Sainsbury K, Marques MM.

The relationship between gluten free diet adherence and depressive symptoms in adults with coeliac disease: A systematic review with meta-analysis.

***Appetite* 2017**

DOI: https://doi.org/10.1016/j.appet.2017.10.017

Accepted Manuscript

The relationship between gluten free diet adherence and depressive symptoms in adults with coeliac disease: A systematic review with meta-analysis

Kirby Sainsbury, Marta M. Marques

PII: S0195-6663(17)31023-1

DOI: 10.1016/j.appet.2017.10.017

Reference: APPET 3649

To appear in: *Appetite*

Received Date: 15 July 2017

Revised Date: 11 October 2017

Accepted Date: 11 October 2017

Please cite this article as: Sainsbury K. & Marques M.M., The relationship between gluten free diet adherence and depressive symptoms in adults with coeliac disease: A systematic review with meta-analysis, *Appetite* (2017), doi: 10.1016/j.appet.2017.10.017.

Title: The relationship between gluten free diet adherence and depressive symptoms in adults with treated coeliac disease: A systematic review with meta-analysis

Short title: Meta-analysis of GFD adherence and depressive symptoms in CD

Authors:

Kirby Sainsbury[a*], PhD, D.Psych (Clinical)

Marta M. Marques[a,b], PhD

a. Research Associate (Health Psychology); Institute of Health and Society, Newcastle University, Baddiley-Clark Building, Richardson Rd, Newcastle Upon Tyne, NE2 4AX, UK; phone: +44 (0) 191 208 8500; ***corresponding author** email: kirby.sainsbury@ncl.ac.uk; ORCID ID: 0000-0002-6716-8305
b. Research Associate (Health Psychology); Centre for Behaviour Change, University College London, 1-19 Torrington Place, Fitzrovia, London WC1E 7HB, UK; phone: +44 (0) 207 679 1979; email: marta.marques@ucl.ac.uk

Conflicts of interest: none

Funding: none

Word count: 4890 (excluding abstract, tables, figures, references)

Title: The relationship between gluten free diet adherence and depressive symptoms in adults with coeliac disease: A systematic review with meta-analysis

ABSTRACT

Purpose: Depressive symptoms are common in patients with coeliac disease (CD) and may represent a barrier to gluten free diet (GFD) adherence. The aims of this meta-analysis were: (1) to synthesise the evidence on the relationship between depression or depressive symptoms and degree of adherence to a GFD in patients with CD who are already attempting a GFD (i.e., post-diagnosis and onset of GFD), and (2) to summarise the direction of causation of any observed relationship.

Methods: A random effects meta-analysis of 8 cross-sectional studies (N=1644) was conducted. Included studies measured self-reported depressive symptoms and GFD adherence using either a dietitian interview or validated self-report questionnaire that considered unintentional gluten consumption.

Results: There was a moderate association between poorer GFD adherence and greater depressive symptoms (r=0.398, 95% CI=0.321-0.469), with marked heterogeneity in the effects (I^2=66.8%). A sensitivity analysis excluding studies with a moderate/high (k=1) or unclear risk of bias (k=1) did not change the results.

Conclusion: The low number of studies meeting inclusion criteria limits the strength of the conclusions. Available evidence suggests there is an association between poorer GFD adherence and self-reported depressive symptoms; however, studies using longitudinal and prospective designs, and reliable measures, particularly for adherence, are needed to confirm this association. The direction of causation between depression and adherence remains unclear.

KEYWORDS

Gluten free diet adherence; depression; depressive symptoms; coeliac disease; meta-analysis

INTRODUCTION

Coeliac disease (CD) is a chronic autoimmune condition involving intolerance to dietary gluten, for which clinical management involves lifelong strict adherence to a gluten free diet (GFD; Green & Cellier, 2007). Undiagnosed or poorly managed, CD is associated with gastrointestinal and malabsorption symptoms and increased risk of long-term health complications, including intestinal cancers, osteoporosis, and infertility (Green & Cellier, 2007; Green & Jabri, 2003). Depression is also often cited as a symptom of undiagnosed CD (Jackson, Eaton, Cascella, Fasano, & Kelly, 2012), and clinically diagnosed depression and/or depressive symptoms (typically collapsed in reviews) appear to occur with greater frequency and/or severity in CD than healthy samples (Smith & Gerdes, 2012; Zingone et al., 2015).

Meta-analyses conducted in other chronic illnesses (e.g., diabetes) have found significant associations between depression and non-adherence to medication and other treatment components (e.g., diet and physical activity recommendations; DiMatteo, Lepper, & Croghan, 2000; Gonzalez et al., 2008; Grenard et al., 2011), with depressed patients being 1.76-to-3 times less likely to adhere to medical treatment recommendations than non-depressed patients (DiMatteo et al., 2000; Grenard et al., 2011). Further, reduced treatment adherence is one mechanism via which the link between depression and many preventable chronic illnesses may be explained (Katon, 2011). That is, depression may act as a barrier to good self-care (e.g., resulting in poor diet and physical inactivity – or in the case of CD, poor management of the GFD) via deficits in energy and memory, which leads to the development of risk factors such as obesity, which prompt or exacerbate the symptoms of chronic illness (e.g., diabetes), and, in turn, become further barriers to good adherence (Katon, 2011).

The literature on a comparable relationship between depression and GFD adherence in CD patients has yet to be synthesised, with existing reviews on the incidence of depression

in CD (Smith & Gerdes, 2012; Zingone et al., 2015) being methodologically unable to answer the more specific question regarding GFD adherence. Firstly, evidence for the depression-adherence association comes primarily from studies assessing differences in depression between newly-diagnosed patients and those already being managed on a GFD, without assessment of the adequacy of dietary adherence in the established gluten free patients. Secondly, amongst studies that have specifically measured GFD adherence, conclusions have been drawn without due consideration of the impact of unreliable measurement of GFD adherence. Finally, the absence of a healthy control group (inclusion criteria for both previous reviews) meant that many studies relevant to answering this more specific question were excluded.

Debate exists on the optimal way to measure GFD adherence (Leffler et al., 2007; Ludvigsson et al., 2014; Vahedi et al., 2003), resulting in large variation in definitions and measurement across studies (Hall, Rubin, & Charnock, 2009). Intentional gluten consumption in patients with CD appears rare, with unintentional non-adherence (e.g., due to cross contamination or errors in label reading) representing the most common reason for lapsing from the GFD (Hall, Rubin, & Charnock, 2013; Sainsbury et al., 2013a). Commonly used adherence measures, such as single-item self-report questions (e.g., 'how strictly do you adhere to your GFD?' with Likert or visual analogue response scales from 'not at all' to 'very strictly') and serological analyses, are unreliable at detecting incomplete adherence, particularly with increased time on a GFD (Leffler et al., 2007). These methods also do not correlate well with dietitian-rated assessments (Fera, Cascio, Angelini, Martini, & Guidetti, 2003; Leffler et al., 2007; Vahedi et al., 2003), the method currently deemed the 'gold standard' (Leffler et al., 2007; Ludvigsson et al., 2014). The dietitian assessment involves completion of a 3-day food record (prior to the session), a food ingredient quiz, and a dynamic clinical interview in which an experienced dietitian evaluates the food record with

the patient to identify any gluten consumption or sources of cross-contamination that may compromise adherence. Regarding simple self-report measures, the discrepancy with dietitian assessments probably results from their failure to consider unintentional gluten consumption, which, by definition, occurs outside of conscious awareness, as well as inaccuracies in patient understanding and knowledge of the GFD (Leffler et al., 2007; Silvester, Weiten, Graff, Walker, & Duerksen, 2016). Serological results adequately indicate gluten-related damage at diagnosis; however, once on a GFD, produce frequent false negative results in known partially adherent individuals (Leffler et al., 2007; Vahedi et al., 2003).

To fill the gap in the availability of valid and reliable tools for assessing GFD adherence and provide a feasible measure within the research context, several questionnaires that do account for unintentional gluten exposure have been designed. These include the Coeliac Dietary Adherence Test (CDAT; Leffler et al., 2009), which was developed in consultation with an expert panel, and has demonstrated psychometric properties. The CDAT has acceptable sensitivity and specificity when compared against a dietitian assessment, and is superior to serological analysis in predicting dietitian-rated adherence categories (Leffler et al., 2009). The Biagi GFD score (Biagi et al., 2009) and the Morisky Medication Adherence Scale (Morisky, Ang, Krousel-Wood, & Ward, 2008; Morisky, Green, & Levine, 1986), adapted to GFD adherence (Casellas et al., 2008) have also been proposed and undergone some psychometric evaluation, although neither have been validated against the gold standard. While measures that consider unintentional gluten exposure are an advancement over single-items that rely on accurate patient recall, truly reliable assessment of GFD adherence is difficult and remains a challenge in both research and clinical practice.

Current guidelines on the management of CD (Ludvigsson et al., 2014; NICE, 2015) and other chronic physical health problems (NICE, 2009) specify that mild-to-severe depression and subclinical depressive symptoms should be recognised and treated. A

potential association between depression and GFD adherence in CD would therefore have important implications for the assessment and treatment of patients, including the goals of optimising adherence and both physical and psychological wellbeing. In this context, inadequate adherence may contribute to the development of depressive symptoms via physiological mechanisms (e.g., malabsorption of nutrients; Hallert, Astrom, & Sedvall, 1982; Hallert, Astrom, & Walan, 1983; Hallert, Svensson, Tholstrup, & Hultberg, 2009). Conversely, the presence of depressive symptoms may limit an individual's ability to achieve and maintain adequate adherence. Building on previous research in CD and informed by that in other chronic illnesses, the primary aim of this meta-analysis was to synthesise the available evidence on the relationship between reliably-measured GFD adherence (that is, where an attempt was made to assess unintentional gluten consumption) and either depression or depressive symptoms in adults with CD. The secondary aim was to summarise the available information on the direction of causation of any observed relationship.

METHODS

Eligibility criteria

Studies were eligible for inclusion if they: (1) assessed the degree of adherence to a GFD using a method that considered unintentional gluten consumption, in adults (≥18 years) with treated CD (i.e., post-diagnosis and attempting a GFD; no specific criteria regarding duration of GFD was imposed); (2) used a validated symptom rating scale to assess depressive symptoms; or used a diagnostic interview or other valid method to assess clinically diagnosable depression; and (3) statistically reported the relationship between GFD adherence and depressive symptoms/depression. All group-based study designs (e.g., cross-sectional, prospective, interventions) were eligible; inclusion was limited to studies published in English. Full-text published studies and those ahead of publication were eligible;

conference abstracts and dissertations were eligible if a full-text paper had not been published. Studies were excluded if they: (1) assumed strict adherence solely based on GFD duration; or (2) only included newly-diagnosed CD patients measured at baseline, as the degree of adherence cannot be assessed prior to GFD onset. Prospective studies of newly-diagnosed patients were, however, eligible if, in addition to the baseline measurement, adherence and depression/depressive symptoms were measured again post-diagnosis/onset of the GFD.

Based on current evidence for the adequacy of measurement of GFD adherence and need to capture unintentional consumption, studies that used the CDAT, the Biagi GFD score, the adapted Morisky scale, a dietitian-rated assessment (the gold standard), or a combination were eligible for inclusion. Studies that used an unreliable measure, such as a visual analogue or Likert scale assessing the perceived strictness of adherence or the frequency of (typically only intentional) gluten consumption, were excluded.

Electronic searches

Comprehensive systematic searches were conducted by the first author (KS) in February 2016 in PsycINFO, Web of Science, PubMed, Cinahl, and the Cochrane Library using the following search strategy: (coeliac disease OR celiac disease OR gluten free diet) AND (depression OR depressive symptoms). The reference lists of included papers and previous reviews (Smith & Gerdes, 2012; Zingone et al., 2015) were additionally hand-searched for any relevant papers (KS).

Study selection

Following the manual removal of duplicates using Endnote (version X7.7; Clarivate Analytics, 2016), the first author (KS) completed the title screening. Articles retained at the abstract and full-text screening stages were independently double-screened and categorised as eligible/ineligible by both authors using a pre-specified eligibility form (which included

information pertaining to each of the eligibility criteria: adequate measurement of GFD adherence and depressive symptoms/clinical depression in adults with diagnosis of CD, post-diagnosis and attempting a GFD; statistically reported the relationship between depressive symptoms/depression and GFD adherence). Any disagreements were resolved through discussion between the two authors (KS, MM) until consensus was reached.

Risk of bias

Risk of bias was assessed for each study using accepted tools for cross-sectional study designs (National Institutes of Health, 2014; The Joanna Briggs Institute, 2016; Thomas, Ciliska, Dobbins, & Micucci, 2004). Items were adapted to the current question using pre-decided criteria, and updated, as needed, after the two assessors (KS, MM) had independently rated and discussed the first three papers (alphabetically) until agreement was reached. The items assessed: (1) clearly defined study population and inclusion criteria; (2) the objective/standardised measurement of CD; (3) representativeness; (4) sample size; and (5) appropriate statistical analysis of the bivariate relationship between depressive symptoms and GFD adherence.

Items assessing the validity/reliability of the outcome (GFD adherence) and exposure (depressive symptoms/depression) measures were not included in the quality assessment, as these formed part of the eligibility criteria; all included studies were therefore rated as having a low risk of bias for this domain. Note, that due to the absence of prior information on the direction of causation, outcome and exposure could equally have been defined as depressive symptoms and GFD adherence, respectively. Each item was rated as ‘very likely’ (low risk of bias), ‘somewhat likely’ (moderate), ‘not likely’ (high), or ‘not reported’ (unclear; see Supplementary material). Overall, studies were rated as having a high, moderate, low, or unclear risk of bias.

Risk of bias assessments were conducted independently by the two authors (KS, MM) and any discrepancies were resolved through discussion. Cohen's kappa was used to calculate inter-rater agreement.

Data extraction

Extracted data included: (1) depressive symptoms/depression and GFD adherence measures; descriptive statistics for each; (2) sample size and characteristics of CD patients (e.g., demographics and GFD duration); (3) study design; (4) statistical estimate of the depressive symptoms/depression-GFD adherence relationship. If this relationship was not reported, or a total score rather than individual subscale scores (e.g., depressed mood subscale of the Psychological General Wellbeing Index) was reported, authors were contacted via email and sent a reminder after two weeks, if they had not responded. Data extraction was completed independently by both authors using a pre-specified form. Any disagreements were resolved through consensus discussion.

Data synthesis

Characteristics and findings of all the included studies were tabulated and statistical estimates were quantitatively combined in a meta-analysis using Comprehensive Meta-Analysis (CMA; version 3; Biostat, 2016). To calculate effect sizes, the sample size and value of the bivariate correlation (Pearson's product moment correlation; *r*) was extracted from the original articles. If not reported, effect sizes were computed in CMA based on the statistical estimators reported. Fisher's *z* transformations were used to calculate the weighted average effect sizes (r_+; all calculations were automatically conducted in CMA). Meta-analyses were conducted using the random-effects model (Borenstein, Hedges, Higgins, & Rothstein, 2009), because a range of methods were used to assess both depressive symptoms and GFD adherence. These were interpreted according to Cohen's guidelines (values of r=0.10, 0.30, and 0.50 correspond to weak, moderate, and strong effects respectively; Cohen,

1988). Z-values, 95% confidence intervals, and corresponding p-values indicated the significance of the association; standard residuals were also inspected for outliers (>1.96). Separate analyses were planned for studies measuring self-reported depressive symptoms and those measuring clinical depression.

Heterogeneity in the effects was determined using: (1) Cochran's Q statistic (Cochran, 1954), for which a significant effect (<0.10 when analysis includes a small number of studies) demonstrates heterogeneity between studies; and (2) the I^2 statistic (Higgins, Thompson, Deeks, & Alman, 2003), which ranges from 0-100%, with values of 25%, 50%, and 75% reflecting low, moderate, and high heterogeneity respectively.

A sensitivity analysis, in which studies rated as having a high, moderate, or unclear risk of bias were excluded, was conducted to determine the impact of risk of bias on the pooled effect size. Publication bias was examined by visually inspecting the funnel plot for evidence of asymmetry. The meta-analysis and systematic review reported here followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, Altman, & PRISMA group, 2009). The review was registered in the PROSPERO register of systematic reviews (reference: CRD42016033711).

RESULTS

Study selection

The systematic search identified 1158 potentially relevant papers for inclusion; after duplicates were removed, 641 papers underwent screening (see Figure 1 for PRISMA flow chart). Following title and abstract screening, 84 papers were retained for full-text screening, which resulted in 8 studies (reported in 9 papers) being included in the meta-analysis (Arigo, Anskis, & Smyth, 2012; Edwards-George et al., 2009; Kerswell & Strodl, 2015; Mahadev, Gardner, Lewis, Lebwohl, & Green, 2015; Nachman et al., 2009; Nachman et al., 2010; Sainsbury, Mullan, & Sharpe, 2013a, 2013b; Weiss et al., 2013). All 8 studies measured

depressive symptoms; no studies that measured clinically diagnosable depression and met other criteria were identified. The main reasons for exclusion at the full-text screening stage (see Supplementary material) were: did not measure depressive symptoms/depression and/or GFD adherence, or did the latter using an unreliable measure (e.g., Likert or visual analogue scale). Eight of the 75 excluded papers were eligible based on the measurement of depressive symptoms/depression and GFD adherence (dietitian interview: k=6; CDAT: k=1), but the data needed to assess the relationship of interest were not reported and authors did not respond to requests for additional information (k=6) or were unable to provide the data (k=2).

Study characteristics

The characteristics of included studies are presented in Table 1. Studies were published between 2009 and 2015, and included between 53 and 390 CD patients (total N=1644; median=188). Four studies were conducted in the USA (Arigo et al., 2012; Edwards-George et al., 2009; Mahadev et al., 2015; Weiss et al., 2013), three in Australia (Kerswell & Strodl, 2015; Sainsbury et al., 2013a, 2013b), and one in Argentina (one study published in two papers; Nachman et al., 2009; Nachman et al., 2010). Females were over-represented in all samples (median=83.4% female; range=76.6-100%). The mean age of included participants ranged from 39-57 years (reported for 6 studies); one study reported the median age (47 years; Mahadev et al., 2015), and one did not report age (Nachman et al., 2009; Nachman et al., 2010). When reported, participants were highly educated (47-93% had completed college/university education; 5 studies) and were usually in employment (~70% full-time or part-time; 4 studies).

Only 4 studies reported the mean (range=4.6-6.8 years) or median (4 years) GFD duration (Edwards-George et al., 2009; Mahadev et al., 2015; Sainsbury et al., 2013a; Sainsbury et al., 2013b). The same 4 studies, and one additional study (Kerswell et al., 2015), reported the mean or median age at diagnosis, which ranged from 36 to 45 years. Comparison

of GFD duration with the difference between current age and age at diagnosis suggested that most participants commenced the GFD immediately upon diagnosis, although this was not reported separately in any of the studies.

Seven studies used a cross-sectional design (Arigo et al., 2012; Edwards-George et al., 2009; Kerswell & Strodl, 2015; Mahadev et al., 2015; Sainsbury et al., 2013a, 2013b; Weiss et al., 2013). The remaining study was a prospective study with measurement at diagnosis and after 1- and 4-years on a GFD; the relationship of interest was analysed cross-sectionally at both follow-up time points (Nachman et al., 2009; Nachman et al., 2010).

GFD adherence and depressive symptoms

Five studies measured GFD adherence using the CDAT (Kerswell & Strodl, 2015; Mahadev et al., 2015; Sainsbury et al., 2013a, 2013b; Weiss et al., 2013), two used a dietitian-rated assessment (Edwards-George et al., 2009; Nachman et al., 2009; Nachman et al., 2010), and one used the adapted Morisky adherence scale (Arigo et al., 2012). When reported, the mean scores for GFD adherence indicated that, on average, participants were highly adherent (Arigo et al., 2012; Edwards-George et al., 2009; Mahadev et al., 2015; Sainsbury et al., 2013a, 2013b). The proportion of each sample falling into the inadequate adherence range (21.4-43.4%) suggested lower rates of strict adherence though (Edwards-George et al., 2009; Mahadev et al., 2015; Nachman et al., 2009; Nachman et al., 2010; Sainsbury et al., 2013a, 2013b).

The Depression Anxiety Stress Scale was the most commonly used self-reported depression rating scale (Kerswell & Strodl, 2015; Sainsbury et al., 2013a, 2013b; Weiss et al., 2013). Other measures included the Beck Depression Inventory (Edwards-George et al., 2009; Nachman et al., 2009; Nachman et al., 2010), the Depressed Mood subscale of the Psychological General Wellbeing Inventory (Mahadev et al., 2015), and the Centre for Disease Studies Depression Scale (Arigo et al., 2012). Mean scores for self-reported

depressive symptoms fell in the normal to minimal/mild range (Arigo et al., 2012; Mahadev et al., 2015; Nachman et al., 2009; Nachman et al., 2010; Sainsbury et al., 2013a, 2013b).

Risk of bias assessment

The risk of bias assessment is presented in Table 2. Inter-rater agreement was 'almost perfect' (Cohen's kappa=0.958; Cohen, 1960). No major threats to study quality were apparent on the first two items: all except one study was clear in their description of the target population and inclusion criteria (item 1), and all attempted to ensure that included patients had a biopsy-confirmed, medical diagnosis of CD (item 2). This was done as part of the research procedure (Nachman et al., 2009; Nachman et al., 2010); or confirmed by patient self-report and/or assumed based on coeliac society membership (medical diagnosis required) or being listed in the clinic database (where the diagnosis was made).

Representativeness (item 3) was deemed 'very likely' in two studies, based on recruitment invites being sent to a randomly-selected sample of eligible coeliac society members identified via a database screen (Sainsbury et al., 2013a, 2013b). All bar two of the remaining studies were rated as 'somewhat likely' for using multiple recruitment methods to reduce bias (Arigo et al., 2012; Edwards-George et al., 2009; Mahadev et al., 2015), or an attempt to demonstrate comparability with previous samples of CD patients (Kerswell & Strodl, 2015). The one non-cross-sectional (prospective) study included a representative sample at baseline, obtained using a consecutive enrolment approach in newly diagnosed patients (Nachman et al., 2009; Nachman et al., 2010), but was compromised by significant attrition at both follow-up time-points and lack of reporting on differences by drop-out status or an attempt to control for attrition in analyses, which reduced the rating to 'not likely'. The final study (conference abstract only; Weiss et al., 2013) did not explicitly report details for criteria 1-3, although they could be partially inferred from the recruitment method (coeliac support group). Only two studies provided a power calculation (item 4; Arigo et al., 2012;

Sainsbury et al., 2013b), but all others met the minimum sample size according to criteria employed for the type of statistical analysis (i.e., 50 cases for a bivariate correlation, or 30 cases per group when comparing two or more groups).

Appropriate analyses (i.e., correlation/linear regression for continuously measured variables, or t-test/ANOVA/logistic regression/odds ratio for data involving one or both categorical variables; item 5) were used in all cases – reported in six papers and conducted by the review team for the two studies for which raw data was obtained from authors (Mahadev et al., 2015; Sainsbury et al., 2013b). Ratings of 'somewhat likely' for items 1-4 were deemed not to pose a major threat to study quality and were therefore viewed as equivalent to 'very likely' for the purposes of summarising overall risk of bias. Six studies were rated as having a low risk of bias (Arigo et al., 2012; Edwards-George et al., 2009; Kerswell & Strodl, 2015; Mahadev et al., 2015; Sainsbury et al., 2013a, 2013b); one study had a moderate risk of bias (Nachman et al., 2009; Nachman et al., 2010); and the other was rated unclear based on incomplete reporting (Weiss et al., 2013).

Meta-analysis

Effect sizes for five studies were extracted directly from the papers (Arigo et al., 2012; Edwards-George et al., 2009; Kerswell & Strodl, 2015; Sainsbury et al., 2013a; Weiss et al., 2013). Raw data was obtained for two studies in which the relevant information was not reported (Mahadev et al., 2015; Sainsbury et al., 2013b) and the correlations were computed by the first author. In the final study (Nachman et al., 2009; Nachman et al., 2010), means and standard deviations on the depression rating scale for the strictly and partially adherent groups were extracted and entered into CMA, where the effect size was automatically computed. The primary meta-analysis was conducted using only the 4-year data (Nachman et al., 2010) from the study that provided 2 effect sizes, as this was deemed more comparable to the other studies on GFD duration.

The pooled effect size for the 8 included studies was r=0.398 (95% CI=0.321-0.469, z=9.355, p<.001; see Figure 2), based on a total sample size of 1644 participants. The study by Kerswell (Kerswell & Strodl, 2015) showed the strongest association between depressive symptoms and adherence (r=0.510), while the Edwards-George study (Edwards-George et al., 2009) showed the weakest association (r=0.250); all effect sizes were significant. Inspection of the standard residuals indicated no outliers. There was evidence of moderate heterogeneity between studies (Q=21.097, p=.004; I^2=66.819).

Sensitivity analysis and publication bias

A sensitivity analysis excluding the two studies that were rated as having a moderate (Nachman et al., 2009; Nachman et al., 2010) or unclear risk of bias (Weiss et al., 2013) did not substantially change the results (r=0.394, 95% CI=0.303-0.477, z=7.887, p<.001).

The small number of included studies also prevented the use of publication bias estimates (it is not recommended to test for publication bias with <10 studies; Sterne et al., 2011), although visual inspection of the funnel plot did not reveal any asymmetry.

DISCUSSION

This meta-analysis represents the first attempt to synthesise the available evidence on the relationship between GFD adherence and depression/depressive symptoms in adult patients with CD. Only eight eligible studies were identified, all of which measured self-reported depressive symptoms rather than the presence of a clinical diagnosis of depression, and reported cross-sectional analyses. Consistent with the hypothesis, the results showed that higher levels of self-reported depressive symptoms were moderately associated with poorer GFD adherence. The present findings are comparable with research in other illnesses (e.g., diabetes) and varying treatment regimens (e.g., medication, diet and physical activity recommendations), where it has been found that depression acts as a barrier to good adherence (DiMatteo et al., 2000; Gonzalez et al., 2008; Grenard et al., 2011).

In CD, there is a paucity of research assessing the relationship between clinically diagnosable depression and adequately measured GFD adherence. Two studies included in this meta-analysis additionally reported the proportion of CD patients with differing adherence levels who scored above the clinical cut-off indicative of depression on their respective questionnaires, albeit with contrasting results (Arigo et al., 2012; Nachman et al., 2009; Nachman et al., 2010). One additional study was identified but could not be included due not reporting the relationship of interest (GFD adherence measured using the CDAT; van Hees, Giltay, Geleijnse, Janssen, & van der Does, 2014). More research is therefore needed to determine the impact of more severe manifestations of depression on GFD adherence and vice versa. Given the moderate-strength relationship between self-reported depressive symptoms and GFD adherence and the fact that, of those that reported this information, mean scores were suggestive of only low or mild levels of depressive symptoms, it seems likely that diagnosable depressive conditions would also show a relationship with worse GFD adherence.

All but one of the included studies adopted a cross-sectional design, so it remains unclear whether depressive symptoms are the cause or consequence of poor adherence. The one study that utilised a prospective data collection protocol (Nachman et al., 2009; Nachman et al., 2010) unfortunately did not report any prospective analyses. Instead the data was analysed cross-sectionally within each time point, so it too was unable to contribute to answering the question of causation. Evidence suggests that gluten exposure in CD patients triggers a series of physiological mechanisms that are linked to the development of depression (e.g., deficiencies in vitamin B deficiency and serotonin metabolites; Hallert et al., 1982; Hallert et al., 1983; Hallert et al., 2009). In the same way that a significant proportion of refractory CD cases (i.e., failure to achieve symptomatic and histological remission despite treatment with a GFD) are accounted for by unintentional gluten exposure, detected only with

rigorous methods (Abdulkarim, Burgart, See, & Murray, 2002), it may also be the case that ongoing consumption of trace amounts of gluten are responsible for the persistence of depressive symptoms in treated CD patients.

The perception of an increased ability to maintain adherence, despite changes in mood and stress, has also been linked to better GFD adherence (Leffler et al., 2008). Similarly, CD patients do not differ in their *intention* to follow a strict GFD, but depressive symptoms appear to limit the translation of positive intention into strict adherence (Sainsbury et al., 2013a). Together, these findings may suggest that under conditions of depression or low mood, some individuals with CD may be more prone to gluten-related lapses. Consistent with psychological theory (e.g., Kwasnicka, Dombrowski, White, & Sniehotta, 2016), this is likely due to a decrease in the usual level of vigilance and self-regulation (e.g., planning and monitoring) that occurs with low mood and that is required to maintain good adherence. This specific hypothesis, as well as the broader question of directionality, requires testing. Intentional gluten consumption appears to be less common (Hall et al., 2009; Hall, Rubin, & Charnock, 2013). More high-quality research, using prospective and longitudinal designs, is needed to provide a more definitive answer to the question of causality.

Limitations and directions for future research

The main limitation of this meta-analysis was the small number of eligible studies, which limited the power of the primary analysis and prevented the conduct of potential post-hoc moderator and subgroup analyses to determine the impact of other variables (e.g., disease and/or methodological characteristics) on the relationship of interest. While the narrow inclusion criteria adopted here certainly contributed to this lack of power, only by assessing the degree of adherence in CD patients who have already commenced a GFD – rather than the problematic reliance on categorical distinctions (e.g., on a GFD vs. yet not started a GFD) – and by excluding studies with inadequate measurement of adherence, can the question of

interest be answered. In support of the latter point, none of the five studies that were excluded based on unreliable measurement of GFD adherence (i.e., Likert or visual analogue scales assessing the frequency of gluten consumption) found a significant relationship with depressive symptoms (Barratt, Leeds, & Sanders, 2011; Ciacci, Iavarone, Mazzacca, & De Rosa, 1998; Ford, Howard, & Oyebode, 2012; Hauser, Janke, Klump, Gregor, & Hinz, 2010; van Hees, Van der Does, & Giltay, 2013).

Regarding publication bias, although there was no evidence of asymmetry in the visual inspection of the funnel plot, this could not be tested formally (e.g., using Egger's test) due to the small number of studies (less than 10 studies were included; Sterne et al., 2011). The eight studies that were eligible based on having measured the two constructs of interest but were excluded as the statistical relationship between them was not reported, may have altered the pattern observed here. It therefore cannot be ruled out that the inclusion of a larger number of studies would have changed the results or led to the identification of real asymmetries in the data.

There was significant heterogeneity in the effects between studies, which may be accounted for by potential moderators. As stated previously, the conduct of moderator analyses was not feasible here due to the limited number of studies meeting inclusion criteria. Previous literature and the characteristics observed in the included studies do, however, suggest that the following variables may worthy candidates for investigation in future research. Despite the known gender imbalance in CD diagnoses (male: female ratio of 1:2-3; Green et al., 2001), more than half the studies recruited ≥80% females (Arigo et al., 2012; Kerswell & Strodl, 2015; Nachman et al., 2009; Nachman et al., 2010; Sainsbury et al., 2013a, 2013b; Weiss et al., 2013), suggesting an additional gender bias that warrants further consideration. In contrast, the included samples appeared representative of the known CD population regarding age of diagnosis (Green & Jabri, 2006). Adjustment to the diagnosis and

the GFD could, however, plausibly differ according to both age of diagnosis and duration of GFD, suggesting that these variables may also moderate the adherence-depression relationship. CD patients with comorbid IBS-type symptoms (Hauser, Musial, Caspary, Stein, & Stallmach, 2007), diabetes and autoimmune conditions (Garud et al., 2009), and elevated thyroid anti-bodies (Carta et al., 2002), have higher rates of depression than patients with CD alone, suggesting that controlling for the presence of comorbidities may also help to explain differences in effects. Representativeness and recruitment or selection bias are common problems in CD research, and it is possible that differences in motivation and dietary vigilance exist according to membership of a coeliac society (Hall et al., 2009), which may change the results if more diverse recruitment methods were used. Methodologically, it also possible that the diverse measures (self-reported questionnaire versus interview-based) used to assess both depressive symptoms/depression and GFD adherence impacted the results, and attention should be given to study and measurement issues in future research.

Finally, the lack of prospective and longitudinal study designs is a clear limitation in the literature and prevented any descriptive or statistical analyses from being conducted to achieve the second aim of this review. This means that the direction of causation between depressive symptoms and GFD adherence remains unclear. Future research, using more rigorous designs and reliable measurement, is needed to achieve clarity on this point. To gain a comprehensive understanding of the full spectrum of factors that may impact the relationship between depression and adherence, a synthesis of the existing literature on the variables that predict either depression/depressive symptoms or GFD adherence in CD patients is needed.

Conclusion

The existing evidence for a relationship between depressive symptoms and/or depression and GFD adherence in adults with CD is limited. Nonetheless, what is available

suggests a moderate-strength relationship. From a clinical point of view, regardless of the additional factors that may be implicated in the relationship and its direction, identification of, and support to improve, poor adherence and depression/depressive symptoms in CD should be considered to reduce the burden of illness associated with deficiencies in both physical and mental health. Based on these tentative findings, there may be a role for psychological services in addition to dietetic input in the ongoing management and follow-up of GFD adherence for affected CD patients (Ludvigsson et al., 2014; NICE, 2015), even in cases of low-level, subclinical depressive symptoms (NICE, 2009). Online and face-to-face interventions using both individual and group-based formats have shown promise in improving GFD adherence and psychological wellbeing in CD (Addolorato et al., 2004; Ring Jacobsson, Friedrichsen, Goransson, & Hallert, 2012; Sainsbury, Mullan, & Sharpe, 2013c), and could help to achieve needed improvements in both directions.

ACKNOWLEDGEMENTS: none

CONFLICTS OF INTEREST: none

FUNDING: none

REFERENCES

Abdulkarim, A. S., Burgart, L. J., See, J., & Murray, J. A. (2002). Etiology of nonresponsive celiac disease: Results of a systematic approach. *American Journal of Gastroenterology, 97*(8), 2016-2021. doi:10.1016/s0002-9270(02)04274-0

Addolorato, G., De Lorenzi, G., Abenavoli, L., Leggio, L., Capristo, E., & Gasbarrini, G. (2004). Psychological support counselling improves gluten-free diet compliance in coeliac patients with affective disorders. *Alimentary Pharmacology & Therapeutics, 20*(7), 777-782. doi:10.1111/j.1365-2036.2004.02193.x

Arigo, D., Anskis, A. M., & Smyth, J. M. (2012). Psychiatric comorbidities in women with celiac disease. *Chronic Illness, 8*(1), 45-55. doi:10.1177/1742395311417639

Barratt, S. M., Leeds, J. S., & Sanders, D. S. (2011). Quality of life in coeliac disease is determined by perceived degree of difficulty adhering to a gluten-free diet, not the level of dietary adherence ultimately achieved. *Journal of Gastrointestinal and Liver Diseases, 20*(3), 241-245.

Biagi, F., Andrealli, A., Bianchi, P. I., Marchese, A., Klersy, C., & Corazza, G. R. (2009). A gluten-free diets score to evaluate dietary compliance in patients with coeliac disease. *British Journal of Nutrition, 102*(6), 882-887. doi:10.1017/S0007114509301579

Biostat, Inc. (2016). Comprehensive Meta-Analysis (CMA) Verson 3. Englewood, NJ.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-analysis.* Chichester, United Kingdom: Wiley & Sons.

Carta, M. G., Hardoy, M. C., Boi, M. F., Mariotti, S., Carpiniello, B., & Usai, P. (2002). Association between panic disorder, major depressive disorder and celiac disease - A possible role of thyroid autoimmunity. *Journal of Psychosomatic Research, 53*(3), 789-793. doi:10.1016/s0022-3999(02)00328-8

Casellas, F., Rodrigo, L., Vivancos, J. L., Riestra, S., Pantiga, C., Baudet, J. S., . . . Malagelada, J. R. (2008). Factors that impact health-related quality of life in adults with celiac disease: A multicenter study. *World Journal of Gastroenterology, 14*(1), 46-52. doi:10.3748/wjg.14.46

Ciacci, C., Iavarone, A., Mazzacca, G., & De Rosa, A. (1998). Depressive symptoms in adult coeliac disease. *Scandinavian Journal of Gastroenterology, 33*(3), 247-250. doi:10.1080/00365529850170801

Clarivate Analytics. (2016). Endnote (Version X7.5.3 for Mac). Philadelphia, PA, USA: Thompson Reuters.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*(1), 101-129.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46. doi:10.1177/001316446002000104

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.

DiMatteo, M. R., Lepper, H. S., & Croghan, T. W. (2000). Depression is a risk factor for noncompliance with medical treatment: Meta-analysis of the effects of anxiety and depression on patient adherence. *Archives of Internal Medicine, 160*, 2101-2107. doi:10.1001/archinte.160.14.2101

Edwards-George, J., Leffler, D. A., Dennis, M. D., Franko, D. L., Blom-Hoffman, J., & Kelly, C. P. (2009). Psychological correlates of gluten-free diet adherence in adults with celiac disease. *Journal of Clinical Gastroenterology, 43*(4), 301-306. doi:10.1097/MCG.0b013e31816a8c9b

Fera, T., Cascio, B., Angelini, G., Martini, S., & Guidetti, C. S. (2003). Affective disorders and quality of life in adult coeliac disease patients on a gluten-free diet. *European Journal of Gastroenterology & Hepatology, 15*(12), 1287-1292. doi:10.1097/01.meg.0000085512.01212.c5

Ford, S., Howard, R., & Oyebode, J. (2012). Psychosocial aspects of coeliac disease: A cross-sectional survey of a UK population. *British Journal of Health Psychology, 17*(4), 743-757. doi:10.1111/j.2044-8287.2012.02069.x

Garud, S., Leffler, D., Dennis, M., Edwards-George, J., Saryan, D., Sheth, S., . . . Kelly, C. P. (2009). Interaction between psychiatric and autoimmune disorders in coeliac disease patients in the Northeastern United States. *Alimentary Pharmacology & Therapeutics, 29*(8), 898-905. doi:10.1111/j.1365-2036.2009.03942.x

Gonzalez, J. S., Peyrot, M., McCarl, L. A., Collins, E. M., Serpa, L., Mimiaga, M. J., & Safren, S. A. (2008). Depression and diabetes treatment nonadherence: A meta-analysis. *Diabetes Care, 31*(12), 2398-2403. doi:10.2337/dc08-1341

Green, P. H., & Cellier, C. (2007). Celiac disease. *New England Journal of Medicine, 357*, 1731-1743. doi:10.1056/NEJMra071600

Green, P. H., & Jabri, B. (2003). Coeliac disease. *Lancet, 362*(9381), 383-391. doi:10.1016/S0140-6736(03)14027-5

Green, P. H., & Jabri, B. (2006). Celiac disease. *Annual Review of Medicine, 57*, 207-221. doi: 10.1146/annurev.med.57.051804.122404

Green, P. H., Stavropoulos, S. N., Panagi, S. G., Goldstein, S. L., McMahon, D. J., Absan, H., & Neugut, A. L. (2001). Characteristics of adult celiac disease in the USA: Results of a national survey. *American Journal of Gastroenterology, 96*(1), 126-131. doi:10.1111/j.1572-0241.2001.03462.x

Grenard, D. L., Munjas, B. A., Adams, J. L., Suttorp, M., Maglione, M., McGlynn, E. A., & Gellad, W. F. (2011). Depression and medication adherence in the treatment of chronic diseases in the United States: A meta-analysis. *Journal of General Internal Medicine, 26*(10), 1175-1182. doi:10.1007/s11606-011-1704-y

Hall, N. J., Rubin, G., & Charnock, A. (2009). Systematic review: Adherence to a gluten-free diet in adult patients with coeliac diease. *Alimentary Pharmacology & Therapeutics, 30*, 315-330. doi:10.1111/j.1365-2036.2009.04053.x

Hall, N. J., Rubin, G., & Charnock, A. (2013). Intentional and inadvertent non-adherence in adult coeliac disease: A cross-sectional survey. *Appetite, 68*(1), 56-62. doi:10.1016/j.appet.2013.04.016

Hallert, C., Astrom, J., & Sedvall, G. (1982). Psychic disturbances in adult coeliac disease III. Reduced central monoamine metabolism and signs of depression. *Scandinavian Journal of Gastroenterology, 17*, 25-28. doi:10.3109/00365528209181039

Hallert, C., Astrom, J., & Walan, A. (1983). Reversal of psychopathology in adult coeliac disease with the aid of pyridoxine (vitamin B6). *Scandinavian Journal of Gastroenterology, 18*(2), 299-304. doi:10.3109/00365528309181597

Hallert, C., Svensson, M., Tholstrup, J., & Hultberg, B. (2009). Clinical trial: B vitamins improve health in patients with coeliac disease living on a gluten-free diet. *Alimentary Pharmacology & Therapeutics, 29*(8), 811-816. doi:10.1111/j.1365-2036.2009.03945.x

Hauser, W., Janke, K. H., Klump, B., Gregor, M., & Hinz, A. (2010). Anxiety and depression in adult patients with celiac disease on a gluten-free diet. *World Journal of Gastroenterology, 16*(22), 2780-2787. doi:10.3748/wjg.v16.i22.2780

Hauser, W., Musial, F., Caspary, W. F., Stein, J., & Stallmach, A. (2007). Predictors of irritable bowel-type symptoms and healthcare-seeking behavior among adults with celiac disease. *Psychosomatic Medicine, 69*, 370-376. doi:10.1097/PSY.0b013e318050d6bb

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Alman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj, 327*(7414), 557-560. doi:10.1136/bmj.327.7414.557

Jackson, J. R., Eaton, W. W., Cascella, N. G., Fasano, A., & Kelly, D. L. (2012). Neurologic and psychiatric manifestations of celiac disease and gluten sensitivity. *Psychiatric Quarterly, 83*(1), 91-102. doi:10.1007/s11126-011-9186-y

Katon, W. J. (2011). Epidemiology and treatment of depression in patients with chronic medical illness. *Dialogues in Clinical Neuroscience, 13*(1), 7-23.

Kerswell, N., L., & Strodl, E. (2015). Emotion and its regulation predicts gluten-free diet adherence in adults with coeliac disease. *Health Psychology and Behavioural Medicine, 3*(1), 52-68. doi:10.1080/21642850.2015.1010534

Kwasnicka, D., Dombrowski, S.U., White, M., Sniehotta, F.F. (2016). Theoretical explanations for maintenance of behaviour change: A systematic review of behaviour theories. *Health Psychology Review, 10*(3), 277-96. doi: 10.1080/17437199.2016.1151372

Leffler, D. A., Dennis, M., Edwards-George, J., Jamma, S., Magge, S., Cook, E. F., . . . Kelly, C. P. (2009). A simple validated gluten-free diet adherence survey for adults with celiac disease. *Clinical Gastroenterology and Hepatology, 7*, 530-536. doi:10.1016/j.cgh.2008.12.032

Leffler, D. A., Edwards-George, J., Dennis, M., Cook, E. F., Schuppan, D., & Kelly, C. P. (2007). A prospective comparative study of five measures of gluten-free diet adherence in adults with coeliac disease. *Alimentary Pharmacology & Therapeutics, 26*, 1227-1235. doi:10.1111/j.1365-2036.2007.03501.x

Leffler, D. A., Edwards-George, J., Dennis, M., Schuppan, D., Cook, F., Franko, D. L., . . . Kelly, C. P. (2008). Factors that influence adherence to a gluten-free diet in adults with celiac disease. *Digestive Diseases and Sciences, 53*, 1573-1581. doi:10.1007/s10620-007-0055-3

Ludvigsson, J. F., Bai, J. C., Biagi, F., Card, T. R., Ciacci, C., Ciclitira, P. J., . . . Authors of the BSG Coeliac Disease Guidelines Development Group. (2014). Diagnosis and management of adult coeliac disease: Guidelines from the British Society of Gastroenterology. *Gut*. doi:10.1136/gutjnl-2013-306578

Mahadev, S., Gardner, R., Lewis, S. K., Lebwohl, B., & Green, P. H. (2015). Quality of life in screen-detected celiac disease patients in the United States. *Journal of Clinical Gastroenterology, 50*, 393-397. doi:10.1097/mcg.0000000000000433

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA group. (2009). The PRISMA reporting statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology, 62*(10), 1006-1012. doi:10.1016/j.jclinepi.2009.06.005

Morisky, D. E., Ang, A., Krousel-Wood, M., & Ward, H. J. (2008). Predictive validity of a medication adherence measure in an outpatient setting. *Journal of Clinical Hypertension, 10*(5), 348-354. doi:10.1111/j.1751-7176.2008.07572.x

Morisky, D. E., Green, L. W., & Levine, D. M. (1986). Concurrent and predictive validity of a self-reported measure of medication adherence. *Medical Care, 24*(1), 67-74. doi:10.1097/00005650-198601000-00007

Nachman, F., Maurino, E., Vazquez, H., Sfoggia, C., Gonzalez, A., Gonzalez, V., . . . Bai, J. C. (2009). Quality of life in celiac disease patients: Prospective analysis on the importance of clinical severity at diagnosis and the impact of treatment. *Digestive and Liver Disease, 41*(1), 15-25. doi:10.1016/j.dld.2008.05.011

Nachman, F., Plancer del Campo, M., Gonzalez, A., Corzo, L., Vazquez, H., Sfoggia, C., . . . Bai, J. C. (2010). Long-term deterioration of quality of life in adult patients with celiac disease is associated with treatment noncompliance. *Digestive and Liver Disease, 42*(10), 685-691. doi:10.1016/j.dld.2010.03.004

National Institutes of Health. (2014). Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. Retrieved from https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/cohort

NICE. (2009). Depression in adults with a chronic physical health problem: Recognition and management. Clinical Guideline 91. Retrieved from https://www.nice.org.uk/guidance/CG91/chapter/introduction

NICE. (2015). Coeliac disease: Recognition, assessment and management. NICE guideline 20. Retrieved from https://www.nice.org.uk/guidance/ng20

Ring Jacobsson, L., Friedrichsen, M., Goransson, A., & Hallert, C. (2012). Does a coeliac school increase psychological well-being in women suffering from coeliac disease, living on a gluten-free diet? *Journal of Clinical Nursing, 21*, 766-775. doi:10.1111/j.1365-2702.2011.03953.x

Sainsbury, K., Mullan, B., & Sharpe, L. (2013a). Gluten free diet adherence in coeliac disease: The role of psychological symptoms in bridging the intention-behaviour gap. *Appetite, 61*(1), 52-58. doi:10.1016/j.appet.2012.11.001

Sainsbury, K., Mullan, B., & Sharpe, L. (2013b). Reduced quality of life in coeliac disease is more strongly associated with depression than gastrointestinal symptoms. *Journal of Psychosomatic Research, 75*(2), 135-141. doi:10.1016/j.jpsychores.2013.05.011

Sainsbury, K., Mullan, B., & Sharpe, L. (2013c). A randomized controlled trial of an online intervention to improve gluten free diet adherence in celiac disease. *American Journal of Gastroenterology, 108*, 811-817. doi:10.1038/ajg.2013.47

Silvester, J. A., Weiten, D., Graff, L. A., Walker, J. R., & Duerksen, D. R. (2016). Is it gluten-free? Relationship between self-reported gluten-free diet adherence and knowledge of gluten content of foods *Nutrition, 32*(7-8), 777-783. doi:10.1016/j.nut.2016.01.021

Smith, D. F., & Gerdes, L. U. (2012). Meta-analysis on anxiety and depression in adult celiac disease. *Acta Psychiatrica Scandinavica, 125*, 189-193. doi:10.1111/j.1600-0447.2011.01795.x

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., . . . Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot assymetry in meta-analyses of randomised controlled trials. *Bmj, 343*, d4002. doi:10.1136/bmj.d4002

The Joanna Briggs Institute. (2016). Checklist for analytical cross sectional studies. Retrieved from http://joannabriggs.org/research/critical-appraisal-tools.html

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions.

Worldviews on Evidenced Based Nursing, 1(3), 176-184. doi:10.1111/j.1524-475X.2004.04006.x

Vahedi, K., Mascart, F., Mary, J. Y., Laberenne, J. E., Bouhnik, Y., Morin, M. C., . . . Matuchansky, C. (2003). Reliability of antitransglutaminase antibodies as predictors of gluten-free diet compliance in adult celiac disease. *American Journal of Gastroenterology, 98*(5), 1079-1087. doi:10.1111/j.1572-0241.2003.07284.x

van Hees, N. J., Giltay, E. J., Geleijnse, J. M., Janssen, N., & van der Does, W. (2014). DHA serum levels were significantly higher in celiac disease patients compared to healthy controls and were unrelated to depression. *Plos One, 9*(5), e97778. doi:10.1371/journal.pone.0097778

van Hees, N. J., Van der Does, W., & Giltay, E. J. (2013). Coeliac disease, diet adherence and depressive symptoms. *Journal of Psychosomatic Research, 74*(2), 155-160. doi:10.1016/j.jpsychores.2012.11.007

Weiss, G., Lackner, J., Keller, C., Gudleski, G., Hauck, C., & Sitrin, M. (2013). Barriers to compliance to gluten-free diet in celiac disease: The role of depression and motivation. *American Journal of Gastroenterology, 108*, S468-S468.

Zingone, F., Swift, G. L., Card, T. R., Sanders, D. S., Ludvigsson, J. F., & Bai, J. C. (2015). Psychological morbidity of celiac disease: A review of the literature. *United European Gastroenterology Journal, 3*, 136-145. doi:10.1177/2050640614560786

Table 1. Characteristics of included studies (k = 8)

Study	Study design	N, % female, mean age, education and employment	Type of CD patients, mean GFD duration, age at diagnosis, recruitment	GFD adherence measure and results	Depression measure and results	Relationship between GFD adherence and depression
Arigo et al. (2012)	Cross-sectional	N = 177 100% female Age = 39.24 Education and employment not reported	Unselected CD patients; GFD duration/age at diagnosis not reported; recruited via CD organisations, online support networks, and online newsletters	Morisky Medication Adherence Scale, adapted for GFD adherence (M = 4.47, SD = 1.06; maximum possible score of 5; higher scores indicate better adherence)	Centre for Disease Studies – Depression Scale (M = 14.96, SD = 10.9; higher scores indicate worse depression)	Higher depression scores were associated with poorer GFD adherence (r = -0.28, p < .01); women who scored at or above the clinical cut-off for depression reported poorer GFD adherence (M = 4.29, SD = 1.26) than women who scored below the cut-off for depression (M = 4.63, SD = 0.82, t = 2.12, p < .05).
Edwards-George et al. (2009)	Cross-sectional	N = 154 76.6% female Age = 50.35 Education: 68% had a bachelor's degree or higher Employment: 70% full time or part time	CD patients on a GFD for at least 3 months; GFD duration (mean) = 4.9 years; age at diagnosis = 44.79; recruited via clinic appt.'s, flyer mailed to patients who had previously attended clinic, local CD support groups, and CD newsletters	Expert dietitian evaluation (analysis of 3-day food records, food ingredient quiz, and a clinical interview) (M = 1.92, SD = 1.12; observed range = 1-6, higher scores indicate poorer adherence; 44.2% excellent adherence: 34.4% good; 21.4% inadequate)	BDI-II (no descriptive statistics reported; higher scores indicate worse depressive symptoms)	Higher levels of non-adherence were associated with higher depression scores (r = 0.25, p = .002); depression was not a significant independent predictor of adherence (dichotomised into good vs. inadequate) in a logistic regression model.
Kerswell & Strodl (2015)	Cross-sectional	N = 253 91.4% female Age = 42.34 Education:	Unselected CD patients; GFD duration not reported; age at diagnosis = 35.72;	CDAT (no descriptive statistics reported; higher scores indicate poorer adherence)	DASS (no descriptive statistics reported; higher scores	Higher depression scores were associated with poorer adherence (r = 0.51, p < .003).

		58% undergraduate or post-graduate degree Employment: 69% full time or part time	recruited via coeliac society (Facebook page and monthly online newsletter)		indicate worse depressive symptoms)	
Mahadev et al. (2015)	Cross-sectional	N = 211 78% female Age (median) = 47 Education: 93% college or graduate school Employment not reported	Screen- and symptom-detected CD patients; GFD duration (median) = 4 years; age at diagnosis = 39; recruited via clinic appt.'s, patient support conferences, and mail/email invites sent to patients in the clinic database	CDAT (M = 12.0, SD = 3.9; observed range = 7-25; higher scores indicate worse adherence; 55.5% excellent or very good; 26.1% moderate; 9.5% fair to poor)	PGWB – depressed mood subscale (M = 15.7, SD = 2.8; observed range = 5-18, higher scores indicate less severe depressive symptoms)	Higher depression scores were associated with poorer adherence (r = -0.473, p < .001).
Nachman et al. (2009)/ Nachman et al. (2010)	Repeated measurement in same CD participants (depressive symptoms-adherence relationship analysed cross-sectionally)	2009 (1-year): N = 84 84% female Age, education, and employment not reported 2010 (4-years): N = 53 90.6% female Age,	Newly diagnosed CD patients assessed after 1-year and 4-years on a GFD; age at diagnosis not reported; consecutively enrolled/recruited at diagnosis from clinic	Combination of dietitian interview, opinion of treating physician, 4-day food diary, and patient self-report (1-year: 70.2% strict adherence, 29.8% partial adherence; 4-years: 50.9% strict, 49.1% partial)	BDI (higher scores indicate worse depressive symptoms)	1 year: Depression scores did not differ between the CD patients who were strictly adherent (M = 7.9, 95% CI = 4.8-11.0) and those who were partially adherent (M = 6.3, 95% CI = 3.6-9.5; p = ns; r = 0.068, p = 0.532). 4 years: Partially adherent CD patients (M = 11.3, 95% CI = 7.6-15) had significantly

		education, and employment not reported				higher depression scores than CD patients with strict GFD adherence (*M* = 5.8, 95% CI = 2.1-9.5, *p* = 0.03; *r* = 0.278, *p* = .034). There was no difference in the proportion of CD patients scoring above the clinical cut-off for depression according to adherence category (strict vs. partial).
Sainsbury et al. (2013a)	Cross-sectional	N = 390 82.8% female Age = 44.2 Education: 57% undergraduate or post-graduate degree Employment: 70% full time or part time/ casual	CD patients, on a GFD for at least 3 months; GFD duration (mean) = 6.8 years; age at diagnosis = 37.4; recruited via coeliac society (email sent to randomly selected members meeting the inclusion criteria, based on database screening)	CDAT (M = 12.31, SD = 3.17; observed range = 7-31, higher scores indicate poorer adherence; 56.7% excellent or very good; 37.2% moderate; 6.2% fair to poor)	DASS (M = 6.2, SD = 8.2; observed range = 0-42, higher scores indicate worse depression)	Higher depression scores were associated with poorer adherence (*r* = 0.33, *p* < .001).
Sainsbury et al. (2013b; study 2 only)	Cross-sectional (baseline relationships in a sample of CD patients enrolled in an intervention to improve GFD	N = 189 87.3% female Age = 46.5 Education: 47% undergraduate or post-graduate	CD patients, on a GFD for at least 3 months; GFD duration (mean) = 4.6 years; age at diagnosis = 42.1; recruited via coeliac society (email sent to randomly selected	CDAT (M = 12.2, SD = 3.44; observed range = 7-28, higher scores indicate poorer adherence; 58.9% excellent or very good; 33.2% moderate; 7.9% fair to poor)	DASS (M = 5.9, SD = 7.4; range = 0-42; higher scores indicate worse depression)	Higher depression scores were associated with poorer adherence (*r* = 0.477, *p* < .001).

	adherence)	Employment: 69% full time or part time/ casual	members meeting the inclusion criteria, based on database screening)			
Weiss et al. (2013)	Cross-sectional	N = 186 79.6% female Age = 56.5 Education and employment not reported	Unselected CD patients; GFD duration/age at diagnosis not reported; recruited via GFD support group (no details provided)	CDAT (no descriptive statistics reported)	DASS (no descriptive statistics reported)	Higher depression scores were associated with poorer adherence (r = 0.48, p < .001).

Note: CD = coeliac disease; GFD = gluten free diet; CDAT = coeliac dietary adherence test; BDI = Beck Depression Inventory; DASS = Depression, Anxiety, Stress Scale; PGWB = Psychological General Wellbeing Index

Table 2. Risk of bias assessment

Study	Population/ inclusion	Objective criteria for CD	Representativeness	Sample size	Statistical analysis	Overall risk of bias
Arigo (2012)	VL	SL	SL	VL	VL	Low
Edwards-George (2009)	VL	SL	SL	SL	VL	Low
Kerswell (2015)	VL	SL	SL	SL	VL	Low
Mahadev (2015)	VL	SL	SL	SL	NR#	Low
Nachman (2009, 2010)	VL	VL	NL*	SL	VL	Moderate
Sainsbury (2013a)	VL	SL	VL	SL	VL	Low
Sainsbury (2013b)	VL	SL	VL	VL~	NR#	Low
Weiss (2013)	NR	NR	NR	SL	VL	Unclear

VL = very likely; SL = somewhat likely; NL = not likely; NR = not reported

* Not likely rating based on significant attrition from baseline to both 1-year and 4-year follow-up, a lack of details reported on differences between drop-outs and those who remained in the study, and failure to account for attrition in follow-up analyses (i.e., per-protocol analysis rather than intention-to-treat)

~ Power analysis reported in linked paper on same sample (randomised controlled trial of behavioural intervention to improve GFD adherence; Sainsbury, Mullan, & Sharpe, 2013c)

Statistics for the relationship between depressive symptoms and GFD adherence were not reported – raw data obtained from study authors and analyses conducted by first author

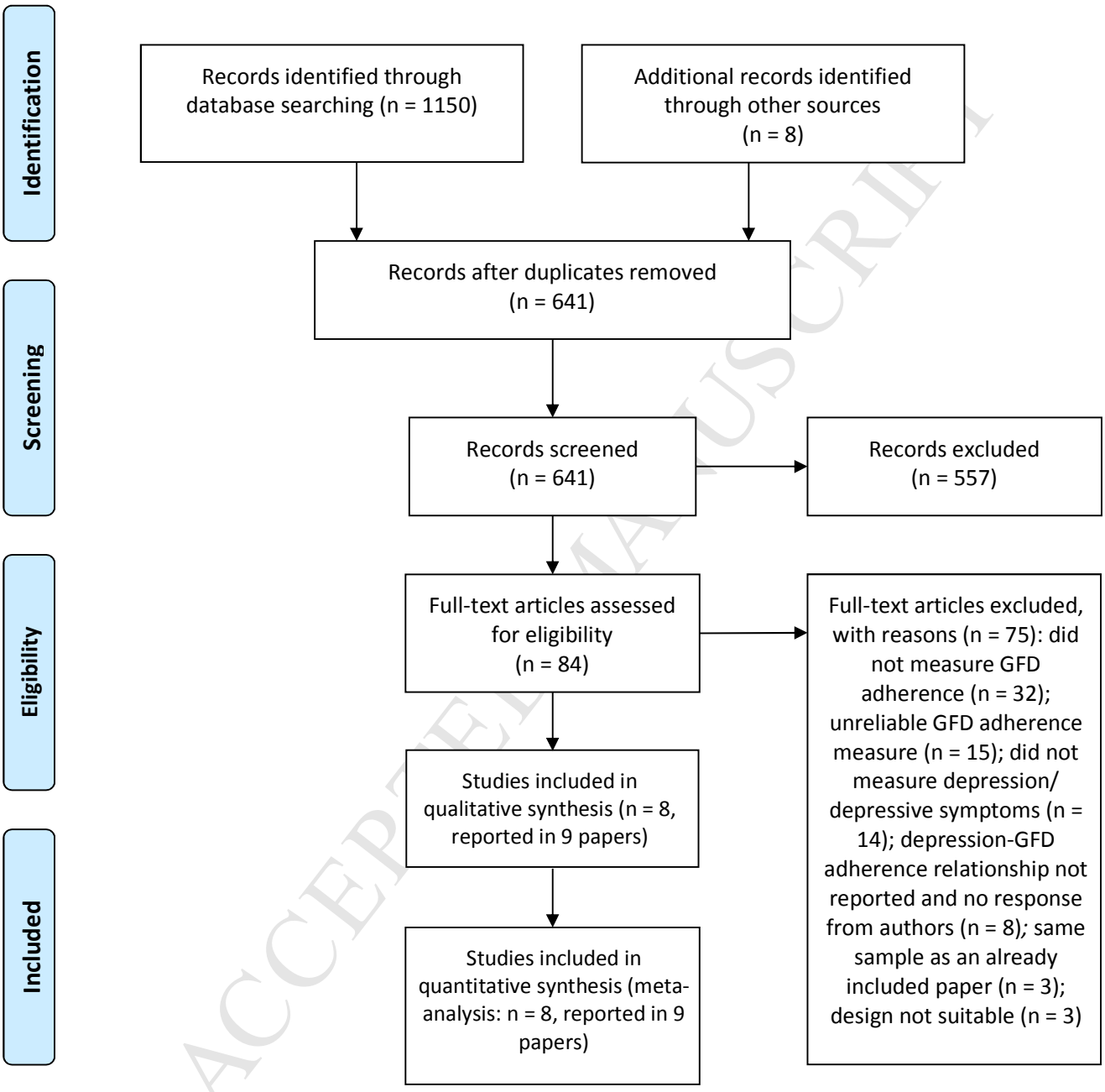

Figure 1. PRISMA flow chart

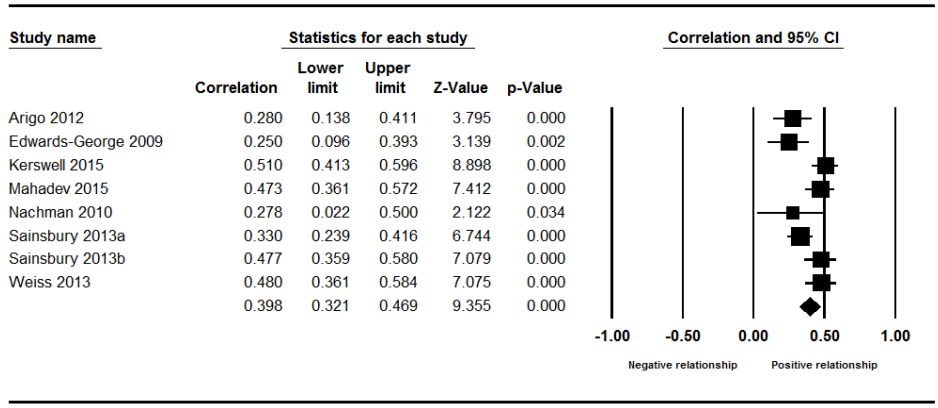

Figure 2. Meta-analysis of the relationship between depressive symptoms and GFD adherence

Note: Nachman et al., 2010 (assessment at 4-years post-diagnosis and GFD onset) was included rather than Nachman et al., 2009 (assessment at 1-year post-diagnosis and GFD onset) due to similarity with other included studies on GFD duration (mean/median > 4 years).