Computers in Biology and Medicine 65 (2015) 369–378

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Trends in celiac disease research

Edward J Ciaccio [a,*], Govind Bhagat [a,b], Suzanne K Lewis [a], Peter H Green [a]

[a] Department of Medicine – Celiac Disease Center, Columbia University, New York, USA
[b] Department of Pathology and Cell Biology, Columbia University, New York, USA

ARTICLE INFO

Article history:
Received 25 November 2014
Accepted 24 May 2015

Keywords:
Celiac disease
Inflammatory bowel disease
MEDLINE
Research
Trends

ABSTRACT

Background: To improve diagnosis and treatment of celiac disease (CD), research efforts are being made in many different areas. However, the focus, trend, and direction of such efforts require clarity, so that future efforts and directions can be appropriately planned.
Method: In this study, MEDLINE was used to search for trends in CD research. The keyword 'celiac disease' and its variants were searched in tandem with keywords commonly associated with CD. This search was done for each year from 1960 to 2013. Year of first instance of the associated keyword, linear regression coefficient, and trend in terms of the slope of the regression line were tabulated. For perspective, the same keywords were searched in tandem with 'inflammatory bowel disease' (IBD).
Results: CD appeared in the medical literature prior to 1960, and IBD first appeared in 1964. However, IBD overtook CD in terms of the number of research papers published per year, beginning in 1988. Keywords with strong positive trends ($r^2 > 0.7$) in association with CD were: 'diagnosis', 'gluten', 'serology', 'autoimmune', 'treatment', 'gluten-free diet', 'endoscopy', 'villous atrophy', 'wasting', 'inflammation', and 'microbiome'. The keyword 'malabsorption' had the sole strong negative trend in association with CD. Keywords with strong positive trends ($r^2 > 0.7$) in association with IBD also had strong positive association with CD: 'autoimmune', 'treatment, 'inflammation', and 'microbiome'.
Conclusions: The MEDLINE search approach is helpful to show first instance, association, and trend of keywords that are affiliated with CD in published biomedical research, and to compare CD research trends with those of other diseases.

1. Introduction

The diagnosis and management of celiac disease (CD) is a subject of intense study, with recent guidelines appearing in the medical literature [1]. The research in CD is diverse, involving pathogenesis, epidemiology, clinical manifestations, diagnosis and therapy. Characterization of CD is potentially extensive, as clinically the manifestations are diverse [2]. The quantitation of CD research data is desirable, including the potential for validation of a clinical prediction score [3]. However, it is difficult to assess severity of disease status because there is no severity index [4]. Population-based studies are useful in determining the risk of disease complications [5,6]. Clinical aspects of CD in areas of the world where it has not received great attention previously, including the Asia–Pacific region, is increasing [7]. In previous work, it has been shown that CD studies are increasingly common in the medical research literature [8]. The sharpest increase was found to have occurred since 1995. In terms of the global research effort for CD, countries including Italy, USA, and UK have the greatest volume of published studies [8]. Generally, the CD studies published in the medical research literature are of high quality and complexity [9]. Although CD research is gaining attention, research directions tend to be unfocused. Hence, it would be helpful to characterize trends in ongoing CD research so that areas of foci can be identified for future research.

To improve diagnosis and treatment of CD, research efforts are being made in many different areas. However, it is not entirely clear where efforts are being made and therefore, what further efforts and directions are needed for future work. In this study, we sought to determine the trends in CD research in terms of the prevalence of readily identifiable keywords in the CD literature. As a comparison, we determined the association of these same keywords with inflammatory bowel disease (IBD), since IBD is similar to CD in its clinical complexity and management difficulties.

2. Method

To search for trends in CD research, the MEDLINE database was used. The MEDLINE database indexes the instance of a searched

* Correspondence to: Harkness Pavilion 934, Columbia University, 180 Fort Washington Avenue, New York, NY 10032, USA.
Tel.: +212 305 5447; fax: +212 342 0447.
E-mail address: ciaccio@columbia.edu (E. Ciaccio).

http://dx.doi.org/10.1016/j.compbiomed.2015.05.023
0010-4825/

word in the title, abstract, and list of keywords of a paper. The search tool can be used to search during single years, which is how it was utilized for this study. CD was searched using the terms shown in the top row in Table 1. The keyword 'celiac disease', and variants 'coeliac disease' and 'celiac sprue' were all included in the search. Also, 32 associated keywords and their similar terms (variants) were searched in association with CD (Table 1). The list was developed from a Google query of the 32 most common terms associated with celiac disease, and are listed in alphabetical order. The Google query was used for this purpose due to ease of procurement. Medical Subject Headings (MeSH) terms were not used, as this is a controlled vocabulary that may not reflect all of the terms affiliated with celiac disease. Although 'autism' and 'Down syndrome' are different disorders, they were grouped together, as they are often associated with CD.

To search for a keyword in association with CD, we used for example, the following syntax –

(*'celiac disease' or 'coeliac disease' or 'celiac sprue'*)
and (*videocapsule or 'video capsule'*) (1)

When a keyword consisted of more than one word, it was bracketed by quotation marks in the MEDLINE form. The logical 'or' was used to combine the occurrence of several variants of a keyword. Thus for the example above,

(*'celiac disease' or 'coeliac disease' or 'celiac sprue'*) (2)

was used to determine any and all research papers containing one or more keywords 'celiac disease', 'coeliac disease', 'celiac sprue' which are three common ways to refer to CD.

Similarly, (videocapsule or 'video capsule')

(*videocapsule or* | *video capsule*|) (3)

are two common ways to refer to this keyword. Query (1) is formed by adding a logical 'and' between queries (2) and (3), to identify research papers in which at least one of the three keywords referring to CD and at least one of the two keywords referring to 'videocapsule' appeared together.

The keywords for CD were searched using the logical 'and' function with each of the associated keywords, for each year from 1960 to 2013. The year 1960 was used as a starting point because at that time, it became widely promulgated that patients improved on a gluten-free diet, an important aspect in disease treatment [10]. Linear regression analysis was used to show trends in association. A trend was defined as a major upward or downward swing in the year-by-year query graph, excluding small deflections, with the endpoint being defined as 2013, the last year in which complete publishing information was available at the time of this study. The starting point of the regression line was defined to be the year that the upward or downward swing began. The starting point year, the slope of the regression line, and the correlation coefficient r^2 were tabulated. The linear regression was constructed using SigmaPlot (Systat Software, Inc., San Jose, CA, 2005).

The keywords in association with CD were normalized with respect to the number of CD papers published in the medical literature during each year, i.e., for any given year,

$$Normalized\_association = \frac{\text{Number of studies in which affiliated keyword appeared with CD}}{\text{Total number of studies in the literature with CD}}$$

which was calculated and then graphed. The above fraction ranges from 0 to 1, and can be expressed as a percentage by multiplying by 100%. Affiliated keywords were categorized using decade ranges according to whether their prevalence with CD in the

Table 1
List of search terms.

Keyword	Variants
Celiac disease	Celiac sprue, celiac disease
Autism	Down syndrome
Autoimmune	Autoimmunity, immune
Biopsy	
Diagnosis	Diagnostic, diagnostics, screening
Dental	Enamel
Dermatitis	Dermatologic, herpetiformis, rash, skin
Diarrhea	
DQ2	DQ8, HLA
Endoscopy	
Fatigue	Apathy, inactivity, lethargy, tiredness, torpidity, torpor, weariness
Female	
Gluten	Barley, gliadin, wheat, rye
Gluten-free diet	Gluten free diet
Heart	Cardiac
Infertility	Fertile, fertility, infertile, pregnancy, pregnant
Inflammation	
Lymphocytic Colitis	
Malabsorption	
Male	
Malignancy	Cancer, lymphoma, neoplasia
Malnourishment	Malnutrition, nourishment, nutrient, nutrients, nutrition
microbiome	
Microscopic Colitis	
Motility	
Neurology	EEG, electroencephalogram, neurologic, neuropathy, seizure, seizures
Osteoporosis	Bone, fracture
Psychiatric	Depression, obsessive compulsive disorder
Serology	Antibodies, IgA, IgG, serological, transglutaminase
Treatment	Therapeutic, therapies, therapy
Videocapsule	Video capsule
Villous Atrophy	
Wasting	Growth, stature, stunted, weight loss

research literature was common (> 10% of CD papers in 2013), moderate (1–10% of CD papers in 2013), or slight (< 1% of CD papers in 2013).

For perspective, the same keywords were searched in tandem with 'inflammatory bowel disease' (IBD). To compare CD and IBD, the following search was done –

('*celiac disease*' or '*coeliac disease*' or '*celiac sprue*')
and (*inflammatory bowel disease*) (4)

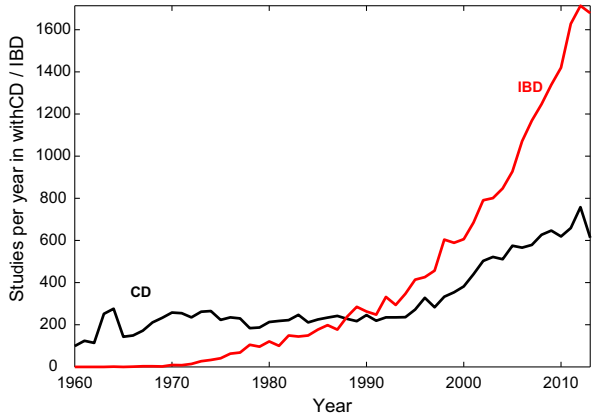

Fig. 1. Relationship of celiac disease (black trace) versus inflammatory bowel disease (red trace). Shown are published papers per year from 1960 to 2013. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

MEDLINE was queried for IBD in association with all CD keywords, for example –

(*inflammatory bowel disease*) *and* (*therapy or therapies or therapeutic*) (5)

The same parameters of first instance of associated keyword, linear regression coefficient, and trend in terms of the slope of the regression line were tabulated.

3. Results

The direct comparison of CD and IBD is shown in Fig. 1. The ordinate axis shows the total number of papers associated with each keyword from 1960 to 2013. Although IBD was first referred to in the medical literature in 1964, it is evident based on Fig. 1 that there has been a much sharper increase in its occurrence in the literature as compared to CD. For reference, the linear regression in each curve was determined, beginning at the crossover point, which was year 1988. For CD, the linear regression trend revealed an increase on average of 21.94 papers referring to CD per year from 1988 to 2013 (r^2=0.939). For IBD, the linear regression trend revealed an increase on average of 60.88 papers referring to IBD per year from 1988 to 2013 (r^2=0.937). Thus the rate of increase of IBD scientific studies appearing in the medical research literature was about three times as great as compared with CD studies during this time interval.

In Figs. 2–7, a comparison of trends for the 22 keywords most affiliated with CD as ordered from greatest to least association in

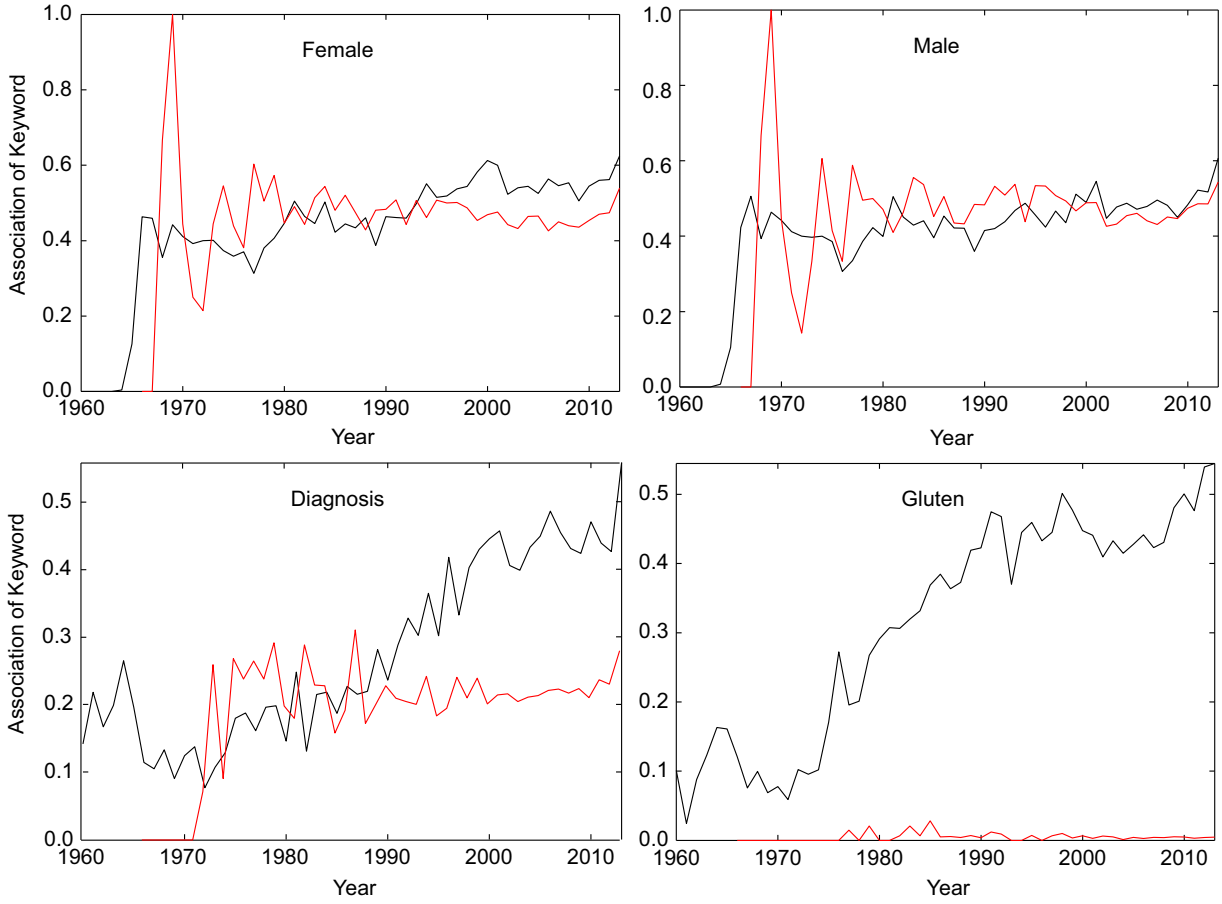

Fig. 2. Keywords commonly associated with celiac disease studies (range of association > 0.5 or 50% in 2013). Shown are changes from 1960 to 2013 (black trace). For reference, changes of each keyword with respect to inflammatory bowel disease is also shown (red trace). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

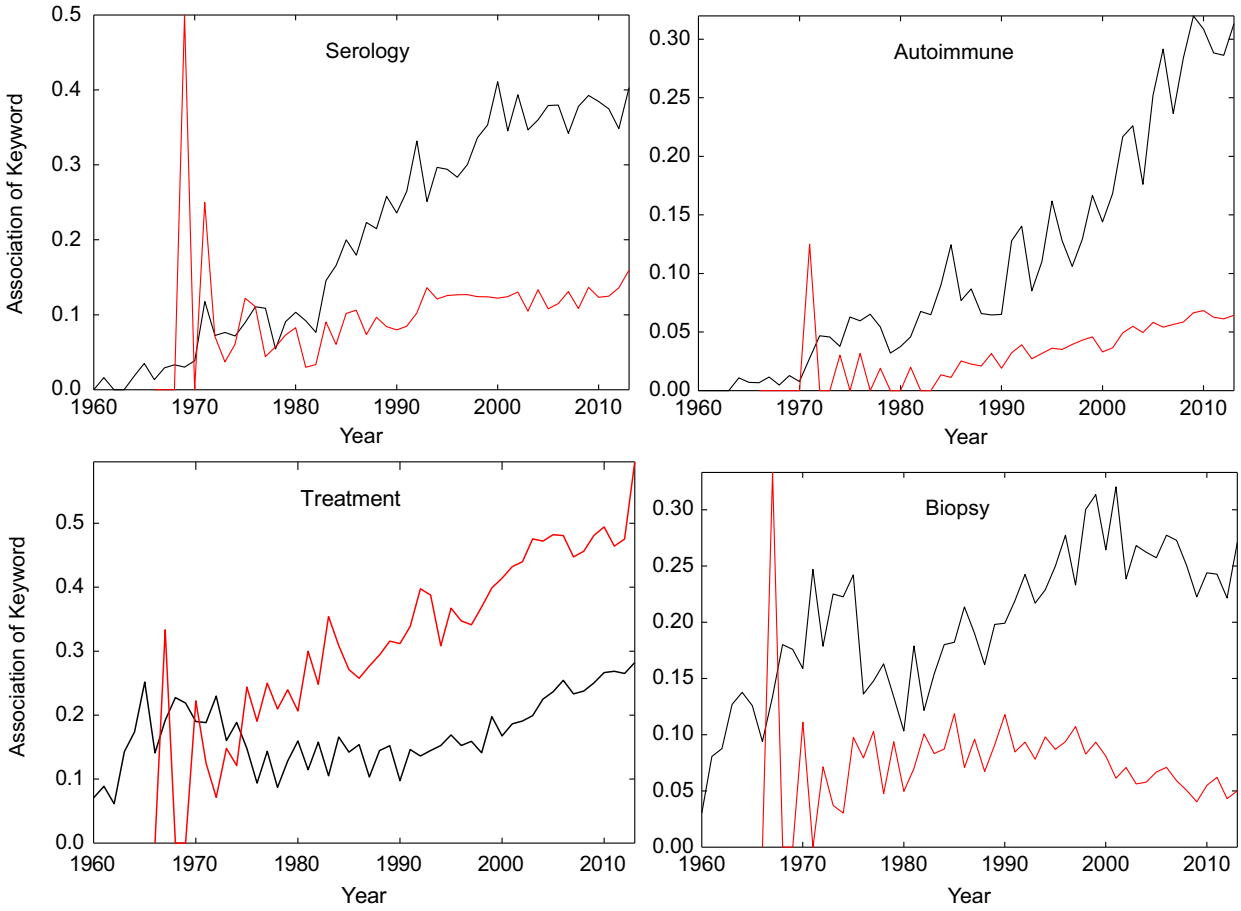

Fig. 3. Keywords commonly associated with celiac disease studies (range of association 0.4–0.27 or 40–27% in 2013). Shown are changes from 1960 to 2013 (black trace). For reference, changes of each keyword with respect to inflammatory bowel disease is also shown (red trace). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2013 are shown (black traces). The most common keyword affiliated with CD is 'female', which appeared in approximately 0.63 (63%) of the CD studies in 2013 (Fig. 2, top left panel). The next most common keyword in association with CD was 'male', which appeared in approximately 0.61 (61%) of studies in 2013 (Fig. 2, top right panel). These keywords are also associated with IBD (red traces). There is a slight increase in the presence of both 'male' and 'female' in CD but not IBD papers (top panels, Fig. 2). This observation is interesting, as CD is diagnosed more frequently in females, despite screening studies that show equal gender prevalence [11]. There has been a sharp increase in interest in keyword 'diagnosis' in association with CD since approximately 1972 (lower left panel in Fig. 2) but interest in 'diagnosis' with IBD has remained relatively flat. The term 'gluten' has shown various levels of increase in association with CD since approximately 1974. From the graph, there was some interest in 'gluten' in association with IBD around 1980, though of much lesser degree as compared with CD, which is not surprising as a gluten-free diet is the sole therapy for CD. In Fig. 3, it is evident that a sharp increase in the appearance of keyword 'autoimmune' and some increase of 'treatment' with CD has occurred since the early 1970s. The interest in 'biopsy' with CD has shown a bimodal distribution, peaking in about 1973 and 2000. Although there was initially increasing interest in the relationship of 'serology' to CD, with the timeline reflecting the generation of serologic markers, there has been no increase since 2000. From Fig. 4, it is evident that 'DQ2' is of great interest in association with CD, and although there has been some interest with respect to IBD, that interest is slowly waning. There has also been a substantial increase in interest in 'endoscopy' and 'villous atrophy' with CD. Keyword 'gluten-free diet' is of strong interest in CD but of almost no interest with respect to IBD. There are similar associations of keywords 'malignancy' and 'wasting' with CD and with IBD (Fig. 5). There is a sharp increase over time in the affiliation of the keyword 'malignancy' with CD and IBD, which appears to a similar degree for each disease. Keyword 'inflammation' exhibits a strong increase over time with respect to IBD, and also an increase with CD but only since approximately 1994. Keyword 'dermatitis' peaked from about 1976–1990 for CD and at 1982 for IBD; since then there has been a mostly downward trend in the association of this keyword with both diseases. In Fig. 6, there is a similar association of keyword 'diarrhea' for both CD and IBD. The diminishment of keyword 'malabsorption' in association with CD, and of keyword 'malnourishment' with IBD, is striking. There has been some interest in 'osteoporosis' in association with both CD and IBD, but this interest appears to have peaked in 2004. Although quite variable from year to year, overall there has been a sharply increasing interest in keyword 'neurological' with CD since approximately 1985 (Fig. 7). The association of keyword 'autism' with CD has increased since approximately 1990, but there is substantial fluctuation from year to year.

For all keywords, the order of prevalence in association with CD in 2013 is given in Table 2. The keywords were grouped in terms of common ($>10\%$), moderate ($>1\%$ and $<10\%$) and slight ($\leq 1\%$) association with CD. Since the maximum possible normalized value was 1.0 (i.e., if a keyword was present in all celiac disease

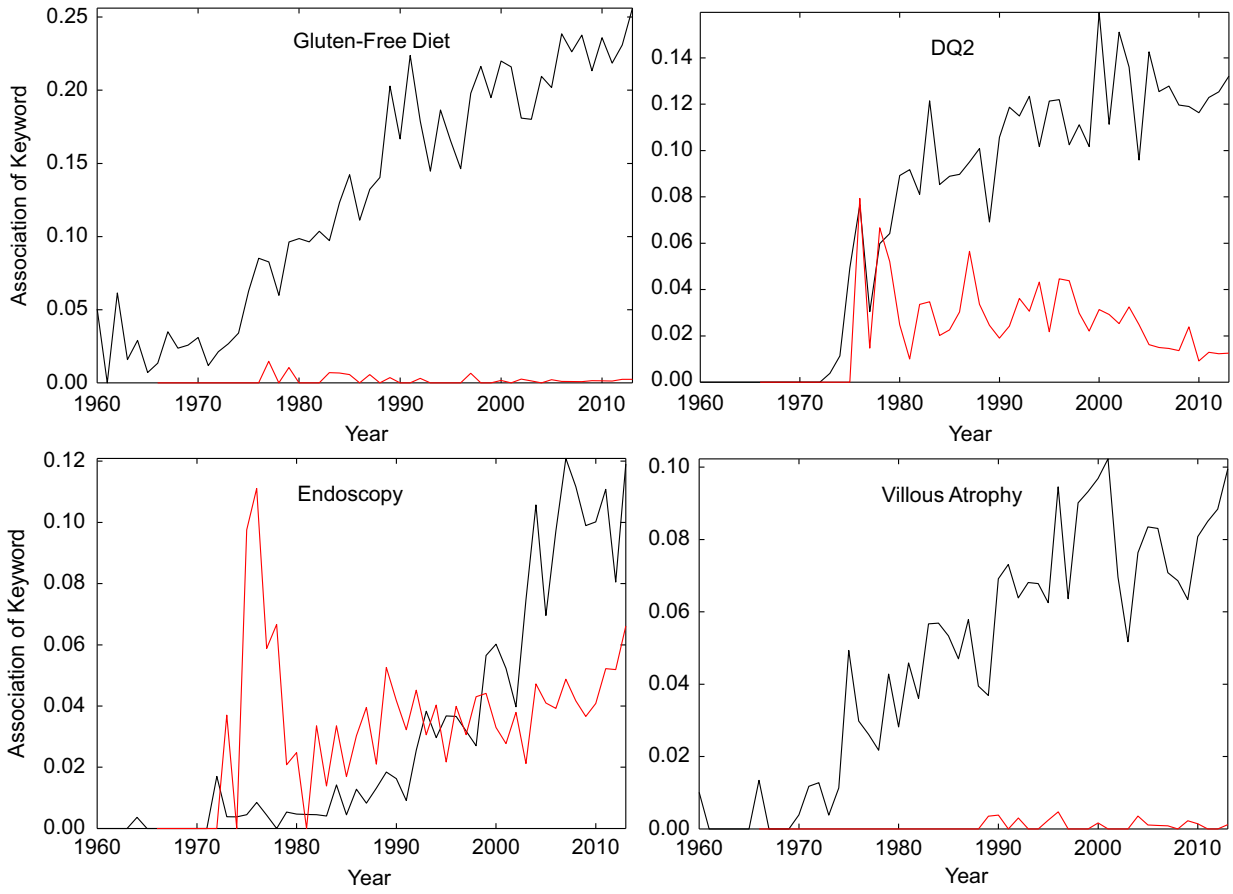

Fig. 4. Keywords commonly associated with celiac disease studies (range of association 0.25–0.1 or 25–10% in 2013). Shown are changes from 1960 to 2013 (black trace). For reference, changes of each keyword with respect to inflammatory bowel disease is also shown (red trace). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

studies), groups were also separated by order of magnitude changes (cutoffs at 0.1 and 0.01).

In Fig. 7, bottom, diagnostic factors ('biopsy', 'serology', 'villous atrophy' and 'videocapsule') and factors pertaining to disease mechanism ('inflammation' and 'autoimmunity' and 'microbiome') are graphed to show the relationships with CD. Since approximately the year 2000, there has been little or no increased interest in diagnostic factors for CD, except that 'videocapsule' has shown a slight uptick since its inception in 2004. There has however, been steadily increasing interest in factors pertaining to autoimmune function in association with CD.

Linear regression parameters for all keywords are provided in Table 3. Shown are the keyword and the linear regression parameters for both the association with CD and with IBD. These are ordered from highest to lowest association with CD in 2013. The strongest trends, in terms of r^2 value above 0.7, are shaded in bold font. For example, for keyword 'diagnosis', the first year of appearance with CD was 1972 and the first year of appearance with IBD was also 1972. There has been a rather strong positive trend of 0.0098 ($\sim$1.0%) increase per year of keyword 'diagnosis' in association with CD in the literature, with a regression coefficient (r^2 value) of 0.904, meaning it is a very linear increase over time, since the closer the r^2 value is to 1.0, the closer the upward trend is to a straight line. As can be observed for the 'diagnosis' keyword graph (Fig. 2, lower left panel) there is an upward trend for association with CD (black trace) since 1972. The starting point for the regression line is 1972 because this is the beginning of the upward trend.

Twelve of the celiac keywords in association with CD had strong trends (bold font): 'diagnosis', 'gluten', 'serology', 'autoimmune', 'treatment', 'gluten-free diet', 'endoscopy', 'villous atrophy', 'wasting', 'inflammation', 'malabsorption', and 'microbiome'. Four keywords in association with IBD had strong trends: 'autoimmune', 'treatment, 'inflammation', and 'microbiome' and these also had a strong association with CD. The keyword 'diagnosis' had a very strong upward trend in association with CD but not IBD. Keyword 'malabsorption' was the only keyword to have a strong downward trend, and it was in association with CD.

Overall, the mean onset time of a keyword trend in association with CD was 1972 while for IBD it was 1980 (Table 3). The mean r^2 value for all trends was 0.516 for keywords in association with CD, and 0.260 for keywords in association with IBD. The mean trend was +0.0021 (0.2% increase per year) for keywords in association with CD, while the mean trend was +0.0005 (0.05% increase per year) for keywords in association with IBD.

4. Discussion

4.1. Summary

In this study the association of 32 keywords typically identified with CD in the medical literature were quantified. For each year from 1960 to 2013, the number of research papers in which each keyword appeared with CD was determined. The data was normalized by scaling by the number of papers in which CD appeared. For reference, the association of the same keywords with IBD was determined, although not all keywords associated with CD are relevant to IBD, for example 'gluten'. There was found to be a sharper increase in IBD papers in the literature, with the number

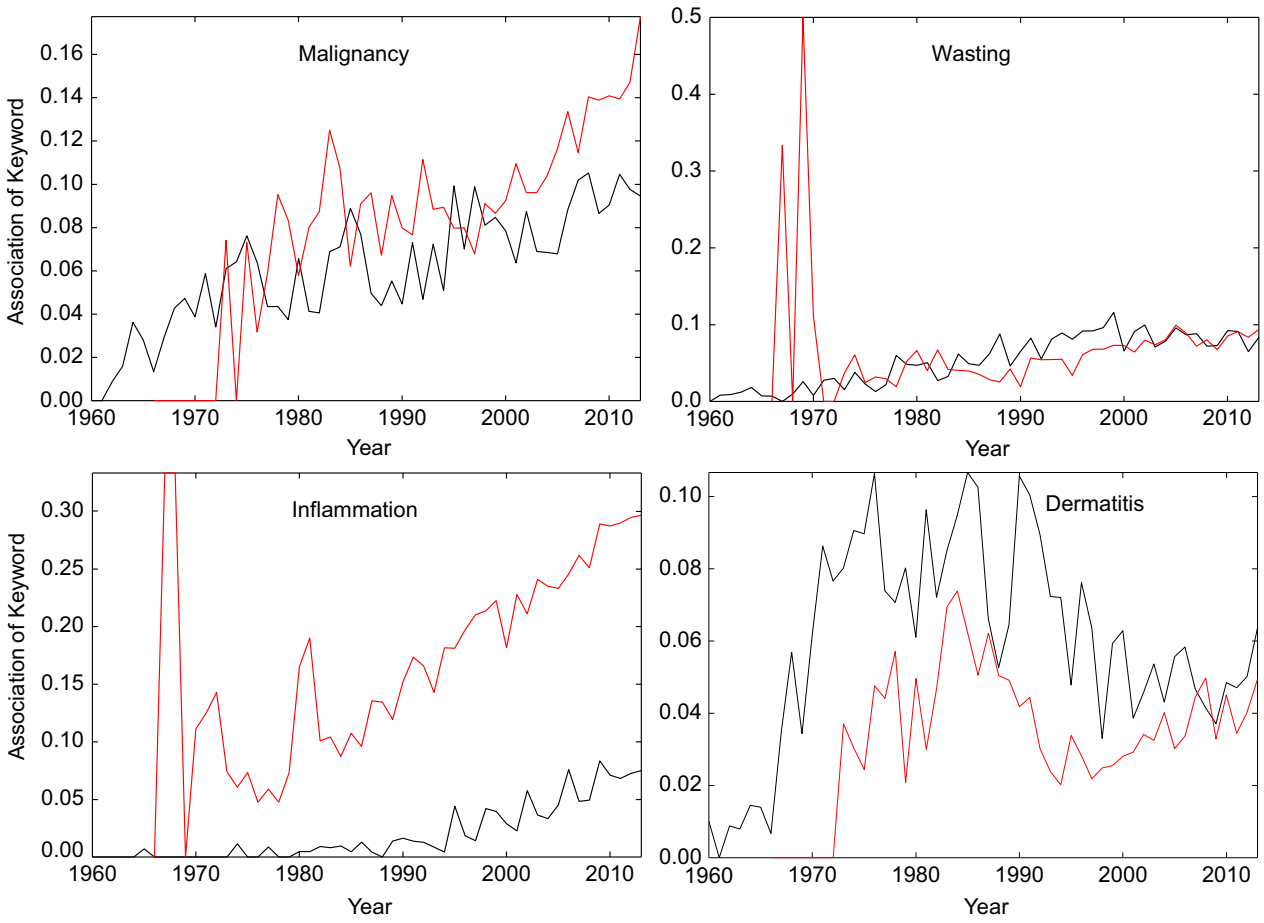

Fig. 5. Keywords moderately associated with celiac disease studies (range of association 0.095–0.06 or 9.5–6%) in 2013. Shown are changes from 1960 to 2013 (black trace). For reference, changes of each keyword with respect to inflammatory bowel disease is also shown (red trace). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of IBD papers published per year overtaking CD papers in 1988. In 2013, there were approximately three times as many IBD papers as CD papers appearing in the medical literature. This finding is somewhat surprising, as CD occurs in about 1% of the population compared to $<0.5\%$ of the population for IBD [12,14].

Other results were graphed as the fraction of CD papers containing an affiliated keyword versus year from 1960 to 2013, and the fraction of IBD papers containing an affiliated keyword versus year from 1960 to 2013 (top 22 keywords). Strong trends in association of a particular keyword with CD or IBD were considered to be those with a trend ≥ 0.002 (0.2%) in magnitude and an r^2 value >0.7. Strong upward trends occurred for 10 of 32 keywords in association with CD. A strong downward trend occurred for keyword 'malabsorption' in association with CD (Table 1), which is not surprising as the clinical face of CD has changed from the malabsorption syndrome to the more atypical presentations [20]. Strong upward trends occurred for 4 of 32 keywords in association with IBD (Table 1). Keywords 'autoimmune', 'treatment, 'inflammation', and 'microbiome' had strong upward trends in association with IBD, and also with CD, consistent with their immune-based pathogenesis. Although keyword 'microbiome' has only been in use for less than a decade, it is of sharply increasing interest to CD and IBD research efforts.

4.2. Evident research trends

There has been a marked increase in published research in both CD and IBD in recent years, as is evident in Fig. 1. The prevalence rate of CD is approximately 1% worldwide [13], while the prevalence of IBD is $<0.5\%$ [12,14]. Thus although IBD is of lesser prevalence than CD, IBD is currently of greater research interest in the medical literature in terms of papers published. Yet, as the rate of diagnosis increases there is still great potential for improvement in diagnosis and treatment of CD [15].

Based on the graphs in Figs. 2–7, the following observations can be made. Affiliated keywords with greatest incidence in conjunction with CD in 2013 (Figs. 2–4 and Table 2, common association keywords) also tend to have linear, strong positive trends (Table 3). These results suggest that the terms are of special interest to CD research, and are becoming more prevalent. From Fig. 2, it appears that gender is clearly of increasing concern in the research of CD, as both terms 'female' and 'male' exhibit positive trends for CD, but not for IBD. This increase may be reflective of gender differences that have been found to exist in celiac disease [16]. The topic of number of specimens submitted during duodenal biopsy is also currently of interest because of the relationship to diagnostic yield [17]. There are significant positive trends for the terms 'autoimmunity', 'treatment, 'inflammation', and 'microbiome' in association with IBD (Table 3). There has been increasing insight into the autoimmune mechanism [18], which likely reflects the strong association of the keyword with IBD. Keyword 'endoscopy' likely exhibited a positive trend in association with IBD (Table 3) since it plays an increasingly important role in diagnosis and management of the disease [19].

There has been a surge of recent interest in some of the keywords that were moderately associated with CD ('malignancy',

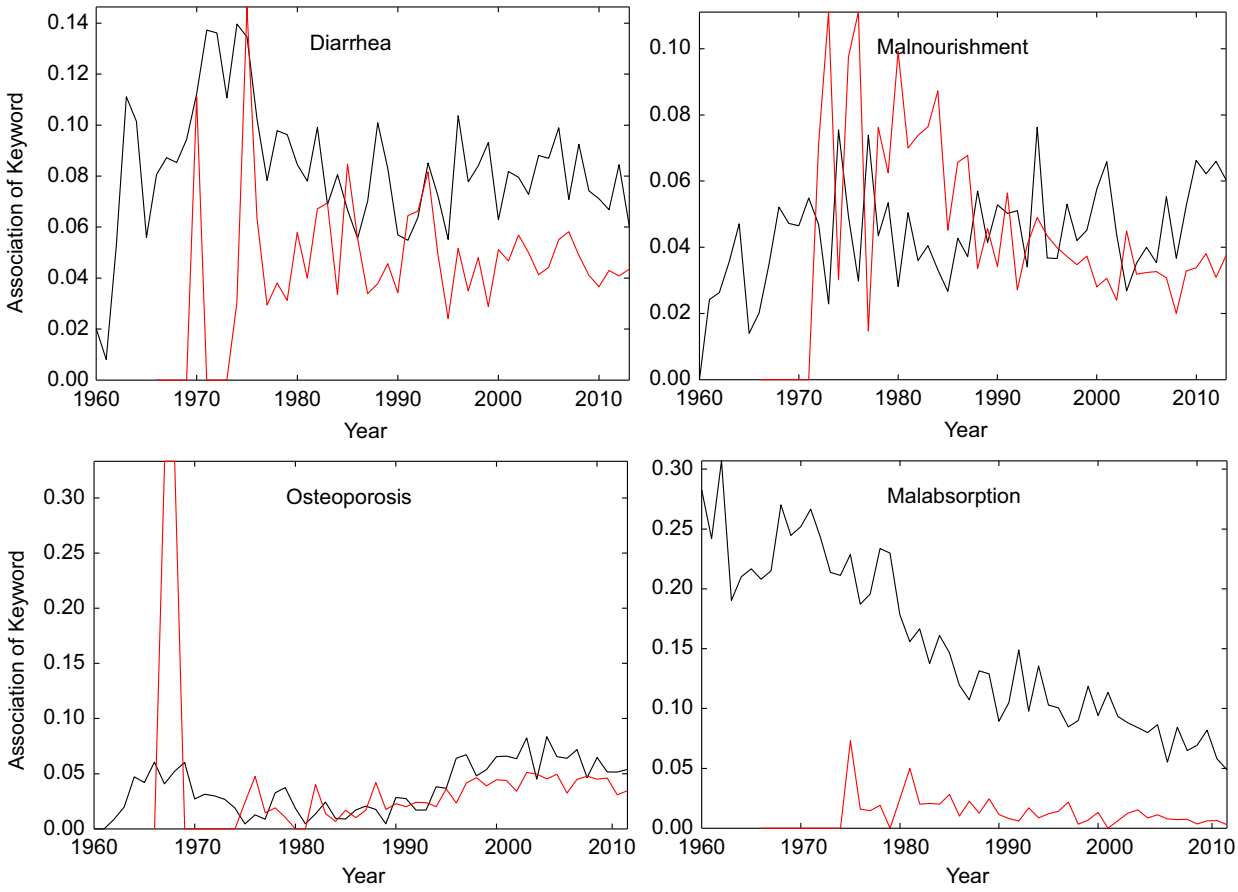

Fig. 6. Keywords moderately associated with celiac disease studies (range of association 0.06–0.05 or 6–5%) in 2013. Shown are changes from 1960 to 2013 (black trace). For reference, changes of each keyword with respect to inflammatory bowel disease is also shown (red trace). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

'malnourishment', 'osteoporosis', 'neurology', 'autism', 'psychiatric', 'fatigue', 'microscopic colitis', 'heart', 'videocapsule'). Several of these keywords exhibited substantial upward trends, which suggests that they are becoming hot topics of CD research. There was a slight negative trend for moderate keyword 'diarrhea' and a strong negative trend for 'malabsorption' in association with CD (Table 3). This is not surprising as the clinical picture of CD has changed, with an increasing proportion of patients lacking the classic features of malabsorption or diarrheal illness [20,21].

Regarding keyword 'microbiome', this term has only appeared in recent years [22,23], so inclusion of subsequent year data points would be helpful to confirm the trend. With respect to keywords with slight association to CD, it is somewhat surprising that 'motility' and 'infertility' are among them (Table 2), as these topics are related to the disease symptoms and sequelae [24–27]. Further investigation is needed in these areas.

4.3. *Practical use of the study findings*

The information contained in this study is best summarized in Table 3. For someone looking for hot topics in CD research, the keywords associated with the bold panels are the ones to note. These show the highly correlated keywords affiliated with CD research (large r^2 values), the magnitude of the trend, and the year that the trend first began leading to the value in 2013. Thus the information provided using the paradigm described in this study is quantitative and unbiased rather than being qualitative and subjective. Based upon the quantitative results, it is evident that some trends began in the 1960s, and have been continually trending in the same direction since then, while others began trending in their present direction only during the last few years. The correlation coefficient r^2 shows the linearity of the trend, and therefore it is a measure of stability. The rate of increase or decrease noted in the trend column is useful to see how rapidly a particular keyword is becoming more or less affiliated with CD research. For several keywords there is a strong correlation both for 'celiac disease' and 'inflammatory bowel disease'. This suggests that scientists doing research on these diseases feel that these particular keywords are affiliated with autoimmune and/or bowel diseases in general, rather than being tied to a specific disease.

This study sought to show relationships and trends between celiac disease and particular areas of research interest. The development of quantitative, computerized information for characterization of CD is a new area of focus that is likely to yield important results in future analyses on celiac disease published research. For example, in a recent review paper by Hegenbart, Uhl, and Vécsei, computer-assisted, quantitative diagnosis of celiac disease from endoscopic imagery is discussed [28]. The first such computer-assisted diagnosis paper was published in 2008, and these studies are becoming increasingly important to improve the diagnosis and monitoring of celiac disease. Many of the quantitative papers focus on image representations, and they have been assistive for improving diagnosis efficacy [28]. However, the small sample sizes in most of the quantitative studies should be validated with larger subject numbers. Furthermore, lens distortion, as well as blur and large camera distances, can contribute to incorrect results, and must be remedied with compensatory measures to further improve diagnostic accuracy [28]. Combining spatio-temporal and textural features may be useful to enhance the robustness of the quantitative systems [28].

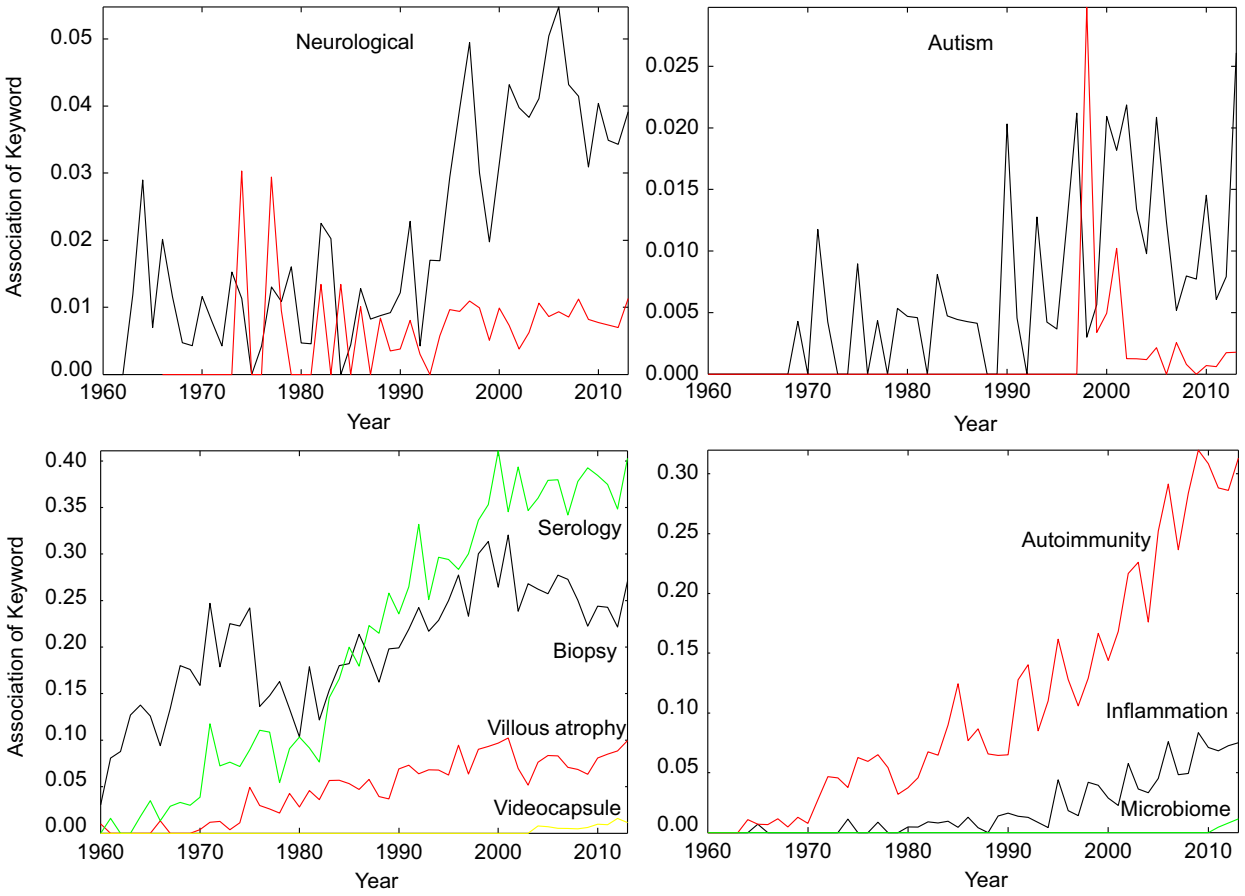

Fig. 7. Top: Keywords moderately associated with celiac disease studies (range of association 0.04–0.025 or 4–2.5%) in 2013. Bottom: Graphs of relationships between diagnostic keywords in association with CD (left) and immune function keywords in association with CD (right).

Table 2
Keywords association with celiac disease (2013).

Common association		Moderate association		Slight association	
Female	0.625	Malignancy	0.095	Motility	0.010
Male	0.607	Wasting	0.083	Dental	0.008
Diagnosis	0.558	Inflammation	0.075	Infertility	0.005
Gluten	0.545	Dermatology	0.064	Lymphocytic colitis	0.005
Serology	0.403	Diarrhea	0.060		
Autoimmune	0.313	Malnourishment	0.060		
Treatment	0.282	Osteoporosis	0.054		
Gluten-free diet	0.256	Malabsorption	0.049		
DQ2	0.132	Neurology	0.039		
Endoscopy	0.119	Autism	0.026		
Biopsy	0.271	Psychiatric	0.020		
Villous atrophy	0.100	Fatigue	0.016		
		Microscopic colitis	0.015		
		Heart	0.013		
		Microbiome	0.011		
		Videocapsule	0.011		

4.4. Limitations

Our study was done using the MEDLINE search tool. Other tools that are available for perusing the medical research literature, such as PubMed and Science Citation Index, were not incorporated. Use of the other search tools may generate slightly different results, as some journals and articles not reflected in one index may appear in another. The data used in this study was measured from year to year. In early years from 1960, there were often few instances of each keyword in affiliation with CD and also in affiliation with IBD; thus fluctuation in the prevalence of these keywords tended to occur early in the studied time interval. The 32 associated keywords were selected as those typically identified with CD, but the list is not necessarily exhaustive, and other keywords may commonly be associated with CD as well. For verification, the work should be repeated in a larger study in which all keywords appearing in the medical literature with CD are quantified.

Although the keyword search presented herein was not complex, it provides a basis for future, more complicated searches. One way to improve on the study would be to include other diseases that may have an association with CD, and to determine the degree of correlation. Another level of complexity would be added

Table 3
Linear regression parameters.

Keyword (includes variants)	Year CD	r^2 CD	Trend CD	Year IBD	r^2 IBD	Trend IBD
Female	1964	0.587	0.0058	1968	0.028	−0.0013
Male	1964	0.370	0.0039	1968	0.003	−0.0005
Diagnosis	**1972**	**0.904**	**0.0098**	1972	0.018	0.0005
Gluten	**1960**	**0.859**	**0.0093**	1979	0.131	−0.0002
Serology	**1961**	**0.937**	**0.0088**	1973	0.563	0.0020
Autoimmune	**1965**	**0.868**	**0.0063**	**1984**	**0.909**	**0.0018**
Treatment	**1990**	**0.929**	**0.0071**	**1971**	**0.903**	**0.0090**
Gluten-free diet	**1960**	**0.906**	**0.0049**	1983	0.123	−0.0001
DQ2	1973	0.617	0.0022	1976	0.283	−0.0007
Endoscopy	**1972**	**0.807**	**0.0029**	1977	0.124	0.0005
Biopsy	1960	0.614	0.0032	1973	0.077	−0.0005
Villous atrophy	**1970**	**0.727**	**0.0018**	1989	0.095	−0.0001
Malignancy	1962	0.657	0.0013	1975	0.562	0.0019
Wasting	**1961**	**0.775**	**0.0018**	1973	0.619	0.0015
Inflammation	**1980**	**0.791**	**0.0023**	**1970**	**0.830**	**0.0053**
Dermatology	1976	0.465	−0.0013	1985	0.115	−0.0005
Diarrhea	1978	0.010	−0.0001	1974	0.052	−0.0004
Malnourishment	1961	0.120	0.0003	1972	0.461	−0.0014
Osteoporosis	1962	0.276	0.0008	1975	0.438	0.0009
Malabsorption	**1960**	**0.862**	**−0.0041**	1975	0.318	−0.0007
Neurology	1963	0.560	0.0008	1977	0.009	0.0001
Autism	1971	0.272	0.0003	1998	0.310	−0.0009
Psychiatric	1968	0.115	0.0002	1983	0.139	0.0002
Fatigue	1990	0.313	0.0005	1990	0.147	0.0002
Microscopic colitis	1993	0.521	0.0006	1985	0.074	0.0001
Heart	1967	0.059	0.0001	1982	0.016	0.0001
Microbiome	**2011**	**1.000**	**0.0034**	**2006**	**0.795**	**0.0035**
Videocapsule	2004	0.469	0.0008	2004	0.001	−0.0001
Motility	1965	0.176	−0.0002	1983	0.035	−0.0001
Dental	1972	0.009	0.0001	1981	0.001	−0.0001
Infertility	1974	0.268	0.0003	1978	0.089	−0.0005
Lymphocytic colitis	1989	0.068	0.0001	1989	0.024	0.0001
Mean	1971.97	0.516	0.0021	1980.22	0.260	0.0005
Standard Deviation	12.98	0.311	0.0031	9.58	0.294	0.0017

A value in trend of $+0.0001$ designates $\leq +0.0001$.
A value in trend of -0.0001 designates ≥ -0.0001.
Bold values are considered significant.
Year: The year in which a continuous trend leading to the value in 2013 began.

by determining the affiliation of two or more keywords with 'celiac disease' rather than just using one keyword. For example, it would be interesting to look at the relationship between keywords 'celiac disease' and both 'autism' and 'dermatitis', and their affiliated keywords, in terms of the number of research studies that have been done, and any trend in this research effort. Thirdly, the importance of an association could be ranked not just by correlation, but by weighting the impact factors of journals in which the studies were published.

5. Conclusions

The MEDLINE search technique was shown to be helpful to determine the relationship between keywords typically identified with CD in the medical literature. A technique that was developed to quantify first instance, association, and trends of keywords with CD in literature studies was successful for characterization of the relationships. From the graphs in Figs. 2–7, and the linear regression parameters of Table 3, it is evident that strong associations and strong trends exist in the affiliation of CD with certain keywords, and there are even a few strong trends and associations of some of these same keywords with IBD. By using this method of querying research topics in CD, one can see the differences in the clinical presentation of CD versus that of a malabsorption syndrome. One can also observe the trends in diagnostic methods and pathogenesis. The research also shows similarities between CD and IBD. Although there is approximately three times the volume of research in IBD compared to CD, the differences are quite prominent. There is more research on 'treatment' of IBD (Fig. 3), reflecting a major issue. Currently the gluten-free diet is the only therapy available for CD. However there is an interest among patients with CD for a drug therapy [29], and there are several drugs being developed [30]. One should see an increase in published research as a result. A surprising similarity was seen in the level of published research for keyword 'malignancy' in terms of similar significant trend (Table 3) for both CD and IBD. In clinical practice however, IBD patients, but not CD patients, are frequently screened for malignancy [31].

Conflicts of interest

None of the authors have conflicts of interest.

References

[1] J.F. Ludvigsson, J.C. Bai, F. Biagi, T.R. Card, C. Ciacci, P.J. Ciclitira, P.H. Green, M. Hadjivassiliou, A. Holdoway, D.A. van Heel, K. Kaukinen, D.A. Leffler, J.N. Leonard, K.E. Lundin, N. McGough, M. Davidson, J.A. Murray, G.L. Swift, M.M. Walker, F. Zingone, D.S. Sanders, Authors of the BSG Celiac Disease Guidelines Development Group, Diagnosis and management of adult celiac disease: guidelines from the British Society of Gastroenterology, Gut 63 (2014) 1210–1228.

[2] S. Smukalla, B. Lebwohl, J.G. Mears, L.A. Leslie, P.H. Green, How often do hematologists consider celiac disease in iron-deficiency anemia? Results of a national survey, Clin. Adv. Hematol. Oncol. 12 (2014) 100–105.

[3] E.D. Braunstein, R. Rosenberg, F. Gress, P.H. Green, B. Lebwohl, Development and validation of a clinical prediction score (the SCOPE score) to predict sedation outcomes in patients undergoing endoscopic procedures, Aliment. Pharmacol. Ther. 40 (2014) 72–82.

[4] H. Abu Daya, B. Lebwohl, S. Smukalla, S.K. Lewis, P.H. Green, Utilizing HDL levels to improve detection of celiac disease in patients with iron deficiency anemia, Am. J. Gastroenterol. 109 (2014) 769–770.

[5] B. Lebwohl, F. Granath, A. Ekbom, K.E. Smedby, J.A. Murray, A.I. Neugut, P. H. Green, J.F. Ludvigsson, Mucosal healing and risk for lymphoproliferative malignancy in celiac disease: a population-based cohort study, Ann. Intern. Med. 159 (2013) 169–175.

[6] B. Lebwohl, H. Eriksson, J. Hansson, P.H. Green, J.F. Ludvigsson, Risk of cutaneous malignant melanoma in patients with celiac disease: A population-based study, J. Am. Acad. Dermatol. 71 (2014) 245–248.

[7] G.K. Makharia, C.J. Mulder, K.L. Goh, V. Ahuja, J.C. Bai, C. Catassi, P.H. Green, S.D. Gupta, K.E. Lundin, B.S. Ramakrishna, R. Rawat, H. Sharma, A. Sood, C. Watanabe, P.R. Gibson, World Gastroenterology Organization-Asia Pacific Association of Gastroenterology Working Party on Celiac Disease. Issues associated with the emergence of celiac disease in the Asia–Pacific region: a working party report of the World Gastroenterology Organization and the Asian Pacific Association of Gastroenterology, J. Gastroenterol. Hepatol. 29 (2014) 666–677.

[8] D. Narotsky, P.H. Green, B. Lebwohl, Temporal and geographic trends in celiac disease publications: a bibliometric analysis, Eur. J. Gastroenterol. Hepatol. 24 (2012) 1071–1077.

[9] S. Master, B. Lebwohl, J.F. Ludvigsson, P.H. Green, Bibliometric study of the quality of celiac disease research publications, J. Pediatr. Gastroenterol. Nutr. 57 (2013) 527–528.

[10] Celiac Disease. Nutrition Reviews 17, 1959, pp. 296–297.

[11] B. Lebwohl, C.A. Tennyson, J.L. Holub, D.A. Lieberman, A.I. Neugut, P.H. Green, Sex and racial disparities in duodenal biopsy to evaluate for celiac disease, Gastrointest. Endosc. 76 (4) (2012) 779–785.

[12] E.V. Loftus Jr., Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences, Gastroenterology 126 (2004) 1504–1517.

[13] J.M. Choi, B. Lebwohl, J. Wang, S.K. Lee, J.A. Murray, M.V. Sauer, P.H. Green, Increased prevalence of celiac disease in patients with unexplained infertility in the United States, J. Reprod. Med. 56 (2011) 199–203.

[14] P.L. Lakatos, Recent trends in the epidemiology of inflammatory bowel diseases: up or down? World J. Gastroenterol. 12 (2006) 6102–6108.

[15] R.P. Anderson, M.J. Henry, R. Taylor, E.L. Duncan, P. Danoy, M.J. Costa, K. Addison, J.A. Tye-Din, M.A. Kotowicz, R.E. Knight, W. Pollock, G.C. Nicholson, B.H. Toh, M.A. Brown, J.A. Pasco, A novel serogenetic approach determines the community prevalence of celiac disease and informs improved diagnostic pathways, BMC Med. 11 (2013) 188.

[16] D. Bai, P. Brar, S. Holleran, R. Ramakrishnan, P.H. Green, Effect of gender on the manifestations of celiac disease: evidence for greater malabsorption in men, Scand. J. Gastroenterol. 40 (2005) 183–187.

[17] B. Lebwohl, R.C. Kapel, A.I. Neugut, P.H. Green, R.M. Genta, Adherence to biopsy guidelines increases celiac disease diagnosis, Gastrointest. Endosc. 74 (2011) 103–109.

[18] C.H. Lin, S. Kadakia, M. Frieri, New insights into an autoimmune mechanism, pharmacological treatment and relationship between multiple sclerosis and inflammatory bowel disease, Autoimmun. Rev. 13 (2014) 114–116.
[19] R. Rameshshanker, N. Arebi, Endoscopy in inflammatory bowel disease when and why, World J. Gastrointest. Endosc. 4 (2012) 201–211.
[20] S.D. Rampertab, N. Pooran, P. Brar, P. Singh, P.H. Green, Trends in the presentation of celiac disease, Am. J. Med. 119 (4) (2006) 355, e9–14.
[21] K.D. Katz, S. Rashtak, B.D. Lahr, L.J. Melton 3rd, P.K. Krause, K. Maggi, N.J. Talley, J.A. Murray, Screening for celiac disease in a North American population: sequential serology and gastrointestinal symptoms, Am. J. Gastroenterol. 106 (2011) 1333–1339.
[22] X.C. Morgan, T.L. Tickle, H. Sokol, D. Gevers, K.L. Devaney, D.V. Ward, J.A. Reyes, S.A. Shah, N. LeLeiko, S.B. Snapper, A. Bousvaros, J. Korzenik, B.E. Sands, R.J. Xavier, C. Huttenhower, Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment, Genome Biol. 13 (2012) R79.
[23] M. Sellitto, G. Bai, G. Serena, W.F. Fricke, C. Sturgeon, P. Gajer, J.R. White, S.S. Koenig, J. Sakamoto, D. Boothe, R. Gicquelais, D. Kryszak, E. Puppa, C. Catassi, J. Ravel, A. Fasano, Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants, PLoS One 7 (2012) e33387.
[24] A. Tursi, Gastrointestinal motility disturbances in celiac disease, J. Clin. Gastroenterol. 38 (2004) 642–645.
[25] L. Elli, M.T. Bardella, Motility disorders in patients with celiac disease, Scand. J. Gastroenterol. 40 (2005) 743–749.
[26] H.J. Freeman, Reproductive changes associated with celiac disease, World J. Gastroenterol. 16 (2010) 5810–5814.
[27] S. Soni, S.Z. Badawy, Celiac disease and its effect on human reproduction: a review, J. Reprod. Med. 55 (2010) 3–8.
[28] S. Hegenbart, A. Uhl, A. Vécsei, Survey on computer aided decision support for diagnosis of celiac disease, Comput. Biol. Med. 65 (2015) 348–358.
[29] C.A. Tennyson, S. Simpson, B. Lebwohl, S. Lewis, P.H. Green, Interest in medical therapy for celiac disease, Therap. Adv. Gastroenterol. 6 (5) (2013) 358–364.
[30] C.A. Tennyson, S.K. Lewis, P.H. Green, New and developing therapies for celiac disease, Therap. Adv. Gastroenterol. 2 (2009) 303–309.
[31] H. Tavakkoli, S. Haghdani, H. Adilipour, H. Daghaghzadeh, M. Minakari, P. Adibi, K. Ahmadi, M.H. Emami, Serologic celiac disease in patients with inflammatory bowel disease, J. Res. Med. Sci. 17 (2) (2012) 154–158.