Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Follow-up duodenal biopsy status and long-term outcomes in adults with celiac disease: a systematic review of prognostic factor studies

Mohamed H. E. Ebrahim
mhebrahim@horus.edu.eg

Horus University
Mazin M. Eladaroussy
Horus University
Kamal A. Kamaleldein Rizk
Horus University
Mohamad Mahfouz
HCA Midwest Centerpoint Medical Center
Amira Al Husseini
Lebanese University
Ramsha Asif Hussain
Mohammed Bin Rashid University of Medicine and Health Sciences
Afrin Jubair Ali
Lancaster University
Ibrahim Mahmoud Eldesouki
Horus University
Aman Kshetrapal
University of Birmingham

Research Article

Keywords: celiac disease, persistent villous atrophy, mucosal healing, follow-up biopsy, prognosis, systematic review

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11705764/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Persistent villous atrophy (PVA) on follow-up duodenal biopsy is common in adults with celiac disease on a gluten-free diet. It is unclear whether it predicts severe complications, death or bone outcomes, and guidelines disagree on whether to repeat the biopsy.

Methods

We searched five databases (January 2019 to 14 October 2026) and the reference list of the previous systematic review. We included cohort studies of adults with biopsy-confirmed celiac disease that compared PVA with mucosal recovery on follow-up biopsy. The primary outcome was a composite of severe complications (refractory celiac disease, enteropathy-associated T-cell lymphoma, other lymphoproliferative malignancy and small-bowel adenocarcinoma). Secondary outcomes were all-cause mortality, fractures, and osteoporosis or bone mineral density. We assessed risk of bias with QUIPS and certainty with GRADE, grouped overlapping reports by cohort, and synthesised findings narratively.

Results

Of 3291 records, 10 reports from six independent cohorts were included. None had low risk of bias (four moderate, six high). Two reports gave a composite complication estimate: adjusted hazard ratio (HR) 9.53 (95% CI 4.77 to 19.04) in a multicentre adult cohort and unadjusted HR 7.61 (3.18 to 17.64) in a conference abstract. Certainty was very low, partly because refractory celiac disease is itself defined by PVA. PVA was associated with lymphoproliferative malignancy in a Swedish registry cohort (adjusted HR 2.26, 1.18 to 4.34; low certainty). Mortality estimates disagreed (HR 1.01, 0.86 to 1.19 in Sweden; 7.69, 0.94 to 50.0 at the Mayo Clinic; 2.93, 1.43 to 6.02 in the multicentre cohort; very low certainty). Any fracture was not associated with PVA (HR 0.93, 0.82 to 1.06; low certainty), whereas hip fracture was (HR 1.67, 1.05 to 2.66; very low certainty).

Conclusions

PVA on follow-up biopsy may mark a higher risk of lymphoproliferative malignancy and complications, but the evidence is of low to very low certainty and comes from few independent cohorts, mostly at high risk of bias. It supports targeted, not routine, repeat biopsy. Prospective adult cohorts with standardised histology and lagged analyses are needed.

Systematic review registration:

PROSPERO CRD420261560295

Background

Celiac disease is a chronic immune-mediated enteropathy triggered by dietary gluten in genetically predisposed people, and the duodenal mucosa is central to its diagnosis and follow-up [1, 2]. Treatment is a lifelong gluten-free diet (GFD). Active disease shows villous atrophy on histology. When villous atrophy persists on a GFD, it is called persistent villous atrophy (PVA); mucosal recovery means that the damaged mucosa has normalised or improved to a non-atrophic state [2]. Symptoms and negative serology do not reliably show that the mucosa has healed [1, 3].

PVA is common and can be silent. In a retrospective Mayo Clinic cohort, confirmed mucosal recovery was estimated at 34% after 2 years and 66% after 5 years, and clinical response did not predict histological recovery [4]. In the CADER cohort, which had structured dietary surveillance, 53% of adults had PVA at the 2-year scheduled biopsy, and most were asymptomatic [5]. Guidance on repeat biopsy reflects this uncertainty. The American College of Gastroenterology guideline allows a follow-up biopsy in asymptomatic adults after 2 years on a GFD through shared decision-making, and advises biopsy when symptoms persist or return [1]. An international guideline on monitoring states that follow-up biopsy is not mandatory in asymptomatic patients on a GFD who have no features suggesting a higher risk of complications [6]. The 2025 European Society for the Study of Coeliac Disease (ESsCD) update

also states that follow-up biopsy is not required routinely in all adults [7]. In United States claims data, a second biopsy was recorded in 19.4% of patients [8]. Rates of healing and recommendations therefore differ, and it is unclear what histological reassessment adds for predicting later outcomes.

Studies that link follow-up histology to outcomes are cohort-specific and report different endpoints in different populations. A Swedish registry cohort found no association between PVA and all-cause mortality [9], and a Mayo Clinic cohort found a non-significant association between confirmed mucosal recovery and lower mortality [4]. The same Swedish registry linked PVA to lymphoproliferative malignancy [10], and a nationwide Swedish cohort of small-bowel adenocarcinoma reported few events and a wide confidence interval [11]. For bone outcomes, the Swedish registry found no association with fractures overall but a higher risk of hip fracture [12].

The only dedicated systematic review, by Szakács and colleagues, included 31 studies and synthesised them without meta-analysis [13]. It searched to September 2019, and further primary studies have since appeared. It included children and conference abstracts, combined several reports from the same Swedish registry in one row, and did not report a GRADE certainty assessment [13]. An updated synthesis is needed that focuses on adults, on direct comparisons of PVA with mucosal recovery, on overlap between cohorts and on the certainty of the evidence.

In this systematic review of prognostic factor studies, we compared adults with persistent villous atrophy and adults with mucosal recovery on follow-up duodenal biopsy for a composite of severe complications (refractory celiac disease, enteropathy-associated T-cell lymphoma, other lymphoproliferative malignancy and small-bowel adenocarcinoma), and reported each component separately where data allowed. Secondary outcomes were all-cause mortality, fractures, and osteoporosis or bone mineral density.

Methods

Protocol and registration

We reported this review according to the PRISMA 2020 statement [14] (Additional file 2). The protocol was registered in PROSPERO (CRD420261560295) before the searches were run.

Eligibility criteria

We set the criteria with the PICOS framework [15] in the protocol. We included adults (18 years or older) with biopsy-confirmed celiac disease (Marsh 3 at diagnosis) who followed a GFD for at least 6 months and had at least one follow-up duodenal biopsy. The prognostic factor was histological status on follow-up biopsy: PVA or mucosal recovery. PVA was defined as Marsh-Oberhuber grade 3a to 3c [16, 17], Corazza-Villanacci grade B1 or B2 [18], Alexander grade III to IV, or a villous height to crypt depth (Vh:Cd) ratio below the study's threshold. Mucosal recovery was Marsh 0 to 2, any non-atrophic histology, or a Vh:Cd ratio above the study's threshold.

The primary outcome was at least one severe complication: refractory celiac disease, enteropathy-associated T-cell lymphoma (EATL), other lymphoproliferative malignancy or small-bowel adenocarcinoma. Secondary outcomes were all-cause mortality, fractures, and osteoporosis or bone mineral density (BMD) measured after the follow-up biopsy. We included prospective and retrospective cohorts, registry cohorts, nested case-control studies and post-hoc analyses of cohorts with comparative data. Cohorts that mixed children and adults were included, and adult-only estimates were extracted when reported. Conference abstracts were eligible if they reported outcomes separately for PVA and recovery. We excluded case reports or series of fewer than 10 patients, reviews and editorials, studies without a clear follow-up biopsy result, studies that selected patients with refractory celiac disease (making it the exposure rather than the outcome), and paediatric-only studies. Studies that measured BMD at the time of the biopsy were described separately as cross-sectional evidence.

Search strategy

We searched PubMed (MEDLINE), Embase, Cochrane CENTRAL, Web of Science Core Collection and Scopus from January 2019 to 14 October 2026, with no language restriction. The start date followed the previous systematic review by Szakács et al. [13], whose search ended in September 2019. Rather than repeat that search, we screened its reference list. We limited the review to published reports and did not contact authors. The full search strategy is given in Additional file 1.

Study selection

We managed records in Covidence. Seven reviewers screened titles and abstracts, and two independent reviewers then screened full texts. A third reviewer resolved conflicts. We piloted the criteria on the first 50 records. We removed duplicates both automatically in Covidence and by hand.

Data extraction

Two reviewers extracted data into a prespecified form based on the CHARMS-PF framework [19], and a further reviewer checked each extraction. Disagreements were resolved by discussion. We extracted study descriptors (report type, country, setting, data source or cohort name, design, recruitment period), participant characteristics, definitions of PVA and recovery, timing of the follow-up biopsy, GFD adherence, and outcomes. Where a study reported mucosal recovery compared with PVA, we took the reciprocal of the published hazard ratio (HR) and its confidence limits, so that every estimate compares PVA with recovery. We recorded whether follow-up was counted from the follow-up biopsy or from diagnosis.

Risk of bias assessment

We assessed each study with the Quality in Prognosis Studies (QUIPS) tool [20]. It has six domains: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting. Each domain was rated low, moderate or high risk of bias. Because follow-up biopsy is selective and its timing varies, the protocol added three review-specific questions under confounding and analysis: whether follow-up was counted from the date of re-biopsy, whether outcomes in the first 6 to 12 months were excluded or examined in a lag analysis, and whether age, sex and symptoms at diagnosis were accounted for. The overall rating was high if any domain was high, moderate if no domain was high and at least one was moderate, and low if all domains were low. Two reviewers rated each study, and two further reviewers checked the ratings. We drew the risk-of-bias plots with robvis [21].

Data synthesis

Several reports used the same patients. We assigned each report to an overlap group by data source, recruitment period and any information on shared patients. The four Swedish ESPRESSO reports (Lebwohl 2013a, Lebwohl 2013b, Lebwohl 2014 and Emilsson 2020) formed one group [9–12]. Kaukinen 2007 and Pekki 2015 formed the Tampere group [22, 23]. The other reports each formed their own group: Valdimarsson 1994 (Linköping) [24], Rubio-Tapia 2010 (Mayo Clinic) [4], Schiepatti 2023 (Sheffield, Pavia and Boston) [25] and Saidani 2024 (Tunis) [26]. Reports in the same group were not counted as independent confirmation of one another.

We grouped studies by outcome and described the direction and size of each association in a structured narrative synthesis that follows the SWiM guideline [27]. The protocol required at least three independent cohorts with comparable estimates before any quantitative combination. No outcome met this threshold. Mortality came closest, with adjusted estimates from three independent cohorts, but these differed in the ages included, the timing of the biopsy and the variables adjusted for, so we did not pool them. This differs from the registered protocol, which also planned leave-one-out and subgroup analyses by definition of recovery, timing of biopsy, age group and risk of bias. Without a pooled estimate these analyses had nothing to test, so we did not run them. We report lagged and time-stratified results where the included studies gave them (Additional file 1: Table S1).

Certainty of evidence

For each outcome we rated the certainty of evidence with GRADE for prognostic factors [28]. We considered risk of bias, inconsistency, indirectness, imprecision and publication bias, and the possibility of rating up. We did not assess publication bias because fewer than 10 studies contributed to any outcome.

Results

Study selection

The database searches retrieved 3238 records, and the reference list of the earlier review [13] added 53. After 1575 duplicates were removed, 1716 titles and abstracts were screened and 23 full texts were assessed (Fig. 1). Ten reports from six independent cohorts were included, and the reasons for excluding the other 13 are given in Fig. 1.

Study characteristics

Table 1 describes the 10 reports. Four came from one Swedish registry [9–12] and two from one Finnish centre [22, 23]. Valdimarsson 1994 [24], Rubio-Tapia 2010 [4], Schiepatti 2023 [25] and Saidani 2024 [26] each came from a separate cohort, and Saidani 2024 is a conference abstract. Definitions of PVA and the time from diagnosis to follow-up biopsy differed between studies (Table 2).

Table 1
Characteristics of the included studies

Study	Country	Study design	Population	Patients with follow-up biopsy, n	PVA / recovery, n	Age at diagnosis, years	Female, %	Follow-up, years	Risk of bias
Valdimarsson 1994 [24]	Sweden	Cross-sectional	Adults	30	13 / 17	NR	73.3	NA	High
Kaukinen 2007 [22]	Finland	Cohort with selected comparison group	Adults	31	13 / 18	NR	77.4	4 to 5	High
Rubio-Tapia 2010 [4]	USA	Retrospective cohort	Adults	241	151 / 90	47.0 (median)	73.0	NR	Moderate
Lebwohl 2013a [10]	Sweden	Retrospective cohort (national registry)	All ages	7625	3308 / 4317	25.0 (median)	63.0	8.9	Moderate
Lebwohl 2013b [9]	Sweden	Retrospective cohort (national registry)	All ages	7648	3317 / NR	28.4 (mean)	63.0	9.9	Moderate
Lebwohl 2014 [12]	Sweden	Retrospective cohort (national registry)	All ages	7146	3105 / 4041	23.0 (median)	64.0	8.6	Moderate
Pekki 2015 [23]	Finland	Prospective cohort	15 years and older	263	85 / 178	45.0 (median)	68.1	NR	High
Emilsson 2020 [11]	Sweden	Retrospective cohort (national registry)	All ages	9532	2787 / 6745	31.6 (mean)	62.7	NR	High
Schiepatti 2023 [25]	UK, Italy, USA	Ambispective cohort (multicentre)	Adults	694	157 / 537	44.0 (mean)	70.7	9.2	High
Saidani 2024 [26]	Tunisia	Retrospective cohort	Mixed ages	151	37 / 114	NR	NR	NR	High

NA = not applicable; NR = not reported; PVA = persistent villous atrophy; QUIPS = Quality in Prognosis Studies tool (overall risk of bias). Lebwohl 2013a, Lebwohl 2013b, Lebwohl 2014 and Emilsson 2020 share one Swedish registry; Kaukinen 2007 and Pekki 2015 share one centre. Median from diagnosis. Conference abstract. Calculated from the reported percentage.

Table 2
Histological definitions and timing of the follow-up biopsy

Study	Classification	Definition of PVA	Definition of recovery	Time from diagnosis to biopsy	Reason for biopsy	Assessment of diet adherence	Good adherence, PVA vs recovery, %
Valdimarsson 1994 [24]	Alexander	Grade III–IV for at least 4 years	Normal mucosa for at least 4 years	NR	NR	Combined methods	7.7 vs 100
Kaukinen 2007 [22]	Vh:Cd	No recovery within 2 years on a strict GFD (Vh:Cd < 2.0)	Good histological recovery (threshold NR)	NR	Mixed	Dietitian	100 vs 100
Rubio-Tapia 2010 [4]	Vh:Cd	Vh:Cd below 3:1	Vh:Cd 3:1 or higher	Within 2 years in 68.0%	NR	Dietitian	61.0 vs 75.0
Lebwohl 2013a [10]	Marsh	Marsh 3	Marsh 0–2	Median 1.3 years	Mostly routine	Not assessed	NR
Lebwohl 2013b [9]	Marsh	Marsh 3	Marsh 0–2	Median 1.3 years	Mostly routine	Not assessed	NR
Lebwohl 2014 [12]	Marsh	Marsh 3	Marsh 0–2	Within 2 years in 71.2%	Mostly routine	Not assessed	NR
Pekki 2015 [23]	Vh:Cd	Vh:Cd < 2.0	Vh:Cd > 2.0	1 year (by design)	Routine	Dietitian	87.0 vs 97.0
Emilsson 2020 [11]	Marsh	Marsh 3	Marsh 0–2	6 months to 5 years	NR	Not assessed	NR
Schiepatti 2023 [25]	Marsh	Marsh 3a or above	No villous atrophy	Median 2.7 years	Clinically driven	Combined methods	38.2 vs 96.8
Saidani 2024 [26]	Marsh	Marsh 3a or above	NR	NR	NR	NR	NR

GFD = gluten-free diet; NR = not reported; PVA = persistent villous atrophy; Vh:Cd = villous height to crypt depth ratio.

Risk of bias

No report had low overall risk of bias (Additional file 1: Figs. S1 and S2). Four were rated moderate [4, 9, 10, 12] and six high [11, 22–26].

Outcomes

Table 3 gives every estimate by outcome with its certainty, and hazard ratios compare PVA with recovery. Certainty was low for lymphoproliferative malignancy and any fracture and very low for all other outcomes.

Table 3
Outcomes in patients with persistent villous atrophy compared with mucosal recovery

Study	Events, PVA vs recovery	Adjusted HR (95% CI)	Adjusted for	Start of follow-up
Severe complications ●○○○				
Schiepatti 2023 [25]	33/157 vs NR	9.53 (4.77–19.04)	Age, sex, presentation, centre	NR
Saidani 2024 [26]	7 vs 2	7.61 (3.18–17.64)	None (unadjusted)	NR
Refractory celiac disease ●○○○				
Kaukinen 2007 [22]	2/13 vs 0/18	NR	None	Study visit
Pekki 2015 [23]	1/85 vs NR	NR	None	NR
Schiepatti 2023 [25]	34 in total	NR	None	NR
Saidani 2024 [26]	3 in total	NR	None	NR
Lymphoproliferative malignancy ●●○○				
Lebwohl 2013a [10]	41/3308 vs 12/4317	2.26 (1.18–4.34)	Age, sex, year of biopsy, education, disease duration	Follow-up biopsy
Pekki 2015 [23]	1/71 vs 2/134	NR	None	NR
T-cell lymphoma ●○○○				
Lebwohl 2013a [10]	10 vs 2	3.51 (0.75–16.34)	Age, sex, year of biopsy, education, disease duration	Follow-up biopsy
Kaukinen 2007 [22]	1/13 vs 0/18	NR	None	Study visit
Schiepatti 2023 [25]	4 in total	NR	None	NR
B-cell lymphoma ●○○○				
Lebwohl 2013a [10]	4 vs 3	0.97 (0.21–4.49)	Age, sex, year of biopsy, education, disease duration	Follow-up biopsy
Schiepatti 2023 [25]	1 in total	NR	None	NR
Saidani 2024 [26]	1 in total	NR	None	NR
Small-bowel adenocarcinoma ●○○○				
Emilsson 2020 [11]	5/2787 vs 1/6745	5.56 (0.62–50.0)□	Age, sex, time between biopsies, education	Follow-up biopsy
Schiepatti 2023 [25]	1 in total	NR	None	NR
Any malignancy ●○○○				
Kaukinen 2007 [22]	3/13 vs 0/18	NR	None	Study visit
Pekki 2015 [23]	8/71 vs 8/134	NR	None	NR
All-cause mortality ●○○○				
Lebwohl 2013b [9]	606 in total	1.01 (0.86–1.19)	Age, sex, calendar period, education, disease duration	Follow-up biopsy
Rubio-Tapia 2010 [4]	10 vs 1	7.69 (0.94–50.0)□	Age, sex	Diagnosis
Schiepatti 2023 [25]	32 in total	2.93 (1.43–6.02)	Age, sex, presentation, centre	NR
Kaukinen 2007 [22]	2/13 vs 1/18	NR	None	Study visit

CI = confidence interval; GRADE = Grading of Recommendations Assessment, Development and Evaluation; HR = hazard ratio; NR = not reported; PVA = persistent villous atrophy. Hazard ratios compare PVA with recovery. Certainty of evidence is the GRADE rating: ●●○○ low, ●○○○ very low. □ Reciprocal of the published hazard ratio for recovery compared with PVA.

Study	Events, PVA vs recovery	Adjusted HR (95% CI)	Adjusted for	Start of follow-up
Severe complications ●○○○				
Pekki 2015 [23]	10/71 vs 12/134	NR	None	NR
Any fracture ●●○○				
Lebwohl 2014 [12]	483/3105 vs 492/4041	0.93 (0.82–1.06)	Age, sex, calendar period, education, disease duration	Follow-up biopsy
Likely osteoporotic fracture ●○○○				
Lebwohl 2014 [12]	96 vs 124	1.11 (0.84–1.46)	Age, sex, calendar period, education, disease duration	Follow-up biopsy
Hip fracture ●○○○				
Lebwohl 2014 [12]	63/3105 vs 26/4041	1.67 (1.05–2.66)	Age, sex, calendar period, education, disease duration	Follow-up biopsy
Osteoporosis or low bone mineral density ●○○○				
Kaukinen 2007 [22]	7/12 vs 4/18	NR	None	Study visit
Pekki 2015 [23]	8/71 vs 16/134	NR	None	NR
Valdimarsson 1994 [24]	7/13 vs NR (low BMD)	NR	None	Study visit

CI = confidence interval; GRADE = Grading of Recommendations Assessment, Development and Evaluation; HR = hazard ratio; NR = not reported; PVA = persistent villous atrophy. Hazard ratios compare PVA with recovery. Certainty of evidence is the GRADE rating: ●●○○ low, ●○○○ very low. Reciprocal of the published hazard ratio for recovery compared with PVA.

Severe complications. The multicentre cohort found a higher hazard of complications with PVA (adjusted HR 9.53, 95% CI 4.77 to 19.04), with events in 33 of 157 patients with PVA [25]. The Saidani 2024 abstract reported an unadjusted HR of 7.61 (3.18 to 17.64) [26]. Four reports gave counts of refractory celiac disease alone, and none gave an effect estimate [22, 23, 25, 26].

Malignancy. In the Swedish registry, lymphoproliferative malignancy was more frequent with PVA (adjusted HR 2.26, 1.18 to 4.34), and the estimate was similar when follow-up started 1 to 12 months after the biopsy (Additional file 1: Table S1) [10]. The HRs were 3.51 (0.75 to 16.34) for T-cell lymphoma and 0.97 (0.21 to 4.49) for B-cell lymphoma [10]. Small-bowel adenocarcinoma occurred in 5 of 2787 patients with PVA and 1 of 6745 with recovery (HR 5.56, 0.62 to 50.0) [11]. Two small Finnish reports gave counts of malignancy without estimates [22, 23].

Mortality. Estimates differed between cohorts: HR 1.01 (0.86 to 1.19) in the Swedish registry [9], 7.69 (0.94 to 50.0) at the Mayo Clinic [4] and 2.93 (1.43 to 6.02) in the multicentre cohort [25]. The Swedish estimate was unchanged in adults only and in time-stratified analyses (Additional file 1: Table S1) [9].

Fractures and bone density. In the Swedish registry, PVA was not associated with any fracture (HR 0.93, 0.82 to 1.06) or likely osteoporotic fracture (1.11, 0.84 to 1.46), but it was associated with hip fracture (1.67, 1.05 to 2.66), and the hip fracture signal was clear only more than 5 years after the biopsy (Additional file 1: Table S1) [12]. Of three small studies, two reported lower bone density or more osteoporosis with PVA [22, 24] and one found no difference [23]. None gave an adjusted estimate, and bone density was measured around the time of biopsy in some of them.

Follow-up biopsy compared with no follow-up biopsy. In the Swedish registry, mortality was lower after a follow-up biopsy (HR 0.88, 0.80 to 0.96), but not when deaths in the first 6 months were excluded (HR 0.97, 0.89 to 1.06) [9]. In the multicentre cohort, survival was no longer higher after adjustment (HR 0.70, 0.44 to 1.12), and complications were far more frequent in patients who had a follow-up biopsy [25]. We did not grade these comparisons.

Discussion

Adults whose duodenal biopsy still showed villous atrophy on a gluten-free diet did not clearly do worse than those whose mucosa had recovered, with two exceptions. Lymphoproliferative malignancy was more frequent with persistent atrophy in the Swedish registry (HR 2.26) [10], and a later report from the same cohort agreed (HR 1.87, 1.11 to 3.16) [29]. Hip fracture was also more frequent (HR 1.67), while fractures overall were not (HR 0.93) [12]. Certainty was low at best.

Mortality is where the studies disagree. The two Swedish registry reports found no difference (HR 1.01 and 0.99) [9, 30], and the multicentre referral cohort found almost a threefold higher risk (HR 2.93) [25]. In that cohort, patients who had a follow-up biopsy were older at diagnosis and developed far more complications than those who did not [25]. Selection into repeat biopsy may therefore explain the gap better than biology does.

The high complication estimates come mostly from refractory celiac disease. In the multicentre cohort, 34 of 44 complications were refractory disease [25], and its definition requires persistent villous atrophy [7, 25]. Part of that association therefore follows from how the outcome is defined. Saidani 2024 reported a similar signal (HR 7.61), but its estimate was unadjusted [26].

The lymphoma signal may also run the other way. In one series, 26 of 50 patients with refractory type II disease developed lymphoma [31], so atrophy found shortly before such a diagnosis may be an early sign of the lymphoma, not its cause. Time-stratified analyses are the right test and were reported for only some outcomes [10, 12]. The hip-only fracture signal fits the explanation given by its authors: thinner subcutaneous tissue and falls, not bone loss [12].

This matters for practice. The 2025 European guideline states that repeat biopsy is not routinely necessary for all adults, advises a personalised approach, for example after diagnosis above 45 years or with a severe first presentation, and rates the evidence low [7]. An international monitoring guideline also recommends against routine re-biopsy [6]. Our results support this position. A biopsy showing persistent atrophy does not yet justify changing management on its own, because we cannot say that it causes harm.

To our knowledge, this is the first systematic review since Szakács and colleagues [13] to examine follow-up histology as a prognostic factor in adults. That review included children and conference abstracts, used quality indicators adapted from QUIPS, and did not report a GRADE assessment [13]. We restricted the review to adults and long-term outcomes, counted each cohort once, rated risk of bias by QUIPS domain, graded certainty with GRADE for prognostic factors, and added the evidence published after 2019.

Strengths and limitations. We followed a registered protocol, searched five databases, and handled overlapping cohorts explicitly. Six cohorts is a small base, and four of the ten reports share one registry, so their agreement is not independent confirmation. Studies defined persistent atrophy differently and re-biopsied at different times, and dietary adherence was rarely measured objectively. Repeat biopsies are done in a minority of patients, who are not representative [8, 25]. One included study is a conference abstract with limited detail, which we rated as high risk of bias [26]. We did not pool results because definitions and adjustment sets differed too much.

Conclusions

Persistent villous atrophy on follow-up biopsy was linked to lymphoproliferative malignancy and hip fracture, but not consistently to mortality or to other cancers. Much of the complication signal comes from refractory disease and from selection into repeat biopsy, so it may not be causal. Certainty is low or very low. The evidence supports targeted, not routine, repeat biopsy. Prospective cohorts with prespecified biopsy timing and objective diet monitoring are needed.

Declarations

Ethics approval and consent to participate:

Not applicable. This is a systematic review of published data.

Consent for publication:

Not applicable.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This work received no specific funding.

Author Contribution

M.H.E.E. conceived the idea, designed the protocol, ran the searches, divided the screening among reviewers, prepared the tables and figures, and finalised and revised the manuscript. The review team screened titles, abstracts and full texts, and M.M. resolved conflicts. I.M.E. and A.K. extracted the data, and A.A.H., M.M.E. and K.A.K.R. checked the extraction. I.M.E. and A.K. made the first risk-of-bias assessments, and M.M.E. and K.A.K.R. revised them. A.A.H. wrote the Background. A.J.A. wrote the Methods and the first half of the Results. R.A.H. wrote the Results by outcome and the Discussion. All authors reviewed the manuscript.

Acknowledgements:

None.

Data Availability

All data were taken from the published reports listed in the references. The search strategies and risk-of-bias plots are in Additional file 1. The extracted data are available from the corresponding author on reasonable request.

References

1. Rubio-Tapia A, Hill ID, Semrad C, Kelly CP, Greer KB, Limketkai BN, et al. American College of Gastroenterology guidelines update: diagnosis and management of celiac disease. Am J Gastroenterol. 2023;118:59–76. https://doi.org/10.14309/ajg.0000000000002075.
2. Al-Toma A, Zingone F, Branchi F, Schiepatti A, Malamut G, Canova C, et al. European Society for the Study of Coeliac Disease (ESsCD) 2025 updated guidelines on the diagnosis and management of coeliac disease in adults. Part 1: diagnostic approach. United Eur Gastroenterol J. 2025;13:1855–86. https://doi.org/10.1002/ueg2.70119.
3. Balaban DV, Enache I, Balaban M, David RA, Vasile AD, Popp A, et al. Outcomes in adults with celiac disease following a gluten-free diet. J Clin Med. 2025;14:5144. https://doi.org/10.3390/jcm14145144.
4. Rubio-Tapia A, Rahim MW, See JA, Lahr BD, Wu TT, Murray JA. Mucosal recovery and mortality in adults with celiac disease after treatment with a gluten-free diet. Am J Gastroenterol. 2010;105:1412–20. https://doi.org/10.1038/ajg.2010.10.
5. Fernández-Bañares F, Beltrán B, Salas A, Comino I, Ballester-Clau R, Ferrer C, et al. Persistent villous atrophy in de novo adult patients with celiac disease and strict control of gluten-free diet adherence: a multicenter prospective study (CADER study). Am J Gastroenterol. 2021;116:1036–43. https://doi.org/10.14309/ajg.0000000000001139.
6. Elli L, Leffler D, Cellier C, Lebwohl B, Ciacci C, Schumann M, et al. Guidelines for best practices in monitoring established coeliac disease in adult patients. Nat Rev Gastroenterol Hepatol. 2024;21:198–215. https://doi.org/10.1038/s41575-023-00872-2.
7. Al-Toma A, Branchi F, Zingone F, Schiepatti A, Malamut G, Canova C, et al. European Society for the Study of Coeliac Disease (ESsCD) 2025 updated guidelines on the diagnosis and management of coeliac disease in adults. Part 2: management, follow-up, and complex disease courses. United Eur Gastroenterol J. 2026;14:e70195. https://doi.org/10.1002/ueg2.70195.
8. Venkat MV, Chen L, Wright JD, Lebwohl B. Prevalence and predictors of follow-up endoscopic biopsy in patients with celiac disease in the United States. J Clin Gastroenterol. 2025;59:155–60. https://doi.org/10.1097/MCG.0000000000001995.
9. Lebwohl B, Granath F, Ekbom A, Montgomery SM, Murray JA, Rubio-Tapia A, et al. Mucosal healing and mortality in coeliac disease. Aliment Pharmacol Ther. 2013;37:332–9. https://doi.org/10.1111/apt.12164.
10. Lebwohl B, Granath F, Ekbom A, Smedby KE, Murray JA, Neugut AI, et al. Mucosal healing and risk for lymphoproliferative malignancy in celiac disease: a population-based cohort study. Ann Intern Med. 2013;159:169–75. https://doi.org/10.7326/0003-

4819-159-3-201308060-00006.

11. Emilsson L, Semrad C, Lebwohl B, Green PHR, Ludvigsson JF. Risk of small bowel adenocarcinoma, adenomas, and carcinoids in a nationwide cohort of individuals with celiac disease. Gastroenterology. 2020;159:1686–e942. https://doi.org/10.1053/j.gastro.2020.07.007.
12. Lebwohl B, Michaëlsson K, Green PHR, Ludvigsson JF. Persistent mucosal damage and risk of fracture in celiac disease. J Clin Endocrinol Metab. 2014;99:609–16. https://doi.org/10.1210/jc.2013-3164.
13. Szakács Z, Gede N, Gyöngyi Z, Solymár M, Csupor D, Erőss B, et al. A call for research on the prognostic role of follow-up histology in celiac disease: a systematic review. Front Physiol. 2019;10:1408. https://doi.org/10.3389/fphys.2019.01408.
14. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71.
15. Amir-Behghadami M, Janati A, Population. Intervention, Comparison, Outcomes and Study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. Emerg Med J. 2020;37:387. https://doi.org/10.1136/emermed-2020-209567.
16. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). Gastroenterology. 1992;102:330–54.
17. Oberhuber G, Granditsch G, Vogelsang H. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. Eur J Gastroenterol Hepatol. 1999;11:1185–94. https://doi.org/10.1097/00042737-199910000-00019.
18. Corazza GR, Villanacci V, Zambelli C, Milione M, Luinetti O, Vindigni C, et al. Comparison of the interobserver reproducibility with different histologic criteria used in celiac disease. Clin Gastroenterol Hepatol. 2007;5:838–43. https://doi.org/10.1016/j.cgh.2007.03.019.
19. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ. 2019;364:k4597. https://doi.org/10.1136/bmj.k4597.
20. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158:280–6. https://doi.org/10.7326/0003-4819-158-4-201302190-00009.
21. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods. 2021;12:55–61. https://doi.org/10.1002/jrsm.1411.
22. Kaukinen K, Peräaho M, Lindfors K, Partanen J, Woolley N, Pikkarainen P, et al. Persistent small bowel mucosal villous atrophy without symptoms in coeliac disease. Aliment Pharmacol Ther. 2007;25:1237–45. https://doi.org/10.1111/j.1365-2036.2007.03311.x.
23. Pekki H, Kurppa K, Mäki M, Huhtala H, Sievänen H, Laurila K, et al. Predictors and significance of incomplete mucosal recovery in celiac disease after 1 year on a gluten-free diet. Am J Gastroenterol. 2015;110:1078–85. https://doi.org/10.1038/ajg.2015.155.
24. Valdimarsson T, Toss G, Ross I, Löfman O, Ström M. Bone mineral density in coeliac disease. Scand J Gastroenterol. 1994;29:457–61. https://doi.org/10.3109/00365529409096838.
25. Schiepatti A, Maimaris S, Raju SA, Green OL, Mantica G, Therrien A, et al. Persistent villous atrophy predicts development of complications and mortality in adult patients with coeliac disease: a multicentre longitudinal cohort study and development of a score to identify high-risk patients. Gut. 2023;72:2095–102. https://doi.org/10.1136/gutjnl-2023-329751.
26. Saidani R, Dabbebi H, Cherif D, Yacoub H, Hassine H, Kchir H, et al. Persistent villous atrophy in celiac disease: risk stratification and long-term outcomes [abstract 248]. United Eur Gastroenterol J. 2024;12:231. https://doi.org/10.1002/ueg2.12614.
27. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020;368:l6890. https://doi.org/10.1136/bmj.l6890.
28. Foroutan F, Guyatt G, Zuk V, Vandvik PO, Alba AC, Mustafa R, et al. GRADE Guidelines 28: use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. J Clin Epidemiol. 2020;121:62–70. https://doi.org/10.1016/j.jclinepi.2019.12.023.
29. Lebwohl B, Green PHR, Emilsson L, Mårild K, Söderling J, Roelstraete B, et al. Cancer risk in 47,241 individuals with celiac disease: a nationwide cohort study. Clin Gastroenterol Hepatol. 2022;20:e111–31. https://doi.org/10.1016/j.cgh.2021.05.034.
30. Lebwohl B, Green PHR, Söderling J, Roelstraete B, Ludvigsson JF. Association between celiac disease and mortality risk in a Swedish population. JAMA. 2020;323:1277–85. https://doi.org/10.1001/jama.2020.1943.
31. Al-Toma A, Verbeek WH, Hadithi M, von Blomberg BM, Mulder CJ. Survival in refractory coeliac disease and enteropathy-associated T-cell lymphoma: retrospective evaluation of single-centre experience. Gut. 2007;56:1373–8.

https://doi.org/10.1136/gut.2006.114512.

Figures

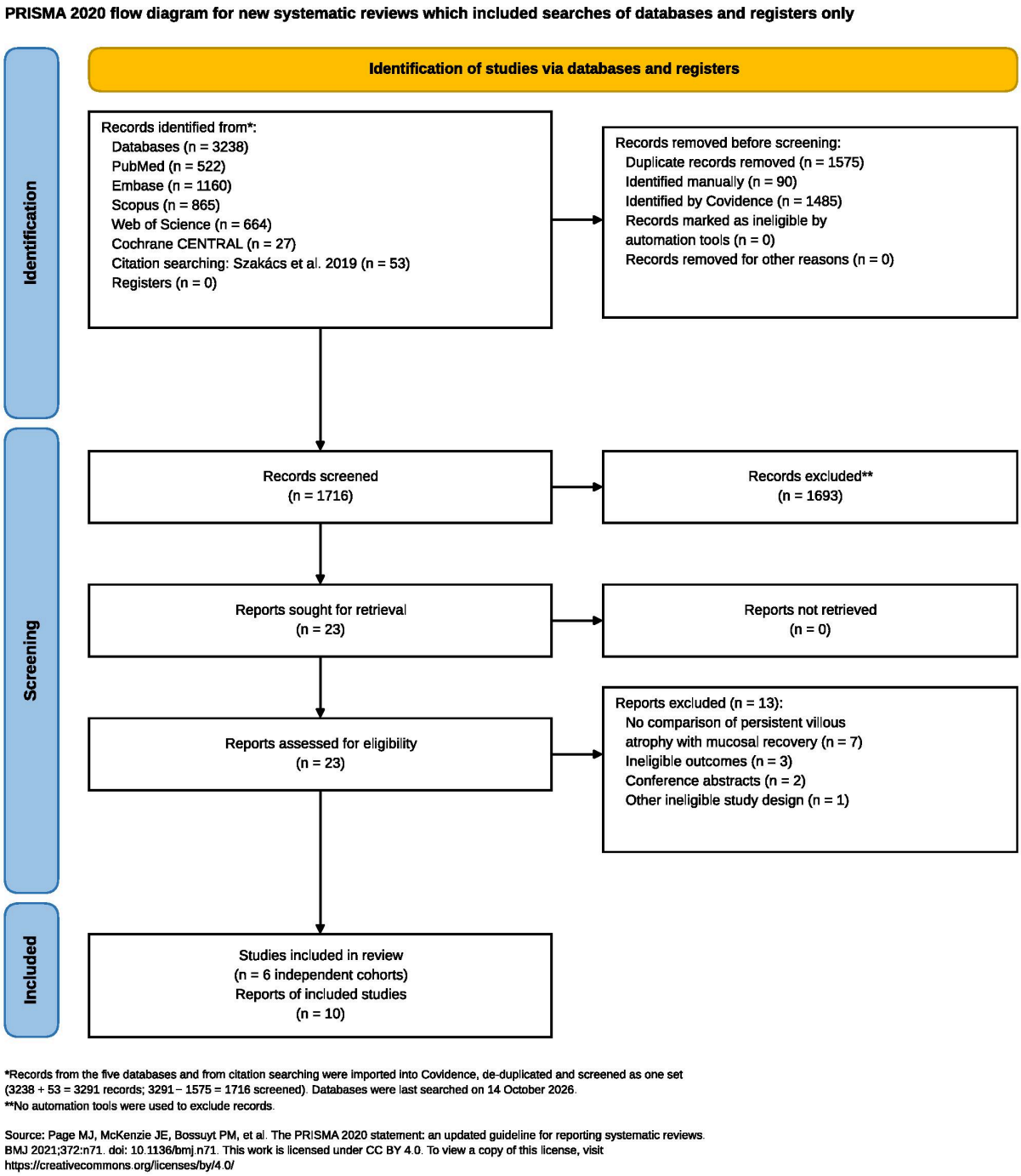

Figure 1

PRISMA 2020 flow diagram of study selection. Records from five databases (n = 3238) and from the reference list of the previous review (n = 53) were de-duplicated and screened together. Ten reports from six independent cohorts were included.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile2.docx
- Additionalfile1.docx