recognition that the number of subclinical and silent CD cases may be eight times that of classically symptomatic cases, it is important to determine if the clinical outcomes vary according to type of clinical presentation. Where possible, results of the analysis according to type of clinical presentation are presented.

Part A

Most papers included in the consequences of testing for CD dealt with patients (who were newly diagnosed) after they initiated a GFD. Most studies evaluating the consequences of nutritional status were before/after studies. In total, 15 studies dealing with either nutritional status, weight, body mass index (BMI) and body composition, were identified.[346-350,352,357,359,361,363-365,369-371]

Seven studies were case control,[347-349,352,357,364,369] one a cohort study,[346] and in seven studies, the patients acted as their own control group.[350,359,361,363,365,370,371]

Eight studies were based on children with CD,[347,349,350,352,357,359,369,370], three studies were based on adolescents with CD[363-365] and four studies were based on adults with CD[346,348,361,371].

There were five studies that evaluated costs of screening as a consequence.[360,366,379,380,382]

Type 1 diabetes and CD. Four studies evaluated diabetes and CD in children.[347,357,359,370] Three studies were from Europe (UK,[347] Hungary,[359] and Finland[370]) and one was from Australia.[357] Two were case control studies[347,357] and two studies had patients with CD act as their own controls.[359,370] All the studies assessed the effect of a GFD diet (range 3-12 months) on the diabetic control of type 1 diabetes.

The UK study[347] evaluated 230 children with type 1 diabetes who were screened for CD with serology. Those children with positive serology were biopsied. Eleven children were diagnosed with CD and followed longitudinally. The control subjects were the children diagnosed with type 1 diabetes with negative serology. The controls were matched for age, sex and duration of diabetes in a 2:1 ratio (22 controls:11 cases). At baseline, the weight (standard deviation score;

SDS), BMI SDS and HbA1c of the cases were statistically lower than the controls. No statistical difference was noted for height SDS, C-peptide level and insulin requirements. Also, the cases (type I diabetes with positive CD serology) received significantly less intensive insulin regimens compared with controls. Six type 1 diabetic children with CD participated in the GFD. After 12 months of a GFD, the differences seen in the BMI SDS was reversed between the cases and controls. HgA1c levels did not improve significantly on a GFD. Insulin dose requirements increased for both cases and controls, but still did not significantly differ from each other. Insulin regimens were not statistically different between cases and controls after a GFD.

The Australian study[357] included children and adolescents with coexisting type 1 diabetes and CD, which were identified from a database of the Diabetes Center at the Royal Alexandra Hospital for Children. CD had to be biopsy-proven. Twenty patients (5M:15F) were enrolled out of 36 patients identified on the database. Forty control patients from the same database were matched for age, sex and duration of IDDM. No immediate criteria on screening from the database was given in the study. At baseline, the current height SDS, current weight SDS, BMI SDS and HbA1c were not significantly different from controls. Compliance with a GFD was based on dietary records classifying patients to: no detectable gluten; trace of gluten; and, gluten containing. For compliance, 30% of patients were classified as adhering to a strict GFD, 30% consumed trace amounts of gluten, and 40% had a significant amount of gluten in their diet. No differences were detected in growth parameters or HbA1c according to compliance to a GFD.

The Hungarian study[359] included 205 children with type 1 diabetes that were randomly selected from screening for CD. None of these patients had suspicion for CD. Twenty-four children were positive for EMA and 17 (7 boys and 10 girls) had subtotal villous atrophy. The height of the children with CD and type 1 diabetes were normal compared with children with only type 1 diabetes at baseline. But the BMI of the 17 children was significantly lower (14.2 vs 16.3 kg/m^2) compared to controls. After three months of a GFD, BMI significantly increased (14.2 vs 16.8 kg/m^2). Furthermore, significant increases in insulin requirements (0.64 U/kg vs 0.48 U/kg) occurred after a GFD. The percentage of HbA1c did not change on a GFD compared with baseline (7.82% versus 7.67%).

The study from Finland by Saukkonen et al.,[370] retrospectively screened 776 children with type 1 diabetes over a 2.7 year period with serology and, if positive, jejunal biopsy. Eighteen children (2.3%) had confirmed CD. HbA1c levels did not change after introduction of a GFD. Correlation of height SDS and mean weight for height were not compared post-GFD.

Body composition and anthropometrics. Six studies specifically detailed body composition after a GFD.[348-350,352,369,371] Of these studies, four examined children,[349,350,352,369] and two included adults.[348,371]

Of the studies conducted in adult patients with CD, one was from Italy[348] and the other from Argentina.[371] In the Italian case-control study, 212 treated patients with histologically-confirmed CD were assessed. Of these, 71 (33.4%) (51 women and 20 men) were asymptomatic, had maintained a constant body weight during the previous 6 months, and were on a strict GFD. Forty-three of the patients were diagnosed as children (28 women and 15 men; average age 5.2 years) and 28 were diagnosed as adults (23 women and 5 men; average age 28 years). The average consumption of a GFD was ≥ 2 years. For each patient, there were two sex- and age-matched healthy controls (142 controls). Body composition was calculated by means of DEXA. The weight and BMI of female CD patients were lower than the controls (55.5 kg vs 58.7 kg,

p=0.004 and 20.9 kg/m^2 vs 22.4 kg/m^2, p=0.03). The height and BMD were not significantly different, although BMD for those diagnosed as adults was lower than controls. Fat mass (22.9% vs 27.5%, p<0.05) and lean mass (38.8% vs 40.5%, p<0.03) were also significantly lower in cases versus controls. The weight (69.2 kg vs 73.3 kg, p=0.03), height (175 cm vs 178 cm, p=0.05) and BMI (21.9 kg/cm^2 vs 23.5 kg/cm^2, p=0.05) of male patients were significantly lower than in controls. Fat mass (13.9% versus 16.8%, p<0.05) and lean mass (55.5% versus 56.7%, p<0.03) were also significantly lower than in controls.

The study from Argentina by Smecuol et al.,[371] enrolled 47 (41 females, 6 males) unselected, consecutive patients with newly diagnosed CD (diagnosed between Sept 1991 and Oct 1993). Twenty-five patients were re-evaluated in 1995 (24 females and 1 male). The diagnosis of CD was based on clinical features of classic and atypical symptoms, with positive small bowel biopsy and positive serology. Three patients were asymptomatic, the rest had classical features of CD. After 12 months, all patients on an initial GFD, improved. In the study, the patients acted as their own control—15 patients adhered strictly to the GFD, while ten were on a partial GFD. Patients on a strict GFD consumed less calories than patients who were poor compliers (p<0.05). After treatment, fat mass (18.2 kg, p<0.0001) and bone mass (2 kg/m^2, p<0.002) increased significantly. Lean tissue mass did not increase. Body weight (55.7 kg, p<0.0001), BMI (22.2 kg/m2, p<0.001) and triceps skinfold thickness (15.8, p<0.0001) were increased significantly; mid-arm muscle circumference and muscle mass did not change. Patients who more strictly adhered to the GFD tended to demonstrate greater increases, although the trend was not significant.

Of the four studies that evaluated children, two were from Italy[349,352], one was from the Netherlands,[350] and one was from India.[369] Both Italian studies were case-control studies, whereas, in the Netherlands study, the patients acted as their own control. In one of the Italian studies by Barera et al.,[349] 29 consecutive children (14 boys and 15 girls) with a diagnosis of CD were enrolled (mean age 9.54 ± 3.42 yr). Diagnosis was according to ESPGAN criteria. Four patients had classic symptoms, while the rest had atypical CD. The patients were studied over 1.02 ± 0.15 years of GFD. Each patient was age- and sex-matched to a healthy control patient (n=29). At baseline, children with CD weighed less than the controls (28.3 ± 11 kg vs 34.5 ± 14.1 kg, p=0.04), had lower lean mass of limbs (8.4 ± 4.8 kg vs 10.8 ± 4.7 kg, p=0.0013), less fat mass (4.6 ± 3.5 kg vs 7.5 ± 4.9 kg, p=0.006), less percentage of fat mass (17.4 ± 8.3% vs 23.7 ± 8.4%, p=0.002) and lower bone mineral content (1067.2 ± 451.3 g vs 1317 ± 553.8 g, p=0.006). Height, BMI, lean mass, and ratio of lean mass to height, did not differ from controls at baseline. After an average of 1 year on a GFD in 23 children, no significant differences were found in weight, height, BMI, lean mass, lean mass to height, lean mass of limbs, fat mass, percentage of fat mass or bone mineral content (BMC), compared with controls. Compliance was good in all patients as assessed by EMA (only three subjects were still positive).

The second Italian study by Rea et al.,[352]enrolled 23 children (8 boys and 15 girls, mean age 4.7 ± 0.76 yr) from Jan 1992 to Dec 1994, according to ESPGAN criteria. They were sex- and age-matched to healthy controls from the ambulatory clinic. At baseline, the height, BMC, arm muscle area (AMA), triceps skinfold (TSF), subscapular skinfold (SSSF), and fat area index (FAI), were significantly lower than controls. The BMI and weight for height index (WHI) were not different. After GFD, all the parameters improved when compared with patients to before GFD. Height, BMC, AMA, BMI, TSF, SSF, FAI and WHI all significantly improved. If patients post-GFD were compared with controls, the height was still significantly lower (p=0.01) but the rest of the values were not significant. After a GFD, the blood chemistry of these patients

was assessed. The hemoglobin, iron, protein, albumin triglycerides, calcium, and zinc levels were significantly different from the baseline value; however, transferrin, cholesterol, phosphorus and alkaline phosphatase levels were not different.

The study from the Netherlands by Boersma et al.,[350] enrolled 28 children (9 boys and 19 girls) with newly diagnosed CD (between Jan 94 to Jan 95). All children had classic symptoms and had positive small bowel biopsies. After 3 years of a GFD, the BMI SDS and height SDS improved significantly ($p<0.0001$ for both). The initial improvement of BMI SDS was seen in the initial 6 months with subsequent gradual improvement. The height SDS improved continuously over the 3 year period, and the improvement was significant.

In a study from India by Poddar et al.,[369] 104 children evaluated for CD between Sept 1997 to Dec 1998 were included. All children had diarrhea, failure to thrive or pallor as a clinical presentation. Fifty-seven were diagnosed as having CD (by modified ESPGAN score) and the remaining 47 were controls. Seven children who did not respond to a GFD and were excluded, were diagnosed with other diseases. The mean follow-up of patients after starting a GFD was 19.6 ± 8 months (range 4-36 months). The remaining 50 children had a dramatic response to the GFD. Symptoms subsided in 16±9.8 days (range 4-30) and all showed significant weight gain (66% ± 14% vs 86% ± 11% of expected, $p<0.001$). Height gain improved, but was not significant (88 ± 5% vs 94 ± 5% of expected, p=not significant). Seventeen percent of the children had poor compliance to the GFD. No attempt at subdividing patients into poor versus good compliance was made.

Nutritional status. Two studies looked at nutritional status with biochemical markers.

In the study from Finland by Kemppainen[346] nutritional status of newly diagnosed patients with CD before and after GFD was reported. Forty patients with CD diagnosed between Nov 1988 to Dec 1990 were included. All had abdominal symptoms. Diagnosis was made on presence of partial villous atrophy (eight patients), subtotal villous atrophy (17 patients) or total villous atroph (15 patients). On mean histomorphometric index, there was a statistically significant trend ($p=0.004$) comparing partial villous atrophy (0.018 ± 0.003), subtotal villous atrophy (0.0015 ± 0.002) and total villous atrophy (0.013 ± 0.002). When biochemical measurements were examined according to grade of villous atrophy, significant differences were seen for ferritin ($p<0.01$) and transferrin ($p<0.05$). Serum ferritin was still significantly lower in total villous atrophy, as was erythrocyte folate levels if sex was standardized in an analysis of variance. Severity of villous atrophy also correlated with ferritin, erythrocyte folate, and serum vitamin B12. Abnormal values of serum protein, vitamin A, and vitamin B12, were low. There were no abnormal vitamin E levels. Villous atrophy improved in all patients within 12 months of a GFD. Two patients had subtotal villous atrophy, 29 had partial villous atrophy and three had normal villi after a GFD. Six patients withdrew from the study. BMI increased after a GFD, as did most of the biochemical measurements. One patient with subtotal villous atrophy still had a low hemoglobin value. Of the 29 patients with partial villous atrophy, three had low folate levels, seven had low hemoglobin, one had low vitamin B12, one had low protein, five had low vitamin A, five were low in ferritin, five had low iron, and ten patients had low zinc levels. Only one patient (out of three) who had normal villi also had low hemoglobin levels.

In the study from Italy, by Bardella et al.,[361] 26 adults (five male and 21 female, mean age 42.2, range 22-81) with malabsorption and biopsy-confirmed CD were enrolled. They were followed for a mean of 55.4 months (range 13-137 months) on a GFD. Eight patients remained in good health with normal blood tests. The remaining 18 patients had abnormalities despite

GFD. No correlation was noted with severity of symptoms of malabsorption and biochemical abnormalities. Iron deficiency was found in five patients. Abnormal calcium, phosphorus, alkaline phosphatase and/or bone density was found in seven patients. Macrocytic anemia was found in four patients. Clinical symptoms were seen in 11 patients. No correlations between abnormal values and grade of histology on biopsy were found.

Compliance. Three studies were identified that looked at compliance, [363-365]. All studies were conducted in Italy and assessed an adolescent population.

In the first study of adolescents that looked at dietary compliance, Fabiani et al.[363] evaluated 28 biopsy-proven CD patients (17 females and 11 males). These 28 adolescents were selected from a group of 6,315 students, age 11 to 14 years, who had previously been screened for CD. All were advised to start a GFD. Twenty-three of the 28 patients participated in this study. The mean follow-up duration was 23 ± 7 months (range 9-3 months). Fifty-two percent (12/23) were on a strict GFD and 47% (11/23) partially adhered to the diet. Improvement in most patients was seen after starting a GFD. Weight gain was reported in 12 patients (52%)—11 had increased height velocity and appetite, eight had disappearance of symptoms of abdominal pain, six had resolution of diarrhea, five had disappearance of anemia and three had disappearance of recurrent aphthous stomatitis. Three patients did not demonstrate any change.

The second study, also by Fabiani,[364] was a 5-year case-control study that enrolled two groups of patients. The first group (group A) included subjects between the ages of 11 and 14 years, who were diagnosed as a result of a mass screening program. The second group (group B) were patients diagnosed due to typical symptoms of CD between 1985 to 1986. All patients had biopsy-proven CD according to ESPGAN criteria. All patients were followed for 5 years and advised to start a GFD. Twenty-seven patients were in group A and 22 agreed to participate; 24 patients were in group B and 22 agreed to participate. There were no differences between the patients in group A and group B in terms of BMI and height SDS. No difference was found between the two groups in terms of symptoms. Adherence to the treatment was significantly lower in patients from group A compared with group B. There were a significantly greater proportion of patients in group B that demonstrated strict adherence to a GFD (15/22; 68%) compared with patients in group A (5/22; 23%).

The third study to look at compliance looked at 306 teenage patients with CD (mean age 15.9 yr; range 10-27 yr) recruited consecutively from a CD clinic.[365] Of the patients, 186(60%) were female and 120 were male. Diagnosis of CD was biopsy confirmed. Recall questionnaire was used to evaluate diet and compliance. Compliance was recorded in three categories: 1) strict gluten diet (n=223 [73%]); 2) occasional relapse (n=46) 15%; and, 3) gluten-containing diet (n=37) 12%. Eighty percent of the female patients, compared with 64.2% of the male patients, adhered to a strict diet (p=0.012). Compliance also varied with age, with older age associated with less compliance (p=0.05). Growth status was grouped according to compliance to a GFD—the mean standardized height, the relative weight for age, and the relative weight for height, did not differ significantly between the compliance groups. Symptom scores were relatively good among all groups. No statistically significant differences were noted. School performance was not significantly different between good versus poor compliers.

Costs. Five studies included an assessment of costs involved in different screening strategies.[360,366,379,380,382]

Harewood et al.[366] performed a decision analysis to compare costs of serological testing versus small bowel biopsy (AGA vs EMA versus small bowel biopsy) for diagnosis of CD. The analytic technique used was a cost minimization and the viewpoint was third-party payer. A sensitivity analysis was conducted. The authors demonstrated that initial screening with EMA is the least costly strategy for diagnosis in a low to medium risk population.

Gomez et al.[382] evaluated a screening algorithm for CD in 1,000 consecutive subjects who were screened while attending a central laboratory. Gomez and colleagues compared two screening protocols: (1) three-level screen–IgG/IgA-AGA antibodies at the first level, then IgA-EMA, and finally intestinal biopsy versus screening, and (2) tTG-GP and total IgA as first-line screen, and EMA for positive patients followed by intestinal biopsy. The analytic framework and viewpoint were not stated. In this study, a comparative cost analysis was performed. They found that the combination of a highly-sensitive test at the first step with a highly-specific test at the second step appears to be a more reliable screening mechanism.

Zaccari et al.,[379] in an Italian model, proposed a four-level screening protocol for children at least 15 months of age, including: 1) AGA, 2) EMA, 3) intestinal permeability, and 4) small bowel biopsy. In this study, they evaluated only the total costs at each level of screening.

Atkinson et al.,[360] in a Canadian study, evaluated the operating costs of EMA in the diagnosis of CD using a cost-minimization model with a decision analytic approach with three strategies. The analytic perspective used was the societal viewpoint, and costs were discounted at 5% per annum. A one-way sensitivity analysis of all probability and cost estimates was performed. Incremental costs of the GFD were estimated from a survey of 25 patients which resulted in a lifetime incremental cost of $44,000. If a small bowel biopsy was performed initially, the cost was $997; for EMA followed by small bowel biopsy, the cost was $866. The total cost was $3,714, which resulted in an incremental cost savings of $2,177 if small bowel biopsy had been performed first. In the sensitivity analysis, the specificity of EMA would have to be greater than 95% to make EMA least expensive.

Part B

There were 27 studies that examined the response of various endpoints to a GFD.

One Italian study,[354] used a case-control design to evaluate the effect of a GFD on thyroid status. The study by Annibale et al.,[358] evaluated the impact of a GFD on anemia and iron deficiency in newly diagnosed CD cases identified from screening of adults with IDA in Italy. In a case-control study, Ciacci et al.[351] investigated the impact of a GFD on pregnancy outcomes, and Addolorato et al.[374] evaluated the impact of a GFD on anxiety and depression in a population of CD patients in Italy. Mortality was evaluated in seven cohort studies.[331,335,336,343,362,367,368] Seventeen studies assessed either change in BMD or fracture as an endpoint in individuals with CD.

Thyroid study. In the Italian study,[354] 241 consecutive adults with biopsy-confirmed CD were enrolled between Jan 1996 and July 1998 (177 women and 64 men). Forty percent of patients had classical symptoms, 44% had atypical symptoms and 16% had silent CD. Two hundred and twelve patients, matched for age, sex and ethnic origin, were used as controls. All newly-diagnosed CD patients were started on a GFD and patients with hypo- or hyperthyroidism were

started on appropriate medical therapy. Thyroid dysfunction was found in 73 (61 women and 12 men) of 241 patients with CD, and in 24 (19 women and 5 men) of the 212 patients in the control group ($p<0.0005$). The difference was statistically significant for women when divided by sex ($p<0.0005$). Hypothyroidism was diagnosed in 31 patients (12.9%) and nine controls (4.2%) ($p<0.003$); it was subclinical in 29 CD patients and eight controls and overt in the remainding patients. The difference was only significant for women ($p=0.0045$). Twenty-one patients and four controls had non-autoimmune hypothyroidism. Ten patients and five controls had autoimmune hypothyroidism. Hyperthyroidism was diagnosed in three patients and seven controls; it was subclinical in two patients and five controls. Autoimmune thyroid disease with euthyroidism was present in 39 patients and eight controls. The difference was only statistically significant in women ($p<0.0005$). At diagnosis, the BMI, hemoglobin, iron, and albumin levels were similar between patients with thyroid disease and those without. After 1 year of a GFD, 128 patients were reassessed. Ninety-one patients had normal thyroid function, whereas, 37 had some impairment. Compliance to diet was not different between the two groups. Subclinical hypothothyroidism improved in 10/14 patients with non-autoimmune hypothyroidism. Three of five patients with autoimmune hypothyroidism shifted to autoimmune thyroid disease with euthyroidism; four out of five patients with no improvement in thyroid function had poor compliance with diet. Significant improvement in nutritional indices was also seen with BMI in females, HBG in both sexes, and serum albumin and serum iron in both sexes.

Iron deficiency. In this Italian prospective study,[358] 190 consecutive patients (160 women and 30 men) who were referred to the GI department from the hematology for IDA between Jan 1994 to May 1997, were examined. Twenty-six patients were diagnosed with CD (24 women and 2 men); average age 31.3 years (range 20 -72). Seventy-seven percent of patients had total villous atrophy and 23% had subtotal atrophy; repeat endoscopy with biopsy specimens were taken after 6 months. After GFD, 20 patients (18 women and 2 men) were followed for 24 months. After 6 months, 14 of the 18 female patients (77%) recovered from IDA. Only 5/18 reversed from iron deficiency as defined by normal ferritin levels. At 12 months, 17/18 recovered from IDA. Nine patients reversed from iron deficiency. After 24 months, the same patient still did not reverse from IDA. Ten patients (55%) reversed their iron deficiency. Of the two males, at 6 months of a GFD, only one recovered from anemia but not from iron deficiency (low ferritin). At 12 months, both patients reversed their anemia and iron deficiency. At 24 months, further increases in ferritin were observed. In a subgroup of patients that had repeat small bowel biopsies at 6 and 12 months, there was a significant inverse correlation between increases in Hb concentrations and decreases in histological scores of duodenitis. This study demonstrated that recovery from IDA occurs within the first 6 to 12 months, but reversal from iron deficiency occurs in 50% of cases (predominantly premenopausal women). Long-term follow-up of ferritin results and small bowel biopsies in subjects with CD would be helpful to determine if iron deficiency resolves completely.

Pregnancy outcomes. In this case-control study from Italy by Ciacci et al.,[351] 297 women with CD were enrolled. Three types of analyses were used. Analysis A was a case-control study between untreated women (n=94; at least one pregnancy when symptoms of CD were present and lead to eventual diagnosis) and treated CD women (n=31; at least one pregnancy after 1 year of a GFD). At baseline, weight, height and body mass index were the similar between the two groups. However, the treated group was significantly younger than the untreated group (37.3 ±

12 yrs vs 22.4 ± 1.6 yrs, $p<0.01$), which may have biased the results. The number of pregnancies per woman was also lower for the treated group (2.72 ± 0.16 vs 1.6 ± 0.11, $p<0.0001$). The number of abortions per woman (0.489 ± 0.085 vs 0.032 ± 0.032, $p<0.0001$), as well as the abortion to pregnancy ratio, was much lower for the treated group compared with the untreated group(0.153 ± 0.027 vs 0.024 ± 0.024, $p<0.005$). Subgroup analysis taking into account the age at diagnosis, demonstrated that for those women diagnosed at age 30 years or less (n=27), the number of abortions per woman was 0.556 ± 0.156 and the abortion to pregnancy ratio was 0.234 ± 0.066. The prevalence of abortion in pregnancies was 17.8% in untreated CD patients, compared with 2.4% in treated patients ($p<0.001$). The RR of abortion was 8.9. Low-birth-weight baby to pregnancy ratio (0.126 ± 0.037 vs 0.024 ± 0.024, $p<0.03$) was significantly lower in the treated group. The duration of breast feeding was significantly longer for the treated group (2.77 ± 0.52 vs 7.03 ± 1.17, $p<0.0003$). The threatened abortion to pregnancy ratio and premature delivery to pregnancy ratio was not significantly different from untreated to treated CD women. For the subgroup of women <30 years (n=27), birth weight, baby to pregnancy ratio, and duration of breast feeding, did not alter the statistical significance. The prevalence of low birth weight babies in nonabortive pregnancies was 12.7% for untreated patients and 2.4% for treated patients ($p<0.05$). The RR of low birth weight babies was 5.84 times greater in the untreated group compared with the treated group.

In Analysis B, women with CD were all untreated and then analyzed depending on whether diarrhea was present or not. The authors found that the abortion to pregnancy ratio and the premature delivery ratio were found to be lower in CD women without diarrhea compared with those women with diarrhea, although the difference was not statistically significant.

In Analysis C, the effect of a GFD on pregnancy outcome was analyzed. The study examined 12 women with CD after 1 year of a GFD (own control); there was at least one pregnancy without treatment. All outcomes were better in the group of women on the GFD: number of pregnancies 2.5 ± 1.24 versus 1.08 ± 0.29 ($p<0.003$); number of abortions per woman 1.08 ± 1.16 versus 0.08 ± 0.28 ($p<0.02$); abortion to pregnancy ratio 0.405 ± 0.140 versus 0.074 ± 0.280, $p<0.02$); and, low birth weight baby to pregnancy ratio 0.292 ± 0.129 versus 0 ($p=0.05$). The threatened abortion to pregnancy ratio, premature delivery to pregnancy ratio, and duration of breast feeding, were not significantly different between the two groups. The prevalence of abortion was 43.3% for the untreated group, compared with 7.7% for the treated group of CD women ($p<0.01$). The RR of abortion was 9.18. There were no low birthweight babies born to women in the GFD group, whereas, the prevalence of low weight babies was 29.4% in the untreated group (RR=11).

One of the limitations of the Ciacci et al. study was that it did not include an external control group or control for confounders. A historical cohort population-based study of the Danish Medical Birth Registry by Norgard, 1999[393] evaluated birth outcomes in women with CD. This study included 211 newborns born to 127 mothers with CD from 1977-1992 and compared them with 1,260 control deliveries. Women with CD were identified from hospital discharge diagnoses. Discharge records were linked to Medical Birth Registry which contained information on relevant outcomes. Outcomes included birthweight, low birthweight (<2500 g) pre-term birth (<37 wk), intrauterine growth retardation (birthweight <2500 g and gestational age ≥37 wk of pregnancy), and perinatal mortality. Potential confounders including maternal age, infant's gender, parity, and gestational age, were adjusted for in the analyses. The investigators could not control for other confounders such as smoking. Another potential limitation is that the date of diagnosis of CD was the initial time of discharge from hospital with CD. It is possible

that women may have been initially diagnosed in the ambulatory care clinic. Details about the clinical presentation of the women with CD and biopsy findings were not available. The mean age at time of delivery was 27.5 years for women with CD and 26.3 years for control women.

Norgard et al.,[393] found that before women were hospitalized for CD, they were at an increased risk of low birthweight babies (adjusted OR=2.6 [95% CI: 1.3-5.5]), and intrauterine growth retardation (12.3% vs 4.8% of controls; adjusted OR=3.4 [95% CI: 1.6-7.2]). After women with CD were first hospitalized, there was no increased risk of low birthweight babies (6% post diagnosis) or intrauterine growth retardation, when compared with controls. The results of this study have implications for women with undiagnosed (atypical or silent) CD.

Anxiety and depression. The study from Italy by Addolorato et al.,[374] enrolled 43 newly-diagnosed adult patients affected with classic CD, selected from 234 adult CD patients from an outpatient clinic between June 1995 and Oct 1998. No psychiatric disorders other than anxiety and/or depression were allowed. The diagnosis of CD was based on positive serology and biopsy. Of the 43 enrolled patients, eight dropped-out leaving 35 (14 males and 21 females, mean age 29.8 ± 7.4 yr) patients for analysis. After a period of 12 months of GFD treatment, the patients were analyzed. The adherence to a GFD was evaluated based on patient self-report and family member interview. A group of 59 healthy asymptomatic controls (27 males and 32 females, age 31.7 ± 6.9 yr) were matched for gender, age, residence, employment, socioeconomic and marital status. The psychological assessment was performed using a self-rating psychometric test for anxiety (State and Trait Anxiety Inventory test) and another for depression (SDS Zung self rating depression scale). Both tests were administered before and after GFD. Of the 59 controls, 23.7% showed high levels of anxiety, 15.2% showed trait anxiety, and 9.5% were positive for depression. Of the 35 untreated CD patients, 71.4% had high levels of anxiety, 25.7% showed trait anxiety and 57.1% were positive for depression. After 1-year of GFD, 25.7% had high levels of anxiety, 17.1% had trait anxiety, and 45.7% were still depressed. The levels of high anxiety (71.4% vs 23.7%, $p<0.0001$) and levels for depression (57.1% vs 9.6%, $p<0.0001$) were significantly higher in the CD patients than in the controls. The proportion of untreated CD patients with trait anxiety did not differ from controls. After a 1-year GFD, a significant decrease in high-state anxiety (71.4% vs 25.7%, $p<0.001$) was found when treated patients were compared with the untreated group. No significant differences were found for trait anxiety or depression.

Fractures. We identified six controlled studies that addressed the outcome of fractures in a CD population[385,388-390,394]and two reviews.[381,391] The study by Cook et al.[395] was not included since it did not have a comparison or control group. The study characteristics and methods for each study are summarized in Evidence Tables 12 (Appendix H).

All six studies were retrospective and there were two cohort studies[385,388] . Two studies included individuals that had biopsy-confirmed CD. All studies included controls as a comparator, and in three studies the controls appeared to be population-based.[385,388,394] With regards to the ascertainment of the outcome of fracture, data was obtained from self-report data from administrative databases,[394] patient register,[385,388,394] or from interview/case reports.[389,390,392] Only two studies mentioned inclusion of asymptomatic subjects.[389,392] Bone histology was mentioned as an outcome in a subset of patients in one study.[390]

The case-control study by Fickling and colleagues,[390] compared individuals with CD attending a GI outpatient department and/or members of local celiac societies. The authors found a higher prevalence of past history of fractures in the CD patients (21%[16/765]) compared with a control group (3% [2/75]; RR 7.0). There was no difference in BMD T-score results between those with and without a history fracture, although those patients with a fracture history were older ($p<0.02$). Limitations of this study include the fact that they did not identify whether CD was biopsy-confirmed, and a potential for selection bias.

Thomason et al.,[373] in a case-control study, used self-report data for 244 patients with biopsy-proven CD and found that fractures were not significantly increased in those with CD compared with controls (OR 1.05, 95% CI: 0.68-1.02), although there did seem to be a trend to increased wrist fractures (OR 1.21, 95% CI: 0.66-2.25). The mean age of these patients was older (60.2) and the mean BMI was higher (23.9) than that reported in other studies. However, this study may have been limited by potentially not having adequate power to detect fractures. In addition, all the fracture data was self-reported.

Vasquez et al.,[389] in a retrospective case-control study, found that 25% (41/165) of CD patients had one to four fractures, compared with 8% in age- and sex-matched controls. The majority of fractures occurred prior to diagnosis of CD and the most common fracture site was the wrist (OR 3.5, 95% CI: 1.8-7.2). Potential sources of bias for this study include the fact that the cases were from a malabsorption clinic and may therefore represent patients with more severe disease (mean BMI=21.4). The OR for vertebral fractures was 2.8 (95% CI: 0.7-1.15), although there was incomplete ascertainment of X-rays, since not all X-rays were of adequate quality. This was the only study to include an assessment of the proportion of patients on a strict versus a reduced GFD.

Two studies were population-based.[385,388] Vestergaard et al.,[388] evaluated all individuals with CD in Denmark captured from hospital discharge data, and did not find an increase in fractures requiring hospitalization in patients with CD (n=1,021; 7,774 patient years) relative to controls (n=23; 316 patient years) with an independent independent relative risk (IRR) at pre-diagnosis of 0.70 (95% CI: 0.45-1.09) for all fractures. For spine, the IRR pre-diagnosis was 2.14 (95% CI: 0.70-6.57) and 1.07 (95% CI: 0.39-2.95) for rib and pelvis. There are significant limitations to this study since the diagnosis of fractures was hospital-based and therefore, fractures that did not require hospitalization would be missed and could lead to under-reporting. In addition, the diagnosis of CD was only validated in a sample of nine cases (with a validity of 78%), and all cases of CD had to be hospitalized to be included.

West et al.,[385] in the largest analysis of fractures in CD patients identified from the UK GPRD primary care database, found an increase in fractures in CD patients relative to controls. The mean age at diagnosis was 43.5 years, and the ascertainment of fractures was from an administrative database. For any fracture, the hazard ratio was 1.3 (95% CI: 1.16-1.46; 137.9/10,000 patient years vs 105.9/10,000 patient years in controls]). The hazard ratio for hip fracture was 1.9 (95% CI: 1.2-3.02) and the hazard ratio for wrist fracture was 1.77 (95% 1.35-2.34). The absolute difference in the overall fracture rate was 3.2/1,000 person years and 0.97/1,000 for hip fractures in those older than age 45. In contrast to earlier studies, the authors did not find a difference in the risk of fracture after CD diagnosis compared with before diagnosis.

A recent case-control cross-sectional study by Moreno et al.,[392] compared fractures in 148 CD patients (53% classically symptomatic, 36% subclinical CD, and 11% silent CD-detected by screening} to 296 controls (functional GI disorders). The fracture data was self-report obtained

by interview/and pre-designed questionnaire. Moreno et al. found an increased number of fractures in the peripheral skeleton for classically symptomatic subjects compared with controls, but did not find an increased number of fractures in the subjects with subclinical or silent CD.

BMD. BMD is a surrogate outcome for fracture, and it is easier to evaluate in short-term studies. Previous studies of osteoporosis therapies in postmenopausal osteoporosis have shown that there may not, however, be a direct correlation between fracture reduction and increases in BMD. Osteoporosis/osteopenia may be a sign of subclinical CD and persisting osteopenia/osteoporosis in a patient with known CD may be a sign that the mucosa has not normalized.

BMD is an areal two-dimensional measure of bone mass and does not give a true volumetric measure and, therefore, may not be an accurate reflection of bone mass in children.

We found 11 articles that addressed the outcome of BMD/BMC in newly diagnosed subjects with CD.[348,352,353,355,356,375-378,386,387] The study characteristics are summarized in the Evidence Tables (see Appendix H).

The majority of these studies assessed BMD at baseline and the percentage change after a variable follow-up period (1 to 5 years in duration). Two studies evaluated the BMD of children with CD,[352,377] one study evaluated a mixed population,[348] and the remaining studies evaluated adults. All studies included individuals with biopsy-proven CD and in most of the studies BMD was compared with a control population. Only two studies had patients with CD act as their own controls.[353,376] The female to male prevalence ratio in CD is 2:1, and in these studies the proportion of females varied from 50% to 80%.

Five studies included assessments of dietary compliance to a GFD and three studies included data on whether subjects were on co-interventions (e.g., vitamin D or calcium), which may have impacted the BMD results. Only two studies[356,376] looked at the potential relationship between the change in histological grade on small bowel biopsy and change in BMD.

Prevalence of osteoporosis/osteopenia. The studies consistently found that BMD results were lower in untreated subjects with CD compared with controls. Regarding the prevalence of osteopenia/osteoporosis in newly diagnosed patients with CD, the estimates varied. Satgena-Guidetta et al.[353] noted a mean Z-score of -1.5 at lumbar spine, and -1.8 at the femoral neck, with 34% of subjects having normal BMD, 40% having osteopenia and 26% osteoporosis. Valdimarsson et al.[355] found the prevalence of severe osteopenia, as defined by a Z-score less thatn -2, to be 15% at the spine, 9% at the femoral neck, and 22% at the forearm. The prevalence of mild osteopenia (defined as $-2 \leq Z < -1$) was 23% at the lumbar spine and 24% at the forearm. There was not any difference in lumber spine BMD between those patients who presented with malabsorption, compared with those patients without malabsorption. Valdimarsson et al., found that 27% of subjects had secondary hyperparathyroidism. After 1 year on a GFD, the prevalence of those with severe osteopenia decreased from 23% to 14%.

In a recent review the authors pooled prevalence results and found that patients with untreated CD had a mean Z-score of -1.42, and a hip Z-score of -1.14.[381]

Valdimarsson et al.,[356] in a prospective study of 105 newly-diagnosed CD patients, performed follow-up small bowel biopsies. Of the 105 subjects, 28 had secondary hyperparathyroidism. They found a greater reduction in BMD in individuals who had secondary hyperparathyroidism (PTH>65). In this group, the BMD increased significantly, but did not completely normalize after 3 years of a GFD. In contrast, in those with normal PTH at diagnosis, the baseline BMD was not as low and there was a 2.5% increase after 1 year with the

BMD normalizing after 2 years of a GFD. Valdimarsson also noted that 22 patients with stage III-IV had lower median Z-scores than 76 patients with mucosal changes grade I-II. In this study, compliance with the GFD was 100% in those with high PTH, and lower at 87% in those with normal PTH levels.

Kemppainen et al.,[376] in a 5-year cohort study of 28 patients in which the cases served as own controls, found that BMD increased or remained stable in 69% of patients at the lumbar spine and in 67% of patients at the femoral neck. In this study, the authors did not notice an effect of the grade of villous atrophy on the mean BMD values or percentage change in BMD. They also did not observe any correlation between adherence to the GFD and the change in BMD.

Bai,[375] in a small cohort of 45 (25 completed) newly-diagnosed CD patients, assessed compliance with the GFD and found that 84% of patients increased their lumbar spine BMD (mean increase of 12%) and total body BMD (mean increase of 7.3%), compared with 151 control subjects. The greatest increase in BMD was noted within the first year. Bai[375] documented prior fractures in two patients, but did not report any fractures during the 4-year follow-up period.

Sategna-Guidetti et al.,[353] in a longitudinal study of 86 CD patients, noted a similar proportion of patients (83.7%) increased their spine BMD after 1 year, with an increase of 5.3% in LS BMD after 1 year (change in Z-score of 0.5 at the spine).

Ciacci et al.,[386] in a retrospective cohort of 41 consecutively diagnosed patients with CD, noted a significant increase in BMD (14% lumbar spine, and 10.4% femoral neck), after 1 year on a GFD. The authors also found that pretreatment BMD predicted response to treatment.

Mustalahati et al.,[378] noted a significant increase in lumber spine and femoral neck BMD with treatment after 1 year compared with controls, and noted that the BMD was lower in symptom-free patients (n=15), suggesting patients with silent CD may have mucosal lesions for longer periods of time.

Bardella,[348] in a case-control study of 71 CD patients (43 who had started a GFD in childhood and 28 who were diagnosed as adults and were on a GFD and in remission), found that the BMD of the adult CD patients was significantly lower than the control value (0.9 g/cm^2 vs 1.1 g/cm^2, $p<0.01$).

McFarlane et al.,[387]in a case control study of 21 biopsy-confirmed subjects with CD, documented that the baseline lumbar spine BMD was 85% of that seen in controls, and the increase in lumbar spine BMD over the first year was 6.6% (95% CI: 3.1-10.1) and 5.5% in the femoral neck.

Children/adolescents. Mora et al.,[377] in a study of 19 patients (211 controls), noted a lower BMD in CD patients versus controls at baseline, and an increase in total body BMD (using DXA) during the first year when compared with controls (15.2%).

Rea et al.,[352] noted an improvement in forearm Z-score after 1 year on a GFD in 23 newly diagnosed children with CD.

Mortality. There were seven cohort studies that addressed mortality data in CD. Two were Italian studies, [335,362] one was from Denmark,[343] one from Sweden,[331] and three were from the UK.[336,367,368] All seven were cohort studies.

Corraro et al.,[362] identified 1,072 biopsy-proven CD subjects from the records of 11 GI units between Jan 1962 to Dec 1994. The inclusion criteria were complete records and reliable

diagnosis of CD. The ratio of men to women was 1 to 3, the mean age at diagnosis was 35.7 years, mean follow-up was 6.0 years and median diagnostic delay was 17 months. Forty-five percent of the population had mild (39%) or asymptomatic disease, and 50 patients were lost to follow-up. Data were collected over accumulated 6,444 patient years of follow-up, with a mean follow-up of 6 years. Adherence to a GFD was assessed. Fifty-three CD patients died compared with 25.9 expected deaths. An increase in mortality was noted in the entire cohort population (SMR 2.0 [95% CI: 1.5-2.7]). The overall SMR did not differ by sex, age of diagnosis, or year of presentation. Diagnostic delay by more than 1 year significantly increased the SMR (2.6 [95% CI: 1.6-4.1]). There was significant mortality among patients presenting with malabsorption (SMR 2.5 [95% CI: 1.8-3.4]). No excess mortality was seen with patients with mild or asymptomatic CD. Significant mortality was also seen when patients did not adhere to a GFD on clinical records (SMR 10.7 [95% CI: 6.0-17.1]) and on patient interview (SMR 6.1 [95% CI: 4.2-8.6]). The causes of death showed an excess of death from malignancy (24 observed cases, SMR 2.6 [95% CI: 1.7-3.9]) and diseases of the respiratory (SMR 3.6 [95% CI: 1.1-8.4]) and digestive tracts (SMR 6.1 [95% CI: 3.0-10.9]). NHL was seen in two-thirds of the malignant cases (n=16). The other malignancies included gastric (n=2), small intestinal (n=1), liver (n=2), pancreatic (n=1), pleura (n=1), and leukemia (n=1). (Table 45)

Cottone et al.[335] evaluated mortality in a prospective cohort study of 228 biopsy-proven CD subjects in Sicily. Mortality was ascertained by reviewing hospital medical records and pathology specimens. Records were incomplete for 5% of patients. The mean age at diagnosis was 34.7 years and 100% of patients were on a GFD. Seventy-six percent were females. The clinical presentation was anemia in 60% of cases, malabsorption in 20% of cases, and asymptomatic in another 10% of cases. The mean follow-up was 73 months. Twelve deaths were observed, with 3.12 deaths expected and the SMR from all causes was 3.8 (95% CI: 1.9-6.7). The mortality rate was increased within the initial 4 years from diagnosis, giving an SMR of 5.8 (95% CI: 2.5-11.5).

Nielsen et al.[343] from Denmark, conducted a retrospective cohort study of 98 CD patients between 1964-1982. Sixty-one percent of patients were females and the median age at diagnosis was 41 years (range 2 to 74 yrs). Twenty-four percent of patients had unclassified CD and were treated with prednisone, since they did not respond to a GFD and had probable refractory CD. Twenty-three deaths occurred during the study (four due to malignancy). Nielsen et al. found that the 5-year survival rate was 88%, the 10-year survival rate 68.5%, and that mortality exceeded that of age- and sex-matched controls in the general population by a factor of 3.4 ($p<0.025$). There was no difference in mortality between males and females (2.7 and 2.3, respectively). Subjects who responded to a GFD had an extra mortality factor of 2.2 ($p<0.025$), and those who did not respond to a GFD had an extra mortality factor of 5.8 ($p<0.005$). Causes of death were poorly documented.

Peters et al.,[331] in a retrospective cohort study, compared 10,032 symptomatic subjects with CD who had been discharged at least once from hospital, to controls who were age/sex-matched for the calendar period cancer incidence rate. Fifty-nine percent were females. Mean follow-up was 9.8 years. Mortality was ascertained from a national death register. There were 828 deaths, with 419.3 expected, resulting in a SMR of 2 (95% CI: 1.8-2.1). Mortality risk decreased slightly with increasing number of years of follow-up (p for trend, 0.004). Mortality risks were increased for patients with NHL, cancer of the small intestine, autoimmune diseases (RA), allergic disorders, inflammatory bowel disorders, diabetes, and tuberculosis.

The first UK study was conducted in Birmingham, by Holmes et al.[367] Series I included 202 patients with idiopathic steatorrhea or CD, followed from 1965-1975. Ten patients had a positive biopsy for CD. Eleven patients could not be traced. In the 10-year period, 20 deaths were seen, with ten due to malignancy. Series II (1989) had 210 patients (94 males and 116 females) with biopsy-proven CD. Seventy patients were on a normal diet and 134 were on a GFD for more than 12 months at the end of the survey. Forty-three patients had died from all causes (expected was 20.82 deaths, $p<0.001$); 21 deaths were due to malignancy—13 reticulum cell sarcomas, six GI tract cancers and two other malignancies. Of the 21, 13 had a GFD for a mean of 41 months. Deaths from all malignancies, irrespective of diet, were statistically increased as a whole (expected 5.048 vs observed 21, $p<0.001$) and divided by sex (men expected 2.878 vs observed 12, $p<0.001$ and women expected 2.170 vs observed 9, $p<0.001$). Patients taking a normal diet were at increased risk of developing a malignant tumor ($p<0.05$). Clinical response did not predict the risk of developing malignancy.

Johnston et al.[368] examined CD in subjects from Northern Ireland using the Belfast MONICA project. MONICA I was the first survey, and began in Oct 1983 with 1,204 subjects. Of the subjects, 102 (52 males and 50 females, mean age 58.1 years) had positive serology, 72 consented to follow-up (34 males and 38 females) for 11.6 years (range 11.3-11.9 years), and 20 of the 72 gave consent to biopsy. Three subjects had villous atrophy. Thirteen subjects in MONICA I (seven males and six females) died (mean age at death 67.3 yrs; range 56-75 yr). Cause of death was obtained from death certificates from the General Register Office or General Practitioner records. Four patients died with malignant disease–pancreas, stomach, bile duct lymphoma and metastatic melanoma. None of the patients had CD, but all had positive serology. The number of cancer-related deaths and all cause mortality in the MONICA I follow-up study did not show an excess number of deaths compared with the general population of Northern Ireland.

Logan et al.[336] followed a prospective cohort of 653 patients with CD in Edinburgh between 1979 and 1981. All patients had biopsy-proven CD and mortality was ascertained from death certificates. Sixty percent of the patients were females and the mean follow-up was 13.5 years. Six percent of subjects were lost to follow-up. Clinical presentation was not reported. The subjects with CD were compared with age/sex-matched controls. There were 115 deaths from all causes; the expected number was 61.8 for a SMR of 1.9 (95% CI: 1.5-2.2). The increased mortality was greatest during the initial year after diagnosis and declined over time. The mortality rate for those diagnosed during childhood was similar to that of the general population.

Quality Assessment

The majority of studies included in this objective were single group “before–after” studies, although some studies also included a comparative healthy control group. We could not identify any quality instruments for this type of study design and in general, this type of study is considered weak, particularly in the absence of a control group. Overall, however, the strength of the evidence for this objective was fair to good (Appendix J, Tables 6-8).

Celiac 5: Promoting or Monitoring Adherence to a GFD

Out of 502 citations identified by the search strategy for the Celiac 5 objective, 189 met level 1 screening criteria (Appendix F). Of these, 86 met level 2 screening criteria and 20 studies met level 3 inclusion criteria.[396-415]

Of the included studies, eight studies offered correlation between serology and mucosal histological grade,[397,398,403,404,407,409,413,415] and eight reported on serology only.[396,399-402,408,410,412] Four studies focused on histologic changes without serology.[405,406,411,414] Nine of the included studies were conducted in an adult population, six in a pediatric or adolescent population, and five studies in mixed populations consisting of adults and children.

Included articles were divided by study population (adult/children/mixed), antibody type (IgG or IgA), and by antibody methodology (e.g., ME or HU).

None of the identified studies directly assessed the efficacy of a specific intervention on the promotion of adherence to a GFD. Six studies hint at interventions that could potentially be effective.[416-421] Four of these studies were applicable to a pediatric population and two studies were applicable to adults.

Monitoring Adherence to a GFD

Biopsy. To evaluate serology in assessing adherence, some information regarding mucosal recovery on GFD must first be known. Although mucosal recovery is generally assumed to occur within 6 to 12 months after starting GFD, there is evidence that recovery may be slower and more incomplete than previously assumed.

In a mixed population, Wahab et al.[405] followed the histologic profiles of 158 patients after institution of a GFD. Histological recovery, defined as the absence of villous atrophy (Marsh 0-II), was seen in only 65% of the patients within 2 years. Within 5 years, 85.3% of patients showed recovery, and an incremental improvement to 89.9% occurred after 5 years. Of the 10.1% of patients not achieving histological recovery during the follow-up period, 11 had symptoms of CD and were therefore, considered to have refractory CD (7% of all patients). Patients with Marsh IIIb and IIIc histology initially had lower rates of recovery, compared with those with Marsh IIIa histology. In a subgroup analysis of 25 children, recovery seemed to occur faster—96% showed histological recovery within 2 years ($p<0.01$ vs adults) and 100% recovered in long-term follow-up. It is important to point out that the validity defining a Marsh II lesion as histological recovery is uncertain. If these patients were not included, rates of histological recovery would be even slower. Nonetheless, clinical improvement was seen despite the slow histological improvement.

An early study by McNicholl et al.,[406] is consistent with the finding of more complete mucosal recovery in children. Thirty-six children on a GFD for a mean of 5.8 years underwent duodenal biopsy. Mucosal morphology was normal in 16 (44%) patients, while the remainder of the patients had minimal changes. Villous atrophy was not seen. IEL counts were normal in 30 (83%) patients. A subsequent gluten-challenge confirmed the diagnosis in all 36 children.

Lee et al.,[411] in a retrospective cohort of 39 adult patients, also found incomplete mucosal recovery. After a mean duration of a GFD for 8.5 years (range 1 to 14 years), histology was normal in only 21% of patients, and partial and total villous atrophy was seen in 69% and 10% of patients, respectively. These patients were felt not to have refractory CD since they had a good

clinical response to the GFD. Also of concern were the results of serologic testing at the time of follow-up biopsy in 31 patients. Despite the relatively high number of patients with some degree of villous atrophy, IgG-AGA, IgA-AGA and IgA-EMA were negative in the majority of patients. In fact, 77% of the 31 patients having serologic tests were negative for all the listed serological tests. The exact number of these 31 patients who had some degree of villous atrophy was not reported, but would be expected to be similar to the overall numbers listed above.

Selby et al.[414] investigated whether the failure of mucosal recovery was due to noncompliance with a GFD. Eighty-nine adult patients with CD on a GFD for a mean in excess of 8 years underwent dietary assessment by a dietician, questionnaire and food diary. They were then classified as either Codex GFD, which allows up to 0.03% of protein from a gluten source, or no-detectable gluten GFD (NDG-GFD). Villous atrophy persisted at high rates in both groups, with 46% of those on Codex GFD and 40% of those on NDG-GFD having persistent villous atrophy. The patients in this study did not have clinical features of refractory sprue. Based on the fact that there were similar histologic profiles in both groups, the authors postulate that persisting mucosal abnormalities may be unrelated to gluten non-compliance. Of course, gluten intake in the NDG-GFD group undetected by study protocols cannot be ruled out.

Serology. The studies assessing the utility of serology in monitoring adherence can be divided into those with,[397,398,403,404,407,409,413,415] and those without[396,399-402,408,410,412] biopsy correlation. The studies without biopsy correlation are reviewed first. They establish an association between serologic positivity and patient compliance.

Bartholomeusz et al.[396] demonstrated higher rates of IgA-AGA positivity in non-compliant as compared with compliant CD patients in a mixed population. How compliance was ascertained is not described. Three of the 17 (17.6 %) patients compliant with a GFD for greater than 6 months were IgA-AGA positive as compared with 11 of 12 (91.6%) non-compliant patients. The PPV for non-compliance was calculated to be 78.5%.

Burgin-Wolff et al.[400] showed that, as expected, serology becomes positive with gluten challenge. One hundred and thirty-four children with CD underwent gluten challenge and were assessed for IgA-AGA and IgA-EMA-ME. At baseline, the rate of serologic positivity was 23% for AGA and 13% for EMA. Within 3 months of gluten challenge, 97% of children were positive for AGA and 65% positive for EMA. Between 3 months and 1 year, 85% of children were positive for AGA and 84% positive for EMA.

In a mixed population, Fabiani et al.[408] demonstrated significantly higher IgA-tTG-GP values in patients deemed to be non-compliant with a GFD as compared with compliant patients.

Bardella et al.[399] demonstrated that the positivity of various serologic markers falls in adults with duration on a GFD (Evidence Tables, Appendix I). The five groups in this study were untreated CD, poor GFD compliance, GFD less than 2 years, GFD greater than 2 years, and a control group. As expected, IgA-AGA, IgA-EMA-ME and IgA-tTG-GP were positive in virtually all untreated CD patients. Also, as expected, there was a low rate of positive serology in the control group, with a higher percentage being IgA-AGA positive than either IgA-EMA-ME or IgA-tTG-PG. In the poorly-compliant CD group, all were positive for all three serologic tests. In patients on a GFD less than 2 years, the rates of positive AGA, EMA and tTG were 40.9%, 54.5%, and 63.6%, respectively. In patients on a GFD for more than 2 years, the rates were 16.2%, 9.5% and 14.2%, respectively. The overlap of CIs intervals was such that no differences between the serologic tests could be determined.

Vahedi et al.[402] studied IgA-EMA and IgA-tTG in adult CD patients. Based on dietary inquiry, patients were divided into those on a strict GFD, those with minor transgressions and those with major transgressions. It was not reported whether the EMA was ME or HU, nor was it reported whether tTG was GP or HR. The median duration of GFD was 75 months. Among those on a strict GFD, 2.5% and 3% were IgA-EMA and IgA-tTG positive, respectively. Among those with minor transgressions, positivity was only 37% and 31%, respectively. Among those with major transgressions, positivity was 86% and 77%, respectively. The sensitivity of IgA-EMA for any dietary transgression was 66%, and for minor transgression it was 37%. For IgA-tTG, the sensitivities were 52% and 31%, respectively. No statistically significant differences were detected between the two serologic tests.

In a mixed population, Scalaci et al.[401] showed a low reliability for IgA-EMA in picking up dietary transgressions reported at interview. It is not reported whether ME or HU was used. In patients on a GFD for at least 6 months, only 11.1% those patients reporting one dietary transgression per month were positive, and only 19% reporting one dietary transgression per week were positive.

Fabiani et al.[410] showed a similarly low rate of serologic detection of non-compliance in screen-detected adolescents. Of 6,315 screened students, 28 biopsy-proven CD patients were found. Of these, 23 agreed to participate in a follow-up study. The mean duration of GFD was 23 months. IgG-AGA, IgA-AGA and IgA-EMA were measured. Whether EMA was ME or HU was not reported. Of the 11 patients reporting any dietary transgression, only two patients (19%) were positive for any of the serologic tests.

Pacht et al.,[412] in a similar study, showed different results. Seventeen children deemed compliant with GFD for at least 1 year were all IgA-EMA-ME-negative, whereas, 22 children deemed non-compliant were IgA-MA-ME-positive. This study suggests a much higher sensitivity for EMA than in other studies.

A number of further studies include serology and biopsy correlation. These are reviewed below.

Sategna-Guidetti et al.[413] looked at 47 adults with CD. All were IgA-EMA-ME positive at diagnosis. After 8 to 30 months of GFD, a second biopsy was taken and IgA-EMA-ME was remeasured. Total AGA was also measured in 39 patients. No patient in which the mucosa recovered to normal had a positive EMA. Only one patient with normal histology had a positive AGA (2.6%). EMA was positive in only five of 23 patients with partial villous atrophy, three of 13 patients with subtotal villous atrophy, and one of two patients with total villous atrophy. AGA was positive in only seven of 20 patients with partial villous atrophy, five of ten patients with subtotal villous atrophy, and two of two patients with total villous atrophy. The PPV of EMA for abnormal histology was 100%, but the NPV was only 23%. The PPV-AGA (total) for abnormal histology was 93.8%, whereas the NPV was only 25%. There was a clear inability of serology to adequately reflect the mucosal state in this study, and serology was negative in a significant number of patients with villous atrophy.

Valentini et al.[407] also found a significant rate of negative serology despite the presence of villous atrophy. In an adult population on a GFD for a mean of 9.9 months (range 6-12 months), 24 patients were IgA-EMA-ME negative on a GFD. Seventeen of these 24 patients (71%) had varying degrees of villous atrophy on biopsy (14 had partial villous atrophy and three had subtotal villous atrophy).

Dickey et al.[409] also showed that disappearance of IgA-EMA-ME did not necessarily indicate mucosal recovery. In adults on GFD for 1 year, IgA-EMA-ME was positive in only two of 22 (9%) with partial villous atrophy, and three of ten (30%) with subtotal/total villous atrophy.

Mengozzi et al.[403] investigated adult CD patients on a GFD for 1 year. Most (95%) had a Marsh III histology at diagnosis. In general agreement with the prior studies, only 12% had normal histology at follow-up biopsy 1 year later. Fifty percent were Marsh I and 38% were Marsh II or III (individual results for Marsh II and III were not reported). IgA-EMA-ME, IgA-tTG-HR (four different assays: DRG Diagnostics, Eurospital, Immunodiagnostik, and Celikey), and IgA-tTG-GP were measured. Taking complete mucosal recovery as a negative biopsy and all other biopsies as positive, the authors looked at concordance of serology to biopsy results. Concordance for EMA, tTG1, tTG2, tTG3, tTG4 and tTG5-PG were 48%, 29%, 65%, 14%, 16%, 19%, respectively. The validity of a Marsh I or perhaps Marsh II histology being classified as positive is unclear, and it would have been interesting to know the corresponding concordance rates if Marsh 0-I and Marsh 0-II were considered normal.

Kaukinen et al.[398] similarly found a lack of correlation between IgA-EMA-HU, IgA-tTG-GP and histologic state. Of 87 adult patients on a GFD for a median of 1 year, 27 still had a Marsh III villous atrophy. Among those with Marsh III villous atrophy, EMA was negative in 74% and tTG was negative in 59% of patients. Furthermore, of 11 patients admitting regular dietary lapses, 55% were EMA and tTG negative. The sensitivity, specificity, PPV, and NPV of EMA for Marsh III villous atrophy was 26%, 93%, 63%, and 74%, respectively. The values for tTG were 41%, 88%, 61% and 77%, respectively.

The issue arises as to whether serology might more accurately reflect mucosal state in long-term follow-up. In patients on GFD over 5 years,[398] two of four patients with Marsh III villous atrophy were EMA and tTG negative, and five of nine patients (56%) admitting dietary transgressions were EMA and tTG negative. In this study, there was no clear advantage of tTG over EMA.

One study by Fotoulaki et al.[397] did show a good correlation between serology and mucosal state. In a mixed population of 30 patients, IgG AGA, IgA AGA and IgA-EMA-ME was measured after 12 months of GFD. Contrary to the preceeding studies, all patients had either a Marsh I or II biopsy on a GFD, and all were IgA AGA and IgA EMA negative, while 40% were still IgG-AGA positive. The age range of patients in this study was much younger (1 to 24 years).

Troncone et al.[415] demonstrated that serology could miss dietary transgressions in children. Twenty-three adolescents were divided into four groups, depending on assessment of gluten intake. IgA-EmA-ME was present in seven of seven patients assessed to be taking >2 g/day of gluten. All seven also had villous atrophy. Conversely, four patients on a strict GFD, had normal histology and negative EMA. For patients with intermediate levels of gluten intake, one of six patients with a gluten intake of less than 0.5 g/d had a positive EMA. This patient also had partial villous atrophy. Three patients in this group had lesser mucosal abnormalities (increased IELs) and negative serology. For patients ingesting 0.5 to 2 g/d of gluten, three had a positive EMA; two of these had villous atrophy. Five patients had increased numbers of IELs.

Interventions to Promote Adherence to a GFD

Anson et al.[416] investigated 43 Jewish Israeli children with CD, and their parents. Thirty-one of the children (70%) were judged compliant based on a combination of clinical symptoms,

biopsy and AGA. It is unclear if serology and biopsy was performed in all children to assess compliance. Parental knowledge was studied using a structured questionnaire. A significant positive correlation between the father being a professional and compliance was found ($p<.01$). Parental level of education was also significantly correlated with compliance. Significant differences in parental ability to choose GFD items from a specific menu were found. Ninety three percent of parents of compliant children were able to pick all five GFD items out of an eight-item menu. This compared with only 67% of parents of non-compliant children ($p<.05$).

In another parental questionnaire, Jackson et al.[418] found that 30 of 50 (60%) parents reported their children to be on a strict GFD. Dietary compliance correlated with membership in the Celiac Society ($p<0.0001$). It also correlated with parental score on an eight-question test related to knowledge of CD ($p<0.001$).

Ljungman et al.[420] found self-reported GFD compliance in children to be positively associated with knowledge of CD. In this study of 47 Swedish children, those deemed compliant scored 14.03 out of 15 on a knowledge test related to CD. This compared with an average score of 12.44 in the non-compliant group.

Lamontagne et al.[419] surveyed 617 past and present members of the Quebec Celiac Foundation. A final sample size of 234 was obtained. Self-reported compliance difficulty with a GFD was inversely correlated with a high level of confidence in treatment information from gastroenterologists and dieticians ($p<.005$).

Hogberg et al.[421] looked at the effect age of diagnosis might have on compliance. In a study population of 29 adults with CD, 15 were deemed compliant with a GFD on the basis of a questionnaire and serology (IgA EMA, IgG EMA and IgA tTG). Eighty percent of patients diagnosed prior to age 4 were GFD compliant compared with 36% of patients diagnosed after age 4 ($p<.05$). A drawback of this study is that serologic markers were collected about 3 years prior to the dietary questionnaire. This risks misclassification of patients if their compliance varied over time.

In an important study with relevance to outcomes of population screening, Fabiani et al.[417] showed a lower compliance in 22 adolescents identified by a mass screening program as compared with 22 age-matched controls with identified CD on the basis of symptoms. All patients had been prescribed a GFD for more than 5 years. Twenty-three percent of screen-detected patients reported being on a strict GFD as compared with 68% of those diagnosed with CD on the basis of symptoms. Patients in the screen-detected group were diagnosed at a later age (mean 14.0 yrs) versus patients identified on the basis of symptoms (mean 4.3 yrs).

A colouring book intervention has been developed to promote GFD compliance,[422] but the effectiveness of this intervention has not been assessed in children with CD.

Quality Assessment

The majority of studies in this objective were of a “before–after” design. In this setting, this design may not pose a major limitation for monitoring studies, since the purpose of the study was to assess the change in serology and histology after introduction of a GFD. In this regard, the strength of the evidence for monitoring adherence to a GFD was fairly good. However, there is almost a complete absence of studies of interventions for the promotion of adherence to a GFD.