This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

AHRQ is the lead Federal agency charged with supporting research designed to improve the quality of health care, reduce its cost, address patient safety and medical errors, and broaden access to essential services. AHRQ sponsors and conducts research that provides evidence-based information on health care outcomes; quality; and cost, use, and access. The information helps health care decisionmakers—patients and clinicians, health system leaders, and policymakers—make more informed decisions and improve the quality of health care services.

Evidence Report/Technology Assessment
Number 104

Celiac Disease

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-02-0021

Prepared by:
University of Ottawa Evidence-based Practice Center, University of Ottawa, Ottawa, Canada

Co-directors: David Moher, PhD and Howard M. Schachter, PhD

Investigators
Alaa Rostom,* MD, MSc, FRCPC
Catherine Dubé,* MD, MSc, FRCPC
Ann Cranney,*† MD, MSc, FRCPC
Navaaz Saloojee,* MD, FRCPC
Richmond Sy,* MD, FRCPC
Chantelle Garritty, BA, DCS
Margaret Sampson, MLIS
Li Zhang, MLIS
Fatemeh Yazdi, MSc
Vasil Mamaladze, MD, PhD
Irene Pan, MSc
Joanne McNeil,* RN
David Moher, PhD
David Mack,* MD, FRCPC
Dilip Patel,* MD, FRCPC

Chalmers Research Group; *Gastrointestinal Clinical Research Unit; †Division of Rheumatology

AHRQ Publication No. 04-E029-2
September 2004

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials noted for which further reproduction is prohibited without the specific permission of copyright holders.

Suggested Citation:
Rostom A, Dubé C, Cranney A, Saloojee N, Sy R, Garritty C, Sampson M, Zhang L, Yazdi F, Mamaladze V, Pan I, McNeil J, Moher D, Mack D, Patel D. Celiac Disease. Evidence Report/Technology Assessment No. 104. (Prepared by the University of Ottawa Evidence-based Practice Center, under Contract No. 290-02-0021.) AHRQ Publication No. 04-E029-2. Rockville, MD: Agency for Healthcare Research and Quality. September 2004.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. This report on *Celiac Disease* was requested and funded by the Office of Medical Applications of Research, National Institutes of Health (NIH) for the Consensus Development Conference on Celiac Disease as well as the National Institute of Diabetes and Digestive and Kidney Diseases, NIH. **Marian D. James, Ph.D., served as AHRQ's Task Order Officer in charge of overseeing the report development process.** The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for health care quality improvement projects throughout the Nation. The reports undergo peer review prior to their release.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the healthcare system as a whole by providing important information to help improve health care quality.

We welcome written comments on this evidence report. They may be sent to: Director, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850. **Questions regarding this report should be sent to epc@ahrq.gov.**

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Barnett S. Kramer, M.D., M.P.H.
Director
Office of Medical Applications of Research, NIH

Allen M. Spiegel, M.D.
Director
National Institute of Diabetes and Digestive and Kidney Diseases, NIH

The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services of a particular drug, device, test, treatment, or other clinical service.

Acknowledgments

The authors would like to thank several individuals for their support of the present project: Keith O'Rourke, who helped with the statistical analysis; Karen Patrias, who helped with conducting the literature search; Gabriela Lewin, who assisted with the quality assessment; and Christine Murray and Isabella Steffensen, who assisted in the editing of the report and the generation of evidence tables.

Author Contribution

Dr. Alaa Rostom was the lead investigator. He was involved in all aspects of the study design, management, planning, analysis and write-up, including article screening, data extraction, quality assessment, statistical analysis, and report write-up. Drs. Catherine Dubé and Ann Cranney were the second investigators. Dr. Catherine Dubé was involved in all aspects of study design, and planning and organization, including task management, article screening, data extraction, and quality assessment. She was the lead writer of Celiac 2 and 3. Dr. Ann Cranney was involved in all aspects of study design and planning, including article screening, data extraction, and quality assessment. She was the lead writer of Celiac 4, and oversaw the screening and data extraction for Celiac 3, 4, and 5. Dr. Navaaz Saloojee was involved in study planning, article screening, data extraction, and quality assessment, and was the lead writer of Celiac 5. Dr. Richmond Sy was involved in study planning, article screening, data extraction, and contributed to the writing of Celiac 4. Drs. David Mack and Dilip Patel were involved in study planning and article screening, in addition to being content experts in pediatric and adult celiac disease, respectively. They also reviewed and advised on the report write-up. JoAnne McNeal was involved in article screening and data extraction for Celiac 1 (serology) and Celiac 2 (prevalence).

Dr. David Moher was involved in all aspects of study design, management, planning, analysis, and write-up. He was the methodological content expert and reviewed and advised on all report conduct and documents. Chantelle Garrity was involved in all aspects of project planning and management, including liaison with all key partners. She oversaw the screening progress, document retrieval, and assisted in report management, review and write-up. Margaret Sampson was the lead information specialist and was involved in all aspects of the search strategy/key word design and refinement, in association with the information specialists at the NLM. She was involved in all aspects of article database management, including article retrieval, set-up of the online computerized SRS article screening and extraction system, and the development and write-up of the QUOROM Flow. Li Zhang was the information specialist involved in all aspects of SRS system management, article retrieval, and implementation of the SRS system. Dr. Vasil Mamaladze and Fatemeh Yazdi performed the data extraction of Celiac 1–serology and Celiac 2–prevalence. Irene Pan performed data extraction and quality assessment of Celiac 2–prevalence.

Structured Abstract

Context. Celiac disease (CD) is a disorder of small bowel malabsorption. It is characterized by mucosal inflammation, villous atrophy and crypt hyperplasia that occur upon exposure to gluten, and clinical and histological improvement with withdrawal of gluten from the diet. The classical presentation of CD has now been shown to be less common than silent or atypical presentation, in which patients do not have intestinal symptoms. Untreated CD is associated with multiple important short- and long-term complications including nutritional derangements, anemia, reduced bone density, as well as intestinal lymphoma. In the vast majority of patients, CD is effectively treated with dietary modifications that eliminate gluten. Mounting evidence suggests that CD is actually considerably more common than previously believed and, therefore, this disorder warrants consideration for screening of at-risk patients, as well as possibly the general population.

Objectives. To conduct a comprehensive systematic review on five areas of CD: (1) sensitivity and specificity of serological tests; (2) prevalence and incidence of CD; (3) CD associated lymphoma; (4) consequences of testing for CD; and, (5) interventions for the promotion and monitoring of adherence to a gluten-free diet (GFD).

Data Sources. Staff of the National Library of Medicine performed a series of searches in support of the literature review of CD. Searches were run in the MEDLINE® (1966 to Oct 2003) and EMBASE (1974 to Dec 2003) databases for each of the five objectives and their respective sub-objectives separately.

Study Selection. Study selection for each objective was performed using three levels of screening with predetermined increasingly more strict criteria to ensure that all relevant articles were captured. Following a calibration exercise, two reviewers independently screened all studies using a web-based system allowed automatic identification of review disagreements. These disagreements were resolved by consensus.

Data Extraction. For each CD objective, a detailed and standardized data abstraction form was developed. For each objective, data abstraction was conducted by one reviewer and verified by another. The extracted data was further verified by one of the principal investigators. Quality assessments were performed using specific instruments for each of the included study types.

Data Synthesis. The data obtained from this review fell into several broad categories, which correspond in large part to the individual study objectives. Data for the sensitivity and specificity of each serological marker was considered separately, and studies were further divided according to the age group of the study population. Attempts were made to identify, explain, and minimize clinical and statistical heterogeneity in the included studies. A Pearson's Chi Square with n-1 degrees of freedom, where n represents the number of included studies in an analysis, was calculated to assess statistical heterogeneity. Pooled estimates were only calculated if clinically and statistically appropriate. In situations where pooling was not performed, a qualitative systematic review was conducted.

To produce clinically useful pooled statistics, a weighted mean of the overall sensitivity and specificity from the included studies was calculated, along with 95% confidence intervals (CIs). The pooled estimates for the sensitivity and specificity were compared with a summary receiver operating characteristic (ROC) curve, calculated for the same group of studies as a second check of the estimates.

Results/Conclusions. This report has provided a systematic review of five broad areas (and corresponding sub-areas) of CD. Perhaps one of the most important findings of this report is the significance of how one chooses to define CD in the era of serological testing, and how this apparently clear-cut task has profound implications on all the results presented in this report. Specifically, can CD be diagnosed solely on the basis of serology? Is some degree of villous atrophy necessary for a diagnosis of CD. These questions have important implications downstream of the diagnosis as well. For example, do CD patients without symptoms or villous atrophy have the same risk of complications as those with villous atrophy. Is serological improvement on a GFD sufficient to reduce CD complications, or must there be documented histological improvement, and what degree of histological improvement is necessary?

The results of the Celiac 1 objective suggest that in the era of EMA and tTG antibody testing, AGA antibody testing in both children and adults has a limited role. The sensitivity and specificity of EMA and tTG are quite high (over 95% for sensitivity, and close to 100% for specificity), as are their positive and negative predictive values; however, one has to be aware that the reported diagnostic parameters are taken from studies in which the prevalence of CD was, for the most part, much higher than that seen in usual clinical practice. The positive predictive values reported for these tests will certainly not be as high as that reported when these tests are used to screen the general population. The bulk of the evidence on the diagnostic characteristics of these tests was derived from studies that defined CD as having at least some degree of VA.

HLA DQ2/DQ8 testing appears to be a useful adjunct in the diagnosis of CD. The test has high sensitivity (in excess of 90%-95%), however, since approximately 30% of the general population, and an even higher proportion of “high-risk” subjects (e.g., diabetics and family members) also carry these markers, the specificity of this test is not ideal. The greatest diagnostic utility of this test appears to be its negative predictive value.

Biopsy itself, when used with a strict cut-off requiring villous atrophy, appears to have high specificity, but poor sensitivity. Using a lower grade cut-off clearly improves sensitivity, but because of the wide differential of causes of histological lesions similar to Marsh I to IIIa, the specificity suffers. The use of histomorphometric measures such as quantification of gamma delta positive intraepithelial lymphocytes ($\gamma\delta$+ IELs) are likely to allow for the use of lower grade cut-offs, while maintaining reasonable specificity. Ultimately, a trial utilizing multiple diagnostic tests in an attempt to capture as many CD patients in a clinically-relevant population as possible, along with a time dimension such as a response to a GFD or gluten challenge, is required to fully assess the diagnostic characteristics of biopsy alone. This type of study would be able to characterize the false-positive and false-negative rates, provided that all studied patients are followed forward in time.

The included prevalence studies demonstrated important differences between the studies including, execution, tests for prevalence assessment, and patient sampling. Thus, results have to be interpreted in the light of some of the limitations that have been identified regarding the diagnostic performance of the tests for CD. Nonetheless, the results of this report suggest that

CD is a very common disorder with a prevalence in the general population that is likely close to 1:100 (1%). Several high-risk groups with a prevalence of CD greater than that of the general population have been identified and include: those suspected of having CD; family members of CD patients; type I diabetics; and, those with iron-defiency anemia (IDA) or low bone mineral density (BMD). Additionally, the review identified many other high-risk groups, including those with Down Syndrome, short stature, and infertility, to name a few. Their inclusion was however, beyond the scope of this report

The results of this report confirm that, apart from a few limitations, there is a strong association between CD and GI lymphoma. The report identified standard incidence ratios (SIR) for lymphoma that ranged from 4 to 40, and standard mortality ratios (SMR) that ranged from 11 to 70. A diagnostic delay—in particular a diagnosis of CD in adulthood as apposed to in childhood—is associated with poorer outcomes. Fortunately, several studies suggest that adherence to a GFD reduces the risk of lymphoma in CD patients.

The consequences of testing for CD in at-risk and symptomatic patients appears to be more straightforward, since these patients appear to be more compliant with a GFD and would be expected to benefit from this intervention. The data is less clear for asymptomatic screen-identified patients, particularly those who have truly silent CD and/or don't have fully-developed villous atrophy. On the one hand the outcome of such patients has not been extensively studied, and on the other hand compliance with a GFD appears problematic, particularly for those diagnosed in adulthood.

Finally, no specific interventions have been identified that promote adherence to a GFD, but education of patients and family members about CD and about the intricacies of a GFD, and participation in local celiac societies, has been shown to improve compliance. Although somewhat controversial, biopsy monitoring of adherence to a GFD appears to be important, since improvement in histological grade has been associated with improved BMD, IDA, and nutritional status. The serological markers appear to be adequate for detecting gross dietary indiscretion, and respond to a gluten challenge, but appear to have poor sensitivity for detecting lesser degrees of dietary indiscretion, and inadequately correlating with histological improvement at least in the short-term. It should, however, be noted, that we could not identify a controlled study that objectively determined the level of histological improvement that would be associated with improved outcomes, and this is an area for future study. Nonetheless, based on this report it would appear that follow-up biopsy, at least 1 year after a GFD in adults to document improvement of the histological grade, would be valuable.

Contents

Evidence Report

Chapter 1. Introduction 3
Overview 3
Definition of CD 4
General Definitions 5
Report Purpose and Target Population 6
Methodological Considerations 6

Chapter 2. Methods 9
Overview 9
Key Questions Addressed in This Report 9
Study Criteria Used in this Review 10
Histological 10
Populations 11
HLA DQ2/DQ8 11
Analytical Framework 12
Study Identification 13
Search Strategy 13
Study Selection and Eligibility Criteria 14
Data Abstraction 17
Quality Assessment 19
Data Synthesis and Analysis 19

Chapter 3. Results 21
Celiac 1: Sensitivity and Specificity of Tests for CD 21
Serology 21
HLA DQ2/DQ8 55
Biopsy 65
Quality Assessment 65
Celiac 2: Incidence and Prevalence of CD 66
Incidence of CD in the General Population 67
Prevalence of CD in the General Population—Different Geographic and Racial/Ethnic Populations 71
Prevalence of CD in Patients with Suspected CD 79
Prevalence of CD in with Type I Diabetes 81
Prevalence of CD in Relatives of Patients with CD 89
Prevalence of CD in Patients with IDA 92
Prevalence of CD in Patients with Low Bone Mineral Density (BMD) 95
Quality Assessment 96
Celiac 3: Risk of Lymphoma in CD 97
Literature Search 97
Measures of Risk 99

Study Characteristics 99
Types of Lymphomas 99
Incidence of Lymphoma and Related Mortality Data 100
Role of a GFD 102
Risk of Lymphoma Versus Symptoms 102
Impact of the Age at Diagnosis of CD 102
Risk of Lymphoma in Refractory CD 103
Quality Assessment 103
Celiac 4: Consequences of Testing for CD 104
Part A 104
Part B 109
Quality Assessment 117
Celiac 5: Promoting or Monitoring Adherence to a GFD 118
Monitoring Adherence to a GFD 118
Interventions to Promote Adherence to a GFD 121
Quality Assessment 122

Chapter 4. Discussion 123
Celiac 1: Sensitivity and Specificity of Tests for CD 123
Serology 123
HLA DQ2/DQ8 128
Biopsy 129
Celiac 2: Incidence and Prevalence of CD 139
Incidence in the General Population—Different Geographic and Racial/Ethnic Populations 139
Prevalence in the General Population—Different Geographic and Racial/Ethnic Populations 140
Prevalence of CD in Patients with Suspected CD 140
Prevalence of CD in Patients with Type I Diabetes 141
Prevalence of CD in Relatives of Patients with CD 141
Prevalence of CD in Patients with Anemia 142
Prevalence of CD in Patients with Low BMD 142
Celiac 3: Risk of Lymphoma in CD 143
Celiac 4: Consequences of Testing for CD 146
Expected Outcomes of Treatment of CD 148
Fractures/BMD/Osteoporosis/Osteopenia 149
Mortality 150
Celiac 5: Promoting or Monitoring Adherence to a GFD 150
Monitoring Adherence to a GFD 150
Interventions to Promote Adherence to a GFD 152
Strength of the Body of Evidence 153
Celiac 1 153
Celiac 2 154
Celiac 3 154
Celiac 4 154

Celiac 5 154
Future Research.....155
Conclusion.....156
References and Included Studies 159
Abbreviations and Acronyms.....183

Tables

Table 1: Inclusion/exclusion criteria by level of screening 16
Table 2: Included studies for IgA-AGA in children 23
Table 3: Included studies for IgA-AGA in adults.....24
Table 4: Included studies for IgA-AGA in studies including both children and adults.....25
Table 5: Included studies for IgG-AGA in adults.....27
Table 6: Included studies for IgG-AGA in children 28
Table 7: Included studies for IgG-AGA in studies including both children and adults.....29
Table 8: Included studies for IgA-EMA-ME in adults 32
Table 9: Included studies for IgA-EMA-ME in children.....33
Table 10: Included studies for IgA-EMA-ME in studies including both children and adults 34
Table 11: Included studies for IgG-EMA-ME in adults 36
Table 12: Included studies for IgG-EMA-ME in children.....36
Table 13: Included studies for IgA-EMA-HU in adults.....37
Table 14: Included studies for IgA-EMA-HU in children 37
Table 15: Included studies for IgA-EMA-HU in studies including both children and adults 38
Table 16: Included studies for IgA-tTG-GP in adults.....41
Table 17: Included studies for IgA-tTG-GP in children 41
Table 18: Included studies for IgA-tTG-GP in studies including both adults and children 41
Table 19: Included studies for IgG-tTG-GP in studies including both children and adults 43
Table 20: Included studies for IgG-tTG-HR in studies including both children and adults 45
Table 21: Included studies for IgA-tTG-HR in adults 45
Table 22: Included studies for IgA-tTG-HR in children.....45
Table 23: Included studies for IgA-tTG-HR in studies including both children and adults 46
Table 24: Included studies for combination IgA and IgG AGA, when either test is positive 49
Table 25: Included studies for combination IgA and IgG tTG-HR, when either test is positive..49
Table 26: Included studies for combination IgA-AGA and IgG-EMA-HU, when either test is positive.....49
Table 27: Weighted pooled estimates with 95% CIs and heterogeneity identified 52
Table 28: HLA studies with biopsied cases and controls.....56
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.....59
Table 30: Sensitivity/specificity (calculated) for HLA DQ2 in case-control studies 60
Table 31: Sensitivity/specificity (calculated) for HLA DQ2 in mixed-design studies 62
Table 32: Sensitivity/specificity (calculated) for HLA DQ8 63
Table 33: Sensitivity/specificity (calculated) for HLA DQ2 or DQ8 64
Table 34: Included studies of incidence of CD in the general population 68
Table 35: Prevalence of CD by country 73

Table 36: Prevalence of CD by serological screening test....76
Table 37: Prevalence of CD by statistical percentiles....77
Table 38: Included studies for prevalence of CD in patients with suspected CD....81
Table 39: Included studies of prevalence of CD in type I diabetes....84
Table 40: Summary of prevalence of CD in type I diabetes by age groups and screening test....88
Table 41: Prevalence of CD in relatives of CD patients....91
Table 42: Included studies of CD in adult patients with anemia....93
Table 43: Summary of prevalence of CD in adult patients with anemia by population and screening test....95
Table 44: Prevalence of CD in patients with low BMD....96
Table 45: Included studies for risk of lymphoma in CD....98
Table 46: Results of study assessing γδ+ IELs in patients with and without CD....132
Table 47: Results of study assessing density of γδ+ IELs in patients with untreated CD, treated CD and control patients....133
Table 48: Results of study comparing density of γδ+ IELs in patients with confirmed CD, those undergoing investigation for CD, and control subjects....133
Table 49: Results of study comparing IEL density and villous/crypt ratio in patients with a suspicion of CD, and 59 biopsy-negative controls with dyspepsia....134
Table 50: Results of study assessing IEL density in routinely stained specimens compared with specimens stained with the CD3 antibody....134
Table 51: Results of study comparing IEL density and villous distribution among patients suspected of CD....135
Table 52: Results of study assessing patients with suspected CD and Marsh I or II, before and after a GFD....137

Figures

Figure 1: Analytic framework....12
Figure 2: IgA-AGA in children with CD....25
Figure 3: IgA-AGA in adults with CD....26
Figure 4: IgA-AGA in adults and children with CD....26
Figure 5: IgG-AGA in adults with CD....29
Figure 6: IgG-AGA in children with CD....30
Figure 7: IgG-AGA in children and adults with CD....30
Figure 8: IgA-EMA-ME in adults with CD....34
Figure 9: IgA-EMA-ME in children with CD....35
Figure 10: IgA-EMA-ME in adults and children with CD....35
Figure 11: IgA-EMA-HU in adults with CD....38
Figure 12: IgA-EMA-HU in children with CD....39
Figure 13: IgA-EMA-HU in adults and children with CD....39
Figure 14: IgA-tTG-GP in adults with CD....42
Figure 15: IgA-tTG-GP in children with CD....42
Figure 16: IgA-tTG-GP in adults and children with CD....43
Figure 17: IgG-tTG-GP in adults with CD....44
Figure 18: IgA-tTG-HR in adults with CD....46

Figure 19: IgA-tTG-HR in children with CD 47
Figure 20: IgA-tTG-HR in adults and children with CD 47
Figure 21: PPV and prevalence from individual studies.....53
Figure 22: NPV and prevalence from individual studies 53
Figure 23: PPV based on the pooled estimates of sensitivity and specificity 54
Figure 24: HLA DQ2 64
Figure 25: HLA DQ2 and DQ8.....65
Figure 26: Frequency distribution of prevalence of CD by serology among included studies.....77
Figure 27: Frequency distribution of prevalence of CD by biopsy among included studies.....78
Figure 28: Prevalence of CD by country.....78
Figure 29: General population prevalence in relation to sample size 79
Figure 30: Prevalence of CD in diabetes by study size.....89
Figure 31: PPV based on pooled estimates of sensitivity and specificity.....147

Appendixes

Appendix A: Histological Data
Appendix B: Search Strategies
Appendix C: Data Assessment and Data Abstraction Forms
Appendix D: Quality Assessment Forms
Appendix E: Summary ROC Curves
Appendix F: Modified QUOROM Flow Chart
Appendix G: Raw Pooled Data
Appendix H: Biopsy Results
Appendix I: Evidence Tables
Appendix J: Quality Assessment
Appendix K: Additional Acknowledgments
Appendix L: List of Excluded Studies

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