Evidence-based Practice Center Systematic Review Protocol

Project Title: Diagnosis of Celiac Disease

I. Background and Objectives for the Systematic Review

Context

Celiac disease (CD) is an immune-mediated disorder triggered in genetically-susceptible individuals by ingestion of foods containing gluten, a family of proteins found in particular grains.[1] The prevalence of CD in the United States has been estimated to be approximately one percent,[2] but appears to be increasing, for reasons that are not clear.[3] Risk factors for CD include positive family history, trisomy 21, Turner syndrome, and Williams's syndrome, as well as several autoimmune diseases.

Clinical signs of CD include aphthous ulcers, low bone density, dermatitis herpetiformis, and other symptoms similar to those of other gastrointestinal disorders among which diarrhea, constipation, nausea, vomiting, and abdominal bloating are common. Its diagnosis therefore is quite complicated because many other factors could be responsible for these symptoms.

CD causes enteropathy of the proximal small intestine resulting in poor absorption of nutrients which may subsequently lead to iron-deficiency anemia, chronic fatigue, failure to thrive, and weight loss.[4] In young children, malabsorption may cause stunted growth and delayed puberty.[5] In women, folate deficiency may lead to poor birth outcomes, including developmental disorders. In the long-term, untreated CD may increase the risk for non-Hodgkin lymphoma.[4]

The only known effective treatment for CD is avoidance of gluten-containing foods, which makes correct and timely diagnosis possibly the most important component in the management of the illness. Many experts believe that diagnosis is important for both symptomatic patients and asymptomatic individuals who are at high risk.

The validity and acceptability of the various diagnostic testing regimens and criteria have remained controversial. A number of diagnostic methods have been developed; they include various serology tests, HLA typing, video capsule endoscopy and endoscopic duodenal biopsy (which is considered to be the gold standard). Serology tests include Anti-gliadin antibodies (AGA), IgA & IgG; Anti-tissue Transglutaminase (tTG), IgA & IgG; Endomysial antibodies (EmA), IgA; and the Deamidated Gliadin Peptide (DGP) Antibodies, IgA & IgG. These tests are often used by providers as a panel in order to increase specificity and prevent false positives that are associated with tests that don't work well under varying circumstances.

Rationale for proposed systematic review

Several systematic reviews and guidelines on diagnosis of CD have been published in the past decade, often with contradictory findings and recommendations. There presently exist at least five published guidelines regarding the diagnosis of CD by recognized research/academic/medical bodies such as the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition (NASPGHAN)[6] and European Society for Pediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN)[7] as well as different systematic reviews by government agencies such as the US Agency for Healthcare Research and Quality (AHRQ)[8] and the UK National Institute for Clinical Excellence (NICE).[9]

The clinical practice guidelines are complex and recommend different approaches to diagnosis. For example, some guidelines propose a different sequence of tests for diagnosing population groups such as symptomatic vs non-symptomatic patients at increased risk (e.g. ESPGHAN). In addition, although some guidelines (ACG and World Gastroenterology Organization (WGO)) still uphold endoscopic biopsy as the gold standard for confirming diagnosis[5, 10], other guidelines (ESPGHAN) explore the use of other tests to serve as substitutes for biopsy even though it can occasionally lead to diagnostic error e.g. pathologists' misinterpretation, over-interpretation, and inadequate biopsy sample numbers.

The diagnosis of celiac disease is further complicated by the limitations of the various tests, lack of provider knowledge, and variability in laboratory cut-off levels to indicate "positive" results. On the other hand, as general population screening is not recommended by any professional society, 85% of the population still remains undiagnosed[2]There may be a high level of underdiagnoses in some populations (for example, low SES) where CD symptoms may be considered "normal" and not receive medical attention. It is also unknown whether different racial, ethnic or other demographic subgroups may be incorrectly diagnosed or underdiagnosed. In addition, false positives and false negatives may lead to significant consequences, including huge lifestyle changes involved for positive diagnosis as well as potential health harm (nutrient malabsorption, osteoporosis, and lymphoma) from undiagnosed CD.

Therefore, this proposed systematic review will investigate which diagnostic methods (including combinations and sequences) are most accurate for various patient types and assess if methods of diagnosis are associated with improvement in clinical and patient-centered outcomes. This project will incorporate results of previous SRs on test accuracy and update them to reflect results of new studies. It will address test combinations and sequences, and accuracy in specific population groups beyond those addressed in previous reviews.

II. The Key Questions

AHRQ posted the Key Questions (KQ) on the Effective Health Care Website for public comment in February, 2014. The EPC refined and finalized the key questions after

review of the public comments, taking into consideration prior input from Key Informants. This input is intended to ensure that the key questions are specific and relevant.

In response to the public posting, one commenter representing an association of celiac researchers requested we add a question on who should be tested for CD. The results of this review may be used by patients, providers, policy makers and guideline developers to help inform clinical decision making; however, making direction recommendations is beyond the scope of AHRQ systematic reviews. Other comments requested language clarification regarding which diagnostic methods were being examined for accuracy, and what is meant by direct and indirect adverse effects. To improve clarity, we revised the language slightly and changed the order of the questions. Finally, in response to a comment about how previously negative testing affects test accuracy, we added "patients who previously tested negative for CD" as a population, for subquestion 3f.

The key questions for this systematic review are thus:

Question 1: What is the comparative effectiveness of the different diagnostic methods (various serological tests, HLA typing, video capsule endoscopy, used individually and in combination) compared with endoscopy with biopsy as reference standard, to diagnose Celiac Disease (CD) in terms of:
a- Accuracy (sensitivity, specificity, LR+, LR-, summary ROCs)
b- Intermediate outcomes such as clinical decision making and dietary compliance
c- Clinical outcomes and complications related to CD
d-Patient-centered outcomes such as QOL and symptoms
Question 2: Does accuracy/reliability of endoscopy with duodenal biopsy vary by:
a- pathologist characteristics, i.e. level of experience or specific training?
b- method, i.e. type or number of specimens?
c- length of time ingesting gluten before diagnostic testing?
Question 3: How do accuracy, (sensitivity, specificity, LR+, LR-, summary ROCs) and outcomes differ among specific populations? (subgroups of KQ1)
a -Symptomatic patients vs. non-symptomatic individuals at risk
b- Adults (over age 18) versus children & adolescents
c- Children under age 24 months vs. older children
d- Demographics, including race, genetics, geography, SES
e- Patients with IgA deficiency
f- Patients previously testing negative for CD
Question 4: What are the direct adverse effects (i.e. bleeding from biopsy) or harms (related to false positives, false negatives, indeterminate results) associated with testing for CD?

In addition, we identify under PICOTS (Populations, Interventions, Comparators, Outcomes, and Timing) the following areas of consideration for the key questions:

Population(s):

- For KQ 1, 2, and 4:
  - All populations tested for CD
- For KQ 3:
  - Patients with signs and symptoms of Celiac Disease, for example:
    - Diarrhea
    - Constipation
    - Dermatitis
    - Malabsorption (anemia, folate deficiency)
  - Asymptomatic individuals at risk of Celiac Disease
    - Family history
    - Type 1 diabetes
    - Auto-immune disease
    - Turner's syndrome
    - Trisomy 21
  - Children, under age 24 months vs older children & adolescents
  - Adults (aged 18+)
  - Ethnic and geographic populations
  - Low socioeconomic status (SES)
  - Patients with IgA deficiency
  - Patients previously testing negative for CD

Interventions:

- For KQ 1, 3:
  - Endomysial antibodies (EmA) IgA test
  - Anti-tissue Transglutaminase (tTG) IgA test
  - Deamidated Gliadin Peptide (DGP) IgA Antibodies
  - EmA IgG, tTG IgG, and DGP IgG tests for IgA deficient individuals
  - HLA typing
  - Video capsule endoscopy
  - Combinations of the above
- For KQ 4:
  - Endoscopy with biopsy
  - Gluten free diet
- For KQ 3:
  - All of the above

Comparators:

- For KQ 1, 3:
  - Endoscopy with duodenal biopsy
- For KQ 2:
  - Repeat biopsy

Outcomes:

- For KQ 1a, KQ2 and 3a-f, for Accuracy

    - Sensitivity
    - Specificity
    - PPV, NPV, FP, FN
    - Positive and negative likelihood ratios
- For KQ 1b, for Clinical decision-making
    - Additional testing for CD
    - Nutritionist advice on gluten-free diet
- Follow up and monitoring by For KQ 1c, for Clinical outcomes and complications
    - Nutritional deficits
    - Persistence of villous atrophy on biopsy
    - Lymphomas
- For KQ 1d, for Patient-centered outcomes
    - Quality of life
    - Discomfort
    - Bloating
    - Abdominal pain
    - Depression
- For KQ 4, for Harms
    - Immediate AEs from biopsy
    - Psychological stress related to false positive results
    - Sequelae of false negatives or indeterminate results

Timing:

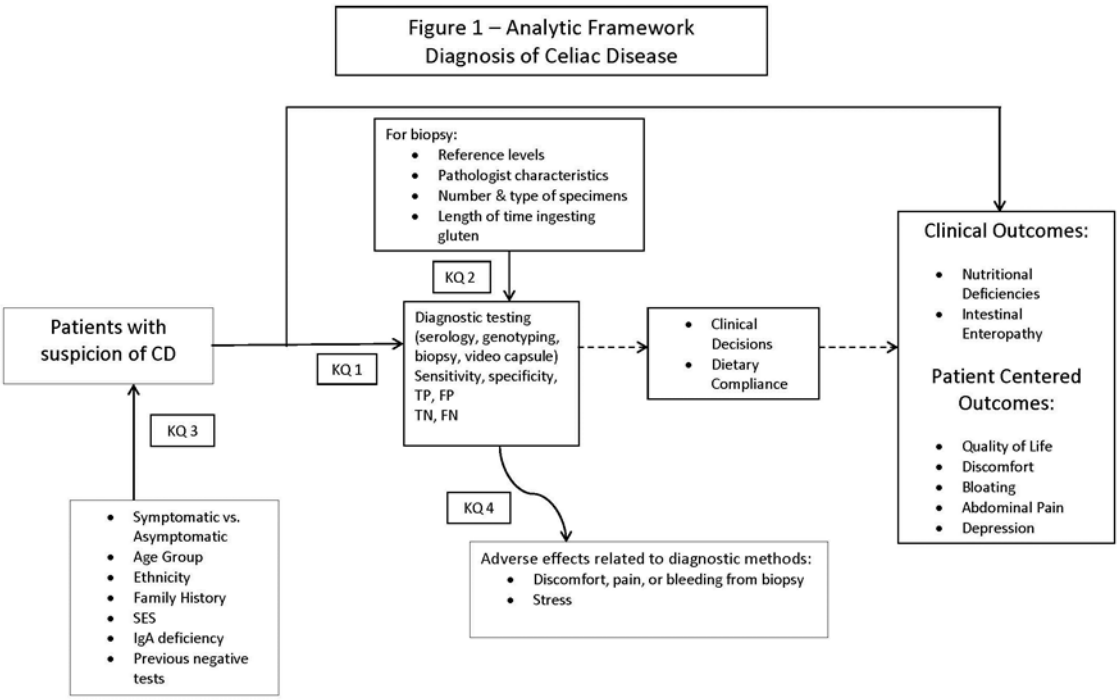
- For KQ 2
    - Length of time ingesting gluten before biopsy

Setting:

- For all KQs
    - Outpatient

III. Analytic Framework.

We provide below an analytic framework to illustrate the population, interventions, outcomes, and adverse effects that will guide the literature search and synthesis for this project.



IV. Methods

In general, this systematic review will follow the procedures of the EPC Methods Guide for Medical Test Reviews.[11]

Criteria for Inclusion/Exclusion of Studies in the Review – are displayed in Table 1 below.

Table 1. Inclusion / Exclusion Criteria

Study design	For KQ1a, we will include high quality systematic reviews (SRs) that address the accuracy of the included diagnostic methods. We will update findings by adding primary studies identified by our searches. For the other KQs, no SRs were identified; therefore, we

	will include primary studies. Included designs acceptable for primary studies include controlled trials, retrospective and prospective cohort studies, and case series. Case reports will be excluded for all KQs.
Study methodology	Studies of diagnostic accuracy (KQ1a, KQ3a-f) will be included only if all individuals underwent both an intervention diagnostic method and a comparator diagnostic method (as listed in the PICOTs in Section II). For example, studies which follow patients with positive serology but no biopsy will be excluded. Only studies that report sensitivity, specificity, PPV, NPV, FP, FN, LR+, LR- or data that allow our calculation will be included.
Language	Studies must be published in English.
PICOTs	Each study must address a population, intervention, comparator, and outcome listed under PICOTs in Section II of this document.

Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions – Our literature methods are summarized in Table 2.

Table 2. Literature search methods

Publication dates	For KQ1a, high quality SRs according to AMSTAR criteria will be used. In addition, our searches will cover the year before in order to not to miss relevant publications and individual studies published after the last SR. For KQ3, on specific populations, our search starts at January, 1990. This year represents the start date for the current EMA test in the US he other relevant serological tests were developed after that date (for example, the current tTG became available in 2000). For KQ2, on duodenal biopsy, our search also starts at January, 1990. ESPGHAN published revised criteria for diagnosis of celiac disease that year[12] which reduced the suggested number of duodenal biopsies from three to one due to the advent of serological tests. For KQ4, on direct and indirect harms, our search starts at January, 2003. This question was covered by an AHRQ funded systematic review published in 2004.[8]
Search terms	The search strategies are designed by our reference librarian in collaboration with our local content expert and project staff. In brief, they include search terms for celiac disease, combined with general terms for diagnosis or terms representing each diagnostic method, plus terms representing the outcomes listed in the PICOTs. The full search strategy is attached as Appendix A. An update search will be conducted after submission of the draft

	report, while the draft undergoes peer review and public comment.
Electronic databases	PubMed, Embase, Cochrane, Web of Science
Scientific Information Packets (SIPs)	Unpublished data regarding KQs 1, 3, and 4 will be requested directly from manufacturers of all serological tests.
Suggestions from experts	We received studies from Key Informants and project clinicians during the Topic Refinement period. Members of this project's Technical Expert Panel (TEP) may also suggest studies. During review of the draft report, peer reviewers and the public may suggest additional studies.
Reference Mining	The reference lists of included articles will be reviewed for identification of relevant studies. If not already included in the project database, the selected references will be ordered and screened for inclusion.

Existing reviews that we are considering including (in total or in part) as evidence will be assessed for their PICOTS as well as the methods they used to assess risk of bias. If important domains were excluded from either the risk of bias assessment, we will determine the need to assess the original studies included in those reviews along these domains. We will also assess the quality of existing reviews we include as evidence using the AMSTAR[13] instrument described below, clearly report the analytic methods used, assess the applicability to the populations of interest, and include the reviews' complete search strategies in an appendix.

An update search will be conducted after submission of the draft report, while the draft undergoes peer review and public comment.

The DistillerSR software package will be used to manage the search output, screening, and data abstraction. Titles and abstracts identified by the searches will be dually screened and all selections will be accepted without reconciliation for further, full-text review. Two researchers will independently screen each full text article for inclusion in the project, with consensus resolution. The lead investigator or clinical expert will resolve any disagreements. The database can be used to calculate inter-rater reliability statistics of agreement and agreement adjusted for chance (kappa statistic) before resolution of disagreements. Assessment of inter-rater reliability can be used to guard against selection bias in choosing the articles for further review.

Data Abstraction and Data Management – Data collection forms will be designed by the project team in Distiller SR, piloted by the reviewers, further modified, and then the final forms piloted with a random selection of included studies to ensure agreement of interpretation. Studies based on large prospective cohorts will be identified in their Distiller records to ensure data are not duplicated. Study-level data abstracted will include: date study conducted; sample size; subjects' demographic characteristics, symptoms, and risk factors; study inclusion/exclusion criteria; study design including analytic methods used; diagnostic methods including test manufacturer and reference standard; and any other potential confounders. Each study will be abstracted

independently by two experienced researchers who will meet to reconcile any differences. Any irreconcilable differences will be discussed with the lead investigator at weekly meetings. At the project's end, all abstracted data will be uploaded to the federally-fund Systematic Review Data Repository.

Assessment of Methodological Risk of Bias of Individual Studies - Risk of bias of each included study will be categorized as high, medium, or low, based on the Methods Guide for Diagnostic Tests, Chapter 5 (Santaguida, 2012).[11] We will use the QUADAS-2 instrument for studies on test accuracy; domains and items are described in Table 3 below. As mentioned above, we will use the AMSTAR instrument,[12] described in Table 4 below, to assess the quality of existing systematic reviews. Each study will be scored individually by two experienced EPC researchers. Discrepancies will be resolved by an experienced methodologist.

Table 3. QUADAS-2 questions for assessing risk of bias in diagnostic accuracy studies

Domain 1: Patient Selection
Was a consecutive or random sample of patients enrolled? (Yes/No/Unclear)
Was a case-control design avoided? (Yes/No/Unclear)
Did the study avoid inappropriate exclusions? (Yes/No/Unclear)
Could the selection of patients have introduced bias? Risk: Low/High/Unclear
Domain 2: Index Test(s) (complete for each index test used)
Were the index test results interpreted without knowledge of the reference standard? (Yes/No/Unclear)
If a threshold was used, was it pre-specified? (Yes/No/Unclear)
Could the conduct or interpretation of the index test have introduced bias? Risk: Low/High/Unclear
Domain 3: Reference Standard
Is the reference standard likely to correctly classify the target condition? (Yes/No/Unclear)
Were the reference standard results interpreted without knowledge of the results of the index test?
(Yes/No/Unclear)
Could the reference standard, its conduct, or its interpretation have introduced bias?
Risk: Low/High/Unclear
Domain 4: Flow and Timing
Was there an appropriate interval between index test(s) and reference standard? (Yes/No/Unclear)
Did all patients receive a reference standard? (Yes/No/Unclear)
Did all patients receive the same reference standard? (Yes/No/Unclear)
Were all patients included in the analysis? (Yes/No/Unclear)
Could the patient flow have introduced bias? Risk: Low/High/Unclear

Table 4. AMSTAR criteria for assessing quality of systematic reviews.

AMSTAR criteria:

01. Was an a priori study design provided?
02. Was there duplicate study selection and data extraction?
03. Was a comprehensive literature search performed?
04. Was the status of publication (gray literature) used as an inclusion criterion?
05. Was a listed of studies (included/excluded) provided?
06. Were the characteristics of the included studies provided?
07. Was the scientific quality of the included studies assessed and documented?
08. Was the scientific quality of the included studies used appropriately in formulating conclusions?
09. Were the methods used to combine the findings of studies appropriate?
10. Was the likelihood of publication bias assessed?

Source: v
Publishe

Data Synthesis – Studies that report the sensitivity, specificity, positive and negative predictive value, receiver-operating characteristics, or provide the data to perform such calculations may be potentially included in a synthesis. If three or more studies of the same diagnostic method and comparator report sensitivity and specificity their results may be pooled. Summary Receiver Operating Characteristic (ROC) curves will be estimated by plotting sensitivity versus 1-specificity and "area under the curve" will be calculated. Studies will be weighted by sample size. Sensitivity analyses will be conducted by performing analyses stratified by population characteristics listed in Section II of this protocol when possible. When pooling is not possible, study results will be described narratively, according to comparisons of interest and study design, and presented in summary tables.

Grading the Strength of Evidence (SOE) for Major Comparisons and Outcomes. We will assess the overall strength of evidence for accuracy outcomes using guidance created by experts in systematic reviews for the AHRQ Effective Health Care Program.[13] This method classifies the grade of evidence as based on five required domains: study limitations, consistency, directness, precision, and publication bias. Three additional domains (plausible confounding, dose-response, and magnitude of effect) will be used if applicable. The grades and their definitions are presented below in Table 5.

Table 5. Strength of Evidence Definitions

High	**We are very confident that the estimate of effect lies close to the true effect for this outcome.** The body of evidence has few or no deficiencies. We believe that the findings are stable, i.e., another study would not change the conclusions
Moderate	**We are moderately confident that the estimate of effect lies close to the true effect for this outcome.** The body of evidence has some deficiencies. We believe that the findings are likely to be stable, but some doubt remains.
Low	**We have limited confidence that the estimate of effect lies close to the true effect for this outcome.** The body of evidence has major or numerous deficiencies (or both). We believe that additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.
Insufficient	**We have no evidence, we are unable to estimate an effect, or we have no confidence in the estimate of effect for this outcome**. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion.

Table 6 below, taken from the AHRQ Methods Guide for Diagnostic Tests, briefly describes the methods used to rate each domain. The rating system was originally designed to assess the body of evidence on health care interventions rather than diagnostic tests; thus, assessing these domains presents unique challenges. For example, in assessing the precision of estimates of test performance, it may be difficult to judge whether a particular confidence interval has any practical clinical implications. In addition, there may be no direct evidence to link a specific test with clinical outcomes.

As this project will include existing high quality systematic reviews, we will report any SOE ratings developed by their authors. (Some funding agencies do not require SOE ratings.) We will discuss their methods of evaluating the domains described below. If we feel that their methods are inadequate or do not address an important domain, we will re-

evaluate per the table below. Of course, if studies applicable to a Key Question are published after the systematic review, we will evaluate the body of evidence as a whole.

Table 6. Required and additional domains and their definitions

Domain	Definition and Elements	Application to Evaluation of Diagnostic Test Performance
Risk of Bias	Risk of bias is the degree to which the included studies for a given outcome or comparison have a high likelihood of adequate protection against bias (i.e., good internal validity), assessed through main elements: • Study design (e.g., RCTs or observational studies) • Aggregate quality of the studies under consideration from the rating of quality (good/fair/poor) done for individual studies	Use one of three levels of aggregate risk of bias: • Low risk of bias • Medium risk of bias • High risk of bias Well designed and executed studies of new tests compared against an adequate criterion standard are rated as "Low risk of bias."
Consistency	Consistency is the degree to which reported study results (e.g., sensitivity, specificity, likelihood ratios) from included studies are similar. Consistency can be assessed through two main elements: • The range of study results is narrow. • Variability in study results is explained by differences in study design, patient population or test variability.	Use one of three levels of consistency: • Consistent (i.e., no inconsistency) • Inconsistent • Unknown or not applicable (e.g., single study) Single-study evidence bases should be considered as "consistency unknown (single study)."
Directness	Directness relates to whether the evidence links the interventions directly to outcomes. For a comparison of two diagnostic tests, directness implies head-to-head comparisons against a common criterion standard. Directness may be contingent on the outcomes of interest.	Score dichotomously as one of two levels of directness: • Direct • Indirect When assessing the directness of the overarching question, if there are no studies linking the test to a clinical outcome, then evidence that only provides diagnostic accuracy outcomes would be considered indirect. If indirect, specific which of the two types of indirectness account for the rating (or both, if this is the case); namely, use of intermediate/surrogate outcomes rather than health outcomes, and use of indirect comparisons, If the decision is made to grade the strength of evidence of an intermediate outcome such as diagnostic accuracy, then the reviewer does not need to automatically "downgrade" this outcome for being indirect.
Precision	Precision is the degree of certainty surrounding an effect estimate with respect to a given outcome (i.e., for each outcome separately).	Score dichotomously as one of two levels of precision: • Precise

	If a meta-analysis was performed, the degree of certainty will be the confidence interval around the summary measure(s) of test performance (e.g., sensitivity, or true positive).	• Imprecise A precise estimate is an estimate that would allow a clinically useful conclusion. An imprecise estimate is one for which the confidence interval is wide enough to include clinically distinct conclusions.
Publications bias	Publication bias indicates that studies may have been published selectively, with the result that the estimate of test performance based on published studies does not reflect the true effect. Methods to detect publication bias for medical test studies are not robust. Evidence from small studies of new tests or asymmetry in funnel plots should raise suspicion for publication bias.	Publication bias can influence ratings of consistency, precision, and magnitude of effect – and, to a lesser degree, risk of bias and directness). Reviewers should comment on publication bias when circumstances suggest that relevant empirical findings, particularly negative or no-difference findings, have not been published or are unavailable.
Dose-response association	This association, either across or within studies, refers to a pattern of a larger effect with greater exposure (including does, duration, and adherence).	The does-response association may support an underlying mechanism of detection and potential relevance for some tests that have continuous outcomes and possibly multiple cutoffs [e.g., gene expression, serum PSA (prostate-specific antigen) levels, and ventilation/perfusion scanning].

Assessing Applicability – Applicability assessment will be based on the similarity of the populations in terms of characteristics listed in the PICOTs. These include age, gender, ethnicity, geographic location, SES, other risk factors, and symptoms. For example, a test may have high sensitivity and specificity in adults but not in small children, due to biological changes during the life course. These issues will be addressed by KQ3. Setting should not affect applicability, as diagnosis of celiac disease almost always occurs in outpatient settings.

V. References

1. Rashtak S, Murray JA. Review article: coeliac disease, new approaches to therapy. Aliment Pharmacol Ther. 2012 Apr;35(7):768-81. PMID: 22324389.

2. Rubio-Tapia A, Ludvigsson JF, Brantner TL, et al. The prevalence of celiac disease in the United States. Am J Gastroenterol. 2012 Oct;107(10):1538-44; quiz 7, 45. PMID: 22850429.

3. Ludvigsson JF, Rubio-Tapia A, van Dyke CT, et al. Increasing Incidence of Celiac Disease in a North American Population. Am J Gastroenterol. 2013 Mar 19PMID: 23511460.

4. See J, Murray JA. Gluten-free diet: the medical and nutrition management of celiac disease. Nutr Clin Pract. 2006 Feb;21(1):1-15. PMID: 16439765.

5. Rubio-Tapia A, Hill ID, Kelly CP, et al. ACG clinical guidelines: diagnosis and management of celiac disease. Am J Gastroenterol. 2013 May;108(5):656-76; quiz 77. PMID: 23609613.

6. Hill ID, Dirks MH, Liptak GS, et al. Guideline for the diagnosis and treatment of celiac disease in children: recommendations of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition. J Pediatr Gastroenterol Nutr. 2005 Jan;40(1):1-19. PMID: 15625418.

7. Husby S, Koletzko S, Korponay-Szabo IR, et al. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. J Pediatr Gastroenterol Nutr. 2012 Jan;54(1):136-60. PMID: 22197856.

8. Rostom A, Dube C, Cranney A, et al. Celiac disease. Evid Rep Technol Assess (Summ). 2004 Jun(104):1-6. PMID: 15346868.

9. National Institute for Health and Clinical Excellence (NICE). Coeliac disease: Recognition and assessment of coeliac disease. National Institute for Health and Clinical Excellence (NICE). London (UK): 2009 May.

10. Bai J, Fried M, Corazza GR, et al. World Gastroenterology Organisation global guidelines on celiac disease. J. Clin Gastroenterol. 2013;47(2):121-6.

11. Chang SM, Matchar DB, al. e. Methods Guide for Medical Test Reviews. AHRQ. 2012.

12. McNeish AS, Harms HK, Rey J, et al. The diagnosis of coeliac disease. A commentary on the current practices of members of the European Society for Paediatric Gastroenterology and Nutrition (ESPGAN). Arch Dis Child. 1979 Oct;54(10):783-6. PMID: 507902.

13. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10. PMID: 17302989.

VI. Definition of Terms

ACG - American College of Gastroenterology
AGA - Anti-gliadin antibodies
CD - Celiac Disease
DGP - Deamidated Gliadin Peptide (DGP) Antibodies
EmA - Endomysial antibodies

ESPGHAN - European Society for Pediatric Gastroenterology, Hepatology, and Nutrition
HLA - Human Leukocyte Antigen
IgA - Immunoglobulin A
IgG - Immunoglobulin G
SES - Socioeconomic status
tTG - Anti-tissue Transglutaminase
WGO - World Gastroenterology Organization

VII. Summary of Protocol Amendments

If we need to amend this protocol, we will give the date of each amendment, describe the change and give the rationale in this section. Changes will not be incorporated into the protocol. Example table below:

Date	Section	Original Protocol	Revised Protocol	Rationale
This should be the effective date of the change in protocol	Specify where the change would be found in the protocol	Describe the language of the original protocol.	Describe the change in protocol.	Justify why the change will improve the report. If necessary, describe why the change does not introduce bias.

VIII. Review of Key Questions

AHRQ posted the key questions on the Effective Health Care Website for public comment. The EPC refined and finalized the key questions after review of the public comments, and input from Key Informants and the Technical Expert Panel (TEP). This input is intended to ensure that the key questions are specific and relevant.

IX. Key Informants

Key Informants are the end users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will inform healthcare decisions. The EPC solicits input from Key Informants when developing questions for systematic review or when identifying high priority research gaps and needed new research. Key Informants are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Key Informants must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals are invited to serve as Key Informants and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Technical Experts constitute a multi-disciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes and identify particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicting opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design, and methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor do they contribute to the writing of the report. They have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

XI. Peer Reviewers

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. The EPC considers all peer review comments on the draft report in preparation of the final report. Peer reviewers do not participate in writing or editing of the final report or other products. The final report does not necessarily represent the views of individual reviewers. The EPC will complete a disposition of all peer review comments. The disposition of comments for systematic reviews and technical briefs will be published three months after the publication of the evidence report.

Potential Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than $10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

XII. EPC Team Disclosures

EPC core team members must disclose any financial conflicts of interest (COI) greater than $1,000 and any other relevant business or professional conflicts of interest. Related financial conflicts of interest that cumulatively total greater than $1,000 will usually disqualify EPC core team investigators. As of February 28, 2014, all project staff members had completed COI forms; none reported any conflict of interest.

XIII. Role of the Funder

This project was funded under Contract HHSA-290-2012-00006I from the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. The Task Order Officer reviewed contract deliverables for adherence to contract requirements and quality. 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.

XIV. Appendix

CELIAC DISEASE
SEARCH METHODOLOGY

KQ1 (DIAGNOSTIC METHODS):

SEARCH #1 (DIAGNOSTIC ACCURACY)
DATABASE SEARCHED & TIME PERIOD COVERED:
PubMed – 1/1/2010-4/3/2014

LANGUAGE:
English

SEARCH STRATEGY:
"celiac disease "[Mesh] OR "celiac disease"[tiab] OR "coeliac disease"[tiab]
AND
diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR "Transglutaminases"[Mesh] OR "HLA Antigens"[Mesh] OR iga OR ttg OR dgp OR IGG OR mass screening OR diagnosis[mh] OR biopsy[mh] OR biopsies[mh] OR "deamidated gliadin peptide antibodies" OR Human leukocyte antigen* OR "video capsule endoscopy" OR endoscop*
AND
Accura* OR Sensitivity and specificity[mh] OR Sensitivity[tiab] OR Specificity[tiab] OR False positive reactions[mh] OR false positive* OR False negative reactions[mh] OR False negative* OR Predictive value OR predictive value of tests[mh] OR Distinguish* OR Differential* OR Identif* OR Detect* OR valid* OR reliab* OR reproducibility of results

==========

SEARCH #2 (INTERMEDIATE OUTCOMES):
DATABASE SEARCHED & TIME PERIOD COVERED:
PubMed – 1/1/2010-4/3/2014

LANGUAGE:
English

SEARCH STRATEGY:
"celiac disease "[Mesh] OR "celiac disease"[tiab] OR "coeliac disease"[tiab]
AND
diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR "Transglutaminases"[Mesh] OR "HLA Antigens"[Mesh] OR iga OR ttg OR dgp OR IGG OR mass screening OR diagnosis[mh] OR biopsy[mh] OR biopsies[mh] OR "deamidated gliadin peptide antibodies" OR Human leukocyte antigen* OR "video capsule endoscopy" OR endoscop* OR misdiagnos* OR undiagnos*
AND

intermediate outcome* OR decision* OR dietary OR diet OR nutrition* OR eating OR food OR foods OR compliance OR comply OR complying OR patient compliance OR adherence OR (("Decision Making"[Mesh]) OR "Decision Support Systems, Clinical"[Mesh]) OR "Food Habits"[Mesh]

==========

SEARCH #3 (CLINICAL OUTCOMES/COMPLICATIONS):
DATABASE SEARCHED & TIME PERIOD COVERED:
PubMed – 1/1/2010-4/3/2014

LANGUAGE:
English

SEARCH STRATEGY:
"celiac disease "[Mesh] OR "celiac disease"[tiab] OR "coeliac disease"[tiab]
AND
diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR "Transglutaminases"[Mesh] OR "HLA Antigens"[Mesh] OR iga OR ttg OR dgp OR IGG OR mass screening OR diagnosis[mh] OR biopsy[mh] OR biopsies[mh] OR "deamidated gliadin peptide antibodies" OR Human leukocyte antigen* OR "video capsule endoscopy" OR endoscop*
AND
clinical outcome* OR complication* OR adverse event* OR adverse effect* OR harm* OR enteropathy OR "quality of life" OR villous atrophy OR abdominal OR anemia OR anemic OR (deficien* AND (folic acid OR folate)) OR "Outcome and Process Assessment (Health Care)"[Mesh] OR "complications" [Subheading] OR "adverse effects" [Subheading] OR "Quality of Life"[Mesh] OR "Folic Acid Deficiency"[Mesh]

==========

SEARCH #4 (ADD TERMS "MISDIAGNOS* OR "UNDIAGNOS*"):
DATABASE SEARCHED & TIME PERIOD COVERED:
PubMed – 1/1/2010-4/3/2014

LANGUAGE:
English

SEARCH STRATEGY:
"celiac disease "[Mesh] OR "celiac disease"[tiab] OR "coeliac disease"[tiab]
AND
misdiagnos* OR undiagnos*

NOTE – THESE RESULTS WERE INCORPORATED INTO THE PREVIOUS RESULT SETS

==========

SEARCH #1 (DIAGNOSTIC ACCURACY):
DATABASE SEARCHED & TIME PERIOD COVERED:

Web of Science Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH – 1/1/2010-4/15/2014

LANGUAGE:
English

SEARCH STRATEGY: ("ts"= topical search)
ts=("celiac disease" OR "coeliac disease")
AND
ts=(diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR Transglutaminase* OR iga OR ttg OR dgp OR IGG OR mass screening OR "Human leukocyte antigen*" OR endoscop* OR misdiagnos* OR undiagnos*)
AND
ts=(Accura* OR Sensitivity OR Specificity OR false positive* OR False negative* OR Predictive value OR Distinguish* OR Differential* OR Identif* OR Detect* OR valid* OR reliab* OR reproducib*)

==========

SEARCH #2 (INTERMEDIATE OUTCOMES):
DATABASE SEARCHED & TIME PERIOD COVERED:
Web of Science Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH – 1/1/2010-4/15/2014

LANGUAGE:
English

SEARCH STRATEGY:
ts=("celiac disease" OR "coeliac disease")
AND
ts=(diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR Transglutaminase* OR iga OR ttg OR dgp OR IGG OR mass screening OR "Human leukocyte antigen*" OR endoscop* OR misdiagnos* OR undiagnos*)
AND
ts=(intermediate outcome* OR decision* OR dietary OR diet OR nutrition* OR eating OR food OR foods OR compliance OR comply OR complying OR adherence)

==========

SEARCH #3 (CLINICAL OUTCOMES/COMPLICATIONS):
DATABASE SEARCHED & TIME PERIOD COVERED:
Web of Science Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH – 1/1/2010-4/15/2014

LANGUAGE:
English

SEARCH STRATEGY:
ts=("celiac disease" OR "coeliac disease")
AND
ts=(diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR Transglutaminase* OR iga OR ttg OR dgp OR IGG OR mass screening OR "Human leukocyte antigen*" OR endoscop* OR misdiagnos* OR undiagnos*)
AND
ts=("adverse effect*" OR "adverse event*" OR "clinical outcome*" OR complication* OR harm* OR enteropathy OR "quality of life" OR villous atrophy OR abdominal OR anemia OR anemic OR "folic acid" OR folate)

==========

DATABASE SEARCHED & TIME PERIOD COVERED:
Cochrane Databases – 1/1/2010-4/15/2014

LANGUAGE:
English

SEARCH STRATEGY:
"celiac disease" OR "coeliac disease" in Title, Abstract, Keywords

NUMBER OF RESULTS: 65
By database:
Cochrane Reviews (0)
Other Reviews (5)
Trials (54)
Methods Studies (0)
Technology Assessments (3)
Economic Evaluations (3)
Cochrane Groups (0)

==========

KQ2 (ENDOSCOPY WITH DUODENAL BIOPSY)

DATABASE SEARCHED & TIME PERIOD COVERED:
PubMed – 1/1/1990-5/13/2014

LANGUAGE:
English

SEARCH STRATEGY:
"celiac disease "[Mesh] OR "celiac disease"[tiab] OR "coeliac disease"[tiab]

AND
endoscopy
AND
duodenal or duodenum
AND
biopsy OR biopsies

==========

DATABASE SEARCHED & TIME PERIOD COVERED:
Web of Science Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH – 1/1/1990-5/13/2014

LANGUAGE:
English

SEARCH STRATEGY:
TOPIC: ("celiac disease" OR "coeliac disease")
AND
TOPIC: (endoscop* AND (duodenal or duodenum) AND (biopsy OR biopsies))

==========

KQ3 (POPULATION):

DATABASE SEARCHED & TIME PERIOD COVERED:
PubMed – 1/1/1990-5/5/2014

LANGUAGE:
English

SEARCH STRATEGY:
"celiac disease "[Mesh] OR "celiac disease"[tiab] OR "coeliac disease"[tiab]
AND
diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR "Transglutaminases"[Mesh] OR "HLA Antigens"[Mesh] OR iga OR ttg OR dgp OR IGG OR mass screening OR diagnosis[mh] OR biopsy[mh] OR biopsies[mh] OR "deamidated gliadin peptide antibodies" OR Human leukocyte antigen* OR "video capsule endoscopy" OR endoscop* OR misdiagnos* OR undiagnos*
AND
"Ethnic Groups"[Mesh] OR "Minority Groups"[Mesh] OR "Socioeconomic Factors"[Mesh] OR "Continental Population Groups"[Mesh] OR "Demography"[Mesh] OR population* OR symptomatic OR nonsymptomatic OR non-symptomatic OR child OR children OR infant OR infants OR pediatric* OR paediatric* OR demograph* OR race OR racial OR ethnic OR ethnicit* OR minority OR minorities OR genetic* OR geograph* OR region OR regions OR regional OR socioeconom* OR socio-econom* OR economic* OR income OR (iga AND deficien*) OR negative OR country[tiab] OR countries[tiab] OR (prevalence OR prevalen*[tiab]
AND

"outcome assessment health care"[MeSH Terms] OR Accura* OR Sensitivity and specificity[mh] OR Sensitivity[tiab] OR Specificity[tiab] OR False positive reactions[mh] OR false positive* OR False negative reactions[mh] OR False negative* OR Predictive value OR predictive value of tests[mh] OR Distinguish* OR Differential* OR Identif* OR Detect* OR valid* OR reliab* OR reproducibility of results OR outcome OR outcomes OR treatment outcome OR treatment outcomes
NOT
case report* OR case reports[pt])

==========

DATABASE SEARCHED & TIME PERIOD COVERED:
Web of Science Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH – 1/1/1990-5/5/2014

LANGUAGE:
English

SEARCH STRATEGY:
ts=("celiac disease" OR "coeliac disease")
AND
ts=(diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR Transglutaminase* OR iga OR ttg OR dgp OR IGG OR mass screening OR "Human leukocyte antigen*" OR endoscop* OR misdiagnos* OR undiagnos*)
AND
ts=(population* OR symptomatic OR nonsymptomatic OR non-symptomatic OR child OR children OR infant OR infants OR pediatric* OR paediatric* OR demograph* OR race OR racial OR ethnic OR ethnicit* OR minority OR minorities OR genetic* OR geograph* OR region OR regions OR regional OR socioeconom* OR socio-econom* OR economic* OR income OR (iga AND deficien*) OR negative OR country OR countries)
AND
ts=(Accura* OR Sensitivity OR Specificity OR false positive* OR False negative* OR "predictive value" OR Distinguish* OR Differential* OR Identif* OR Detect* OR valid* OR reliab* OR reproducib* OR outcome* OR prevalen*)

==========

KQ4 (ADVERSE EVENTS):

DATABASE SEARCHED & TIME PERIOD COVERED:
PubMed – 1/1/2003-5/14/2014

LANGUAGE:
English

SEARCH STRATEGY:
"celiac disease "[Mesh] OR "celiac disease"[tiab] OR "coeliac disease"[tiab]
AND
diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy

OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR "Transglutaminases"[Mesh] OR "HLA Antigens"[Mesh] OR iga OR ttg OR dgp OR IGG OR mass screening OR diagnosis[mh] OR biopsy[mh] OR biopsies[mh] OR "deamidated gliadin peptide antibodies" OR Human leukocyte antigen* OR "video capsule endoscopy" OR endoscop* OR misdiagnos* OR undiagnos*
AND
adverse effect* OR adverse event* OR harm* OR bleeding OR perforat* OR danger* OR safe*[tiab] OR safety[tiab] OR patient safety OR accident*

==========

DATABASE SEARCHED & TIME PERIOD COVERED:
Web of Science Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH – 1/1/2003-5/14/2014

LANGUAGE:
English

SEARCH STRATEGY:
ts=("celiac disease" OR "coeliac disease")
AND
ts=(diagnosis OR diagnoses OR diagnostic OR diagnose OR diagnosing OR endomysial OR transglutaminase* OR serolog* OR antibody OR antibodies OR leucocyte* OR hla OR biopsy OR biopsies OR test OR tests OR testing OR screen OR screening OR screened OR Transglutaminase* OR iga OR ttg OR dgp OR IGG OR mass screening OR "Human leukocyte antigen*" OR endoscop* OR misdiagnos* OR undiagnos*)
AND
ts=(adverse OR harm* OR danger* OR bleed* OR perforat* OR OR safe* OR accident*)