ORIGINAL ARTICLE

# Disease activity indices in coeliac disease: systematic review and recommendations for clinical trials

Pieter Hindryckx,[1,2] Barrett G Levesque,[1,3] Tom Holvoet,[2] Serina Durand,[3]
Ceen-Ming Tang,[4,5] Claire Parker,[1] Reena Khanna,[1,6] Lisa M Shackelton,[1]
Geert D'Haens,[1,7] William J Sandborn,[1,3] Brian G Feagan,[1,6,8] Benjamin Lebwohl,[9]
Daniel A Leffler,[10] Vipul Jairath[1,6,8]

## ABSTRACT

**Objective** Although several pharmacological agents have emerged as potential adjunctive therapies to a gluten-free diet for coeliac disease, there is currently no widely accepted measure of disease activity used in clinical trials. We conducted a systematic review of coeliac disease activity indices to evaluate their operating properties and potential as outcome measures in registration trials.

**Design** MEDLINE, EMBASE and the Cochrane central library were searched from 1966 to 2015 for eligible studies in adult and/or paediatric patients with coeliac disease that included coeliac disease activity markers in their outcome measures. The operating characteristics of histological indices, patient-reported outcomes (PROs) and endoscopic indices were evaluated for content and construct validity, reliability, responsiveness and feasibility using guidelines proposed by the US Food and Drug Administration (FDA).

**Results** Of 19 123 citations, 286 studies were eligible, including 24 randomised-controlled trials. Three of five PROs identified met most key evaluative criteria but only the Celiac Disease Symptom Diary (CDSD) and the Celiac Disease Patient-Reported Outcome (CeD PRO) have been approved by the FDA. All histological and endoscopic scores identified lacked content validity. Quantitative morphometric histological analysis had better reliability and responsiveness compared with qualitative scales. Endoscopic indices were infrequently used, and only one index demonstrated responsiveness to effective therapy.

**Conclusions** Current best evidence suggests that the CDSD and the CeD PRO are appropriate for use in the definition of primary end points in coeliac disease registration trials. Morphometric histology should be included as a key secondary or co-primary end point. Further work is needed to optimise end point configuration to inform efficient drug development.

## Significance of this study

### What is already known on this subject?
► The treatment of coeliac disease is entering a new era with the development of several new pharmacological agents, including immunomodulators, biologics, vaccines and specific proteolytic enzymes.
► Coeliac disease activity and response to therapy can be measured through any combination of symptoms and/or histology; endoscopy has also been proposed.
► Several scoring indices have been developed to measure these outcomes, but their operating properties have not been systematically evaluated.

### What are the new findings?
► All scoring indices for measurement of coeliac disease activity were identified and their operating properties (validity, reliability, responsiveness, feasibility) assessed.
► No histological, serological or endoscopic indices met all of the assessed evaluative criteria, and overall performance was poorest for the categorical histological indices.
► Two patient-reported outcome measures, the Celiac Disease Symptom Diary and the Celiac Disease Patient-Reported Outcome, and morphometric histology emerged as the most appropriate outcome measures for coeliac disease trials.

### How might it impact on clinical practice in the foreseeable future?
► To facilitate the development of new therapies, it is essential that clinical trial end points are efficient, reliable and responsive to change.
► This systematic review helps to determine the optimal indices for developing and selecting end points in coeliac disease trials.

## INTRODUCTION
Coeliac disease is a gluten-sensitive enteropathy that occurs in genetically predisposed individuals leading to small intestinal villous atrophy. The estimated prevalence is about 1% in most Western populations.[1 2] Aside from the classic symptoms of diarrhoea, weight loss, constipation, bloating and fatigue, nearly half of all patients with coeliac disease present with extraintestinal manifestations,

some of which are a direct consequence of malabsorption. Untreated coeliac disease carries an increased risk of intestinal and extraintestinal malignancy, particularly lymphoma, although recent evidence from large population-based

cohorts suggests that the previously reported risk of malignancy is likely overestimated.[3 4]

The only effective treatment for this lifelong disorder, at present, is a gluten-free diet (GFD) that is highly burdensome and requires long-term patient motivation, compliance and follow-up. Some patients experience persistent symptoms due to inadvertent or deliberate gluten ingestion, a comorbid GI condition or refractory coeliac disease. Furthermore, a large proportion of adult patients fail to achieve complete histological healing of their small bowel mucosa despite adherence to a GFD.[5] The treatment of coeliac disease, however, is entering a new era, with the development of several new pharmacological agents, including immunomodulators, biologics, vaccines and intraluminal therapies such as gluten-specific proteases and endopeptidases, some of which have completed the proof-of-concept stage of clinical development.[6–8] These therapies should address the needs of patients with non-responsive coeliac disease with continuing symptoms and incomplete mucosal healing despite a GFD, but might also decrease the therapeutic burden of a GFD in vulnerable patient groups, such as those with concomitant diabetes mellitus or neurodevelopmental/behavioural conditions.[9–11]

These novel opportunities for drug development also pose substantial challenges for selection of appropriate end points to measure disease activity and response to treatment in clinical trials in order to support labelling claims. These challenges were recently summarised at the US Food and Drug Administration (FDA) Gastroenterology Regulatory Endpoints and the Advancement of Therapeutics 3 (GREAT 3) meeting.[11] Coeliac disease activity and response to therapy can be measured through any combination of symptoms and/or histology, and endoscopic indices have also been proposed. Although several scoring indices have been developed to measure these outcomes, the operating properties of many of these indices have not been systematically evaluated.

The aim of this systematic review was to identify all currently available scoring indices used for the measurement of coeliac disease activity, assess their operating properties and explore their potential utility as outcome measures in clinical trials and drug development programmes.

## METHODS
### Search strategy
We searched MEDLINE (Ovid), EMBASE (Ovid) and CENTRAL (The Cochrane library) databases for records from 1966 to 2015, without language restriction. Search strategies are reported in the online supplementary material. Abstracts from conference proceedings (Digestive Disease Week and United European Gastroenterology Week; 2002–2015) and bibliographies of relevant studies, review articles and meta-analyses were hand-searched to identify additional studies.

### Study selection
Randomised controlled trials (RCTs), cohort, case–control and cross-sectional studies were eligible provided they included patients with confirmed coeliac disease (histologically for adults, and according to the European Society for Paediatric Gastroenterology Hepatology and Nutrition criteria for children[12]), and had employed coeliac disease activity markers in their study outcomes. Two pairs of reviewers (PH+TH and SD +CMT) independently screened abstracts for eligibility. Reasons for ineligibility were recorded. Full-text publications of all potentially eligible articles were retrieved by the same pairs of reviewers to ensure that the inclusion criteria were met.

Disagreements were resolved by consensus or in discussion with a third person (CP).

### Data extraction
All data were extracted into a standard Microsoft Excel spreadsheet and included study details (abstract only or full manuscript, design, sample size), characteristics of the study population (children, adults or both), the type of the end point assessed (patient-reported outcome (PRO), histological or endoscopic index, serological tests, gut permeability assay, genetic tests) and the timing of assessment.

For each evaluative instrument identified, we assessed the following parameters:[13]

1. *Validity*: The ability of the instrument to (a) measure what it intends within the target population ('content validity'); (b) distinguish among groups known to be different, for example, patients and healthy controls ('discriminant construct validity'); and (c) correlate with established markers of disease activity (ie, show 'convergent construct validity').
2. *Reliability*: The inter-item consistency within an instrument and the intraobserver and interobserver agreement.
3. *Responsiveness*: The ability of an instrument to adequately change in response to a treatment of known efficacy.
4. *Feasibility*: The ease of administration and time required for scoring.

All parameters were rated as described in the methodological literature[14] as either + (positive), ? (indeterminate), − (negative) or 0 (no information available) using a priori defined criteria (table 1), based on the FDA guidance document for industry PRO measures.[15]

## RESULTS
### Search results and description of included studies
The search yielded a total of 19 123 citations, of which 7075 were identified as duplicates and removed (figure 1). Of the remaining 12 048 records screened, 347 full-text articles were selected and reviewed for eligibility. A total of 286 studies were considered eligible and included for data extraction. The study types included 50 cross-sectional studies, 66 case-controlled studies, 135 cohort studies and 35 clinical trials (see online supplementary material for the complete reference list). The clinical trials included 24 RCTs, of which 10 were published as an abstract and 14 were published as a full paper A description of the 14 RCTs is outlined in online supplementary table S1. Remarkably, none of the published RCTs clearly defined a meaningful response and seven did not specify a primary outcome.

### Histological coeliac disease activity indices
Histology is the gold standard for diagnosis of coeliac disease in adults and may also help to differentiate coeliac disease-related symptoms from those unrelated to the disease. Six histological indices were identified (table 2), all based on two parameters: intraepithelial lymphocyte (IEL) infiltration and morphological damage, which includes villous atrophy and crypt hyperplasia.[16–21]

The original Marsh classification[16] describes four subtypes: type 1 (infiltrative lesion), type 2 (hyperplastic lesion), type 3 (destructive lesion) and type 4 (atrophic-hypoplastic lesion). The modified Marsh-Rostami classification,[18] similar to the modified Marsh-Oberhüber classification,[17] subcategorises the Marsh type 3 lesion into partial, subtotal and total villous atrophy (3a, 3b and 3c, respectively). The Marsh type 3 lesion subdivision was subsequently somewhat simplified again in the Ensari[20] and

**Table 1**  Criteria to evaluate the operating characteristics of coeliac disease activity instruments

|  | + | − | ? | 0 |
|---|---|---|---|---|
| Content validity | *PROs*: component items based on focus groups, patient interviews or qualitative cognitive interviews<br>*Other evaluative instruments*: component items based on expert consensus | *PROs*: component items not based on focus groups, patient interviews or qualitative cognitive interviews<br>*Other evaluative instruments*: component items not based on expert consensus | Limited or conflicting data preclude firm conclusions | No data available |
| Construct validity | Correlation with established coeliac disease activity markers<br>Able to distinguish among groups that are known to be different | No correlation with established coeliac disease activity markers<br>Unable to distinguish among groups that are known to be different | Limited or conflicting data preclude firm conclusions | No data available |
| Reliability | Acceptable intraobserver and interobserver agreement (intraclass correlation coefficient ≥0.8 and interclass correlation coefficient ≥0.5) | Unacceptable intraobserver and interobserver agreement (intraclass correlation coefficient <0.8 and interclass correlation coefficient <0.5) | Limited or conflicting data preclude firm conclusions | No data available |
| Responsiveness | Adequate and measurable change in response to a gluten challenge or a treatment of known efficacy | Inadequate or no change in response to a gluten challenge or a treatment of known efficacy | Limited or conflicting data preclude firm conclusions | No data available |
| Feasibility | No technical expertise required<br>Favourable cost/benefit ratio<br>High patient acceptability (for PROs, >90% of patients respond to all questions) | Technical expertise required<br>Unfavourable cost/benefit ratio<br>Low patient acceptability (for PROs, <90% of patients respond to all questions) | Limited or conflicting data preclude firm conclusions | No data available |

+, positive; ?, indeterminate; −, negative; 0, no information available; PRO, patient-reported outcome.
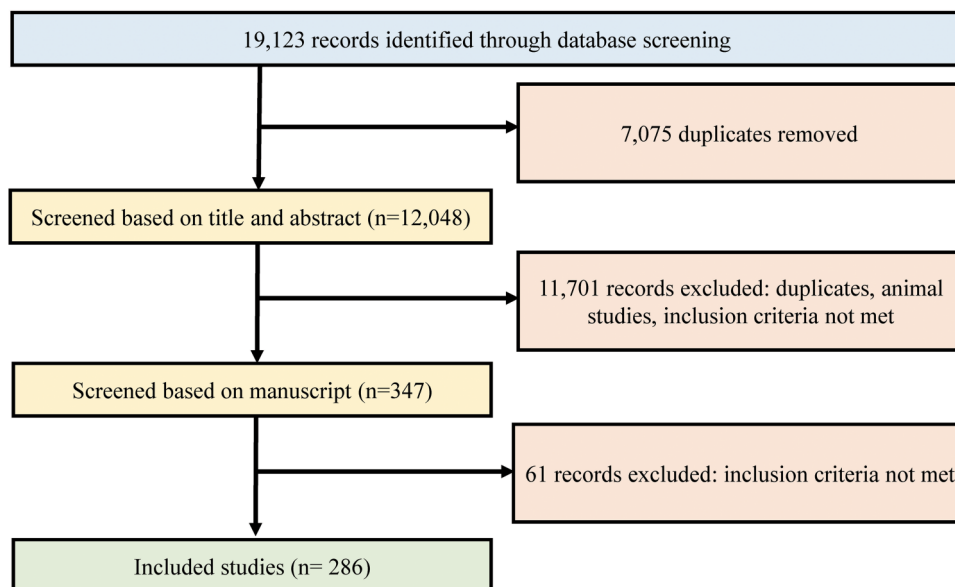
Corazza/Villanacci[19] classifications, both of which eliminated the somewhat subjective distinction between partial (3a) and subtotal (3b) villous atrophy. The morphometric scoring index described by Taavela *et al*[21] is based on quantitative measurements of the villus height-to-crypt depth ratio and the IEL density. It is important to mention that all histological indices identified were derived by a small group of experts, based on face validity. Reliability has only been investigated for four of the six and, with the exception of the morphometric quantitative index, histological responsiveness to gluten challenge or treatment occurs slowly and is not robust (figure 2).

There are logistical challenges to the routine use of histology in clinical trials, including the need for endoscopy, sampling error, patchy distribution of disease and the need for centralised interpretation by experts. Additionally, morphometric analyses are time consuming and highly dependent on optimal orientation of the specimen. Beyond these practical issues, and for

reasons that are not completely understood, rates of histological healing following GFD, generally defined as a complete normalisation of the intestinal villous architecture, vary according to age and may not occur in a significant proportion of adult patients.[5 22 23] There are also conflicting data on the correlation between histology and clinically meaningful disease activity measures, such as patient symptoms. Some of this discordance may be related to the differing timeframes for resolution of symptoms (days) and detection of histological change (months) (figure 2).[24]

### Symptom-based coeliac disease activity indices
Many coeliac disease studies used generic symptoms and symptom-based scales or exploratory coeliac disease-specific subsets of these scales to measure clinical coeliac disease activity. Table 3 shows a list of the individual symptoms and the percentage of studies that reported them in their outcome measures.



19,123 records identified through database screening

7,075 duplicates removed

Screened based on title and abstract (n=12,048)

11,701 records excluded: duplicates, animal studies, inclusion criteria not met

Screened based on manuscript (n=347)

61 records excluded: inclusion criteria not met

Included studies (n= 286)

**Figure 1**  Flow diagram for search strategy to identify relevant studies.

**Table 2**  Operating properties of histological coeliac disease activity indices

| | Classification/grade/items | Content validity | Construct validity | Reliability | Responsiveness | Feasibility |
|---|---|---|---|---|---|---|
| *Histological* | | | | | | |
| Marsh classification[16] | Type 1: increased IEL count<br>Type 2: increased IEL count+crypt hyperplasia<br>Type 3: villus effacement and crypt hyperplasia<br>Type 4: total villous atrophy+crypt hypoplasia | − | ? | − | ? | + |
| Modified Marsh-Oberhuber classification[17] | Type 1: increased IEL count<br>Type 2: increased IEL count+crypt hyperplasia<br>Type 3: partial (a), subtotal (b) or total (c) villous atrophy<br>Type 4: destructive lesion | − | ? | − | ? | + |
| Modified Marsh-Rostami classification[18] | Type 1: increased IEL count<br>Type 2: increased IEL count+crypt hyperplasia<br>Type 3: partial (a), subtotal (b) or total (c) villous atrophy | − | 0 | 0 | 0 | + |
| Corazza/Villanacci classification[19] | Grade A: increased IEL count<br>Grade B1: increased IEL count+villous–crypt ratio <3:1<br>Grade B2: total villous atrophy | − | ? | + | 0 | + |
| Ensari classification[20] | Type 1: increased IEL count<br>Type 2: increased IEL count+crypt hyperplasia+shortened villi<br>Type 3: total villous atrophy | − | ? | 0 | 0 | + |
| Morphometry[21] | Quantitative analysis of IEL count and villous height to crypt depth ratio | − | + | + | + | ? |

+, positive; ?, indeterminate; −, negative; 0, no information available; IEL, intraepithelial lymphocytes.

The most frequently reported symptoms (in >50% of the identified studies) were diarrhoea/loose stools, abdominal pain/cramps and bloating. The most commonly used generic symptom scale was the Gastrointestinal Symptom Rating Scale (GSRS), which was originally developed to rate symptoms in patients with IBS and peptic ulcer disease, and includes five subdomains: diarrhoea, abdominal pain, indigestion, constipation and reflux.[25] In the first RCT with larazotide acetate in coeliac disease, published in 2012, Leffler et al[26] reported on a modified score, the coeliac disease (CeD)-GSRS, that excluded the constipation and reflux subdomains. While not specifically developed as a PRO measure (psychometric instruments consisting of items exclusively reported by patients to quantify symptoms) for coeliac disease, the GSRS has construct validity for this disease due to its correlation with established disease activity markers such as morphometric histopathology scores and serum antibody levels.[27] In addition, the GSRS and CeD-GSRS are rapidly responsive to gluten challenge and therapeutic intervention (figure 2).[26 28]

We also identified five coeliac disease-specific PROs (table 4).[28–32] Two of these (the Celiac Disease Symptom Diary (CDSD) and the Celiac Disease Patient-Reported Outcome (CeD PRO))[28 29] have been developed according to the Study Endpoints and Labeling Development (SEALD) guidelines from the FDA and have been cleared and licensed for use.[15] Importantly, these PROs were developed using adult patients and their applicability to children has not been systematically addressed.

### Endoscopic coeliac disease activity indices

Several non-specific endoscopic features have been identified in coeliac disease.[33] These features include a mosaic, or micronodular appearance of the mucosa, the presence of fissures or grooves between duodenal folds and scalloping, and loss or reduction of these folds. We identified four cohort studies that employed empirically derived endoscopic coeliac disease activity scoring systems (table 5).[34–37]

Tursi et al[35] reported the responsiveness of the score used in their study to a GFD and showed that endoscopic improvement was more rapid (within 6 months) than that of categorical histological indices (within 1 year) (figure 2). The use of endoscopy in coeliac disease clinical trials is limited by the fact that it only detects macroscopic intestinal damage (villous changes) corresponding to more advanced stages of the disease. Advanced technological breakthroughs in endoscopic imaging and interpretation will be required before endoscopy can be a useful assessment tool in coeliac disease.[38]

### Serological coeliac disease activity markers

Several coeliac serology tests, including anti-deamidated gliadin peptide, anti-tissue transglutaminase and anti-endomysial antibodies, have been approved by the FDA as aides in the diagnosis

**Figure 2**  Expected timing for detection of coeliac disease activity improvement with various indices upon initiation of an effective treatment. IEL, intraepithelial lymphocyte.
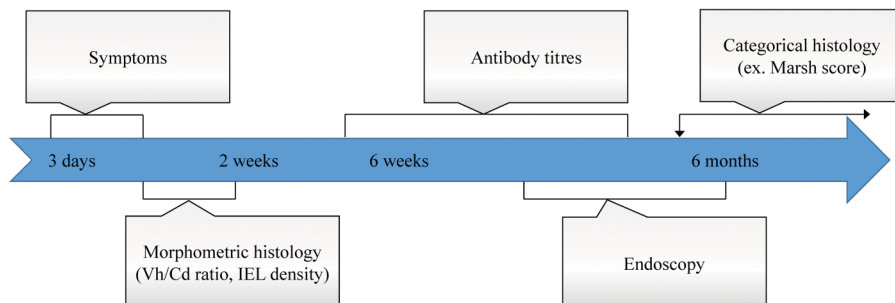
**Table 3** Symptoms identified in generic coeliac disease scales (N=289)

| | <10% of studies | 10–50% of studies | >50% of studies |
|---|---|---|---|
| Flatulence | | X | |
| Diarrhoea/loose stools | | | X |
| Constipation | | X | |
| Abdominal pain/cramps | | | X |
| Nausea | | X | |
| Bloating | | | X |
| Fatigue | | X | |
| Irritability | X | | |
| Steatorrhea | | X | |
| Oedema | X | | |
| Headache | X | | |
| Itchy skin | X | | |
| Reflux | X | | |
| Borborygmi | X | | |
| Weight loss | | X | |

of coeliac disease and are correlated with histological status.[39–41] While they are directly involved in disease pathogenesis and reflect disease activity, their prognostic significance is unknown and they have not been FDA-approved for use as a monitoring tool for coeliac disease activity due to lack of supportive data for this purpose. Antibody titres, therefore, cannot be primary outcome measures in clinical trials but could be used to select or stratify patients with coeliac disease at study inclusion, the latter to better define subpopulations that are more likely to benefit from treatment. They could also potentially be used to monitor for adherence to a GFD in clinical trials as rising levels are nearly diagnostic of increased gluten exposure. For the latter, it is important to note that antibody titres respond more slowly to gluten challenge or a GFD than symptoms or histological morphometrics (figure 2).[42]

### Other markers of coeliac disease activity

A significant proportion of the studies we identified used a change in intestinal permeability, measured by the urinary lactulose/mannitol fractional excretion ratio, as an outcome measure. While an increase in intestinal permeability is a feature of active coeliac disease, it lacks specificity, suffers from high inter-individual variability and has shown poor responsiveness to

treatment in coeliac disease clinical trials.[42 43] It is therefore unlikely to be used as an evaluative instrument for future coeliac disease clinical trials unless the assays are optimised.

Cytokine release assays (to assess the presence of gluten-reactive T cells), a serum or plasma marker of intestinal epithelial damage (intestinal fatty acid-binding protein), faecal calprotectin and a variety of mucosal genetic biomarkers have been described as potential coeliac disease activity markers, but more data are needed to support their use as outcome measures.[44–47]

### DISCUSSION

Coeliac disease affects millions of people worldwide.[2] Adherence to a GFD is challenging and adds social as well as financial burdens to patients. Despite best efforts, inadvertent gluten contamination occurs frequently and some patients will experience symptoms in response to products classified as gluten-free according to the FDA standards (<20 ppm).[48 49] Almost two-thirds of patients with coeliac disease feel that there is a need for alternative therapies, often driven by persistent symptoms or a desire to increase their gluten intake.[10] Several pharmacological therapies are emerging and are in various stages of clinical development in response to this medical need.[7] Although the FDA and European Medicines Agency have increased their collaboration over the years, it is important to note that the two agencies may hold different views on the designation of coeliac disease therapies (eg, drugs vs medical devices), as well as on the criteria for approval of these therapies.[50] To facilitate the development of new therapies, it is essential that clinical trial end points are efficient, reliable and responsive to change.

In this systematic review, we conducted an extensive search and screened >19 000 records to identify all coeliac disease activity indices reported to date and to assess their operating characteristics, with the ultimate goal of proposing a framework for the selection of outcome measures in coeliac disease clinical trials (figure 3). A key finding in this study was that only coeliac disease-specific PROs met all of our stringent evaluative criteria.[15] A recently published systematic review reported four appropriate PROs for use in coeliac disease trials.[51] It should be noted, however, that only two of these, the CDSD and the CeD-PRO, are coeliac disease-specific PROs that were developed according to the FDA SEALD criteria and can thus be used for labelling purposes. No histological, serological or endoscopic indices met all of the assessed evaluative criteria and overall performance was worst for the categorical histological indices.

**Table 4** Operating properties of coeliac disease patient-reported outcome indices

| | Classification/grade/items | Content validity | Construct validity | Reliability | Responsiveness | Feasibility |
|---|---|---|---|---|---|---|
| Celiac Disease Symptom Diary (CDSD, Alvine Pharmaceuticals)[29]* | Diarrhoea, constipation, abdominal pain, bloating, gas, nausea, skin rash, fatigue, headache and difficulty thinking clearly | + | +† | + | + | 0 |
| Coeliac Disease Quality of Life survey[30] | Coeliac disease-related limitations, dysphoria, health concerns and inadequate treatment | + | +† | ? | 0 | + |
| Celiac Disease Questionnaire[31] | Emotional problems, social problems, disease-related worries and GI symptoms | + | +† | + | ? | + |
| Celiac Symptom Index[32] | Coeliac disease-related symptoms and general health | + | +† | + | + | 0 |
| Celiac Disease Patient-Reported Outcome (CeD PRO, Alba Therapeutics)[28]* | Abdominal symptoms, headache and tiredness | + | +† | + | + | 0 |

*Operating characteristics based on the approved US Food and Drug Administration application for use of PROs in clinical trials to support marketing claims.
†Some patients may be clinically asymptomatic but exhibit histological activity.
+, positive; ?, indeterminate; 0, no information available.

**Table 5** Operating properties of endoscopic coeliac disease activity indices

| | Classification/grade/items | Content validity | Construct validity | Reliability | Responsiveness | Feasibility |
|---|---|---|---|---|---|---|
| Lee et al, 2003[34] | 1. Normal (at least threefold per endoscopic field)<br>2. Reduced duodenal folds (<3 folds per endoscopic field)<br>3. Absent duodenal folds, scalloping of folds, fissures, mosaic or nodular appearance of the mucosa | − | + | 0 | 0 | + |
| Tursi et al, 2006[35] | 1. Normal<br>2. Mild alterations (micronodular bulb, granular mucosa in the second part of the duodenum)<br>3. Moderate alterations (scalloping of duodenal folds, reduction of duodenal folds)<br>4. Severe alterations ('mosaic' pattern of the mucosa in the second part of the duodenum and loss of duodenal folds) | − | + | 0 | + | + |
| Cammarota et al, 2007[36] | 1. Villi definitely present<br>2. Villi partially present<br>3. Villi definitely absent using water immersion technique | − | + | 0 | 0 | ? |
| Maiden et al, 2009[37] | 1. Normal<br>2. Mild–moderate changes on video capsule endoscopy<br>3. Moderate–severe changes on video capsule endoscopy<br>Items: micronodular mucosa, mosaicism, scalloping of folds, reduction in number of folds | − | + | + | 0 | − |

+, positive; ?, indeterminate; −, negative; 0, no information available.

In line with FDA recommendations for a paradigm shift from mainly disease-oriented care towards a patient-centred care for chronic medical disorders such as coeliac disease, it is likely that PROs will become primary outcome measures in coeliac disease registration trials.[7 15] A PRO is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response.[15] The obvious advantage of PROs is that they are the most relevant end points from the patient's perspective. However, there are inherent limitations to using a PRO alone as a clinical trial end point for coeliac disease. First, coeliac disease can result in a multitude of mostly non-specific individual symptoms (table 3) that may fluctuate during the course of illness. Second, PRO measurements may be subjective and may result in measurement errors, inefficiency and high placebo responses that can hamper assay sensitivity within clinical trials. Emerging evidence from other therapeutic areas indicates that placebo rates are more pronounced in trials in which disease activity is measured by PROs compared with trials where activity is assessed by physicians.[52]

Third, there is no consensus on a PRO change that constitutes a clinically meaningful response. The largest coeliac disease trial to date[28] illustrates the limitations of a PRO as a solitary end point. This trial included 342 patients with coeliac disease with persisting symptoms despite adherence to a GFD. Patients were randomised to receive larazotide acetate 0.5, 1 or 2 mg three times daily, or placebo. The primary end point was a statistically significant difference in average on-treatment CeD-GSRS. The primary end point was met for the lowest dose only. In addition, the mean CeD-GSRS score in the placebo group decreased by >25% after 16 weeks, resulting in a modest treatment effect size of 0.3 points.[28]

In addition to a definition for a clinically meaningful response, there is also a need to define acceptable values for the number needed to treat (NNT) and number needed to harm (NNH) for coeliac disease drugs. Most existing drugs, as well as those in development for IBS, have an NNT between 5 and 15.[53 54] We therefore believe that an NNT between 5 and 10 would be a reasonable assumption for first-generation coeliac disease drugs
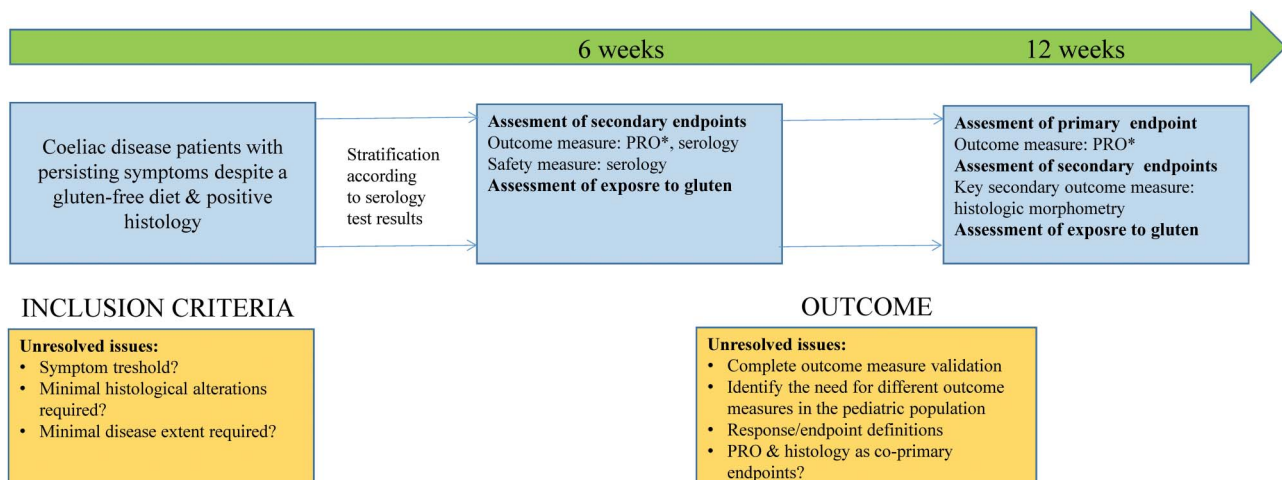
**Figure 3** Evidence-based recommended framework for the conduct and improvement of late-phase coeliac disease clinical trials. *Only the Celiac Disease Symptom Diary and the Celiac Disease Patient-reported Outcome have been accepted by the US Food and Drug Administration. PRO, patient-reported outcome.

used as adjuncts to a GFD. The NNH for chronic conditions such as coeliac disease should be high and not associated with serious harm. Histology or serology might provide more objective trial end points. Histology is arguably a more meaningful target than serology because it is the only parameter associated prognostically with disease course; persistent villous damage is associated with a risk of lymphoproliferative disease[55] and bone fracture.[56] The lack of a clear relationship with symptoms,[24] however, likely precludes the use of histology as a single primary trial end point. This was also concluded at the FDA GREAT 3 meeting, where histology was proposed as a key secondary or safety outcome measure.[11]

With regard to trial design, lessons can be extrapolated from IBD, where it is recognised that the incorporation of objective evidence of inflammation (eg, endoscopy or biochemical markers) in both inclusion criteria and therapeutic end points is important to reduce placebo rates and improve trial efficiency.[57 58] In addition, mucosal healing improves long-term prognosis with reduced need for surgery and fewer hospitalisations.[59 60] Accordingly, the FDA is moving away from their recommendation of the symptom-based Crohn's Disease Activity Index as a primary outcome measure towards co-primary end points comprised of improvement in PROs and endoscopic inflammation.[11 61] The concept of co-primary end points has already been successfully implemented in other disease areas such as psoriasis and rheumatoid arthritis[62 63] and could also be considered in coeliac disease trials through a combination of PROs and histology. This proposal, however, is not without challenges. First, the timeframe for change in these measures can be considerably different and has implications for when they should be assessed (see figure 2). Second, empirical research from trial data sets of treatments of known efficacy would be required to define the optimal testing configuration for these measures (as co-primary end points or within a composite score) and the most efficient index cut points. Given that these choices have profound consequences for trial efficiency, this issue presents a critical challenge to coeliac disease drug development that requires further research. Based on the current best evidence and considerations discussed above, we propose a plan for the conduct and improvement of late-phase coeliac disease clinical trials to help inform this research (figure 3).

The strengths of this systematic review include its scope and size, with identification of all outcome measures used in coeliac disease trials based on an explicit, prespecified and reproducible search strategy. The fact that we have not assessed possible publication bias is a limitation. We attempted to minimise this by also including abstracts from conference proceedings, although this approach may have led to incomplete information on certain coeliac disease activity indices due to the inherent content limits of 'abstract-only' records.

In summary, we have identified all scoring indices for measurement of coeliac disease activity and assessed their operating properties. Some PROs have been validated according to the most rigorous methodological framework, and only the CDSD and the CeD PRO have been accepted for use by the FDA. The performance of categorical histological scales is generally poor, but is not the case for the quantitative scale based on morphometrics. Further work is needed to determine the optimal choice and configuration of end points for coeliac disease trials.

## Author affiliations
[1]Robarts Clinical Trials Inc., University of Western Ontario, London, Ontario, Canada
[2]Department of Gastroenterology, University of Ghent, Ghent, Belgium
[3]Division of Gastroenterology, University of California San Diego, La Jolla, California, USA
[4]Oxford Centre for Clinical Magnetic Resonance Research, University of Oxford, Oxford, UK
[5]Oxford University Clinical Academic Graduate School, John Radcliffe Hospital, Oxford, UK
[6]Department of Medicine, University of Western Ontario, London, Ontario, Canada
[7]Inflammatory Bowel Disease Centre, Academic Medical Centre, Amsterdam, The Netherlands
[8]Department of Epidemiology and Biostatistics, University of Western Ontario, London, Ontario, Canada
[9]Celiac Disease Center, Columbia University, New York, New York, USA
[10]The Celiac Center, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

## REFERENCES

1 Lebwohl B, Ludvigsson JF, Green PH. Celiac disease and non-celiac gluten sensitivity. *Br Med J* 2015;351:h4347.
2 Tack GJ, Verbeek WH, Schreurs MW, *et al*. The spectrum of celiac disease: epidemiology, clinical aspects and treatment. *Nat Rev Gastroenterol Hepatol* 2010;7:204–13.
3 Dewar DH, Ciclitira PJ. Clinical features and diagnosis of celiac disease. *Gastroenterology* 2005;128:S19–24.
4 Ilus T, Kaukinen K, Virta LJ, *et al*. Incidence of malignancies in diagnosed celiac patients: a population-based estimate. *Am J Gastroenterol* 2014;109:1471–7.
5 See JA, Kaukinen K, Makharia GK, *et al*. Practical insights into gluten-free diets. *Nat Rev Gastroenterol Hepatol* 2015;12:580–91.
6 McCarville JL, Caminero A, Verdu EF. Pharmacological approaches in celiac disease. *Curr Opin Pharmacol* 2015;25:7–12.
7 Wungjiranirun M, Kelly CP, Leffler DA. Current status of celiac disease drug development. *Am J Gastroenterol* 2016;111:779–86.
8 Gottlieb K, Dawson J, Hussain F, *et al*. Development of drugs for celiac disease: review of endpoints for Phase 2 and 3 trials. *Gastroenterol Rep (Oxf)* 2015;3:91–102.
9 Shah S, Akbari M, Vanga R, *et al*. Patient perception of treatment burden is high in celiac disease compared with other common conditions. *Am J Gastroenterol* 2014;109:1304–11.
10 Branchi F, Tomba C, Ferretti F, *et al*. Celiac disease and drug-based therapies: inquiry into patient's demands. *Digestion* 2016;93:160–6.
11 Leffler D, Kupfer SS, Lebwohl B, *et al*. Development of celiac disease therapeutics: report of the third Gastroenterology Regulatory Endpoints and Advancement of Therapeutics Workshop. *Gastroenterology* 2016;151:407–11.
12 Husby S, Koletzko S, Korponay-Szabó IR, *et al*. European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 2012;54:136–60.
13 Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27–36.
14 Terwee CB, Bot SD, de Boer MR, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
15 United States Food and Drug Administration. *Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims*. US Department of Health and Human Services, 2009.
16 Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). Gastroenterology 1992;102:330–54.
17 Oberhuber G, Granditsch G, Vogelsang H. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *Eur J Gastroenterol Hepatol* 1999;11:1185–94.
18 Rostami K, Kerckhaert J, von Blomberg BM, *et al*. SAT and serology in adult coeliacs, seronegative coeliac disease seems a reality. *Neth J Med* 1998;53:15–19.
19 Villanacci V, Magazzù G, Pellegrino S, *et al*. Comparison of the Marsh-Oberhuber classification with a new grading system in identifying patients with latent celiac disease. *Minerva Gastroenterol Dietol* 2010;56:371–5.
20 Ensari A. Gluten-sensitive enteropathy (celiac disease): controversies in diagnosis and classification. *Arch Pathol Lab Med* 2010;134:826–36.
21 Taavela J, Koskinen O, Huhtala H, *et al*. Validation of morphometric analyses of small-intestinal biopsy readouts in celiac disease. *PLoS One* 2013;8:e76163.
22 Lanzini A, Lanzarotto F, Villanacci V, *et al*. Complete recovery of intestinal mucosa occurs very rarely in adult coeliac patients despite adherence to gluten-free diet. *Aliment Pharmacol Ther* 2009;29:1299–308.
23 Lebwohl B, Murray JA, Rubio-Tapia A, *et al*. Predictors of persistent villous atrophy in coeliac disease: a population-based study. *Aliment Pharmacol Ther* 2014;39:488–95.
24 Freeman HJ. Pearls and pitfalls in the diagnosis of adult celiac disease. *Can J Gastroenterol* 2008;22:273–80.
25 Svedlund J, Sjödin I, Dotevall G. GSRS—a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci* 1988;33:129–34.
26 Leffler DA, Kelly CP, Abdallah HZ, *et al*. A randomized, double-blind study of larazotide acetate to prevent the activation of celiac disease during gluten challenge. *Am J Gastroenterol* 2012;107:1554–62.
27 Taavela J, Kurppa K, Collin P, *et al*. Degree of damage to the small bowel and serum antibody titers correlate with clinical presentation of patients with celiac disease. *Clin Gastroenterol Hepatol* 2013;11:166–71.
28 Leffler DA, Kelly CP, Green PH, *et al*. Larazotide acetate for persistent symptoms of celiac disease despite a gluten-free diet: a randomized controlled trial. *Gastroenterology* 2015;148:1311–19.
29 Adelman D, Leffler D, Lebwohl B, *et al*. Celiac disease symptom frequency and severity using a disease-specific patient-reported outcome diary: Observations from a psychometric validation study in 202 patients. *Am J Gastroenterol* 2012;107:S603.
30 Dorn SD, Hernandez L, Minaya MT, *et al*. The development and validation of a new coeliac disease quality of life survey (CD-QOL). *Aliment Pharmacol Ther* 2010;31:666–75.
31 Häuser W, Gold J, Stallmach A, *et al*. Development and validation of the Celiac Disease Questionnaire (CDQ), a disease-specific health-related quality of life measure for adult patients with celiac disease. *J Clin Gastroenterol* 2007;41:157–66.
32 Leffler DA, Dennis M, Edwards George J, *et al*. A validated disease-specific symptom index for adults with celiac disease. *Clin Gastroenterol Hepatol* 2009;7:1328–34.
33 Shah VH, Rotterdam H, Kotler DP, *et al*. All that scallops is not celiac disease. *Gastrointest Endosc* 2000;51:717–20.
34 Lee SK, Lo W, Memeo L, *et al*. Duodenal histology in patients with celiac disease after treatment with a gluten-free diet. *Gastrointest Endos* 2003;57:187–91.
35 Tursi A, Brandimarte G, Giorgetti GM, *et al*. Endoscopic and histological findings in the duodenum of adults with celiac disease before and after changing to a gluten-free diet: A 2-year prospective study. *Endoscopy* 2006;38:702–7.
36 Cammarota G, Cuoco L, Cesaro P, *et al*. A highly accurate method for monitoring histological recovery in patients with celiac disease on a gluten-free diet using an endoscopic approach that avoids the need for biopsy: a double-center study. *Endoscopy* 2007;39:46–51.
37 Maiden L, Elliott T, McLaughlin SD, *et al*. A blinded pilot comparison of capsule endoscopy and small bowel histology in unresponsive celiac disease. *Dig Dis Sci* 2009;54:1280–3.
38 Pellisé M, Panés J. Confocal endomicroscopy in celiac disease. *Gastroenterology* 2011;140:1097–9.
39 United States of America Food and Drug Administration. http://www.accessdata.fda. gov/cdrh_docs/reviews/K032571.pdf (accessed 16 Jun 2016).
40 United States of America Food and Drug Administration. http://www.accessdata.fda. gov/cdrh_docs/reviews/k060157.pdf (accessed 16 Jun 2016).
41 United States of America Food and Drug Administration. http://www.accessdata.fda. gov/cdrh_docs/reviews/K113863.pdf (accessed 16 Jun 2016).
42 Leffler D, Schuppan D, Pallav K, *et al*. Kinetics of the histological, serological and symptomatic responses to gluten challenge in adults with coeliac disease. *Gut* 2013;62:996–1004.
43 Daveson AJ, Jones DM, Gaze S, *et al*. Effect of hookworm infection on wheat challenge in celiac disease—a randomised double-blinded placebo controlled trial. *PLoS one* 2011;6:e17366.
44 Ontiveros N, Tye-Din JA, Hardy MY, *et al*. Ex-vivo whole blood secretion of interferon (IFN)-γgamma and IFN-γ-inducible protein-10 measured by enzyme-linked immunosorbent assay are as sensitive as IFN-γ enzyme-linked immunospot for the detection of gluten-reactive T cells in human leucocyte antigen (HLA)-DQ2.5(+)-associated coeliac disease. *Clin Exp Immunol* 2014;175:305–15.
45 Vreugdenhil AC, Wolters VM, Adriaanse MP, *et al*. Additional value of serum I-FABP levels for evaluating celiac disease activity in children. *Scand J Gastroenterol* 2011;46:1435–41.
46 Ertekin V, Selimoğlu MA, Turgut A, *et al*. Fecal calprotectin concentration in celiac disease. *J Clin Gastroenterol* 2010;44:544–6.
47 Capone P, Rispo A, Imperatore N, *et al*. Fecal calprotectin in coeliac disease. *World J Gastroenterol* 2014;20:611–12.
48 United States Food and Drug Administration. http://www.fda.gov/NewsEvents/ Newsroom/PressAnnouncements/ucm363474.htm (accessed 16 Jun 2016).
49 Hollon JR, Cureton PA, Martin ML, *et al*. Trace gluten contamination may play a role in mucosal and clinical recovery in a subgroup of diet-adherent non-responsive celiac disease patients. *BMC Gastroenterol* 2013;13:40.
50 Kashyap UN, Gupta V, Raghunandan HV. Comparison of drug approval process in United States & Europe. J Pharm Sci Res 2013;5:131–6.
51 Canestaro WJ, Edwards TC, Patrick DL. Systematic review: patient-reported outcome measures in coeliac disease for regulatory submissions. *Aliment Pharmacol Ther* 2016;44:313–31.
52 Enck P, Bingel U, Schedlowski M, *et al*. The placebo response in medicine: minimize, maximize or personalize? *Nat Rev Drug Discov* 2013;12:191–204.
53 Enck P, Junne F, Klosterhalfen S, *et al*. Therapy options in irritable bowel syndrome. *Eur J Gastroenterol Hepatol* 2010;22:1402–11.
54 Wall GC, Bryant GA, Bottenberg MM, *et al*. Irritable bowel syndrome: a concise review of current treatment concepts. *World J Gastroenterol* 2014; 20:8796–806.
55 Lebwohl B, Granath F, Ekbom A, *et al*. Mucosal healing and risk for lymphoproliferative malignancy in celiac disease: a population-based cohort study. *Ann Intern Med* 2013;159:169–75.
56 Lebwohl B, Michaëlsson K, Green PH, *et al*. Persistent mucosal damage and risk of fracture in celiac disease. *J Clin Endocrinol Metab* 2014;99:609–16.

57 Hindryckx P, Baert F, Hart A, *et al*. Clinical trials in luminal Crohn's disease: a historical perspective. *J Crohns Colitis* 2014;8:1339–50.

58 Hindryckx P, Baert F, Hart A, *et al*. Clinical trials in ulcerative colitis: a historical perspective. *J Crohns Colitis* 2015;9:580–8.

59 Khanna R, Bressler B, Levesque BG, *et al*. Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. *Lancet* 2015;386:1825–34.

60 Bryant RV, Burger DC, Delo J, *et al*. Beyond endoscopic mucosal healing in UC: histological remission better predicts corticosteroid use and hospitalisation over 6 years of follow-up. *Gut* 2016;65:408–14.

61 Williet N, Sandborn WJ, Peyrin-Biroulet L. Patient-reported outcomes as primary end points in clinical trials of inflammatory bowel disease. *Clin Gastroenterol Hepatol* 2014;12:1246–56.

62 Bachelez H, van de Kerkhof PC, Strohal R, *et al*. Tofacitinib versus etanercept or placebo in moderate-to-severe chronic plaque psoriasis: a phase 3 randomised non-inferiority trial. *Lancet* 2015;386:552–61.

63 Emery P, Breedveld FC, Hall S, *et al*. Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. *Lancet* 2008;372:375–82.