A Call for Research on the Prognostic Role of Follow-Up Histology in Celiac Disease: A Systematic Review

Zsolt Szakács, Noémi Gede, Zoltán Gyöngyi, Margit Solymár, Dezső Csupor, Bálint Erőss, Áron Vincze, Alexandra Mikó, Andrea Vasas, László Szapáry, Dalma Dobszai, Viktória Balikó, Roland Hágendorn, Péter Hegyi, and Judit Bajor

Background: Convincing evidence is lacking on the benefit of follow-up biopsy in celiac disease. Regardless, achieving mucosal recovery (MR) has remained a desirable goal of therapy. We aimed to conduct a systematic review to determine whether MR is a protective factor and persisting villous atrophy (PVA) has negative consequences on long-term outcomes of celiac disease.

Methods: Seven databases were searched for articles discussing celiac patients subjected to a gluten-free diet who had a follow-up biopsy, and clinical and laboratory characteristics were reported by follow-up histology (MR vs. PVA). Outcomes included clinical symptoms, mortality, malignant tumors, nutritional parameters, and metabolic bone disease. Comparative and descriptive studies were included. Since data proved to be ineligible for meta-analysis, the evidence was synthesized in a systematic review.

Results: Altogether, 31 studies were eligible for systematic review. Persisting symptoms were more frequently associated with PVA than with MR, although a lot of symptom-free patients had PVA and a lot of symptomatic patients achieved MR. PVA might be a risk factor of lymphomas, but mortality and the overall rate of malignant tumors seemed independent of follow-up histology. Patients with PVA tended to develop metabolic bone disease more often, although fracture risk remained similar in the groups except in hip fractures of which PVA was a risk factor. Reports on nutritional markers are only anecdotal.

Conclusions: The limited evidence calls for high-quality prospective cohort studies to be arranged to clarify the exact role of follow-up histology in celiac disease.

Keywords: celiac disease, mucosal recovery, persistent villous atrophy, follow-up biopsies, gluten-free diet
INTRODUCTION

Celiac disease (CeD) is a systemic disorder with an increasing worldwide prevalence of ~1% (Di Sabatino and Corazza, 2009; Catassi et al., 2014). In CeD, the immune-mediated destruction of small intestinal villous architecture is triggered by dietary gluten (Reilly et al., 2017).

Rationale of the Study

Although the value of intestinal biopsy at diagnosis is beyond dispute (Bai et al., 2013; Rubio-Tapia et al., 2013; Ludvigsson et al., 2014), the role of follow-up biopsy is a matter of controversy (Pekki et al., 2017). While the restitution of intestinal villi is expected on a strict gluten-free diet, the mucosa fails to restore entirely in a considerable fraction of patients (Szakacs et al., 2017). Recent guidelines recommend a follow-up biopsy if signs and symptoms persist or relapse despite strict adherence to a gluten-free diet (Bai et al., 2013; Rubio-Tapia et al., 2013; Ludvigsson et al., 2014). However, reports proposed that neither the resolution of symptoms (Bardella et al., 2007; Biagi et al., 2014; Fang et al., 2017; Mahadev et al., 2017) nor a strict dietary adherence (Szakacs et al., 2017) guarantees mucosal recovery (MR). One might expect that CeD patients with persistent villous atrophy (PVA) experience a less favorable disease course than those achieving MR (Haines et al., 2008), although convincing evidence is lacking. Yet, achieving MR has remained a desirable goal in CeD. The importance of the topic roots in the burden imposed by the endoscopic procedures and duodenal histological sampling as well as in the subsequent clinical decisions made upon follow-up histology.

Objective of the Study

With this systematic review, we aimed to be the first who systematically collect all available evidence on the impact of follow-up histology (MR vs. PVA) on disease characteristics and clinical course of CeD.

Research Question

Are CeD patients who have not achieved MR exposed to adverse clinical outcomes more frequently than those patients who have achieved MR?

METHODS

Study Design

This systematic review was reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses 2009 (Moher et al., 2009).

Participants and Exposure

Participants of this systematic review include CeD patients subjected to a gluten-free diet. The exposure of interest is the follow-up duodenal histology assessed as either MR or PVA.

Systematic Review Protocol

The study protocol was registered a priori on PROSPERO under CRD42017069815.

Search Strategy

We performed a systematic literature search in MEDLINE (via PubMed), Embase, Cochrane Controlled Register of Trials (CENTRAL), Web of Science, Scopus, WHO Global Index Medicus, and ClinicalTrial.gov from inception up to 14th September 2019 for relevant articles. Free-text terms and Medical Subject Headings were combined into a query, as follows: celiac AND (“mucosal recovery” OR “mucosal healing” OR “mucosal atrophy” OR “intestinal atrophy” OR “duodenal atrophy” OR “villous atrophy” OR “persistent mucosal damage” OR “follow-up biopsy” OR “follow-up duodenal biopsy” OR “follow-up intestinal biopsy” OR “follow-up small intestinal biopsy” OR “follow-up histology” OR “repeated biopsy” OR “repeated histology” OR “control biopsy” OR “control histology”). No filters were imposed upon the search.

Relevant cited articles were explored by reviewing the reference lists of included papers. Citing articles were identified with Google Scholar.

Data Sources, Study Selection, and Data Extraction

Eligibility Criteria

Eligible papers discussed CeD patients subjected to a gluten-free diet with an available record of duodenal follow-up histology, and reported on disease and patients characteristics by follow-up histology. Analytical and descriptive full-text articles or conference abstracts but not case studies were included without language restriction to reduce publication bias.

In our previous meta-analysis, we experienced a substantial variance in the definition of MR across studies (Szakacs et al., 2017). In this systematic review, we defined MR as the resolution of villous atrophy at follow-up biopsy assessed as “non-atrophic” histology based on crypt height:villous depth ratio, Marsh grades 0-2, Marsh-Oberhuber grades 0-2, or Corazza-Villanacci grade A; while PVA was defined as “atrophic” histology, Marsh grade 3, Marsh-Oberhuber grade 3, or Corazza-Villanacci grade B1-B2 (Marsh, 1992; Oberhuber et al., 1999; Corazza and Villanacci, 2005).

Disease and patients characteristics included signs and symptoms, vitamin and mineral levels, anemia, body mass index, metabolic bone disease, malignant tumors and other co-morbid conditions, and long-term mortality.

Selection and Data Collection

Records were combined in a reference manager software (EndNote X7.4, Clarivate Analytics, Philadelphia, PA, the USA) to remove duplicates and overlapping database content. Then,
the standard three-step selection was performed by title, abstract and full-texts. Each step was carried out by two investigators in duplicate. Discrepancies were resolved by third-party arbitration. K-statistics was used to measure the agreement between the investigators after each step.

Numeric and text data were extracted by two investigators onto a pre-defined Excel sheet, discrepancies were resolved by consensus. Although we contacted the authors of original studies for further raw data via email, we discarded these data from the systematic review when we realized that the material is ineligible for meta-analysis due to several reasons (as detailed later).

Design of the Studies Included and Quality Assessment
First, the design of the included papers was identified. Then, quality indicators were chosen based on the Quality in Prognostic Studies tool (Hayden et al., 2006), as follows:
- way of recruitment,
- diagnosis of CeD (only biopsy-verified cases or not),
- the recency of diagnosis (newly diagnosed patients or treated patients were included),
- representativeness of study population to the general CeD population (based on the inclusion and exclusion criteria of the individual studies),
- timing of follow-up biopsy (taken prospectively after enrolment or earlier),
- time elapsed between the diagnosis of CeD and the follow-up biopsy, and that between the follow-up biopsy and the measurement of outcomes,
- definition of PVA (histological classification),
- biopsy sampling site,
- timing of outcome assessment (prospectively after enrolment or earlier),
- definitions of outcomes (with cut-off values if applicable),
- blinding,
- adherence to a gluten-free diet (strict or not), and
- statistical considerations (the analysis directly compared the clinical characteristics by MR and PVA or not and the analysis was adjusted for reasonable confounding factors or not).

FIGURE 1 | Flowchart of the search and selection process.
| Study                    | Setting                      | Recruitment       | Age              | Biopsy-proven CeD? | Representative population? | Newly diagnosed CeD? | Prospectively taken follow-up biopsy? | Villous atrophy and sampling site |
|-------------------------|------------------------------|-------------------|------------------|--------------------|-----------------------------|---------------------|---------------------------------------|----------------------------------|
| Bardella et al. (2007)  | Italy (single-center)        | Consecutive pts   | Adults and children | Yes                | No                          | No                  | No                                    | Marsh-Oberhuber 3−c, duodenum     |
| Cammarota et al. (2007) | Italy (2-center)             | Consecutive pts   | Adults           | Yes                 | Yes                         | No                  | Yes                                   | Marsh-Oberhuber 3−c, duodenum     |
| Carroccio et al. (2008) | Italy (single-center)        | Consecutive pts   | Adults           | Yes                 | Yes                         | No                  | Yes                                   | Marsh-Oberhuber 3−c, duodenum     |
| Ciacci et al. (2002)    | Italy (single-center)        | Consecutive pts   | Adults           | Yes                 | Yes                         | No                  | No                                    | Marsh 3, small intestine          |
| Cornell et al. (2016)   | Poland (single-center)       | Consecutive pts   | Adults           | Yes                 | Yes                         | No                  | Yes                                   | Marsh 3, duodenum                 |
| Dickey et al. (2008)    | Northern Ireland (single-center) | Consecutive pts | Adults           | Yes                 | Yes                         | No                  | Yes                                   | Marsh 3, duodenum                 |
| Fang et al. (2017)      | The USA (Single-center)      | Consecutive pts   | Adults (?)       | No                  | No                          | No                  | No                                    | Corazza-Villanacci B1−2, duodenum |
| Ghazzawi et al. (2014)  | The USA (single-center)      | Consecutive pts   | Children         | Yes                 | Yes                         | No                  | No                                    | Marsh 3, small intestine          |
| Haere et al. (2016)     | Norway (single-center)       | Consecutive pts   | Adults           | Yes                 | Yes                         | No                  | Yes                                   | Marsh-Oberhuber 3<−c, duodenum    |
| Kaukinen et al. (2007)  | Finland (single-center)      | Consecutive pts   | Adults (?=−?)    | Yes                 | No                          | No                  | No                                    | Vv/Cr 3<−c, duodenum, jejunum      |
| Kemppainen et al. (1998)| Finland (single-center)      | Post-hoc analysis of a randomized trial | Adults           | Yes                 | Yes                         | No                  | No                                    | Not reported (“atrophic”), duodenum |
| Koskinen et al. (2010)  | Finland (single-center)      | Consecutive pts   | Adults           | Yes                 | No                          | Yes/No²             | Yes                                   | Vv/Cr 3<−c, jejunum                |
| Lebwohl et al. (2013a,b); Lebwohl et al. (2014, 2015a,b); Emilsson et al. (2018), Ludvigsson et al. (2018), Kurien et al. (2018) | Sweden (28-center) | Consecutive pts (nationwide sample) | Adults and children | Yes | Yes | No | No | Marsh 3, small intestine |
| Leonard et al. (2017)   | Italy (2-center)             | Consecutive pts   | Children         | Yes                 | Yes                         | No                  | No                                    | Marsh-Oberhuber 3<−c, duodenum    |
| Mahadev et al. (2017)   | North-America, Ireland, UK, Norway, and Finland (multicenter) | Post-hoc analysis of a randomized trial | Adults (?)       | No                  | No                          | No                  | No                                    | Vv/Cr 3<−c, duodenum               |
| Pekki et al. (2015)     | Finland (single-center)      | Consecutive pts   | Adults           | Yes                 | (?)                         | Yes                 | Vv/Cr 3<−c, duodenum                 |
| Pekki et al. (2017)     | Finland (single-center)      | Consecutive volunteers | Adults           | Yes                 | Yes                         | No                  | No                                    | Vv/Cr 4<−c, small intestine        |
| Rubio-Tapia et al. (2010)| The USA (single-center)      | Consecutive pts   | Adults           | Yes                 | Yes                         | No                  | No                                    | Vv/Cr 3<−c, duodenum               |

(Continued)
### TABLE 1 | Continued

| Study | Setting | Recruitment | Age | Biopsy-proven CeD? | Representative population? | Newly diagnosed CeD? | Prospectively taken follow-up biopsy? | Definition of villous atrophy and sampling site |
|-------|---------|-------------|-----|--------------------|---------------------------|---------------------|-------------------------------------|-----------------------------------------------|
| Selby et al. (1999) | Australia (single-center) | Symptomatic and asymptomatic pts from a previous food-intolerance study | Adults | Yes | Yes | No | No | Vh/Cr <2, duodenum |
| Souroujon et al. (1982) | Israel (single-center) | Consecutive pts | Adults and Children | Yes | Yes | Yes | Yes | Not reported ("atrophic"), small intestine |
| Thornquist et al. (1993) | Norway (single-center) | Consecutive pts by invitation | Adults | Yes | Yes | No | Yes/No² | Alexander III-IV, jejunum |
| Tuire et al. (2012) | Finland (single-center) | Consecutive volunteers | Adults | Yes | Yes | No | Yes | Marsh 3, duodenum |
| Valdimarsson et al. (1994) | Sweden (single-center) | Consecutive pts with persistent villous atrophy and matched pts with mucosal recovery | Adults | (?) | Yes | No | No | Alexander III-IV, small intestine |
| Walters et al. (1995) | UK (single center) | Consecutive pts by invitation | Adults | Yes | No | No | Yes/No² | Not reported ("atrophic"), small intestine |

* Study population is considered representative to the average CeD population if the study avoided inappropriate exclusions.

² The study observed a group of newly diagnosed CeD patients for the short-term and a group of followed-up patients for the long-term.

³ Not all patients were newly diagnosed.

(?) indicates uncertainty. CeD, celiac disease; pts, patients; Vh/Cr, villous height/crypt depth ratio.

### RESULTS

#### Study Selection and Characteristics

After careful search and selection (Figure 1), 31 papers were included in the systematic review (Souroujon et al., 1982; Thornquist et al., 1993; Valdimarsson et al., 1994; Walters et al., 1995; Kemppainen et al., 1998; Selby et al., 1999; Ciacci et al., 2002; Bardella et al., 2007; Cammarota et al., 2007; Kaukinen et al., 2007; Carroccio et al., 2008; Dickey et al., 2008; Koskinen et al., 2010; Rubio-Tapia et al., 2010; Tuire et al., 2012; Lebwohl et al., 2013a,b, 2014, 2015a,b; Ghazzawi et al., 2014; Pekki et al., 2015, 2017; Cornell et al., 2016; Haere et al., 2016; Fang et al., 2017; Leonard et al., 2017; Mahadev et al., 2017; Emilsson et al., 2018; Kurien et al., 2018; Ludvigsson et al., 2018) (Table 1). Cohen’s κ representing inter-reviewer agreement was 0.72 (substantial), 0.79 (substantial), and 0.92 (almost perfect) for selection by title, abstract, and full-text, respectively.

Out of the 31 papers, eight retrieved data on different outcomes from the Swedish CeD registry; therefore, study populations partially overlap (Lebwohl et al., 2013a,b, 2014, 2015a,b; Emilsson et al., 2018; Kurien et al., 2018; Ludvigsson et al., 2018).

#### Highlights of the Last Five Years

In this subchapter, we summarize the findings of the studies which investigated the prognostic role of follow-up histology and were published after January 2015 (Lebwohl et al., 2013a, 2015a; Pekki et al., 2015, 2017; Haere et al., 2016; Fang et al., 2017; Mahadev et al., 2017; Emilsson et al., 2018; Kurien et al., 2018; Ludvigsson et al., 2018). Eight studies were conducted in Scandinavia, one in the USA, and there was another multicenter study from Europe and North-America. Evidence from univariate analysis suggested a borderline association of PVA with persisting symptoms (OR = 1.656 with CI: 0.949–2.889; \( p = 0.076 \)) favoring mucosal recovery (Fang et al., 2017), which was not confirmed by co-variate-adjusted analysis of the baseline cohort of patients of a multicenter randomized trial on symptomatic CeD patients (Mahadev et al., 2017). Conclusions from three Scandinavian studies on symptom and well-being scores corroborated these findings (Pekki et al., 2015, 2017; Haere et al., 2016). PVA might be associated with lower lumbar T-score measured at 1 year after diagnosis, although neither the risk of osteoporosis at 5 years after diagnosis nor that of fractures at 2 years after diagnosis was higher in this group (Pekki et al., 2015, 2017). Decreased vitamin D and calcium levels might contribute to the impaired bone mineral density (Fang et al., 2017). Regarding other nutrients and minerals, zinc deficiency should be highlighted (39 vs. 14% in patients with PVA and MR, respectively; \( p = 0.0005 \)). The effect of PVA did not manifest itself regarding erythropoiesis: three studies reported similar hemoglobin levels and rate of anemia in patients with PVA and MR (Pekki et al., 2015; Fang et al., 2017; Mahadev et al., 2017). Rate of malignant tumors and...
that of lymphomas were not significantly different in histology groups in two studies 5 and 8 years after diagnosis (Pekki et al., 2015, 2017). Similar neutral associations were found on the rate of cardiovascular diseases (Lebwohl et al., 2015a; Mahadev et al., 2017), serious infections including sepsis, streptococcal, pneumococcal, influenza, herpes zoster, and *Clostridium difficile* infections (Emilsson et al., 2018), and adverse pregnancy outcomes (Lebwohl et al., 2015b). MR seemed to be a protective factor against respiratory and dermatological diseases (Pekki et al., 2017). In contrast, anxiety and depression co-occurred more frequently with MR (Ludvigsson et al., 2018) as did epilepsy (Kurien et al., 2018). With respect to overall mortality, a higher rate was reported with PVA on 8-years follow-up (14 vs. 9%) but the difference was not statistically significant ($p = 0.259$).

**Narrative Review**

In this subchapter, we summarize all available evidence published on the effect of MR and PVA on the outcomes of CeD.

**Symptoms, Symptom Scores and Quality of Life**

In asymptomatic patients, PVA rates ranged from 2 to 62% (Walters et al., 1995; Bardella et al., 2007; Koskinen et al., 2010; Cornell et al., 2016). In symptomatic patients, PVA rate was 38% (Mahadev et al., 2017) (Table 2). Six studies reported symptoms as dichotomies (symptomatic vs. asymptomatic) (Thornquist et al., 1993; Selby et al., 1999; Kaukinen et al., 2007; Carroccio et al., 2008; Fang et al., 2017; Leonard et al., 2017). In univariate analysis, the percentage points of symptomatic patients with PVA and MR were 55 vs. 70% ($p = 0.075$) (Leonard et al., 2017), 69 vs. 57% ($p = 0.076$) (Fang et al., 2017), and 80 vs. 25% ($p < 0.0001$) (Carroccio et al., 2008), respectively (Table 3). Only one study adjusted for reasonable co-variates where the adjusted OR was 3.2, 95% CI: 1.6–6.4 ($p = 0.001$), indicating that patients with PVA tended to be symptomatic more often (Carroccio et al., 2008). Regarding the individual symptoms, only one study adjusted for co-variates where none of the symptoms in focus, that is, diarrhea, bloating, abdominal pain, nausea, tiredness, constipation, heartburn and headache, was significantly associated with PVA (Mahadev et al., 2017) (Table 3).

Five studies (Kaukinen et al., 2007; Tuire et al., 2012; Pekki et al., 2015, 2017; Haere et al., 2016) reported gastrointestinal symptom scores measured with the Gastrointestinal Symptom Rating Scale (GSRS) or its irritable bowel syndrome-adapted version (GSRS-IBS). Four of which performed statistical analysis and reported no association between histology and the scores consistently (Kaukinen et al., 2007; Pekki et al., 2015, 2017; Haere et al., 2016) (Table 4). Results were similar in quality of life measured with the Psychological General Well-Being Index (PGWB) and/or the 36-Item Short Form Health Survey (SF-36) (Kaukinen et al., 2007; Tuire et al., 2012; Pekki et al., 2015, 2017) (Table 4).

**Metabolic Bone Disease and Bone Fractures**

Four studies assessed bone mineral density in univariate analysis (Valdimarsson et al., 1994; Walters et al., 1995; Kaukinen et al., 2007; Pekki et al., 2015) (Table 5). Patients with PVA tended to have lower forearm, femoral and trochanter Z-scores (Valdimarsson et al., 1994); and lower femoral T-score with similar femoral Z- and lumbar T- and Z-scores than those with MR (Pekki et al., 2015). One study showed a reduced bone mineral density with PVA ($OR = 24.5$, $p < 0.0275$ favoring MR) (Walters et al., 1995). Two studies investigated the risk of osteoporosis in the long-term and reported conflicting results: a case-control study reported an increased frequency of osteoporosis with PVA (Kaukinen et al., 2007) but a cohort study failed to confirm (Pekki et al., 2015).

Three studies (Valdimarsson et al., 1994; Lebwohl et al., 2014; Pekki et al., 2017) reported no association between fractures and

---

**TABLE 2 | Persistent villous atrophy in studies including only symptomatic or asymptomatic celiac patients.**

| Study | Symptoms assessed prospectively? | Symptoms assessed at the time of follow-up biopsy? | Only strict GFD? | Symptoms | N\(^0\) of pts with persistent villous atrophy/total (% of total) | Timing of follow-up biopsy (from CeD diagnosis) |
|-------|---------------------------------|-----------------------------------------------|-----------------|----------|-------------------------------------------------|-----------------------------------------------|
| **ASYMPTOMATIC POPULATION** | | | | | | |
| Bardella et al. (2007) | No | Yes | Yes | Not specified | 71 cases/114 pts (82) | 2 years (median) with 1–23 years (range) |
| Walters et al. (1995) | (?) | (?) | Yes | Not specified | 9 cases/17 pts (53) | 2 years (min) |
| Koskinen et al. (2010) | Yes | Yes | (?) | Specified\(^a\) | 21 cases/71 pts (30) | 1 years |
| Corneli et al. (2016) | Yes | Yes | (?) | Not specified | 9 cases/20 pts (45) | 4 m-36 years (range) |
| **SYMPTOMATIC POPULATION** | | | | | | |
| Mahadev et al. (2017) | No | Yes | Yes | Specified\(^b\) | 511 cases/1,345 pts (38) | 4 years (median) with 1 years (min) |

\(^a\)Symptoms included abdominal complaints, malabsorption, and extra-intestinal symptoms.

\(^b\)Symptoms include bloating, tiredness, abdominal pain, diarrhea, nausea, constipation, heartburn, and headache. (?) indicates uncertainty. CeD, celiac disease; GFD, gluten-free diet; m, month; min, minimum; pts, patients.
### TABLE 3 | Presence of symptoms in celiac patients with persistent villous atrophy and mucosal recovery.

| Study                  | Symptoms assessed prospectively? | Symptoms assessed at the time of follow-up biopsy? | Only strict GFD? | Symptoms | N\(^2\) of symptomatic pts/total with persistent villous atrophy (% of total) | N\(^2\) of symptomatic pts/total with mucosal recovery (% of total) | Statistics (atrophy vs. recovery) | Timing of follow-up biopsy (from CeD diagnosis) |
|------------------------|---------------------------------|---------------------------------------------------|-----------------|----------|-------------------------------------------------|-------------------------------------------------|---------------------------|---------------------------------------------|
| **SYMPTOMATIC vs. ASYMPTOMATIC PTS** |                                 |                                                   |                 |          |                                                 |                                                 |                           |                                             |
| Carroccio et al. (2008)| Yes                             | Yes                                               | Yes             | Specified\(^a\) | 36 cases/45 pts (80)                             | 6 cases/24 pts (25)                               | OR = 12 (CI: 3.7–38.9) (p < 0.0001, favors recovery); adjusted OR = 3.2 (CI: 1.6–6.4) (p = 0.001, favors recovery) | 6.5 y and 6.1 y (means for groups) with 3.6 y and 4.6 y (SDs for groups) |
| Fang et al. (2017)     | No                              | Yes                                               | Yes             | Specified\(^c\) | 53 cases/77 pts (69)                             | 116 cases/203 pts (57)                            | OR = 1.656 (CI: 0.949–2.889) (p = 0.076, favors recovery) | 2 y (min)                                      |
| Kaukinen et al. (2007) | Yes                             | No\(^d\)                                          | Yes             | Not specified | 4 cases/13 pts (31)                              | Not reported                                     | Not reported               | 8 y and 10 y (medians for groups) with 3–30 y (range) |
| Leonard et al. (2017)  | No                              | Yes                                               | No              | Not specified | 11 cases/20 pts (55)                             | 58 cases/83 pts (70)                              | p = NS                      | 2.4 y (median) with 1–12 y (range) and 1.4–4 y (Q\(_1\)-Q\(_3\)) |
| Selby et al. (1999)    | (?)                             | Yes                                               | No\(^e\)        | Not specified | 5 cases/26 pts (19)                              | 3 cases/24 pts (13)                               | Not reported               | Not reported                                  |
| Thornquist et al. (1993)| Yes                             | No\(^e\)                                          | Yes             | Specified\(^f\) | 1 case/4 pts (25)                                | 7 cases/9 pts (78)                                | Not reported               | 17 y (mean) with 15–18 y (range?)            |
| **SPECIFIC SYMPTOMs (PRESENT vs. ABSENT)** |                                 |                                                   |                 |          |                                                 |                                                 |                           |                                             |
| Cammarota et al. (2007)| Yes                             | Yes                                               | No              | Diarrhea | 10 cases/49 pts (20)                             | 2 cases/16 pts (13)                               | Not reported               | 6–14 m (range)                               |
|                        |                                 |                                                   |                 | Bloating | 9 cases/49 pts (18)                              | 3 cases/16 pts (19)                               |                           |                                             |
|                        |                                 |                                                   |                 | Abdominal pain | 5 cases/49 pts (10)                             | 7 cases/16 pts (44)                               |                           |                                             |
| Carroccio et al. (2008)| Yes                             | Yes                                               | Yes             | Typical symptoms | 6 cases/45 pts (13)                             | 0 cases/24 pts (0)                                | p = 0.005 (favors recovery) | 6.5 y and 6.1 y (means for groups) with 3.6 y and 4.6 y (SDs for groups) |
|                        |                                 |                                                   |                 | GERD-like symptoms | 9 cases/45 pts (20)                             | 3 cases/24 (13)                                  | p = 0.01 (favors recovery) |                                             |
|                        |                                 |                                                   |                 | lower abdominal symptoms | 21 cases/45 pts (47)                             | 3 cases/21 pts (14)                               | p = 0.001 (favors recovery) |                                             |
|                        |                                 |                                                   |                 | Abdominal pain | 2 cases/5 pts (40)                               | 17 cases/34 pts (50)                              | Not reported               | 2 y (mean) with 22 m (SD) and 4–120 m (range) |
| Ghazzawi et al. (2014) | No                              | Yes                                               | Yes             | Diarrhea | 2 cases/5 pts (40)                               | 5 cases/34 pts (15)                               |                           |                                             |
|                        |                                 |                                                   |                 | Constipation | 0 cases/5 pts (0)                                | 5 cases/34 pts (15)                               |                           |                                             |

(Continued)
| Study            | Only strict GFD | Symptoms assessed prospectively? | Symptoms assessed at the time of follow-up biopsy? |
|------------------|-----------------|----------------------------------|-----------------------------------------------|
| Mahadev et al. (2017) | No              | Yes                               | Yes                                           |
| Statistics (atrophy vs. recovery) | Timing of follow-up biopsy (from Celiac diagnosis) | Mean ± SD | Median (range) |
|                  |                 |                                   |                                               |
| Bloating         | 435 cases/513 pts | 736 cases/834 pts (88) | p = 0.04 (favors atrophy), p = 0.001 (favors atrophy), p = 0.0009 (favors atrophy), and p = 0.04 (favors recovery) for bloating, abdominal pain, nausea, and heartburn, respectively; p = NS for the other comparisons; adjusted p = NS for all comparisons |
| Abdominal pain   | 410 cases/513 pts | 724 cases/834 pts (87) |                                               |
| Tiredness        | 421 cases/513 pts | 708 cases/834 pts (85) |                                               |
| Diarrhea         | 375 cases/513 pts | 643 cases/834 pts (77) |                                               |
| Nausea           | 239 cases/513 pts | 451 cases/834 pts (54) |                                               |
| Constipation     | 263 cases/513 pts | 421 cases/834 pts (50) |                                               |
| Heartburn        | 140 cases/513 pts | 186 cases/834 pts (22) |                                               |
| Headache         | 112 cases/513 pts | 208 cases/834 pts (25) |                                               |

Multivariate analyses are highlighted with bold. *Symptoms included typical (chronic diarrhea, weight loss, anemia), GERD-like, and lower abdominal symptoms (abdominal pain, bloating, nausea, diarrhea, bloating, muscle cramps, and weight loss). Symptoms were assessed 4–5 years after the follow-up biopsy.*

a Symptoms included typical (chronic diarrhea, weight loss, anemia), GERD-like, and lower abdominal symptoms (abdominal pain, bloating, nausea, fatigue, diarrhea, bloating, muscle cramps, and weight loss). Symptoms were assessed after follow-up biopsy (interval undetermined). b Symptoms included gastrointestinal and malabsorptive symptoms, such as diarrhea, bloating, diarrhea, bloating, and abdominal pain. c Symptoms were evaluated after follow-up biopsy (interval undetermined). d Symptoms were assessed 4–5 years after the follow-up biopsy. e Symptoms included gastrointestinal and malabsorptive symptoms, such as diarrhea, bloating, diarrhea, bloating, and abdominal pain. f Symptoms were assessed after follow-up biopsy (interval undetermined). g The analysis was adjusted for age, gender, body mass index, duration of Celiac disease, medications, and laboratory tests. (?) indicates uncertainty. Celiac disease (CeD), celiac disease, CI, confidence interval; GERD, gastroesophageal reflux disease; GFD, gluten-free diet; m, month; min, minimum; OR, odds ratio; pts, patients; NS, non-significant; SD, standard deviation; Q1-Q3, 25 and 75% quartiles; y, year.
| Study                          | Symptoms assessed prospectively? | Symptoms assessed at the time of follow-up biopsy? | Only strict GFD? | Tool            | Persistent villous atrophy | Mucosal recovery | Statistics (atrophy vs. recovery) | Timing of follow-up biopsy (from CeD diagnosis) |
|-------------------------------|---------------------------------|----------------------------------------------------|------------------|----------------|--------------------------|----------------|-------------------------------|-----------------------------------------------|
| Pekki et al. (2017)           | Yes                             | No<sup>a</sup>                                      | No               | GSRS           | 200 Not reported mean: 2.0 (CI: 1.4–2.6) | 276 Not reported mean: 2.0 (CI: 1.7–2.3) | p = NS                        | 1 y                                           |
| Kaukinen et al. (2007)        | Yes                             | Yes                                                 | Yes              | GSRS           | 13 median: 1.5 (Q<sub>1</sub>–Q<sub>3</sub>: 1.3–1.8) | 18 Median: 1.5 (Q<sub>1</sub>–Q<sub>3</sub>: 1.3–2.0) | p = 0.321                     | 8 y and 10 y (medians for groups) with 3–30 y (range) |
| Pekki et al. (2015)           | Yes                             | Yes                                                 | No               | GSRS           | ∑<sup>44</sup> Median: 1.6 (Q<sub>1</sub>–Q<sub>3</sub>: 1.0–4.0) | ∑<sup>44</sup> Median: 1.9 (Q<sub>1</sub>–Q<sub>3</sub>: 1.0–3.0) | p = 0.132                     | 5 y (median)                                |
| Tuire et al. (2012)           | Yes                             | Yes                                                 | Yes              | GSRS<sup>c</sup> | 7 Mean: 1.73 (CI: 1.39–2.07) | 170 Not reported | Not reported | 2–41 y (range)                  |
| Haere et al. (2016)           | Yes                             | No                                                  | No               | GSRS-IBS<sup>d</sup> | 7 Mean: 25.1 (SD: 9.9) | 116 Mean: 26.4 (SD: 9.9) | p = NS                        | 9.3 y (mean) with 5 y (SD) and 2 y (min) |

**GASTROINTESTINAL SYMPTOM SCORES**

**QUALITY OF LIFE SCORES**

---

<sup>a</sup>Scores were assessed years after the follow-up biopsy. <sup>b</sup>The number of patients were not reported for atrophic and recovery groups separately. <sup>c</sup>Subdimensions of GSRS were reported within the original article in detail. <sup>d</sup>A 3-day modified version of the original tool was used. CeD, celiac disease; CI, confidence interval; GFD, gluten-free diet; GSRS, Gastrointestinal Symptom Rating Scale; GSRS-IBS: Gastrointestinal Symptom Rating Scale for Irritable Bowel Syndrome; min, minimum; NS, non-significant; PGWB, Psychological General Well-Being; pts, patients; SF-36, Short Form Health Survey-36; Q<sub>1</sub>–Q<sub>3</sub>, 25% and 75% quartiles; SD: standard deviation.
### TABLE 5 | Bone mineral density in celiac patients with persistent villous atrophy and mucosal recovery.

| Study | Parameter of interest assessed prospectively? | Study Parameter of interest assessed at the time of follow-up biopsy? | Bone mineral density | Persistent villous atrophy | Mucosal recovery | Statistics (atrophy vs. recovery) | Timing of follow-up biopsy (from CeD diagnosis) |
|-------|---------------------------------------------|-----------------------------------------------------------------|----------------------|--------------------------|---------------------|-------------------------------|---------------------------------------------|
|       | Only strict GFD? | Bone mineral density | Scores | N° of pts | Scores | N° of pts | | |
|       | Lumbar (T-score) | Median: −1.1 (Q1−Q3: −2.2 to −0.2) | Σ152a | 85 pts | Median: −0.7 (Q1−Q3: −1.7 to 0.4) | Σ152a | p = 0.117 | 1 y |
|       | Femoral neck (T-score) | Median: −1.1 (Q1−Q3: −1.9 to −0.5) | 85 pts | Median: −0.6 (Q1−Q3: −1.5 to 0.0) | 178 pts | p = 0.024 (favors recovery) | |
|       | Lumbar (Z-score) | Median: −0.5 (Q1−Q3: −1.3 to −0.1) | Σ150a | 85 pts | Median: 0.1 (Q1−Q3: −1.1 to 0.9) | Σ150a | p = 0.376 | |
|       | Femoral neck (Z-score) | Median: −0.5 (Q1−Q3: −0.7 to −0.1) | Σ140a | 85 pts | Median: −0.3 (Q1−Q3: −0.6 to 0.3) | Σ140a | p = 0.323 | |
|       | Valdimarsson et al. (1994) | (?) | No | Forearm (Z-score) | Not reported | 13 pts | Not reported | 17 pts | p < 0.01 (favors recovery) | 8 y and 9 y (medians for groups) with 4–14 y (range) |
|       |       |               |               |               |               |               |               |               | |
|       |       |               |               |               |               |               |               |               | Femoral neck (Z-score) | p < 0.01 (favors recovery) |
|       |       |               |               |               |               |               |               |               | Femoral trochanter (Z-score) | p < 0.05 (favors recovery) |
|       | Kaukinen et al. (2007) | Yes | Yes | Osteoporosis (T-score ≤−2.5 SD) | 7 cases/12 pts (58) | 4 cases/18 pts (22) | p = 0.04 | 8 y and 10 y (medians for groups) with 3–30 y (range) |
|       |       |               |               |               |               |               |               |               | Osteopenia (T-score between −1.0 and −2.4 SD) | 4 cases/12 pts (33) | 8 cases/18 pts (44) | p = NS |
|       | Pekki et al. (2015) | Yes | Noa | Osteoporosis (T- and Z-scores) | 8 cases/71 pts (14) | 16 cases/134 pts (13) | p = 0.850 | 5 y (median) |
|       | Walters et al. (1995) | Yes | Noa | Osteopenia (Z-score) | 7 cases/9 pts (78) | 1 case/8 pts (13) | p < 0.0275 (favors recovery) | 2 y (min) |
|       |       |               |               |               |               |               |               |               | |

**OSTEOPOROSIS AND OSTEOPENIA**

All measurements were performed with dual-energy X-ray absorptiometry except in forearm bone mineral density in the study of Valdimarsson et al. (1994) which used single-photon absorptiometry. a The number of patients were not reported for atrophic and recovery groups separately. b Parameter was assessed years after follow-up biopsy. (?) indicates uncertainty. GFD, gluten-free diet; m, month; min, minimum; NS, non-significant; pts, patients Q1−Q3, 25% and 75% quartiles; SD, standard deviation; y, year.
follow-up histology, except in the frequency of hip fractures being seemingly increased after a 5-years follow-up with PVA (adjusted HR = 2.18, CI: 1.17–4.05) (Lebwohl et al., 2014) (Table 6).

Micro- and Macronutrients
Table 7 summarizes the laboratory findings.

The association of water-soluble vitamins and follow-up histology is understudied. We found the numerical values of vitamins B12 and B9 levels or their laboratory indicators in patients with MR and PVA but the studies did not perform statistical comparison (Dickel et al., 2008). Regarding vitamins B12 and B9 (folic acid) levels, studies found no difference between groups by histology (Pekki et al., 2015; Fang et al., 2017). No reports are available on other water-soluble vitamins.

Reports on fat-soluble vitamins are scarce. Only numerical data are available on vitamin A without statistical comparison made by the authors (Valdimarsson et al., 1994). Patients with PVA tended to have lower 25-hydroxyvitamin D level and total calcium (Fang et al., 2017), though serum parathyroid hormone (Valdimarsson et al., 1994; Pekki et al., 2015), ionized calcium (Valdimarsson et al., 1994; Pekki et al., 2015) and alkaline phosphatase (Valdimarsson et al., 1994) seemed similar in patients with MR and PVA.

Zinc but not copper deficiency may be associated with PVA (39 vs. 14% for zinc deficiency with PVA and MR, respectively, \( p = 0.0005 \)) (Fang et al., 2017).

Anemia
Nine studies reported on hemoglobin, four studies in dichotomies (aemic vs. non-anemic) and five as continuous data (Thorquinst et al., 1993; Kemppainen et al., 1998; Ciacci et al., 2002; Cammarota et al., 2007; Kaukinen et al., 2007; Tuire et al., 2012; Pekki et al., 2015; Fang et al., 2017; Mahadev et al., 2017), of which four reported statistical comparisons (Kaukinen et al., 2007; Pekki et al., 2015; Fang et al., 2017; Mahadev et al., 2017) (Table 8). Hemoglobin levels proved to be not statistically different in two studies (Kaukinen et al., 2007; Pekki et al., 2015). Another two studies examined the frequency of anemia (Fang et al., 2017; Mahadev et al., 2017), of which one found it more frequent with PVA (22 vs. 18%, \( p = 0.04 \)) (Mahadev et al., 2017). However, the association was not confirmed after adjusting for co-variates (\( p = 0.236 \)).

Body Mass Index
Of the six studies reporting body mass index, three provided statistical evidence on having no significant difference between groups while the other three reported numerical values without analysis (Valdimarsson et al., 1994; Kaukinen et al., 2007; Dickey et al., 2008; Tuire et al., 2012; Pekki et al., 2015; Cornell et al., 2016) (Table 9).

Malignant Tumors and Other Co-morbid Conditions
Four studies reported the overall rate of malignant tumors during follow-up (Kaukinen et al., 2007; Tuire et al., 2012; Pekki et al., 2015, 2017), two of which performed univariate analysis where event rates were 4.5 vs. 5.1% (\( p = 0.762 \)) and 11.3 vs. 6% (\( p = 0.116 \)) with PVA vs. MR, respectively (Pekki et al., 2015, 2017) (Table 10).

Incidence of lymphomas was 1.4 vs. 1.6% (\( p = 0.968 \)) with PVA vs. MR in univariate analysis (Pekki et al., 2015). In a registry analysis, PVA had an adjusted HR of 2.26 (CI: 1.18–4.34) for the overall rate of lymphomas, being true for the subgroup of non-Hodgkin lymphomas but not for the subset of T-cell lymphomas (Lebwohl et al., 2013b) (Table 10).

Sporadic records concerned other co-morbid conditions. Cardiovascular diseases including heart failure and atrial fibrillation (Lebwohl et al., 2015a), hypertension (Mahadev et al., 2017), serious infections (including sepsis, streptococcal,

### TABLE 6 | Fractures in celiac patients with persistent villous atrophy and mucosal recovery.

| Study | Events assessed prospectively? | Only strict GFD? | Type of fracture | Persistent villous atrophy | Mucosal recovery | Statistics (atrophy vs. recovery) | Timing of follow-up biopsy (from CeD diagnosis) | Follow-up period |
|-------|-------------------------------|-----------------|-----------------|--------------------------|-----------------|---------------------------------|-----------------------------------------------|---------------|
| Valdimarsson et al. (1994) | Yes | No | Any fracture | 9 fractures in 6 cases/13 pts | 5 fractures in 5 cases/17 pts | \( p = \) NS for the number of patients with fractures | 8 y and 9 y (medians for groups) with 4–14 y (range) | Not reported |
| Pekki et al. (2017) | Yes | No | Any fracture | 53 cases/200 pts | 80 cases/276 pts | \( p = 0.859 \) | 1 y | Years (?) |
| Lebwohl et al. (2014) | No | (?) | Any fracture\(^a\) | 492 cases/34409 PYO | 483 cases/36418 PYO | \( HR = 0.93 \) (CI: 0.82–1.06)\(^b\) | 6 m–5 y (range) | 8.6 y (median) with 6.9–15.7 y (Q\(_1\)–Q\(_3\)) from follow-up biopsy |
| | | | Likely osteoporotic fracture\(^a\) | 124 cases/20617 PYO | 96 cases/11916 PYO | \( HR = 1.11 \) (CI: 0.84–1.46)\(^b\) | | |
| | | | Hip fracture\(^d\) | 26 cases/37925 PYO | 63 cases/39861 PYO | \( HR = 1.67 \) (CI: 1.05–2.66)\(^b\) (favors recovery) | | |

\(^a\) Detailed data are available in the original article. 
\(^b\) The analysis was adjusted for age, gender, duration of celiac disease, calendar period, and educational attainment. (?) indicates uncertainty. CI, confidence interval; GFD, gluten-free diet; HR, hazard ratio (reference group: mucosal recovery); NS, non-significant; PYO, person-years observation; Q\(_1\)–Q\(_3\): 25 and 75% quartiles.
TABLE 7 | Vitamins, minerals, and homocysteine in celiac patients with persistent villous atrophy and mucosal recovery.

| Study                      | Parameter of interest assessed prospectively? | Parameter of interest assessed at the time of follow-up biopsy? | Only strict GFD? | Parameter of interest assessed at the time of follow-up biopsy? | Only strict GFD? | Laboratory study                  | Study Parameter of interest assessed at the time of follow-up biopsy? | Only strict GFD? | Parameter of interest assessed at the time of follow-up biopsy? | Only strict GFD? | Total N of pts | N of pts with deficiency (% of total) | Level | N of pts | N of pts with deficiency (% of total) | Level | Statistics (atrophy vs. recovery) | Timing of follow-up biopsy (from CeD diagnosis) |
|----------------------------|-----------------------------------------------|---------------------------------------------------------------|-----------------|---------------------------------------------------------------|-----------------|-----------------------------------|---------------------------------------------------------------------|-----------------|---------------------------------------------------------------|-----------------|----------------|-------------------------------------|-------|----------------|-------------------------------------|-------|----------------|-------------------------------------|-------|----------------|-------------------------------------|-------|
| **VITAMIN B₂ (RIBOFLAVIN)** |                                               |                                                               |                 |                                                               |                 | EGRAC                             | Vitamin B₂ (RIBOFLAVIN)                                              |                 |                                                               |                 | 24                          | Mean: 1.38 (SD: 0.19)       | 41     | Mean: 1.28 (SD: 0.13)       | Not reported            | 1 y (min)                         |
| Dickey et al. (2008)       | Yes                                           | Yes                                                           | Yes             |                                                               |                 | 24                                | Mean: 69.9 (SD: 24.0)                                               |                 |                                                               |                 | 41                          | Mean: 91.0 (SD: 43.3)       | Not reported | 1 y (min)                         |
| **VITAMIN B₆ (PYRIDOXINE)**|                                               |                                                               |                 |                                                               |                 | Plasma pyridoxal phosphate (nmol/l, vitamin)                        | **VITAMIN B₆ (PYRIDOXINE)**                                        |                 |                                                               |                 | 24                          | Mean: 1.28 (SD: 0.13)       | Not reported | 1 y (min)                         |
| Dickey et al. (2008)       | Yes                                           | Yes                                                           | Yes             |                                                               |                 | 24                                | Mean: 69.9 (SD: 24.0)                                               |                 |                                                               |                 | 41                          | Mean: 91.0 (SD: 43.3)       | Not reported | 1 y (min)                         |
| Kemppainen et al. (1998)   | No                                             | Yes                                                           | (?)             |                                                                |                 | Erythrocyte folic acid (cut-off: not reported)                       | **Kemppainen et al. (1998)**                                      | No             | Yes                                                          | (?)             | 37                          | 3 (8)                  | Not reported | 1 y                                 |
| Pekki et al. (2015)        | Yes                                           | Yes                                                           | No              |                                                               |                 | Erythrocyte folic acid (nmol/l)                                      | **Pekki et al. (2015)**                                            | No             | Yes                                                          | (?)             | 37                          | 0 (0)                  | Not reported | 1 y                                 |
| Thornquist et al. (1993)   | Yes                                           | No                                                            | No              |                                                               |                 | Serum folic acid (nmol/l)                                            | **Thornquist et al. (1993)**                                      | No             | Yes                                                          | (?)             | 4                           | Not reported | Not reported | 17 y (mean) with 15–18 y (range ?) | 2–41 y (range) |
| Tuire et al. (2012)        | Yes                                           | Yes                                                           | Yes             |                                                               |                 | Erythrocyte folic acid (nmol/l)                                      | **Tuire et al. (2012)**                                            | Yes            | Yes                                                          | (?)             | 7                           | 170                  | Not reported | Not reported                       |
| **VITAMIN B₉ (FOLIC ACID)** |                                               |                                                               |                 |                                                               |                 | Erythrocyte folic acid (nmol/l)                                      | **VITAMIN B₉ (FOLIC ACID)**                                        |                 |                                                               |                 | 24                          | Mean: 900 (SD: 574)        | 41     | Mean: 1048 (SD: 791)        | Not reported            | 1 y (min)                         |
| Dickey et al. (2008)       | Yes                                           | Yes                                                           | Yes             |                                                               |                 | Erythrocyte folic acid (nmol/l)                                      | **VITAMIN B₉ (FOLIC ACID)**                                        |                 |                                                               |                 | 24                          | Mean: 23.5 (SD: 22.8)       | 41     | Mean: 25.7 (SD: 31.5)        | Not reported            | 1 y (min)                         |
| Kemppainen et al. (1998)   | No                                             | Yes                                                           | (?)             |                                                               |                 | Erythrocyte folic acid (cut-off: not reported)                       | **Kemppainen et al. (1998)**                                     | No             | Yes                                                          | (?)             | 37                          | 3 (8)                  | Not reported | 1 y                                 |
| Pekki et al. (2015)        | Yes                                           | Yes                                                           | No              |                                                               |                 | Erythrocyte folic acid (nmol/l)                                      | **Pekki et al. (2015)**                                            | No             | Yes                                                          | (?)             | 4                           | Not reported | Not reported | 17 y (mean) with 15–18 y (range ?) | 2–41 y (range) |
| Thornquist et al. (1993)   | Yes                                           | No                                                            | No              |                                                               |                 | Serum folic acid (nmol/l)                                            | **Thornquist et al. (1993)**                                      | No             | Yes                                                          | (?)             | 4                           | Not reported | Not reported | 17 y (mean) with 15–18 y (range ?) | 2–41 y (range) |
| Tuire et al. (2012)        | Yes                                           | Yes                                                           | Yes             |                                                               |                 | Erythrocyte folic acid (nmol/l)                                      | **Tuire et al. (2012)**                                            | Yes            | Yes                                                          | (?)             | 7                           | 170                  | Not reported | Not reported                       |
| **VITAMIN B₁₂ (COBALAMIN)** |                                               |                                                               |                 |                                                               |                 | Serum vitamin B₁₂ (pmol/l)                                           | **VITAMIN B₁₂ (COBALAMIN)**                                        |                 |                                                               |                 | 24                          | Mean: 242.5 (SD: 106.8)     | 41     | Mean: 276.6 (SD: 137.1)     | Not reported            | 1 y (min)                         |
| Dickey et al. (2008)       | Yes                                           | Yes                                                           | Yes             |                                                               |                 | Serum vitamin B₁₂ (pmol/l)                                           | **Dickey et al. (2008)**                                           |                 |                                                               |                 | 24                          | Mean: 242.5 (SD: 106.8)     | 41     | Mean: 276.6 (SD: 137.1)     | Not reported            | 1 y (min)                         |
| Fang et al. (2017)         | No                                             | No                                                            | Yes             |                                                               |                 | Serum vitamin B₁₂ (cut-off: 180 ng/l)                               | **Fang et al. (2017)**                                             | No             | No                                                            | No              | 60                          | 2 (3)                   | 175     | 3 (2)                          | 2 y (min)                              |
| Kemppainen et al. (1998)   | No                                             | Yes                                                           | (?)             |                                                               |                 | Serum vitamin B₁₂ (cut-off: not reported)                            | **Kemppainen et al. (1998)**                                     | No             | Yes                                                          | (?)             | 37                          | 1 (3)                   | 3                   | 0 (0)                          | 1 y                                 |
| Pekki et al. (2015)        | Yes                                           | Yes                                                           | No              |                                                               |                 | Serum vitamin B₁₂ (pmol/l)                                           | **Pekki et al. (2015)**                                            | Yes            | Yes                                                          | No              | 85                          | 2 (3)                   | 175     | 3 (2)                          | 2 y (min)                              |
| Thornquist et al. (1993)   | Yes                                           | No                                                            | No              |                                                               |                 | Serum vitamin B₁₂ (pmol/l)                                           | **Thornquist et al. (1993)**                                      | Yes            | No                                                            | No              | 4                           | Not reported | Not reported | 17 y (mean) with 15–18 y (range ?) | 2–41 y (range) |
|                           |                                               |                                                               |                 |                                                               |                 | Serum vitamin B₁₂ (pmol/l)                                           | **Thornquist et al. (1993)**                                      | No             | No                                                            | No              | 4                           | Not reported | Not reported | 17 y (mean) with 15–18 y (range ?) | 2–41 y (range) |

(Continued)
### TABLE 7 | Continued

| Study                          | Parameter of interest assessed prospectively? | Parameter of interest assessed at the time of follow-up biopsy? | Only strict GFD? | Laboratory study | Persistent villous atrophy | Mucosal recovery | Statistics (atrophy vs. recovery) | Timing of follow-up biopsy (from CeD diagnosis) |
|-------------------------------|---------------------------------------------|---------------------------------------------------------------|------------------|-----------------|-----------------------|----------------|----------------------------------|-----------------------------------------------|
| *VITAMIN A (RETINOL)*         |                                             |                                                               |                  |                 |                       |                |                                  |                                               |
| Kemppainen et al. (1998)      | No                                         | Yes                                                           | (?)              | Serum vitamin A (cut-off: not reported) | 37          | 5 (14)                       | 3               | 0 (0)                          | Not reported                                 | 1 y                                           |
| *VITAMIN D (CHOLECALCIFEROL)* |                                             |                                                               |                  |                 |                       |                |                                  |                                               |
| Fang et al. (2017)            | No                                         | No                                                            | Yes              | Serum total 25-hydroxy vitamin D<sub>2</sub> and D<sub>3</sub> (cut-off: 25 ng/ml) | 55          | 17 (31)                      | 174             | 14 (8)                         | \(p = 0.0001\) (favors recovery)              | 2 y (min)                                    |
| Valdimarsson et al. (1994)    | Yes                                        | (?)                                                           | No               | Serum 25-hydroxy vitamin D                                | 13          | Median: 52 (range: 18–91)     | 17              | Median: 80 (range: 37–95)       | \(p < 0.05\) (favors recovery)               | 8 y and 9 y (medians for groups with 4–14 y (range) |
| *IRON AND FERRITIN*           |                                             |                                                               |                  |                 |                       |                |                                  |                                               |
| Fang et al. (2017)            | No                                         | No                                                            | Yes              | Serum ferritin (cut-off: males: 24 mcg/l, females: 11 mcg/l) | 69          | 7 (10)                       | 165             | 21 (13)                        | \(p = 0.6636\)                               | 2 y (min)                                    |
| Kemppainen et al. (1998)      | No                                         | Yes                                                           | (?)              | Serum total iron (cut-off: not reported)                  | 37          | 15 (41)                      | 3               | 0 (0)                          | Not reported                                 | 1 y                                           |
| Kemppainen et al. (1998)      | No                                         | Yes                                                           | (?)              | Serum ferritin (cut-off: not reported)                    | 37          | 5 (14)                       | 3               | 0 (0)                          | Not reported                                 | 1 y                                           |
| Peikki et al. (2015)          | Yes                                        | Yes                                                           | No               | Serum total iron (μmol/l) \(\sum^{144}\)                  | \(\sum^{144}\) | Median: 19.6 (Q<sub>1</sub>–Q<sub>3</sub>: 13.4–24.4) | \(\sum^{144}\) | Median: 17.6 (Q<sub>1</sub>–Q<sub>3</sub>: 14.0–21.5) | \(p = 0.245\)                               | 1 y                                           |
| Souroujon et al. (1982)       | Yes                                        | Yes                                                           | (?)              | Serum ferritin \(2\)                                     | 34          | Mean: 16.2 (SD: 1.38)        | 34              | Mean: 25.7 (SD: 1.82)          | Not reported                                 | 1 y                                           |
| *CALCIUM*                     |                                             |                                                               |                  |                 |                       |                |                                  |                                               |
| Fang et al. (2017)            | No                                         | No                                                            | Yes              | Serum total calcium (cut-off: 8.9 mg/dl)                   | 65          | 13 (20)                      | 164             | 4 (2)                          | \(p < 0.0001\) (favors recovery)             | 2 y (min)                                    |

(Continued)
| Study                        | Parameter of interest assessed prospectively? | Parameter of interest assessed at the time of follow-up biopsy? | Only strict GFD? | Laboratory study                                                                 | Persistent villous atrophy | Mucosal recovery | Statistics (atrophy vs. recovery) | Timing of follow-up biopsy (from CeD diagnosis) |
|------------------------------|-----------------------------------------------|---------------------------------------------------------------|-----------------|---------------------------------------------------------------------------------|----------------------------|-----------------|---------------------------------|-----------------------------------------------|
| ZINC                         |                                               |                                                               |                 | Serum ionized calcium (mmol/l)                                                   | Median: 1.25 (Q₁-Q₃:    | Median: 1.24 (range:           | p = 0.708         | 1 y                             |
|                              |                                               |                                                               |                 |                                                                                   | 1.21-1.28)                 | 1.16-1.31)          |                                 |                                 |
| Fang et al. (2017)           | No                                            | No                                                           | Yes             | Serum zinc (cut-off: 0.66 mcg/ml)                                                | 18 (39)                   | 20 (14)                     | p = 0.0005 (favors recovery)  | 2 y (min)                        |
| Kemppainen et al. (1998)     | No                                            | Yes                                                           | (?)             | Serum zinc (cut-off: not reported)                                               | 10 (27)                   | 0 (0)                       | Not reported                   | 1 y                             |
| COPPER                       |                                               |                                                               |                 | Serum copper (0.75 mcg/ml)                                                       | 51                         | 4 (3)                       | p = 0.5740                     | 2 y (min)                        |
| Fang et al. (2017)           | No                                            | No                                                           | Yes             | Serum copper (0.75 mcg/ml)                                                       | 0 (0)                     | 4 (3)                       |                                 |                                 |
| HOMOCYSTEINE                  |                                               |                                                               |                 | Plasmatic homocysteine (μmol/l)                                                  | 24                         | 41                          | Not reported                   | 1 y (min)                        |
| Dickey et al. (2008)         | Yes                                           | Yes                                                          | Yes             | Plasmatic homocysteine (μmol/l)                                                  | Mean: 11.4 (SD: 5.5)       | Mean: 10.4 (SD: 2.6)         |                                 |                                 |

*Only the total N⁰ of pts was given. *Parameter was measured years after the follow-up biopsy. *Parameter was measured within 1 month of follow-up biopsy. *Values were within the normal range for all patients and not reported for recovery and atrophic groups separately. (?) indicates uncertainty. CeD, celiac disease; EGRAC, erythrocyte glutathione reductase activation coefficient; GFD, gluten-free diet; m, month; min, minimum; NS, non-significant; Q₁-Q₃, 25% and 75% quartiles; SD, standard deviation; y, year.
## TABLE 8 | Anemia and hemoglobin levels in celiac patients with persistent villous atrophy and mucosal recovery.

| Study                     | Parameter of interest assessed prospectively? | Parameter of interest assessed at the time of follow-up biopsy? | Only strict GFD? | Laboratory study | Persistent villous atrophy | Mucosal recovery | Statistics (atrophy vs. recovery) | Timing of follow-up biopsy (from CeD diagnosis) |
|---------------------------|------------------------------------------------|---------------------------------------------------------------|------------------|-----------------|---------------------|---------------------|----------------------------------|-----------------------------------------------|
| Cammarota et al. (2007)   | Yes                                            | Yes                                                          | No               | Iron-deficiency anemia (Hb cut-off: not reported) | 49                  | 16 (33)             | 16                  | 4 (25)                         | Not reported                  | 6–14 m (range) |
| Ciacci et al. (2002)      | No                                             | Yes                                                          | No               | Hb (g/l)        | 93                  | Mean: 133            | 297                 | Not reported^a                  | Not reported^a                | 7.78 y (mean) |
| Fang et al. (2017)        | No                                             | No^b                                                         | Yes              | anemia (Hb cut-off: males: 135 g/l, females: 120 g/l) | 97                  | 14 (14)             | 249                 | 32 (13)                        | p = 0.7255                    | 2 y (min)       |
| Kaukinen et al. (2007)    | Yes                                            | Yes                                                          | Yes              | Hb (g/l)        | 13                  | mean: 121 (CI: 114–128) | 18                  | Mean: 13.2                      | p = NS                          | 8 y and 10 y (medians for groups) with 3–30 y (range) |
| Kemppainen et al. (1998)  | No                                             | Yes                                                          | (?)              | anemia (Hb cut-off: not reported)               | 37                  | 7 (19)              | 3                   | 1 (33)                         | Not reported                 | 1 y |
| Mahadev et al. (2017)     | No                                             | Yes                                                          | Yes              | anemia (Hb cut-off: not reported)                | 511                 | 114 (22)            | 834                 | 148 (18)                       | unadjusted                      | 4 y (median) with 1 y (min) |
| Pekki et al. (2015)       | Yes                                            | Yes                                                          | No               | Hb (g/l)        | ∑153^f              | Median: 145 (Q1–Q3: 134–150) | ∑153^f              | Median: 147 (Q1–Q3: 147–151) | p = 0.342                      | 1 y |
| Thornquist et al. (1993)  | Yes                                            | No^d                                                         | No               | Hb              | 4                   | Not reported^a      | 9                   | Not reported^a                  | Not reported^a                | 17 y (mean) with 15–18 y (range ?) |
| Türe et al. (2012)        | Yes                                            | Yes                                                          | Yes              | Hb              | 7                   | Median: 140 (range: 128–161) | 170                 | Not reported^a                  | Not reported^a                | 2–41 y (range) |

Multivariate analyses are highlighted with bold. ^a Values were given by Marsh 0 and 1–2 separately. ^b Parameter was assessed within 1-month of follow-up biopsy. ^c The analysis was adjusted for age, gender, body mass index, duration of GFD, medications, laboratory tests, and symptoms. ^d Parameter was measured years after the follow-up biopsy. ^e Values by recovery and atrophy were not reported separately. (?) indicates uncertainty. CeD, celiac disease; CI: confidence interval; GFD, gluten-free diet; Hb, hemoglobin; m, month; min, minimum; NS, non-significant; Q1–Q3, 25% and 75% quartiles; SD, standard deviation; y, year.
pneumococcal, influenza, herpes zoster, and *Clostridium difficile* infections) (Emilsson et al., 2018) and adverse pregnancy outcomes (Lebwohl et al., 2015b) were not influenced by follow-up histology. However, there was a significant reduction in the frequency of respiratory and dermatological diseases (Pekki et al., 2017) and in the rate of elevated aspartate and alanine transaminase levels with MR (Mahadev et al., 2017). In contrast, MR predisposed to developing anxiety and depression (Ludvigsson et al., 2018) as well as epilepsy (Kurien et al., 2014). The effect of follow-up histology on immune-mediated co-morbidities including dermatitis herpetiformis is severely understudied (Valdimarsson et al., 1994; Tuire et al., 2012; Mahadev et al., 2017).

### All-Cause and Cause-Specific Mortality

Four studies reported all-cause mortality during follow-up (Kaukinen et al., 2007; Rubio-Tapia et al., 2010; Lebwohl et al., 2013a; Pekki et al., 2015), two of which adjusted the analysis for significant co-variates (Rubio-Tapia et al., 2010; Lebwohl et al., 2013a). In this regard, MR seemed to be protective in a small retrospective study with a HR of 0.13 (CI: 0.02–1.06) (Rubio-Tapia et al., 2010), not confirmed by a nationwide registry analysis (HR = 1.0, CI: 0.86–1.19 for PVA) (Lebwohl et al., 2013a). In the latter study, results were consistent on cardiovascular, tumor-related and respiratory deaths: the beneficial effect of MR was not confirmed (Lebwohl et al., 2013a) (Table 11).

### DISCUSSION

#### Summary of Findings

Whether the follow-up biopsy in asymptomatic CeD is needed is uncertain, and, to date, no systematic review has addressed this question. Achieving MR is a desirable goal of treatment; however, findings of reports which investigated its beneficial effects on disease course are inconsistent. In line with our previous findings, MR rates ranged from 9 to 97% across the studies included (Szakacs et al., 2017). The fact that a considerable fraction of CeD patients does not achieve MR underlines the importance of investigating the potential prognostic role of follow-up histology.

One important conclusion of this review is that many asymptomatic patients do not achieve MR, and vice versa, many symptomatic patients do achieve MR. Analogously, MR cannot guarantee the symptoms to disappear, whereas PVA is often associated with persisting symptoms, even on a long-term follow-up exceeding 1 year (Tables 2, 3). Persisting symptoms may indicate poor dietary adherence (Abdulkarim et al., 2002; Leffler et al., 2007; Haere et al., 2016) or gluten-independent food intolerance (Carroccio et al., 2008) but other diseases, such as pancreatic insufficiency or small intestinal bacterial overgrowth, might be in the background (Fine et al., 1997). Noteworthy that a few studies included patients with questionable adherence to gluten-free diet, although a top-quality study established a straight relationship between persisting symptoms and PVA in co-variate-adjusted analysis of patients with good adherence (Carroccio et al., 2008). Studies used various definitions for assessing symptoms; some did not report how symptoms were specified at all. Although a smaller study favored MR

| Study | Parameter of interest assessed at the time of follow-up biopsy? | Only strict GFD? | Total N | Body mass index (kg/m²) | Persistent villose atrophy | Mucosal recovery | Time between the diagnosis of CeD and follow-up biopsy (atrophy vs. recovery) |
|-------|---------------------------------------------------------------|----------------|---------|------------------------|-------------------------|-----------------|--------------------------------------------------------------------------------|
| Cornell et al. (2016) | Yes | Yes | 11 | Mean: 22.72 (SD: 3.70) | Yes | 4 m (range) | 8 y and 9 y (medians for groups) with 3–30 y (range) |
| Dickey et al. (2008) | Yes | Yes | 9 | Mean: 24.9 (SD: 4.7) | No | 1 y (min) | 2–41 y (range) |
| Kaukinen et al. (2007) | Yes | Yes | 41 | Mean: 25.9 (SD: 3.9) | Yes | 1 y | 8 y and 10 y (medians for groups) |
| Emilsson et al. (2018) | Yes | Yes | 18 | Mean: 24.3 (SD: 3.3) | No | 13 | 8 y and 9 y (medians for groups) |
| Lebwohl et al., 2013a | Yes | Yes | 136 | Median: 23.9 (Q1–Q3: 22.1–27.1) | No | 13 | 8 y and 9 y (medians for groups) |
| Mahadev et al. (2017) | Yes | Yes | 3,164 | Median: 24.3 (Q1–Q3: 22.0–27.3) | No | 13 | 8 y and 9 y (medians for groups) |
| Pekki et al. (2015) | Yes | Yes | 220 (range: 19–35) | Median: 23.6 (Q1–Q3: 21–27) | No | 13 | 8 y and 9 y (medians for groups) |
| Rubio-Tapia et al., 2010 | Yes | Yes | 2,129 | Median: 23.7 (range: 17–29) | No | 13 | 8 y and 9 y (medians for groups) |
| Lebwohl et al. (2013a) | Yes | Yes | 170 | Median: 23.7 (range: 17–29) | No | 13 | 8 y and 9 y (medians for groups) |
| Valdimarsson et al. (1994) | Yes | Yes | 17 | Median: 23.6 (range: 19–30) | No | 13 | 8 y and 9 y (medians for groups) |
| Tuire et al. (2012) | Yes | Yes | 1 | Median: 24 (range: 19–25) | No | 13 | 8 y and 9 y (medians for groups) |
| Carroccio et al. (2008) | Yes | Yes | 1 | Median: 25 (range: 23–29) | No | 13 | 8 y and 9 y (medians for groups) |

**TABLE 2** | Body mass index in celiac patients with persistent villose atrophy and mucosal recovery.
regarding a set of individual symptoms (Carroccio et al., 2008), the most extensive study including only symptomatic patients found significant associations in univariate analysis (with an unexpected inverse association between PVA and heartburn). Interestingly, all associations proved to be nonsignificant after adjusting for co-variates (Mahadev et al., 2017). Self-reporting of subjective complaints may contribute to the discrepant results. Use of symptom scales (Svedlund et al., 1988) provided a comparable measure with homogeneous results: none of the studies attributed higher scores for those patients with PVA compared to those with MR (Table 4). The same association applies to quality of life; however, based on small sets of patients (Table 4). Altogether, the diagnostic accuracy of persisting symptoms seems insufficient for indicating a follow-up biopsy. Evidence coming from the two prospective studies is not enough to decide whether PVA precisely predicts the long-term persistence of symptoms (Pekki et al., 2015, 2017) (Table 4).

PVA appears to be associated with radiologically detected metabolic bone disease in univariate analysis (Table 5), which translates into an increased risk of hip fractures (as an independent predictor) but not into that of the overall fractures (Table 6). Calcium malabsorption (Fang et al., 2017) together with vitamin D deficiency (Valdimarsson et al., 1994; Fang et al., 2017) are likely to be causative factors. Length of follow-up (not standardized among studies) may be insufficient for restoring bone mineral density (Pekki et al., 2015), especially in those with severe bone impairment at diagnosis. Body mass index, as an important co-variate, seems to be independent of mucosal status (Table 9).

Laboratory studies describing vitamin and mineral levels tend to recover on an adequate diet. Patients on a long-term strict gluten-free diet may suffer from vitamin deficiencies (Hallert et al., 2002, 2009), this phenomenon might be explained by the vitamin- and micronutrient-poor nature of the gluten-free diet as compared to a balanced gluten-containing diet (Thompson, 1999). None of the studies adjusted for dietary adherence and vitamin supplementation which might counteract malabsorption and masked the effects of PVA. Variability in clinical and histological severity at diagnosis might delay villous restitution, thereby affecting vitamin status (not taken into account in the studies) (Kemppainen et al., 1998). Besides, improvement of histology might lag behind the recovery of laboratory values (Pekki et al., 2015), although the length of follow-up exceeded 1 year in all studies (Table 7).

Regarding mortality, the adjusted HRs calculated in the individual studies attributed a neutral effect to MR and PVA (Table 11) (Rubio-Tapia et al., 2010; Lebwohl et al., 2013a). Findings were similar on malignant tumors, except in certain lymphoproliferative diseases based on an extensive registry analysis (noteworthy that data were not adjusted for dietary adherence) (Lebwohl et al., 2013b). A possible explanation could be that the increment in rates of mortality and tumors with PVA

| Study | Only strict GFD? | Malignant tumors | N⁰ of cases/Total (%) | Statistics | Timing of follow-up biopsy (from diagnosis of CeD) | Follow-up period |
|-------|----------------|------------------|----------------------|------------|--------------------------------------------------|-----------------|
| Kaukinen et al. (2007) | Yes | Any malignant tumor⁹ | 3 cases/13 pts (23.1) | Not reported | 8 y and 10 y (medians for groups) (range: 3-30 y) | 4-5 years from follow-up biopsy |
| Lebwohl et al. (2013b) | No | Any lymphoma | 41 cases/4317 pts (0.9) | HR = 2.26 (CI: 2.28-4.34)⁰ | 1.3 (median) (range: 0.1-5 y) | 10.6 y (median) from diagnosis of CeD and 8.9 y (median) from follow-up biopsy |
| Pekki et al. (2015) | No | Any malignant tumor⁹ | 9 cases/200 pts (4.5) | p = 0.762 | 1 y | 8 y (median) from diagnosis of CeD |
| Pekki et al. (2017) | No | Any malignant tumor⁹ | 8 cases/71 pts (11.3) | p = 0.116 | 1 y | 5 y (median) from diagnosis of CeD |
| Tuire et al. (2012) | Yes | Any malignant tumor⁹ | 0 cases/7 pts (0.0) | Not reported | 7 y, 9 y, and 10 y (medians for groups) (range: 2-41 y) | From diagnosis of CeD until follow-up biopsy |

Multivariate analyses are highlighted with bold. ⁹Cases included T-cell lymphoma, small bowel adenocarcinoma, and carcinoid tumor; ¹⁰Cases included the uterus, breast, lung, pancreas, urinary bladder, and prostate tumors; lymphomas, and sarcomas. ¹¹Cases included breast, uterine, and prostate cancer; and lymphoma. ¹²The analysis was adjusted for age, sex, year of follow-up biopsy, education, and duration of celiac disease at the time of follow-up biopsy. CeD, celiac disease; CI, confidence interval; GFD, gluten-free diet; HR, hazard ratio; pts, patients; m, month; y, year.
was counteracted by a decreased risk of ovarian, endometrial and breast cancer, likely due to lower body weight and/or earlier menopause (Ludvigsson et al., 2012; Lebwohl et al., 2013a).

**Strengths and Limitations**

Our study has several strengths. The question we raised is unique without known previous systematic review. We carried out an extensive systematic search with high coverage of patient-important outcomes in a transparent manner by independent (and pre-trained) investigators with good inter-rater agreement. The quality of the included papers was rigorously assessed.

Although the amount of data presented in the tables would be enough for meta-analytical aggregations that we intended to do initially (as declared in the protocol of the work, see the PROSPERO record); finally, our team decided not to do so because of several concerns. The decision-making algorithm is presented in Figure 2.

We are aware that the evidence suffers from several limitations. (1) Most studies did not recruit patients immediately after the diagnosis of CeD (29/31, 94%), did not take the follow-up biopsy after recruitment (23/31, 74%) and did not assess the disease course prospectively. (2) Although the diagnosis of CeD was mainly biopsy-proven, only two papers included newly diagnosed cases and observed them longitudinally until the follow-up biopsy. (3) Details of histological sampling and processing (e.g., sampling site, number of tissue samples, orientation, staining) were not consistently reported across papers, while the procedures often deviated from recent gold standards (e.g., samples were taken from the jejunum, histological classification systems were not specified). Similarly, the definition of villous atrophy varied across the studies. (4) Observational studies are vulnerable to bias: in registry analyses, data of the deceased were often missing (survivorship bias), the definition of villous atrophy varied across the studies. (4) Observational studies are vulnerable to bias: in registry analyses, data of the deceased were often missing (survivorship bias), attendance to regular control visits was incomplete (attrition bias), baseline differences between groups of patients with MR and those with PVA were rarely balanced with co-variate-adjusted analyses (selection bias), study samples were not taken consistently from the general CeD population with consecutive recruitment (selection bias), and investigators assessing the outcomes were rarely blind to mucosal status (performance bias, detection bias). (5) Most evidence came from studies focusing on adults while an inverse association between age and rate of MR was not a criterion in several studies, the length of diet varied. (7) Publication bias cannot be reliably assessed in systematic reviews.

**CONCLUSION**

**Implications for Practice**

The results of publications on the prognostic role of follow-up histology (that is, MR and PVA) are not in agreement. Some adverse outcomes (e.g., persistent symptoms and metabolic bone...
Implications for Research

With a view to the future, prospective cohort studies are urged to be organized to collect decisive evidence on the prognostic role of follow-up histology.

DATA AVAILABILITY STATEMENT

All datasets analyzed this study are included in the manuscript/supplementary files.

AUTHOR CONTRIBUTIONS

ZS and JB conceptualized the study. AM, DD, VB, LS, AV, and RH collected the data. NG performed the formal analysis and designed the figures. ZS assessed the quality of studies. ZG, DC, and MS interpreted the results. ÁV, ZS, JB, and BE drafted the manuscript. ÁV and PH supervised the work. All authors approved the final draft submitted, involved in the study design, edited, read and approved the final manuscript.

FUNDING

This study was supported by an Economic Development and Innovation Operative Programme Grant (GINOP 2.3.2-15-2016-00048), an Institutional Development for Enhancing Intelligent Specialization Grant (EFOP-3.6.2-16-2017-0006) of the National Research, Development and Innovation Office, the ÚNKP-18-3-I and ÚNKP-19-3-I New National Excellence Program of the Ministry of Human Capacities. Sponsors had no effect on study design, data collection, analysis, interpretation of the findings, or preparation of the manuscript.

REFERENCES

Abdulkarim, A. S., Burgart, L. J., See, J., and Murray, J. A. (2002). Etiology of non-responsive celiac disease: results of a systematic approach. Am. J. Gastroenterol. 97, 2016–2021. doi: 10.1111/j.1572-0241.2002.05917.x

Bai, J. C., Fried, M., Coraza, G. R., Schuppan, D., Farthing, M., Catassi, C., et al. (2013). World gastroenterology organisation global guidelines on celiac disease. J. Clin. Gastroenterol. 47, 121–126. doi: 10.1097/MCG.0b013e31827a6f83

Bardella, M. T., Velio, P., Cesana, B. M., Prampolini, L., Casella, G., Di Bella, C., et al. (2007). Coeliac disease: a histological follow-up study. Histopathology 50, 465–471. doi: 10.1111/j.1365-2559.2007.02621.x

Bagi, F., Vattiao, C., Agazzi, S., Balduzzi, D., Schiepatti, A., Gobbi, P., et al. (2014). A second duodenal biopsy is necessary in the follow-up of adult celiac patients. Ann. Med. 46, 430–433. doi: 10.3109/07853890.2014.913378
Cammarota, G., Cuoco, L., Cesaro, P., Santoro, L., Cazzato, A., Montalto, M., et al. (2007). 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 39, 46–51. doi:10.1055/s-2006-945044

Carroccio, A., Ambrosiano, G., Di Prima, L., Pirrone, G., Iacono, G., Florena, A., M., et al. (2008). Clinical symptoms in celiac patients on a gluten-free diet. Scand. J. Gastroenterol. 43, 1315–1321. doi:10.1080/03002590802200044

Catassi, C., Gatti, S., and Fasano, A. (2014). The new epidemiology of celiac disease. J. Pediatr. Gastroenterol. Nutr. 59 (Suppl. 1), S7–S9. doi:10.1097/MPG.0000000000000539

Kurien, M., Ludvigsson, J. F., Sanders, D. S., Zilberberg, H. M., Green, P. H., Sundelin, H. E. K., et al. (2018). Persistent mucosal damage and risk of epilepsy in people with celiac disease. Eur. J. Neurol. 25, 592–e538. doi:10.1111/ene.13564

Lebwohl, B., Emilsson, L., Frobert, O., Einstein, A. J., Green, P. H., and Ludvigsson, J. F. (2015a). Mucosal healing and the risk of ischemic heart disease or atrial fibrillation in patients with celiac disease: a population-based study. PLoS ONE 10:e0117529. doi:10.1371/journal.pone.0117529

Lebwohl, B., Granath, F., Ekborn, A., Montgomery, S. M., Murray, J. A., Rubio-Tapia, A., et al. (2013a). Mucosal healing and mortality in coeliac disease. Aliment. Pharmacol. Ther. 37, 332–339. doi:10.1111/apt.12164

Lebwohl, B., Granath, F., Ekborn, A., Smedby, K. E., Murray, J. A., Neugut, A. I., et al. (2013b). Mucosal healing and risk for lymphoproliferative malignancy in celiac disease: a population-based cohort study. Ann. Intern. Med. 159, 169–175. doi:10.7326/m3-819-159-3-201308060-00006

Lebwohl, B., Michaelsson, K., Green, P. H., and Ludvigsson, J. F. (2014). Persistent mucosal damage and risk of fracture in celiac disease. J. Celiac. Metab. 99, 609–616. doi:10.1016/j.jcm.2014.02.001

Leffler, D. A., Dennis, M., Hyett, B., Kelly, E., Schuppan, D., and Kelly, C. P. (2007). Etiologies and predictors of diagnosis in non-responsive celiac disease. Clin. Gastroenterol. Hepatol. 5, 445–450. doi:10.1016/cgh.2006.12.006

Leonard, M. M., Weir, D. C., Degrootse, M., Mitchell, P. D., Singh, P., Silvester, J., A., et al. (2017). Value of IgA tTG in predicting mucosal recovery in children with celiac disease on a gluten-free diet. J. Pediatr. Gastroenterol. Nutr. 64, 286–291. doi:10.1097/MGP.0000000000001460

Ludvigsson, J. F., Bai, J. C., Biagi, F., Card, T. R., Ciacci, C., Ciclitira, P. J., et al. (2014). Diagnosis and management of adult celiac disease: guidelines from the British Society of Gastroenterology. Gut 63, 1210–1228. doi:10.1136/gutjnl-2013-306578

Ludvigsson, J. F., Lebwohl, B., Chen, Q., Broms, G., Wolf, R. L., Green, P. H. R., et al. (2018). Anxiety after celiac disease diagnosis predicts mucosal healing: a population-based study. Aliment. Pharmacol. Ther. 48, 1091–1098. doi:10.1111/apt.14991

Ludvigsson, J. F., West, J., Ekborn, A., and Stephansson, O. (2012). Reduced risk of breast, endometrial and ovarian cancer in women with celiac disease. Int. J. Cancer 131, E244–250. doi:10.1002/ejic.201454

Mahadev, S., Murray, I. A., Wu, T. T., Chandan, V. S., Torbenson, M. S., Kelly, C. P., et al. (2017). Factors associated with villous atrophy in symptomatic coeliac disease patients on a gluten-free diet. Aliment. Pharmacol. Ther. 45, 1084–1093. doi:10.1111/apt.13988

Marsh, M. N. (1992). Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity (celiac sprue). Gastroenterology 102, 330–354. doi:10.1016/S0016-5085(92)81918-P

Mohr, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339:b2535. doi:10.1136/bmj.b2535

Oberhuber, G., Granditsch, G., and Vogelsang, H. (1999). The histopathology of celiac disease: time for a standardized report scheme for pathologists. Eur. J. Gastroenterol. Hepatol. 11, 1185–1194. doi:10.1080/09541339911000001

Pekhi, H., Kurppa, K., Maki, M., Huhtala, H., Laurila, K., Ilus, T., et al. (2017). Performing routine follow-up biopsy 1 year after diagnosis does not affect long-term outcomes in celiac disease. Aliment. Pharmacol. Ther. 45, 1459–1468. doi:10.1111/1365-2036.13805

Pekhi, H., Kurppa, K., Maki, M., Huhtala, H., Sievanen, H., Laurila, K., et al. (2015). Predictors and significance of incomplete mucosal recovery in celiac disease after 1 year on a gluten-free diet. Am. J. Gastroenterol. 110, 1078–1085. doi:10.1038/ajg.2015.155
Reilly, N. R., Husby, S., Sanders, D. S., and Green, P. H. R. (2017). Coeliac disease: to biopsy or not? Nat. Rev. Gastroenterol. Hepatol. 15:60–66. doi: 10.1038/nrgastro.2017.121
Rubio-Tapia, A., Hill, I. D., Kelly, C. P., Calderwood, A. H., and Murray, J. A. (2013). ACG clinical guidelines: diagnosis and management of celiac disease. Am. J. Gastroenterol. 108, 656–676. doi: 10.1038/ajg.2013.79
Rubio-Tapia, A., Rahim, M. W., See, J. A., Lahr, B. D., Wu, T. T., and Murray, J. A. (2010). Mucosal recovery and mortality in adults with celiac disease after treatment with a gluten-free diet. Am. J. Gastroenterol. 105, 1412–1420. doi: 10.1038/ajg.2010.10
Selby, W. S., Painter, D., Collins, A., Faulkner-Hogg, K. B., and Loblay, R. H. (1999). Persistent mucosal abnormalities in coeliac disease are not related to the ingestion of trace amounts of gluten. Scand. J. Gastroenterol. 34, 909–914. doi: 10.1080/003655299750025390
Souroujon, M., Ashkenazi, A., Lupo, M., Levin, S., and Hegesh, E. (1982). Serum ferritin levels in celiac disease. Am. J. Clin. Pathol. 77, 82–86. doi: 10.1093/ajcp/77.1.82
Svedlund, J., Sjödin, I., and Dotevall, G. (1988). GSRS—a clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. Dig. Dis. Sci. 33, 129–134. doi: 10.1007/BF01535722
Szakacs, Z., Matrai, P., Hegyi, P., Szabo, I., Vincze, A., Balasko, M., et al. (2017). Younger age at diagnosis predisposes to mucosal recovery in celiac disease on a gluten-free diet: a meta-analysis. PLoS ONE 12:e0187526. doi: 10.1371/journal.pone.0187526
Thompson, T. (1999). Thiamin, riboflavin, and niacin contents of the gluten-free diet: is there cause for concern? J. Am. Diet. Assoc. 99, 858–862. doi: 10.1016/S0002-8223(99)00205-9
Thornquist, H., Jacobsen, G. S., Dahl, L. B., and Marhaug, G. (1993). Coeliac disease and gluten-free diet: a following-up study of fifteen young adults. Ann. Nutr. Metab. 37, 293–301. doi: 10.1159/000177780
Tuiri, I., Maria-Leena, L., Teera, S., Katri, H., Jukka, P., Paivi, S., et al. (2012). Persistent duodenal intraepithelial lymphocytosis despite a long-term strict gluten-free diet in celiac disease. Am. J. Gastroenterol. 107, 1563–1569. doi: 10.1038/ajg.2012.220
Valdimarsson, T., Toss, G., Ross, I., Lofman, O., and Strom, M. (1994). Bone mineral density in coeliac disease. Scand. J. Gastroenterol. 29, 457–461. doi: 10.3109/00365529409096838
Walters, J. R., Banks, L. M., Butcher, G. P., and Fowler, C. R. (1995). Detection of low bone mineral density by dual energy x ray absorptiometry in unsuspected suboptimally treated celiac disease. Gut 37, 220–224. doi: 10.1136/gut.37.2.220

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Szakács, Gede, Gyöngyi, Solymár, Csupor, Erőss, Vincze, Mikó, Vásas, Szapáry, Dobszai, Balikó, Hágendorn, Hegyi and Bajor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.