Defining the optimum strategy for identifying adults and children with coeliac disease: systematic review and economic modelling

Martha MC Elwenspoek,[1,2*] Howard Thom,[2] Athena L Sheppard,[1,2,3] Edna Keeney,[2] Rachel O'Donnell,[1,2] Joni Jackson,[1,2] Cristina Roadevin,[2] Sarah Dawson,[2] Deborah Lane,[4] Jo Stubbs,[4] Hazel Everitt,[5] Jessica C Watson,[2] Alastair D Hay,[2] Peter Gillett,[6] Gerry Robins,[7] Hayley E Jones,[2] Sue Mallett[8] and Penny F Whiting[2]

[1]National Institute for Health and Care Research Applied Research Collaboration West, University Hospitals Bristol NHS Foundation Trust, Bristol, UK
[2]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[3]Department of Health Sciences, University of Leicester, Leicester, UK
[4]Patient representative, UK
[5]Primary Care Research Centre, Population Sciences and Medical Education, University of Southampton, Southampton, UK
[6]Paediatric Gastroenterology, Hepatology and Nutrition Department, Royal Hospital for Sick Children, Edinburgh, UK
[7]Department of Gastroenterology, York Teaching Hospital NHS Foundation Trust, York, UK
[8]Centre for Medical Imaging, University College London, London, UK

*Corresponding author Martha.Elwenspoek@bristol.ac.uk

Declared competing interests of authors: Howard Thom reports grants from the National Institute for Health and Care Research (NIHR) [Health Technology Assessment (HTA) 18/134 and NIHR Bristol Biomedical Research Centre], and the Medical Research Council (MRC) (MR/S036709/1); Howard Thom is also a part owner of Clifton Insight and has received consulting fees, paid to Clifton Insight, from Novartis International AG (Basel, Switzerland), F. Hoffmann-La Roche AG (Basel, Switzerland), Pfizer Inc. (New York, NY, USA), Bristol Myers Squibb™ (BMS) (New York, NY, USA), Eisai Co., Ltd (Tokyo, Japan), Argenx (Ghent, Belgium), H. Lundbeck A/S (Copenhagen, Denmark) and Janssen Pharmaceuticals (Beerse, Belgium). He has also received other fees from Clifton Insight. Athena L Sheppard reports grants from the NIHR HTA programme (18/134), payments made to their institution, an employment contract with F. Hoffmann-La Roche (2020), a training grant from the National Centre for Research Methods and a travel grant from the Royal Statistical Society. Edna Keeney reports personal fees from Novartis International AG, F. Hoffmann-La Roche, Pfizer Inc. and BMS. Alastair D Hay reports grants from NIHR (Senior Investigator Award NIHR200151), and membership of the Efficacy and Mechanism Evaluation (EME) Funding Committee (2019 to present) and the National Institute for Health and Care Excellence managing common infections committee (2019 to present). Gerry Robins reports membership of the Trustee Board for Coeliac UK (High Wycombe, UK). Hayley E Jones reports grants from MRC-NIHR (New Investigator Research Grant MR/T044594/1) and consulting fees from Aquarius Population Health (London, UK).

Published October 2022
DOI: 10.3310/ZUCE8371

Scientific summary

Optimium strategy to identify coeliac disease

Health Technology Assessment 2022; Vol. 26: No. 44
DOI: 10.3310/ZUCE8371

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Coeliac disease (CD) is an autoimmune disorder, triggered by the protein gluten, which affects an estimated 1% of the UK population. Some people with CD may have minimal symptoms; and others present with non-specific symptoms, making diagnosis difficult: only one in three is thought to be diagnosed. Treatment for CD is lifetime adherence to a gluten-free diet. Untreated CD may lead to persistent symptoms, anaemia, osteoporosis and, occasionally, lymphoma. Guidelines recommend that adults and children 'at high risk' of CD should be offered testing. However, it is not clear which groups are at sufficiently high risk to justify testing, which symptoms should prompt testing, which tests should be offered or if confirmatory biopsy is necessary.

Objectives

The overall aim of this project was to define at-risk groups and determine the cost-effectiveness of active case-finding in primary care.

We defined the following objectives to address this overall aim:

- systematic review of the accuracy of potential diagnostic indicators for CD
- routine data analysis to develop a prediction model to identify people who should be tested for CD
- systematic review of the accuracy of diagnostic tests for CD
- systematic review of the accuracy of genetic tests for CD
- online survey to identify diagnostic thresholds for testing, starting treatment and referral for biopsy
- economic modelling to identify the cost-effectiveness of different active case-finding strategies, informed by the findings of the previous objectives.

Methods

Accuracy of diagnostic indicators

For the first review, six databases [MEDLINE® (National Library of Medicine, Bethesda, MD, USA), Embase® (Elsevier, Amsterdam, the Netherlands), Cochrane Library, Web of Science™ (Clarivate™, Philadelphia, PA, USA), the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) and the National Institutes of Health Clinical Trials database] were searched from January 1990 to April 2021. Studies investigating diagnostic indicators, such as symptoms or risk conditions, among people with and people without CD were eligible for inclusion. International guidance for systematic review methods was followed and the reviews were registered at PROSPERO. Risk-of-bias assessments were performed using the quality assessment of diagnostic accuracy studies 2 (QUADAS-2) tool. Bivariate random-effects meta-analyses were used to pool sensitivity and specificity across studies.

Prediction model development

For the prediction models, we used three data sets: two primary care databases (Clinical Practice Research Datalink Gold and Aurum) containing routinely collected primary care data and a subcohort of the Avon Longitudinal Study of Parents and Children. We fitted logistic regression models with CD as the outcome and multiple diagnostic indicators as predictors. From the results, we produced estimates of discrimination and calibration of the models, the accuracy of predictions at different thresholds and the percentage of people with CD who were missed at these thresholds.

Accuracy of serological tests

For the second systematic review on the accuracy of tests for CD, seven electronic databases [MEDLINE, Embase, Cochrane Library, Web of Science, Kleijnen Systematic Reviews (KSR) Evidence, the WHO ICTRP and the National Institutes of Health Clinical Trials database] were searched from January 1990 to August 2020. We included diagnostic cohort studies that evaluated serological tests for CD [i.e. immunoglobulin A (IgA) tissue transglutaminase (tTG), immunoglobulin G (IgG) tTG, IgA endomysial antibody (EMA), IgG EMA, IgA deamidated gliadin peptide (DGP), IgG DGP and IgA actin antibody] among people presenting with symptoms suggestive of CD.

Accuracy of genetic tests

The review of the accuracy of genetic tests for CD was based on the same search used for the first review of diagnostic indicators and included studies that provided accuracy on the combination of human leucocyte antigen (*HLA*)-*DQ2*/*DQ8* testing. All reviews followed the same internationally recognised methods for systematic reviews.

Online survey

We developed an online survey in collaboration with patient representatives to identify how confident people want to be in their diagnosis before starting a gluten-free diet or accepting a biopsy. The survey was open for 2.5 months (January–March 2021) and was disseminated using social media.

Economic modelling

The cost-effectiveness of CD testing of patients with pre-test probabilities of CD above certain thresholds was evaluated with long-term economic models. We used a decision tree and discrete-time cohort Markov model to compare the cost-effectiveness of case-finding strategies at different levels of pre-test probability separately for men, women and children.

Results

Accuracy of diagnostic indicators

The review of diagnostic indicators included 183 studies reporting on 25 indicators, which comprised seven symptoms, 17 risk conditions and family history. There was large variation in diagnostic accuracy estimates between studies, and most studies were at high risk of bias. None of the identified diagnostic indicators alone had good sensitivity for detecting CD; however, some showed promise in helping to identify patients who should be offered serological testing. The estimated positive predictive values for migraine, family history of CD, anaemia, type 1 diabetes, osteoporosis and chronic liver disease were all > 2%, with 95% confidence intervals (CIs) lying entirely above the population prevalence of 1%. Individual gastrointestinal symptoms showed poor diagnostic ability. People with a first-degree relative with CD were three times more likely to have CD than the general population.

Prediction model development

We developed prediction models for children, women and men that comprised 24, 24 and 21 predictors, respectively. For children, having type 1 diabetes, Turner syndrome, IgA deficiency or a first-degree relative with CD were estimated to be the strongest predictors (i.e. had the highest estimated coefficients). For women and men, the strongest predictors were having a first-degree relative with CD, or having anaemia. In the development data set, the model showed good discrimination between patients with and patients without CD, as demonstrated by high *c*-statistics of 0.84 (95% CI 0.83 to 0.84) for children, 0.77 (95% CI 0.77 to 0.78) for women and 0.81 (95% CI 0.81 to 0.82) for men. The model discriminated less well between patients with and patients without CD in the external validation data set, for which the *c*-statistics reduced to 0.60 for children, 0.55 for women and 0.62 for men. However, the predictor first-degree relative was not recorded in the validation data set, which was one of the most important predictors, leading to an underestimation of model performance in this data set. The models were poorly calibrated and tended to overestimate the

risk of having CD in all three groups in the development data set and validation data set. The models suggest that individuals with any of the selected predictors have an increased risk of CD of > 50%, and thus warrant testing for CD.

Accuracy of serological tests

The review of test accuracy included 113 studies ($n = 28,338$), all in secondary care populations. A subset of studies was included in meta-analyses because of variations in diagnostic thresholds. The majority of included studies were at high risk of bias. The summary sensitivity and specificity of the IgA tTG test were 91% (95% CI 87% to 93%) and 87% (95% CI 84% to 90%), respectively, for adults (five studies) and 98% (95% CI 91% to 99%) and 70% (95% CI 39% to 90%), respectively, for children (six studies). The summary sensitivity and specificity of the IgA EMA test were 88% (95% CI 75% to 95%) and 99.6% (95% CI 92% to 100%), respectively, for adults (five studies) and 95% (95% CI 89% to 97%) and 94% (95% CI 85% to 98%), respectively, for children (five studies). To select estimates to inform the economic model, we restricted our analyses to studies that had evaluated the two main serological tests of interest (IgA tTG and IgA EMA, alone and in combination) at the same threshold. This was to ensure that estimates used in the economic model were directly comparable. None of the studies that evaluated both tests alone and in combination reported accuracy estimates for the same thresholds. We therefore selected the studies that were judged to have the lowest risk of bias and that had the largest sample sizes. For both adults and children, the IgA tTG test had the highest sensitivity, although estimates for children were very similar, and the IgA EMA test had the highest specificity. There was little improvement in either sensitivity or specificity when the tests were used in combination.

Accuracy of genetic tests

Four studies ($n = 12,087$) evaluated the accuracy of *HLA-DQ2* and/or *-DQ8* genetic variants for diagnosing CD. Three studies were deemed to be at low risk of bias, and one was deemed to be at high risk of bias, as serology alone was used to confirm CD status. The summary sensitivity was 99% (95% CI 83% to 100%) and specificity was 56% (95% CI 50% to 61%), suggesting that it would be a useful test to rule out CD.

Online survey

The survey was completed by 472 people. Of these, 244 (52%) had CD, with the disease confirmed by a blood test and/or biopsy. Among those who completed the demographic questions, the vast majority were white ($n = 264$, 95%) and female ($n = 239$, 86%); most respondents went to university or college ($n = 159$, 58%) and lived in the south-west of England ($n = 98$, 36%). Survey respondents wanted to be 66% [median interquartile range (IQR) 33–90%] certain of the diagnosis before starting a gluten-free diet when they were asked to imagine that they had CD symptoms. Without symptoms, respondents wanted to be more certain, around 90% (median IQR 66–99%), before committing to a gluten-free diet. However, a higher proportion of respondents opted to wait for a confirmation biopsy, if given the option, instead of starting a gluten-free diet immediately, even if a hypothetical blood test gave 75–90% certainty.

Economic modelling

The cost-effectiveness analysis found that, for serological testing alone, testing adult men and women who have a 1% pre-test probability (i.e. testing all adults with a 1% pre-test probability of CD, which is equivalent to population screening) had the highest net benefit, at £20,000 per quality-adjusted life-year (QALY). This resulted in incremental net benefits, relative to no screening, of £24,331 [95% credible interval (CrI) £5080 to £56,493] for men and £24,382 (95% CrI £4829 to £59,154) for women. The serological tests (i.e. IgA EMA and IgA tTG) had similar cost-effectiveness and there was limited benefit to including both IgA EMA and IgA tTG tests. Strategies using both HLA and serological testing with pre-test probabilities of 1–20% had very similar net benefits to each other and to those of IgA tTG testing with 1% pre-test probability, and 95% CrIs were completely overlapping. The probability that any one test had the highest net benefit was < 60% for adult men and 50% for adult women, suggesting uncertainty.

Among children, testing all those with a pre-test probability of ≥ 10% with HLA plus IgA tTG had the greatest net benefit at £20,000 per QALY, with an incremental net benefit of £13,090 (95% CrI £3929 to £36,260), relative to no screening; it also had the highest probability (≈ 80%) of being cost-effective at > £10,000 per QALY. Again, there was limited difference in cost-effectiveness between pre-test probabilities, so long as either IgA EMA plus HLA or HLA plus IgA tTG was used as the testing combination.

There was substantial uncertainty in these results, and a value-of-information analysis indicated that they were sensitive to the probability of diagnosis of CD during routine care and the accuracy of HLA and serological tests. The total population expected value of perfect information was £25.7M for men, £79.0M for women and £18.4M for children, indicating potential value of further research, particularly for women.

Conclusions

Implications for practice

Based on the cost-effectiveness analysis, the most cost-effective strategy for adults, using serological testing alone, appears to be population-based screening (1% pre-test probability) using either the IgA tTG or IgA EMA test alone or both tests combined. However, there is substantial uncertainty in these results, and further research is needed prior to any implementation of screening. Given the wider availability of IgA tTG in UK laboratories, and the more objective nature of the test, IgA tTG is probably the preferred serological test. Decisions to implement population-based screening should not be made based on this economic analysis alone: the proposed screening programme must meet UK National Screening Committee criteria. Although a CD screening programme meets some of these criteria, it does not yet meet all criteria. Additional required criteria are as follows: a consensus on an appropriate threshold for the screening test (i.e. IgA tTG), agreement on further diagnostic workup among those testing positive for IgA tTG and randomised trials showing the effectiveness of the screening programme.

Given that population screening is not considered appropriate, we recommend a strategy for adults that combines HLA testing with IgA tTG among those with at least a 1.5% pre-test probability of having CD. These strategies had nearly identical cost-effectiveness to that of the IgA tTG test with 1% pre-test probability, based on our cost-effectiveness analysis. They also had similar cost-effectiveness to more targeted strategies with pre-test probabilities of 5–20%, and people with lower pre-test probabilities are easier to identify, based on our review of diagnostic indicators and prediction models. For children, the most cost-effective testing strategy is to test those with a 10% pre-test probability of CD (more cost-effective than population screening). Therefore, indicators that should prompt testing are those that increase the risk of CD to at least 1.5% among adults (equivalent to at least one of the identified predictors) and to 10% among children, that is children with certain high-risk predictors (e.g. anaemia) or a combination of lower-risk predictors (e.g. failure to thrive and gastrointestinal symptoms). These are diagnostic indicators identified by our review of diagnostic indicators and through the prediction model. The most predictive indicator in all populations was having a first-degree relative with CD. Other indicators identified by our review, but not currently recommended in existing guidelines, that should prompt testing include migraine and chronic liver disease.

The cost-effectiveness analysis found that HLA testing prior to IgA tTG testing was the most cost-effective ordering of these tests. However, in practice such a strategy may have unintended costs and consequences not captured by the economic model. A strategy whereby serological testing is performed first may therefore be preferable, although this would be likely to lead to a greater number of false-negative and false-positive results overall.

All strategies assumed that biopsy would be recommended if the post-test probability following positive test results remained < 90%. Whether or not this is the case will depend on the pre-test probability of disease, and so it may be difficult to implement such a strategy in practice. The variation among individuals in their preferred diagnostic certainty and attitudes towards having a biopsy or following a gluten-free diet suggests that shared decision-making in which patient preferences are taken into account is important in determining the 'optimum' diagnostic pathway.

Suggested research priorities

Given that one of the most cost-effective strategies based on our cost-effectiveness analysis was population-based screening, future work should consider whether or not population-based screening for CD could meet the UK National Screening Committee criteria.

A value-of-information analysis suggested that future research should focus on the probability of CD diagnosis during routine care and the accuracy of serological and HLA testing.

There is a need for large prospective cohort studies in which all participants receive accurate tests for CD, to provide a more accurate estimate of the diagnostic ability of indicators and to develop a more robust clinical prediction model.

Study registration

This study is registered as PROSPERO CRD42019115506 and CRD42020170766.

Funding

This project was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 26, No. 44. See the NIHR Journals Library website for further project information.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.014

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 4.014 and is ranked 27th (out of 108 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2021 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), Embase (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number NIHR129020. The contractual start date was in January 2020. The draft report began editorial review in November 2021 and was accepted for publication in April 2022. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).