HHS Public Access
Author manuscript
J Clin Gastroenterol. Author manuscript; available in PMC 2022 April 01.

Published in final edited form as:
J Clin Gastroenterol. 2021 April 01; 55(4): 327–334. doi:10.1097/MCG.0000000000001361.

Estimating the Impact of Verification Bias on Celiac Disease Testing

Isabel A. Hujoel, M.D.[1], **Claire L. Jansson-Knodell, M.D.**[2], **Philippe P. Hujoel, Ph.D.**[3], **Margaux L. A. Hujoel, M.A.**[4], **Rok Seon Choung, M.D., Ph.D.**[1], **Joseph A. Murray, M.D.**[1], **Alberto Rubio-Tapia, M.D.**[5]

[1]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, 55905

[2]Division of Gastroenterology and Hepatology, Indiana University, Indianapolis, IN, 46202

[3]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, 98195

[4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02215

[5]Department of Gastroenterology and Hepatology, Cleveland Clinic, Cleveland, OH, 44103

Abstract

Goals: To estimate the impact of verification bias on the diagnostic accuracy of immunoglobulin A tissue transglutaminase (IgA tTG) in detecting celiac disease as reported by an authoritative meta-analysis, the 2016 Comparative Effectiveness Review (CER).

Background: Verification bias is introduced to diagnostic accuracy studies when screening test results impact the decision to verify disease status.

Study: We adjusted the sensitivity and specificity of IgA tTG reported by the 2016 CER with the proportion of IgA tTG positive and negative individuals who are referred for confirmatory small bowel biopsy. We performed a systematic review from January 1, 2007 to July 19, 2017 to determine these referral rates.

Corresponding Author: Isabel A. Hujoel, M.D., Division of Gastroenterology and Hepatology, Mayo Clinic, 200 First Street S.W., Rochester, Minnesota, 55905, (507) 284-2511, hujoel.isabel@mayo.edu.

Authorship Statement

Guarantor of article: Isabel A. Hujoel

Specific author contributions:
IAH: concept, drafted manuscript, data acquisition, interpretation of data
CJK: data acquisition, contributed to manuscript preparation
PPH: statistical analysis, contributed to manuscript preparation
MLAH: statistical analysis, contributed to manuscript preparation
RSC: contributed to manuscript preparation
JAM: input on content and analysis, contributed to manuscript preparation
ART: contributed to manuscript preparation, interpretation of data, input on content
All authors approved the final version of the article.

Financial Support: None

Conflict of Interest: None

Conflict of Interest: None; all authors have nothing to disclose

Funding: None

Results: The systematic review identified 793 articles of which 9 met inclusion criteria (n=36,477). 3.6% (95% confidence interval (CI): 1.1-10.9%) of IgA tTG negative and 79.2.2% (95% CI: 65.0-88.7%) of IgA tTG positive individuals were referred for biopsy. Adjusting for these referral rates the 2016 CER reported sensitivity of IgA tTG dropped from 92.6% (95% CI: 90.2-94.5%) to 57.1% (95% CI: 35.4-76.4%) and the specificity increased from 97.6% (95% CI: 96.3-98.5%) to 99.6% (95% CI: 98.4-99.9%).

Conclusions: The CER may have largely overestimated the sensitivity of IgA tTG due to a failure to account for verification bias. These findings suggest caution in the interpretation of a negative IgA tTG to rule out celiac disease in clinical practice. More broadly, they highlight the impact of verification bias on diagnostic accuracy estimates and suggest that studies at risk for this bias be excluded from systematic reviews.

Keywords

Diagnostic accuracy; verification bias; serology; celiac disease

INTRODUCTION

Guidelines recommend testing for celiac disease (CD) with immunoglobulin A anti-tissue transglutaminase (IgA tTG) and performing confirmatory small bowel biopsy in individuals with a positive serologic test or in those with a negative serologic test but high likelihood for having the disease[1,2]. IgA tTG is recommended as the initial diagnostic test in part because of its perceived high sensitivity. Authoritative reviews can have a significant impact on such guidelines, and it is imperative that they adhere to sound design and analysis. Many sources of bias need to be considered in a meta-analysis of a diagnostic test[3]. One particularly relevant criteria is that disease status be determined in all, or a random selection, of study participants[4]. Verification bias is introduced when disease status is verified in a non-random subset, selected on the basis of screening test results or clinical characteristics of the subjects. An overestimation of the sensitivity and underestimation of specificity can occur when those with a positive screening test are more likely to have their disease status verified than those with a negative result (Supplemental Figure 1). Studies on diseases where the gold standard is invasive or expensive, such as the small bowel biopsy in CD, are at risk for verification bias[5].

The United States Preventive Services Task Force cited a 2016 Comparative Effectiveness Review (CER) to support their statement on the high sensitivity of IgA tTG as a screening test for CD[6-8]. We aimed to assess the impact of verification bias on the CER estimates of diagnostic accuracy, to evaluate whether the use of IgA tTG as a screening test remains reasonable, and to raise the question as to whether systematic reviews on diagnostic tests with an invasive gold standard should explicitly exclude studies at risk for verification bias.

MATERIALS AND METHODS

Overview of Methods

The 2016 CER aimed in part to assess the accuracy of IgA tTG in diagnosing CD. To accomplish this, they performed a systematic review and identified 9 studies which met their

inclusion and exclusion criteria. Their sensitivity and specificity estimates of 92.6% (95% CI: 90.2-94.5%) and 97.6% (95% CI: 96.3-98.5%) respectively were based on these 9 studies. We rated these studies as being at high-, low-, or unclear risk for verification bias. We then adjusted the high-risk studies for verification bias by accounting for the fraction of IgA tTG positive and negative individuals referred to small bowel biopsy for disease confirmation, termed the positive and negative referral rate respectively. We performed a systematic review to determine the referral rates reported in the literature. The adjusted and unadjusted sensitivities and specificities were then pooled (Supplemental Figure 2).

Adjusting Sensitivity and Specificity

The studies in the CER were graded as being high-, low-, or unclear risk for verification bias. A study was graded as being at high-risk of verification bias when a small bowel biopsy was performed in a non-random subset of those who underwent the IgA tTG test. Such non-random selection occurs frequently as guidelines advise clinicians to selectively refer patients to small bowel biopsy based on IgA tTG[1]. This selective referral process can lead to only partial verification of the accuracy of the index test which leads to biased estimates[5]. Studies that included cases of CD diagnosed in those settings were graded as high-risk[9]. Low-risk studies were those where all or a random subset of IgA tTG tested individuals underwent a biopsy. Unclear risk studies were those where the methods did not allow for distinguishing between these two groups, or where it was unclear if IgA tTG results impacted the decision to biopsy. Overall agreement between the two reviewers was quantified using the kappa statistic.

The sensitivities and specificities of high-risk studies were adjusted for verification bias using the Begg and Greenes method[10]. The sensitivities and specificities of low- and unclear- risk studies were not adjusted. Unclear-risk studies were not adjusted in order to provide a conservative estimate of accuracy. A summary estimate of the unadjusted sensitivity and specificity was computed using a bivariate normal model[11]. The variance of adjusted estimates was calculated based on the central limit theorem and delta method (Supplemental Document 1).

The Begg and Greenes method of accounting for verification bias involves adjusting the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts used to calculate sensitivity (TP/(TP+FN)) and specificity (TN/(TN+FP)). Of individuals with a positive IgA tTG there are those with CD (TP) and those without (FP). Similarly, in individuals with a negative IgA tTG, there are those without CD (TN) and those with (FN). To get accurate TP, TN, FP, and FN counts, disease status needs to be verified with small bowel biopsy in all or a random subset of those undergoing IgA tTG. However, if only a non-random subset is verified, the estimates of these counts will be biased (verification bias). Begg and Greenes have proposed a method to adjust TP, FP, TN, FN for the referral rates – the number of those with a positive IgA tTG referred for biopsy (positive referral rate (PRR)) and the number of those with a negative IgA tTG referred (negative referral rate (NRR)). Specifically, the TP and FP are adjusted by dividing these values by the PRR while the TN and FN are adjusted by dividing by the NRR. These adjusted values are then used to

calculate adjusted sensitivity ((TP/PRR)/(TP/PRR+FN/NRR)) and specificity ((TN/NRR)/(TN/NRR+FP/PRR)).

The CER included both pediatric and adult studies in its estimates of sensitivity and specificity. Given the difference in clinical practice in pediatric and adult populations, particularly in relation to decisions to biopsy, we also performed a sub-analysis for these two groups. We applied referral rates drawn from studies on pediatric populations to the high-risk pediatric studies in the CER, and applied referral rates drawn from studies on adult populations to the high-risk adult studies in the CER. Results were then pooled with the unadjusted low- and unclear-risk pediatric and adult studies respectively.

Duplicating the CER Methods

Three of the 9 studies reported 25 estimates of sensitivity and specificity; one study reported twenty[12], one reported two[13], and one reported three[14]. It is unclear if the CER accounted for this correlation, and to evaluate if they had, we duplicated their methods. In our estimates of diagnostic accuracy, we selected one estimate of sensitivity and specificity from each of these three studies with a combined 25 estimates. These estimates were selected to be consistent with the other included studies (Supplemental Document 2).

Search Strategy to Estimate Referral Rates

A comprehensive search of several databases from 2007 to July 19th, 2017 was conducted. The search was limited to articles in English and aimed to identify studies which reported on referral rates after a positive and negative IgA tTG test. The databases included Ovid Medline In-Process & Other Non-Indexed Citations, Ovid MEDLINE, Ovid EMBASE, Ovid Cochrane Central Register of Controlled Trials, Ovid Cochrane Database of Systematic Reviews, and Scopus. The search strategy was based on controlled vocabulary supplemented with keywords (Supplemental Document 3).

Study Selection for Referral Rate Estimation

Two investigators (I.A.H and C.J.K) independently reviewed the identified abstracts. Full articles were reviewed when the abstract suggested that the article would report on the number of individuals tested with IgA tTG and the number who underwent biopsy. The abstract did not have to report referral rate for the full article to be reviewed. Studies were included that detailed total number of individuals tested with IgA tTG, total number of individuals that had positive or negative values of IgA tTG, and positive and negative referral rates. Studies that looked at already diagnosed cases of CD or where all cases were biopsied were excluded.

Data Extraction for Referral Rate Estimation

Data from included studies were extracted using a standardized collection template (Supplemental Table 1). The referral rates for a positive and negative serology were transformed into logits and pooled using the DerSimonian-Laird estimators for random effects models[15]. The Q statistic was calculated to determine heterogeneity[16].

RESULTS

Systematic Review to Estimate Referral Rate

The literature search identified 793 abstracts. 783 were excluded: 415 did not address the research question, 10 were systematic reviews, 59 evaluated already diagnosed cases, 92 reported only on individuals referred for biopsy, 5 did not perform a biopsy, 28 biopsied all participants, 120 only biopsied those with positive serologies, 51 did not provide sufficient information to calculate referral rate, and for 3 the article could not be located. Ten studies which reported referral rates were included[17-25]. Of the 28 studies that biopsied all participants (and thus are not at risk for verification bias), only 1 was included in the CER[26]. This study is not a diagnostic accuracy study, but instead aimed to estimate the prevalence of celiac disease in the cirrhotic population.

Referral Rate Estimates

Referral rates were estimated based on the included ten studies (36,477 total participants). Half were retrospective, with data on referral rates in the community[17,19,23,25,27]. The other half were prospective[20-22,24,28] (Supplemental Table 1). Three studies were performed in the adult population[19,20,23], 4 in children[18,21,24,28], and 3 in individuals of any age[17,22,25]. For 2 of the 10 studies, there was a concern that the studied sample overlapped. We contacted an author for each of these two studies and were unable to exclude a significant overlap in patients[19,23]. These two studies were of similar size and were carried out on adult populations. We therefore estimated adjusted sensitivity and specificity using referral rates computed after randomly excluding one of these two studies[23].

Referral rates for biopsy are shown in Figures 1 and 2. The pooled referral rates were 79.2% (95% CI 65.0-88.7%) for a positive IgA tTG and 3.6% (95% CI 1.1-10.9%) for a negative IgA tTG. There was significant heterogeneity in referral rates, with a Q statistic of 249 ($p<0.0001$) for seropositive and 2,096 ($p<0.0001$) for seronegative individuals.

Overall referral rates for biopsy without exclusion of one of the potentially overlapping studies, as well as separate referral rates for the adult and pediatric populations can be found in Supplemental Document 4.

Calculating an Adjusted Estimate of Sensitivity and Specificity

The CER included 15 unique studies on the diagnostic accuracy of IgA tTG. Their sensitivity and specificity estimates are based on 9 of these 15. The remaining 6 studies were excluded by the CER as they did not provide sufficient information for their results to be pooled[29-34].

We evaluated the risk of verification bias in the 9 studies used for sensitivity and specificity estimates in the CER. Five were rated at high-risk for verification bias, 3 at low-risk, and 1 at unclear risk (Supplemental Document 5). The kappa statistic of 0.82 suggests near perfect agreement between the two reviewers. Only those studies at low risk reported their referral rate (equal to 1). The 6 studies excluded by the CER were not included in our pooled sensitivity or specificity in order to replicate their methods (Table 1).

The sensitivities and specificities of the 5 high-risk studies were adjusted using the calculated pooled referral rates[12-14,35,36]. The 3 low-risk[26,37,38] and 1 unclear risk[39] studies were left unadjusted. The adjusted and unadjusted sensitivities and specificities were then pooled, resulting in a sensitivity of 57.1% (95% CI 35.4-76.4%) and a specificity of 99.6% (95% CI 98.4-99.9%) (Figures 3 and 4).

Using the pooled referral rates estimated after including both potentially overlapping studies led to a sensitivity of 62.0% (95% CI 42.2-78.5%) and specificity of 99.4% (95% CI: 98.1-99.8%).

Of the nine included studies, 3 were performed in individuals of all age groups[12,35,37], 1 in children[13], 3 in adults[14,26,39], and 2 did not specify age[36,38]. Of the 5 high-risk studies, only 4 specified age: 2 were performed in all age groups[12,35], 1 in adults[14], and 1 in children[13]. The 1 high-risk study in adults was adjusted for the adult referral rate and pooled with the 2 other adult CER studies (one being of low risk and one of unclear risk). Using the pooled referral rates after exclusion of the possible duplicate referral study led to an adjusted sensitivity of 47.5% (95% CI 25.3-70.8%) and specificity of 96.6% (95% CI 93.9-98.1%) for adults. Using the pooled referral rates computed including both potentially overlapping studies resulted in an adjusted sensitivity of 49.7% (95% CI 28.0-71.5%) and specificity of 96.6% (95% CI 94.0-98.1%) for adults. The only study exclusively on children was high-risk and was adjusted for the referral rate drawn from pediatric studies, resulting in an adjusted sensitivity of 60.7% (95% CI 18.6-91.3%) and specificity of 99.9% (95% CI 99.2-100.0%) for children.

Pooling the 3 studies included in the CER which were at low risk for verification bias led to a sensitivity of 71.6% (95% CI, 40.6-90.3%) and specificity of 98.3% (95% CI, 85.6-99.8%).

Duplication of the CER

The CER systematic review performed a meta-analysis of the 9 studies (31 estimates, with 3 studies providing 25 estimates) using the reitsma function within the mada package in R[11]. This method accounts for potential correlation between sensitivity/specificity estimates from the same study and estimates the sensitivity and specificity jointly. The method also, *if specified*, can account for correlation between multiple estimates originating from the same study. The CER does not describe if they specified this. To evaluate if they had, we duplicated the CER by using the reitsma function in the mada package. We did not account for the correlation between multiple estimates originating from the same study. Using this technique, we obtained a sensitivity of 92.5% (95% CI, 90.0-94.4%) and specificity of 97.6% (95% CI, 96.3-98.5%). This is within 0.2% of the CER estimate of sensitivity and identical to their specificity estimate. This suggests that the CER likely did not adjust for the correlation of repeated diagnostic test results on the same subjects. The 25 estimates of sensitivity/specificity from the 3 studies had a correlation of 0.19 for the sensitivity and 0.88 for the specificity, suggesting significant correlation.

Using the reitsma function in the mada package, and the one selected estimate from the studies with multiple estimates, the sensitivity and specificity changed to 87.2% (95% CI, 72.5-94.7%) and 97.7% (95% CI, 93.8-99.1%), respectively.

DISCUSSION

Our main finding is that the reported sensitivity and specificity of IgA tTG in the CER are substantially biased due to a lack of adjustment for verification bias. Specifically, adjusting for verification bias decreases the sensitivity of IgA tTG from 92.5 to 57.1%, with a drop in the lower limit of the 95% confidence interval to 35.4%, and an increase in the specificity from 97.9 to 99.6%. The low estimated sensitivity of IgA tTG raises concern on the accuracy of this test, and supports performing a systematic review that accounts for verification bias. Such potentially substantial misrepresentation of the value of a diagnostic test has significant clinical implications for CD and also methodologic implications for how systematic reviews on diagnostic tests should be conducted.

After adjusting for verification bias, the estimated sensitivity of IgA tTG falls to the point where the serologic marker may no longer be clinically useful as a screening test. Failure to account for verification bias leads to an approximately 80% underestimation of the false-negative rate. Many cases of CD will thus remain undiagnosed when initial testing is based on IgA tTG. This may compound the suboptimal detection rate of CD[40,41]. Individuals with symptomatic undiagnosed CD, who are most likely to benefit from detection, may be disproportionately impacted by ignoring verification bias. Their symptoms are commonly atypical or too mild to come to personal or clinical attention, and therefore are unlikely to trigger the high suspicion for CD that guidelines recommend should lead to biopsy in the face of a negative IgA tTG[1,42,43].

It should be emphasized that the Begg and Greenes method has potential for error and provides an estimate of sensitivity and specificity under various assumptions. The method assumes that only the result of the screening test impacts the decision to proceed with biopsy. However, in clinical practice multiple factors impact this decision, such as symptoms, familial history, and existing conditions. While alternative methods have been developed that address this, we did not have the necessary data to use them[44]. Additionally, concerns have been raised about the accuracy of the Begg and Greenes estimation when there are low numbers of false-negatives, as is presumed to be the case with IgA tTG and celiac disease. Small changes in the false-negative rate can lead to dramatic changes in sensitivity estimates[45]. Our results, and the low sensitivity derived by pooling the low-risk CER studies, suggest that verification bias may affect sensitivity estimates significantly, however more accurate estimates need to be determined through a systematic review accounting for this bias.

Verification bias, is not unique to CD. For example, verification bias is present in nearly two-thirds of studies on fecal occult blood testing and hepatitis C antibody testing [46,47]. Given the possibly dramatic impact of not accounting for verification bias, we propose that systematic reviews on diagnostic accuracy studies either exclude or adjust for those studies where verification bias is introduced.

Currently, systematic reviews on diagnostic accuracy do not appear to routinely adjust for verification bias. This may be in part due to guidelines which advise authors to identify and grade the risk of biases, such as verification bias, and to provide this information in the manuscript[48,49]. This assessment of bias is performed by applying the QUADAS-2 criteria, a set of yes and no questions, to each included study. These questions, however, are not sufficiently explicit to allow correct identification of verification bias. As a result, studies at high risk are included in systematic reviews, are incorrectly labeled as being at low risk, and lead to highly biased estimates of diagnostic accuracy which remain unrecognized and unappreciated. Furthermore, given the serious impact of verification bias on study conclusions, it can be argued that merely providing an estimated risk of verification bias (high, low) and not adjusting for this bias, is unacceptable as it will lead to misleading results [3].

Specifically, we propose that the QUADAS-2 criteria are too vague to allow clinical investigators to correctly identify studies at risk for verification bias. The QUADAS-2 screens for verification bias by assessing if “all patients received the reference standard” in the included study[50]. This question does not correctly identify verification bias in studies where the decision to refer to the reference standard is guided by clinical decision making as opposed to study design. Verification bias is created when a non-random subset of patients who undergo the index test are referred to confirmatory testing with the reference standard. Clinicians in practice, and outside of the setting of rigorous clinical testing, often are recommended by guidelines to use the result of the index test to decide who should be referred. Therefore, using a study population where the decision to refer is made in the community likely introduces verification bias. Studies that include cases of CD diagnosed in community practice, can therefore have study subjects who all underwent the reference standard, but still be at high risk for verification bias because referral was non-random. Indiscriminate application of the QUADAS-2 criteria to such studies may mistakenly lead to the conclusion that they are at low risk for verification bias.

The CER illustrates how application of the QUADAS-2 criteria can fail to identify verification bias. For every study, the CER authors correctly indicated that all patients received the reference standard. At face value, this suggests no risk for verification bias. However, 5 included cases identified in the community, and 5 were retrospective with either a clear or unclear impact of the serology results on decision to biopsy. These studies therefore examined populations that were non-randomly referred to the reference standard. Inclusion of these studies led to inflated sensitivity estimates.

In addition to the lack of identification or adjustment for verification bias, the CER included several estimates of sensitivity and specificity originating from the same study. Our duplication of their methods suggest that they did not account for this correlation. Because these estimates are not independent, this led to spuriously narrow confidence intervals for sensitivity. Additionally, the CER included heterogeneous studies with different thresholds for positive serology, and different test manufacturers. To address these limitations, we selected one estimate from each study, and selected those estimates that were most consistent with the other studies in regards to positivity threshold and test manufacturer.

A limitation of our study was the utilization of referral rates drawn from a literature review, as opposed to the unique referral rates from each study. However, only 5 of the CER studies provided referral rates, and unfortunately, the methodology of several of the studies precluded the ability to determine a referral rate.

The heterogeneity of the referral rates drawn from the literature review introduce another limitation to our study. The inclusion of retrospective and prospective studies may have led to different indications to biopsy, and therefore different referral rates. There was also likely variation in clinical practice regarding referral to endoscopy, and in referral in adult and pediatric populations, which were both included. We performed a sub-analysis on the adult and pediatric populations and found substantially different referral rates and estimates of sensitivity in these two populations. Finally, referral rates, with the exception of one study, do not account for patients who may have been recommended biopsy but refused.

In summary, we found that a large number of studies estimating the accuracy of IgA tTG were at high risk for verification bias and that adjusting for this bias led to a substantial decrease in the sensitivity estimate. These findings have considerable clinical implications for CD. Continued use of IgA tTG as an initial diagnostic test may lead to underdiagnosis of CD. If IgA tTG is used, a negative result should be viewed cautiously, with a low threshold for further testing in the appropriate clinical context. Our findings should be validated with an updated systematic review and meta-analysis. While this study focused on CD, verification bias can be found throughout diagnostic accuracy studies. We suggest that the QUADAS-2 criteria be explicit to allow reviewers to correctly identify verification bias. Given the substantial impact of verification bias on the estimates of diagnostic accuracy, we propose that systematic reviews either exclude or adjust for studies at high risk of verification bias.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations:

CD	celiac disease
IgA tTG	immunoglobulin A tissue transglutaminase
CI	confidence interval
CER	Comparative Effectiveness Review

REFERENCES

1. Rubio-Tapia A, Hill ID, Kelly CP, Calderwood AH, Murray JA, American College of G. ACG clinical guidelines: diagnosis and management of celiac disease. American Journal of Gastroenterology. 2013;108(5):656–676; quiz 677.
2. Bai J, Ciacci C. World Gastroenterology Organisation Global Guidelines: Celiac Disease February 2017. Journal of Clinical Gastroenterology. 2017;51(9):755–768. [PubMed: 28877080]

3. Leeflang M, Deeks J, Gatsonis C, Bossuyt P. Systematic Reviews of Diagnostic Test Accuracy. Ann Intern Med. 2010;149(12):889–897.
4. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology. 2003;3:25. [PubMed: 14606960]
5. Cohen J, Korevaar D, Altman D, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6:e012799.
6. Bibbins-Domingo K, Grossman D, Curry S, et al. Screening for Celiac Disease: US Preventive Services Task Force Recommendation Statement. JAMA. 2017;317(12):1252–1257. [PubMed: 28350936]
7. Maglione MA, Okunogbe A, Ewing B, et al. Agency for Healthcare Research and Quality (US). 2016;AHRQ Comparative Effectiveness Reviews, Report No.:15(16)-EHC032–EF.
8. Chou R, Bougatsos C, Blazina I, Mackey K, Grusing S, Selph S. Screening for Celiac Disease: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA. 2017;317(12):1258–1268. [PubMed: 28350935]
9. Biesheuvel C, Vergouwe Y, Oudega R, Hoes A, Grobbee D, Moons K. Advantages of the nested case-control design in diagnostic research. BMC Medical Research Methodology. 2008;8(48).
10. Begg C, Greenes R. Assessment of Diagnostic Tests When Disease Verification is subject to Selection Bias. Biometrics. 1983;39:207–215. [PubMed: 6871349]
11. Reitsma J, Glas A, Rutjes A, Scholten R, Bossuyt P, Zwinderman A. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982–990. [PubMed: 16168343]
12. Van Meensel B, Hiele M, Hoffman I, et al. Diagnostic accuracy of ten second-generation (human) tissue transglutaminase antibody assays in celiac disease. Clin Chem. 2004;50(11):2125–2135. [PubMed: 15388634]
13. Wolf J, Hasenclever D, Petroff D, et al. Antibodies in the Diagnosis of Coeliac Disease: A Biopsy-Controlled, International, Multicentre Study of 376 Children with Coeliac Disease and 695 Controls. PLOS ONE. 2014;9(5):e97853. [PubMed: 24830313]
14. Swallow K, Wild G, Sargur R, et al. Quality not quantity for transglutaminase antibody 2: The performance of an endomysial and tissue transglutaminase test in screening coeliac disease remains stable over time. Clinical and Experimental Immunology. 2013;171(1):100–106. [PubMed: 23199329]
15. DerSimonian R, Laird N. Meta-Analysis in Clinical Trials. Controlled Clinical Trials. 1986;7:177–188. [PubMed: 3802833]
16. Huedo-Medina T, Sanchez-Meca J, Marin-Martinez F. Assessing heterogeneity in Meta-Analysis: Q Statistic or I^2 Index? Psychological Methods. 2006;11(2):193–206. [PubMed: 16784338]
17. Bayram Y, Parlak M, Aypak C, Bayram I, Yilmaz D, Cikman A. Diagnostic accuracy of IgA anti-tissue transglutaminase in celiac disease in Van-Turkey. Eastern Journal of Medicine. 2015;20(1):20–23.
18. Gidrewicz D, Potter K, Trevenen CL, Lyon M, Butzner JD. Evaluation of the ESPGHAN Celiac Guidelines in a North American Pediatric Population. American Journal of Gastroenterology. 2015;110(5):760–767.
19. Kabbani TA, Vanga RR, Leffler DA, et al. Celiac disease or non-celiac gluten sensitivity? an approach to clinical differential diagnosis. American Journal of Gastroenterology. 2014;109(5):741–746.
20. Kratzer W, Kibele M, Akinli A, et al. Prevalence of celiac disease in Germany: a prospective follow-up study. World Journal of Gastroenterology. 2013;19(17):2612–2620. [PubMed: 23674868]
21. Oana B, Otilia M. The usefulness of IgA/IgG DGP/tTG screen assay for celiac disease detection among symptomatic and at risk young children. International Journal of Celiac Disease. 2013;1(1):23–26.
22. Rubio-Tapia A, Van Dyke CT, Lahr BD, et al. Predictors of family risk for celiac disease: a population-based study. Clinical Gastroenterology & Hepatology. 2008;6(9):983–987. [PubMed: 18585974]

23. Pallav K, Kabbani T, Tariq S, Vanga R, Kelly CP, Leffler DA. Clinical Utility of Celiac Disease-Associated HLA Testing. Digestive Diseases and Sciences. 2014;59(9):2199–2206. [PubMed: 24705698]
24. Smarrazzo A, Misak Z, Costa S, et al. Diagnosis of celiac disease and applicability of ESPGHAN guidelines in Mediterranean countries: a real life prospective study. BMC Gastroenterology. 2017;17(1):17. [PubMed: 28109250]
25. Toftedal P, Nielsen C, Madsen JT, Titlestad K, Husby S, Lillevang ST. Positive predictive value of serological diagnostic measures in celiac disease. Clinical Chemistry & Laboratory Medicine. 2010;48(5):685–691. [PubMed: 20201743]
26. Wakim-Fleming J, Pagadala M, McCullough A, et al. Prevalence of celiac disease in cirrhosis and outcome of cirrhosis on a gluten free diet: A prospective study. Journal of Hepatology. 2014;61:558–563. [PubMed: 24842303]
27. Gidrewicz D, Potter K, Trevenen CL, Lyon M, Butzner JD. Evaluation of the ESPGHAN Celiac Guidelines in a North American Pediatric Population. Am J Gastroenterol. 2015;110(5):760–767. [PubMed: 25823767]
28. Al-Hussaini A, Sulaiman N, Al-Zahrani M, Alenizi A, El Haj I. High prevalence of celiac disease among Saudi children with type 1 diabetes: a prospective cross-sectional study. BMC Gastroenterology. 2012;12:180. [PubMed: 23259699]
29. Dahle C, Hagman A, Ignatova S, Strom M. Antibodies against deamidated gliadin peptides identify adult coeliac disease patients negative for antibodies against endomysium and tissue transglutaminase. Alimentary Pharmacology & Therapeutics. 2010;32(2):254–260. [PubMed: 20456302]
30. Vermeersch P, Geboes K, Marien G, Hoffman I, Hiele M, Bossuyt X. Serological diagnosis of celiac disease: comparative analysis of different strategies. Clinica Chimica Acta. 2012;413(21-22):1761–1767.
31. Vermeersch P, Coenen D, Geboes K, Marien G, Hiele M, Bossuyt X. Use of likelihood ratios improves clinical interpretation of IgA anti-tTG antibody testing for celiac disease. Clinica Chimica Acta. 2010;411:13–17.
32. Zanini B, Magni A, Caselani F, et al. High tissue-transglutaminase antibody level predicts small intestinal villous atrophy in adult patients at high risk of celiac disease. Digestive & Liver Disease. 2012;44(4):280–285. [PubMed: 22119616]
33. Basso D, Guariso G, Bozzato D, et al. New screening tests enrich anti-transglutaminase results and support a highly sensitive two-test based strategy for celiac disease diagnosis. Clinica Chimica Acta. 2011;412:1662–1667.
34. Barada K, Habib R, Malli A, et al. Prediction of celiac disease at endoscopy. Endoscopy. 2014;46:110–119. [PubMed: 24477366]
35. Dahlbom I, Karoponay-Szabo I, Kovacs J, Szalai Z, Maki M, Hansson T. Prediction of Clinical and Mucosal Severity of Coeliac Disease and Dermatitis Herpetiformis by Quantification of IgA/IgG Serum Antibodies to Tissue Transglutaminase. Journal of Pediatric Gastroenterology and Nutrition. 2010;50:140–146. [PubMed: 19841593]
36. Srinivas M, Basumani P, Podmore G, Shrimpton A, Bardhan K. Utility of Testing Patients, on Presentation, for Serologic Features of Celiac Disease. Clinical Gastroenterology and Hepatology. 2014;12:946–952. [PubMed: 24262940]
37. Mansour A, Najeeb A. Coeliac disease in Iraqi type 1 diabetic patients. Arab Journal of Gastroenterology. 2011;12:103–105. [PubMed: 21684484]
38. Harrison E, Li K-K, Petchey M, Nwokolo C, Loft D, Arasaradnam R. Selective measurement of anti-tTG antibodies in coeliac disease and IgA deficiency: an alternative pathway. Postgraduate Medical Journal. 2013;89:4–7. [PubMed: 22872871]
39. Emami M, Karimi S, Kouhestani S. Is Routine Duodenal Biopsy Necessary for the Detection of Celiac Disease in Patients Presenting with Iron Deficiency Anemia? International Journal of Preventive Medicine. 2012;3(4):273–277. [PubMed: 22624084]
40. Rubio-Tapia A, Ludvigsson JF, Brantner TL, Murray JA, Everhart JE. The Prevalence of Celiac Disease in the United States. The American Journal of Gastroenterology 2012;107:1538–1544. [PubMed: 22850429]

41. Rubio-Tapia A, Kyle RA, Kaplan EL, et al. Increased Prevalence and Mortality in Undiagnosed Celiac Disease. Gastroenterology. 2009;137(1):88–93. [PubMed: 19362553]
42. Kivelä L, Kaukinen K, Huhtala H, Lähdeaho M-L, Mäki M, Kurppa K. At-Risk Screened Children with Celiac Disease are Comparable in Disease Severity and Dietary Adherence to Those Found because of Clinical Suspicion: A Large Cohort Study. The Journal of Pediatrics. 2017;183:115–121. [PubMed: 28153477]
43. Hujoel I, Van Dyke C, Brantner T, et al. Natural History and Clinical Detection of Undiagnosed Coeliac Disease in a North American Community. Aliment Pharmacol Ther. 2018;47(10):1358–1366. [PubMed: 29577349]
44. Hunink M, Richardson D, Doubilet P, Begg C. Testing for fetal pulmonary maturity: ROC analysis involving covarites, verification bias, and combination testing. Med Decis Making. 1990;10(3):201–211. [PubMed: 2370827]
45. Cronin A, Vickers A. Statistical methods to correct for verification bias in diagnostic studies are inadequate when there are few false negatives: a simulation study. BMC Medical Research Methodology. 2008;75:10.1186/1471-2288-1188-1175. [PubMed: 19014457]
46. Cadieux G, Campbell J, Dendukuri N. Systematic review of the accuracy of antibody tests used to screen asymptomatic adults for hepatitis C infection. CMAJ Open. 2016;4(4):E737–E745.
47. Rosman A, Korsten M. Effect of Verification Bias on the Sensitivity of Fecal Occult Blood Testing: a Meta-analysis. Journal of General Internal Medicine. 2010;25(11):1211–1221. [PubMed: 20499198]
48. Deville W, Buntinx F, Bouter L, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Medical Research Methodology. 2002(2):9. [PubMed: 12097142]
49. Bossuyt P, Leeflang M. Chapter 6: Developing Criteria for Including Studies. In: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4. The Cochrane Collaboration; 2008.
50. Whiting P, Rutjes A, Westwood M, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Annals of Internal Medicine. 2011;155(8):529–536. [PubMed: 22007046]
51. Vermeersch P, Geboes K, Marien G, Hoffman I, Hiele M, Bossuyt X. Diagnostic performance of IgG anti-deamidated gliadin peptide antibody assays is comparable to IgA anti-tTG in celiac disease. Clinica Chimica Acta. 2010;411(13-14):931–935.

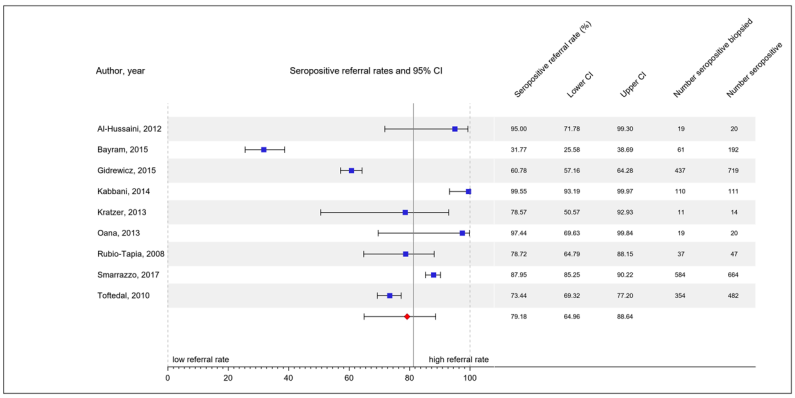

Figure 1:

Referral rates to upper endoscopy and duodenal biopsy after an abnormal IgA tissue transglutaminase antibody test from the 9 studies identified in the systematic review, as well as a pooled referral rate. CI: confidence interval. LCI: lower bound of 95% confidence interval. UCI: upper bound of 95% confidence interval.

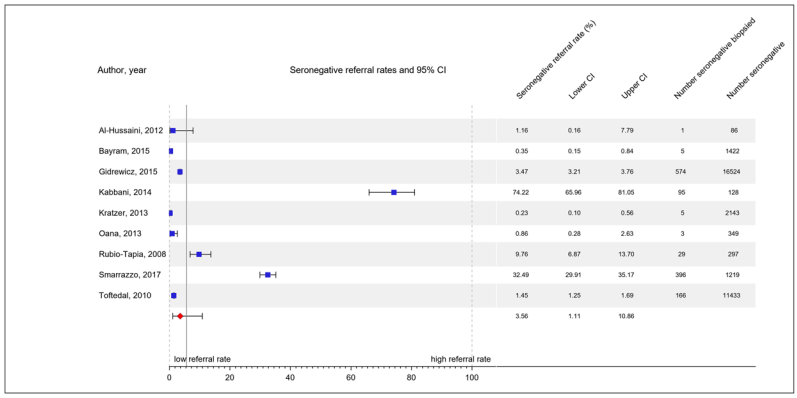

Figure 2:

Referral rates to upper endoscopy and duodenal biopsy after a normal IgA tissue transglutaminase antibody test from the 9 studies identified in the systematic review, as well as a pooled referral rate. CI: confidence interval. LCI: lower bound of 95% confidence interval. UCI: upper bound of 95% confidence interval.

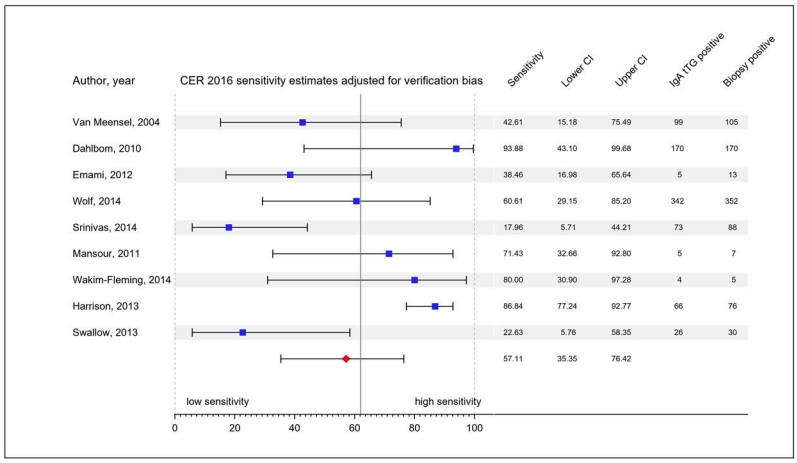

Figure 3:

Sensitivity estimates from the 9 studies included in the 2016 CER's pooled estimate of sensitivity of IgA tTG in celiac disease diagnosis, with the sensitivities from the 5 studies at high-risk for verification bias adjusted for this bias. Of the three studies with multiple estimates, only one has been selected in order to remove non-independent estimates. CI: confidence interval. Lower CI: lower bound of 95% confidence interval. Upper CI: upper bound of 95% confidence interval.

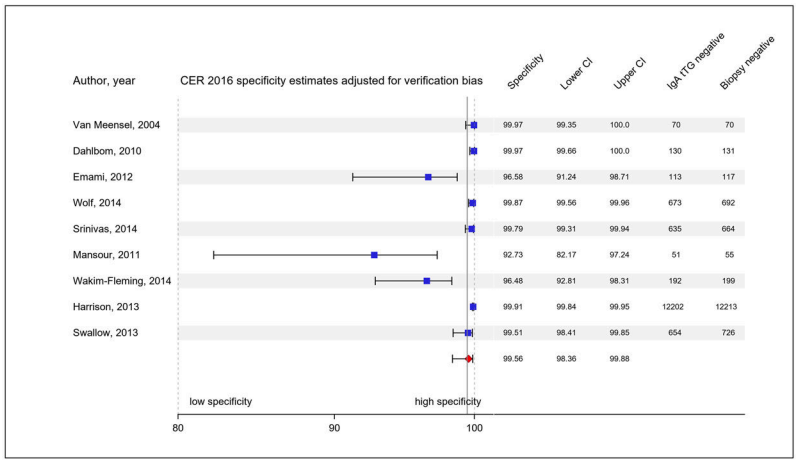

Figure 4:

Specificity estimates from the 9 studies included in the 2016 CER's pooled estimate of specificity of IgA tTG in celiac disease diagnosis, with the sensitivities from the 5 studies at high-risk for verification bias adjusted for this bias. Of the three studies with multiple estimates, only one has been selected in order to remove non-independent estimates. CI: confidence interval. Lower CI: lower bound of 95% confidence interval. Upper CI: upper bound of 95% confidence interval.

Table 1:

Characteristics of 15 unique studies reported in 2016 CER; 9 were included by the 2016 CER in pooled calculations of sensitivity and 6 were excluded from calculations due to insufficient data.

Reference	Threshold	Risk for Verification Bias	Sensitivity X%(number of estimates with that sensitivity)	Specificity X% (number of estimates with that specificity)	Included/ excluded in pooled calculations[a]
Van Meensel, 2004[12][b]	2.64, 3.13, 3.69, 4, 4.43, 5, 7, 7.16, 7.98, 9.73, 10, 15, 19.05, 20, 20.47, 40, 50, 56.9 kilounits/L	High	91% (2), 93% (5), 96% (6), 97% (4), 99%	93%, 96% (2), 99% (6), 100% (9)	Included
Dahlbom, 2010[35]	3 U m/L	High	100%	99%	Included
Wolf, 2014[13][b]	10 ULN, 20 U/mL	High	88%, 97%	97% (2)	Included
Srinivas, 2014[36]	10 IU/mL	High	83%	96%	Included
Swallow, 2013[14]	Not reported	High	71%, 87% (2)	89%, 90% (2)	Included
Vermeersch, 2010[51][b]	7 U/mL	High	86%, 95%	93%, 95%	Excluded
Vermeersch, 2012[30][b]	7, 20 U/mL	High	81%, 84%	96%, 99%	Excluded
Mansour, 2011[37]	15 U/mL	Low	71%	93%	Included
Wakim-Fleming, 2014[26]	20 IU	Low	80%	96%	Included
Harrison, 2013[38]	Not reported	Low	87%	100%	Included
Barada, 2014[34]	Not reported	Low	72%	98%	Excluded
Dahle, 2010[29]	5 U/mL	Low	76%	95%	Excluded
Emami, 2012[39]	10 AU/mL	Unclear	38%	97%	Included
Basso, 2011[33][b]	17.5, 20, 24, 75.6, 100, 909.3 U/mL	Unclear	63%, 76%, 91%, 94%, 95%, 96%	81%, 97% (3), 100% (2)	Excluded
Zanini, 2012[32][b]	7, 8, 16, 21, 24, 35, 40, 48, 80 U/mL	Unclear	10%, 38%, 43%, 59% (2), 70%, 88%, 89%, 95%	43%, 59%, 76%, 88%, 92%, 97%, 99%, 100% (2)	Excluded

[a]Studies that did not provide true positive, false negative, false positive, and true negative information were excluded in the pooled sensitivity and specificity both in the original systematic review as well as in this study.

[b]These studies provided multiple estimates of sensitivity and specificity