



# Trends in gluten research and its relationship to autoimmune and allergic diseases



Edward J. Ciaccio<sup>a,\*</sup>, Govind Bhagat<sup>a,b</sup>, Suzanne K. Lewis<sup>a</sup>, Peter H. Green<sup>a</sup>

<sup>a</sup> Celiac Disease Center, Department of Medicine, Columbia University Medical Center, NY, USA

<sup>b</sup> Department of Pathology and Cell Biology, Columbia University Medical Center, NY, USA

## ARTICLE INFO

### Article history:

Received 13 April 2016

Received in revised form

16 May 2016

Accepted 25 May 2016

Available online 7 June 2016

### Keywords:

Celiac disease

Gluten

Gliadin

Medline

Transglutaminase

## ABSTRACT

**Introduction:** Gluten is a protein found in certain grains, and causes an autoimmune response in celiac disease patients. Although the subject of considerable research, gluten research foci and future directions are largely unknown.

**Methods:** The MEDLINE search tool was used to evaluate research trends. For perspective, yearly publications on gluten and other celiac disease reactive proteins were compared to food allergy proteins research. Then the relationship of gluten publications to affiliated keywords was determined. The affiliated keywords belonged to one of several groups: grains, vitamins and minerals, interaction, autoimmunity, genetics, or enzymes. The yearly number of publications in the peer-reviewed medical literature was determined for each relationship from years 1960–2013. The relationships were graphed, and linear regression analysis was used to determine the rate of change in publications per year, and the coefficient of determination.

**Results:** Among celiac disease reactive proteins (gluten, gliadin, and glutenin) and selected food allergy proteins (ovalbumin, lysozyme, ovomucoid, and Ara h), gluten showed the greatest rate of increase in published medical research (+20.01 studies/year since 1996,  $r^2=0.97$ ). Additionally, there were sharp increases in the rate of gluten research publications per year in association with keywords 'autoimmunity' (+7.69 studies/year since 1997), 'wheat' (+6.08 studies/year since 1999), and 'transglutaminase' enzyme (+5.05 studies/year since 1995). The longest running moderate trend was research on 'gluten' and 'antibodies' (+2.50 studies/year,  $r^2=0.92$  since 1971).

**Conclusions:** Research on gluten as a reactive protein is of rapidly growing interest in the medical literature. MEDLINE is helpful to determine foci and future directions.

© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Contents

1. Introduction.....	8
2. Method.....	8
3. Results.....	10
3.1. Summary statistics.....	12
4. Discussion.....	13
4.1. Summary.....	13
4.2. Use of the selected keywords.....	14
4.3. Other studies.....	14
4.4. Limitations and future directions.....	14
Acknowledgement.....	14
References.....	14

\* Correspondence to: P&S 7-446, Columbia University, 630 West 168th Street, New York, NY 10032, USA.

E-mail address: [ciaccio@columbia.edu](mailto:ciaccio@columbia.edu) (E.J. Ciaccio).

### 1. Introduction

Celiac disease is a common health concern throughout the world, affecting approximately 1% of the population [1]. It is an autoimmune disease [2], with the reactive substrate being specific proteins found in the grains wheat, rye, and barley. When these grains are consumed, there is an autoimmune response, at the level of the small intestinal mucosa, to the protein gluten, and its component proteins gliadin and glutenin [3]. Research on gluten and celiac disease is important to determine the mechanisms of the disease and possible treatments [4].

In prior work we have used the MEDLINE search tool to determine research directions and foci in the area of celiac disease from the peer-reviewed medical literature [5]. In the current study we analyze the major reactive protein in celiac disease, gluten, to determine the foci and future directions of the research, with respect to other facets of celiac disease. This is done by pairing keyword ‘gluten’ with associated keywords used in celiac disease research. For perspective, the published literature on gluten is compared with its component proteins which are also reactive in celiac disease (gliadin and glutenin) and to selected proteins reactive in food allergies, which are an increasing clinical problem and of increasing interest in medical research.

### 2. Method

The MEDLINE search tool was used for all comparisons and analyses [6]. The search checks the title of the paper, the abstract, and the keywords selected by the authors of the paper. The MEDLINE tool was initialized to include multimedia references, and so as not to map the term to the subject heading. The limits used involved only the publication year. First the total number of publications from the years 1960–2015 inclusive was determined, for the following keywords (each noted by single quotation marks):

Proteins reactive in celiac disease - ‘gluten’, ‘gliadin’, ‘glutenin’  
 Proteins reactive in selected food allergies - ‘ovalbumin’, ‘lysozyme’, ‘ovomucoid’, and ‘Ara h’

‘Ara h’, a protein reactive in peanut allergy, was searched as: (‘Ara h’ or ‘Ara h1’ or ‘Ara h2’ or ‘Ara h3’).

MEDLINE treats each of these as distinct entities. The number of research publications per year for all reactive proteins was then determined via a MEDLINE search for each year, and graphed for the years 1960–2013. The year 2013 was used as an endpoint because it was the last year at the time of the study for which all research publications had been accounted for. Once the graphs of reactive protein publications were constructed, linear regression analysis (SigmaPlot, 2013, Systat Software, San Jose, CA) was used to determine the rate of change in the number of publications per year, i.e., the slope of the regression line, and also the coefficient of determination or  $r^2$  value (fit to a straight line). The first data point in each keyword graph used for regression analysis was the last year in which the number of publications which included the keyword was equal to zero. If for a particular keyword, there were research publications for all years from 1960 to 2013, then the year 1960 was used as the starting point for regression analysis. An  $r^2$  value of less than 0.8 was used to indicate a lesser fit to the straight line. For any such keywords, if a later portion of the graph in time had an evident better fit to a straight line, it was used as a separate linear regression model for analysis.

The degree of cross-publication between major reactive protein keywords was also determined, i.e., the following searches were made for the years 1960–2013:

**Table 1**  
Total number of studies involving reactive protein and their association with terms used in celiac disease.

Protein	Total	Wheat	Rye	Barley	Oats	Rice	Vitamins	Calcium	Iron	def.	Sensitivity	Intolerance	Allergen	Inflammation	Antibodies	Autoimmune	B cell	T cell	IgA	IgE	tTG	Enzyme	HLA
Gluten	8078	1704	295	340	135	136	96	230	325		911	561	305	851	2019	1589	87	663	991	104	906	923	840
Gliadin	3082	944	157	152	62	39	10	43	73		555	143	194	299	1650	716	38	331	866	143	605	696	423
Glutenin	614	547	32	37	6	28	0	0	1		18	2	46	5	57	20	3	17	6	40	13	56	13
Ovalbumin	21,468	101	8	7	2	18	20	478	32		744	21	5674	5371	6236	5889	996	5191	492	2774	18	2651	108
Lysozyme	20,477	61	2	21	1	27	33	546	60		1035	4	198	1129	2484	2029	548	821	518	102	19	4066	139
Ovomucoid	835	42	1	4	0	2	2	18	0		34	0	145	7	131	46	7	15	7	103	1	243	4
Ara h	393	1	0	0	0	3	0	2	2		53	0	265	6	115	49	6	27	4	252	2	77	4

(gluten or gliadin or glutenin) and (“Ara h” or “Ara h1” or “Ara h2” or “Ara h3”)  
 (gluten or gliadin or glutenin) and (ovalbumin or lysozyme or ovomucoid)  
 (“Ara h” or “Ara h1” or “Ara h2” or “Ara h3”) and (ovalbumin or lysozyme or ovomucoid)

Then keywords commonly associated with gluten, categorized according to one of several topics, were used for comparison of

publication trends. The following 21 keywords were searched in conjunction with (gluten or gliadin or glutenin), separated into six categories:

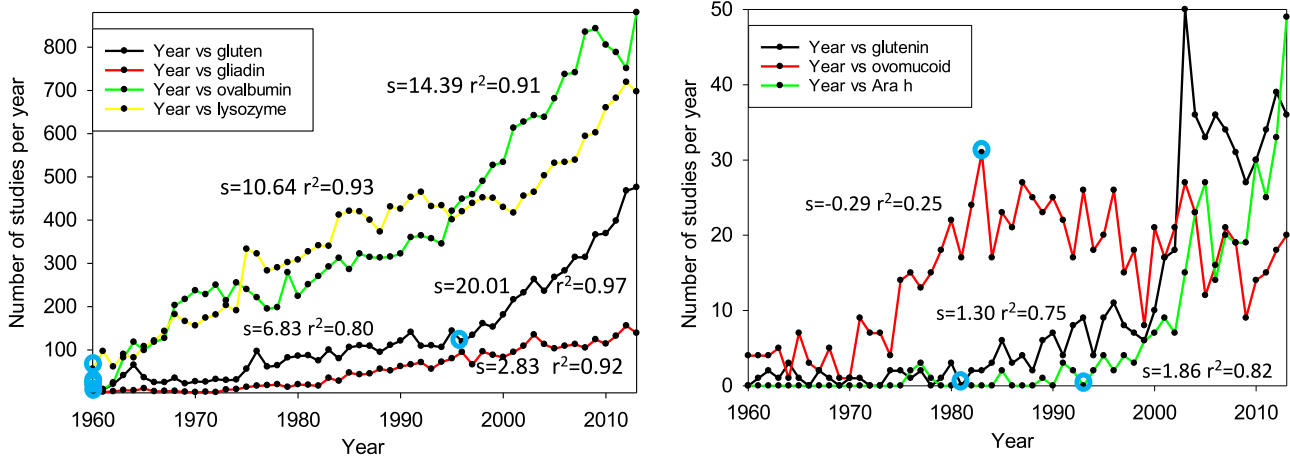
Grains – wheat, rye, barley, oats, rice

Vitamins and minerals – vitamins, calcium, ‘iron deficiency’

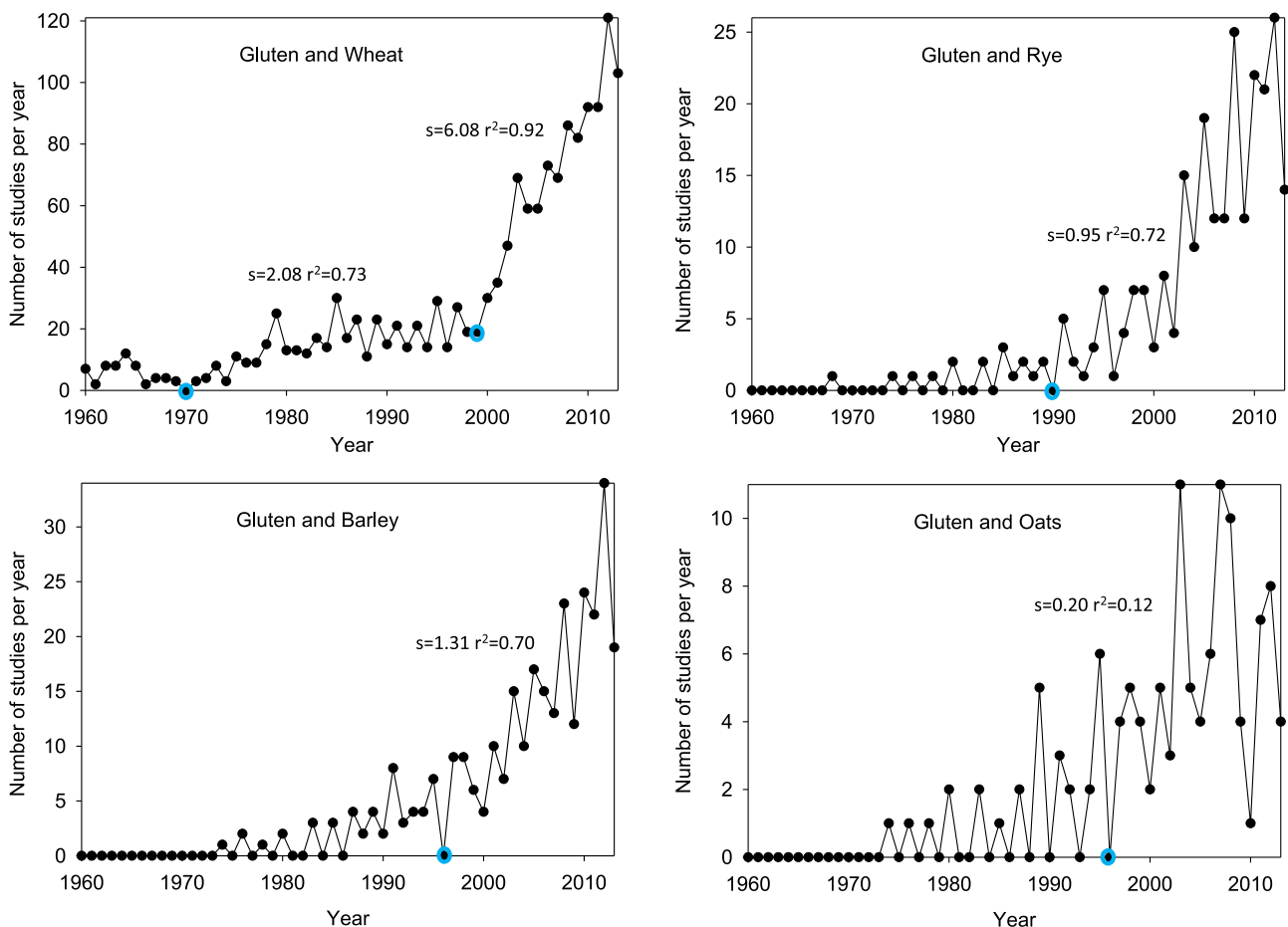
Interaction – sensitivity, intolerance, allergen, inflammation

Immunology – antibodies, autoimmune, B cell, T cell, IgA, IgE

Enzymology – transglutaminase, enzyme



**Fig. 1.** Reactive proteins – publications per year. The rate of change in publications per year ( $s$ ) and goodness of fit to a straight line ( $r^2$ ) are noted. Onset point for each linear regression model is noted with a blue circle. Left panel. proteins with large changes per year. Right panel. Proteins with lesser changes per year, magnified on ordinate axis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 2.** The number of biomedical publications per year for research on gluten in combination with several grains: wheat, rye, barley, and oats.

## Genetics – HLA

where HLA is the human leukocyte antigen. The variants that were searched for included both the singular and plural versions of each associated keyword, when applicable, because MEDLINE distinguishes between these entities. Furthermore, for several keywords, multiple variants of the above list were used in the search term, i.e.:

(anemia or anemic or 'iron deficiency').  
 (transglutaminase or tTG).  
 (allergy or allergies or allergic or allergen).  
 (inflammation or inflammatory).  
 (autoimmune or autoimmunity or immune).

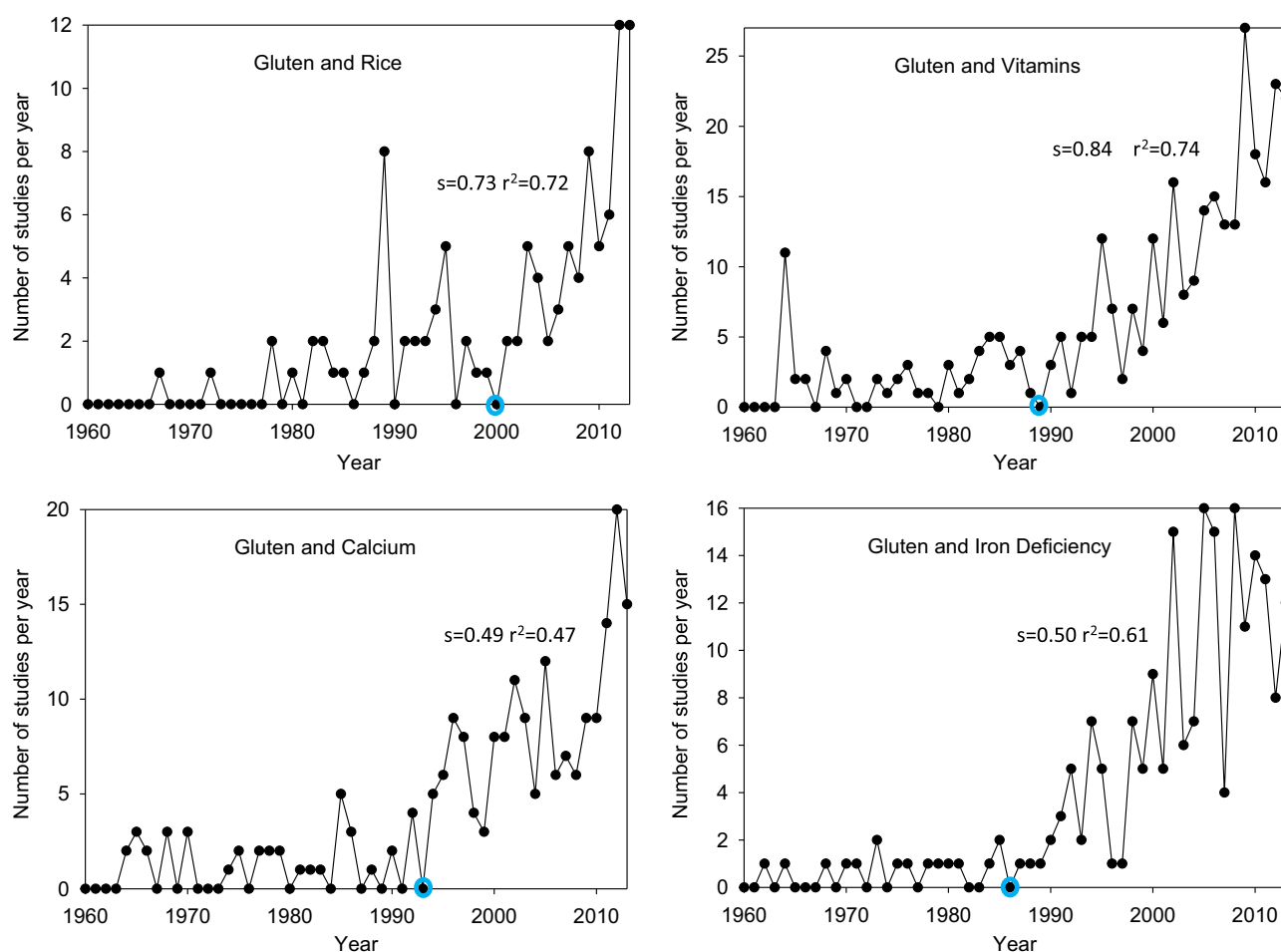
The rate of change in the number of publications per year, the goodness of fit to a straight line, and the starting year of the trend from each graph were tabulated for comparison.

## 3. Results

In Table 1 is shown the total number of publications with keyword instances from years 1960–2015 inclusive. The total number of publications for selected reactive proteins that were analyzed in the study are noted at left. Also shown are the total number of instances of all associated keywords affiliated in

combination with each of the proteins. Thus for example, affiliated keyword 'allergen' is found in combination with the celiac disease reactive proteins gluten (305 total publications), gliadin (194) and glutenin (46) but also with egg allergy reactive proteins ovalbumin (5674 studies), lysozyme (198 studies) and ovomucoid (145 studies) and peanut allergy proteins (265 studies). The most published research has been on the reactive proteins in egg allergies 'ovalbumin' (21,468) and 'lysozyme' (20,477), followed by the major celiac disease protein gluten (8078) and its component protein gliadin (3082), with lesser research studies being published for the celiac disease component protein glutenin (614), egg allergy protein ovomucoid (835), and peanut allergy proteins Ara h (393). Although 'wheat' is normally associated with celiac disease reactive proteins gluten (1704 studies), and component proteins gliadin (944 studies) and glutenin (547 studies) it also has associations with egg reactive proteins ovalbumin (101 studies), lysozyme (61 studies), and ovomucoid (42 studies), as well as peanut reactive proteins Ara h (1 study). Throughout the table, it is evident that many research publication keywords affiliated with celiac disease are also affiliated with food allergies, albeit to a lesser extent in some instances.

The graphs of reactive protein publications per year are provided in Fig. 1. Two graphs are shown; in panel A those proteins with the largest publication rates of change are shown and in panel B, those proteins with lesser rates of change are given for clarity. Next to each trace is the rate of change, or slope ( $s$ ), and the goodness of fit to a straight line ( $r^2$ ). A perfect straight line fit



**Fig. 3.** The number of biomedical publications per year for research on gluten in combination with rice, vitamins, calcium, and iron deficiency (includes auxiliary keywords anemic and anemia).

would yield an  $r^2$  value of 1.0. For none of the major reactive protein traces (gluten, gliadin, ovalbumin, and lysozyme) were there zero publications for any year; thus the starting point for the regression line of each is the year 1960. However, the  $r^2$  value for the gluten trace, 0.80, was substantially less than that of the other traces, which were all above 0.90. Thus another data point was used to determine if a more linear model could be obtained for the increase in gluten publications, with a begin time of 1996. The model was substantially more linear ( $r^2=0.97$ ) and the rate of increase in publications per year increased to 20.01, indicating the greatest rate of increase for this protein. For those proteins with lesser rate of change, shown in panel B, trends were more difficult to ascertain owing to the lesser number of publications per year, and the relatively large degree of fluctuations per year. The last year with zero publications for proteins glutenin and Ara h was 1981 and 1993, respectively. For the protein ovomucoid, there was actually a decreasing trend, so that the maximum peak was used as the starting point for the regression analysis model.

A comparison of the main keyword “gluten” with affiliated keywords is shown in the remaining figures. The results for several grains in association with “gluten” are provided in Fig. 2. There was a marked increase in research on gluten and wheat in combination since 1970, with the initial linear regression model showing an increase of +2.08 publications per year. However, the goodness of fit was only 0.73 for this model. There has been an even steeper increase in publications since 1999, with a rate of +6.08 publications per year ( $r^2=0.92$ ). Lesser increases in publications per year occurred for grains rye (0.95/year), barley (1.31/year), and oats (0.20/year). The association of gluten and wheat

had the most total publications for any year, 120, followed by a moderate number of publications for rye and barley, 26 and 34, and a lesser number for oats, 12. In Fig. 3A, it is evident that there has been an increase in publications on gluten and rice in combination, up to a maximum of 12 in 2012 and 2013. Regarding the category of vitamins and minerals, shown in Fig. 3B–D, there has been a lesser and fluctuating increase in publications on vitamins, calcium, and iron deficiency in combination with gluten. Still, there is an increase in each case beginning about 1990 (Fig. 3).

Moderate increases in publications on gluten in combination with associated keywords sensitivity, intolerance, allergen, and inflammation, have occurred (Fig. 4) of between +1 and +4 increase in number of publications per year. There have been steady increases in the number of publications on gluten in combination with keywords in the immunology group (Fig. 5). The increase in publications on gluten and antibodies has been linear ( $r^2=0.92$ ) and relatively large (+2.50/year) since 1971. An even greater increase in publications involving gluten and autoimmunity has occurred since 1997 (+7.69/year,  $r^2=0.96$ ). There has been a steady increase in research papers on gluten and IgA since 1970, while there has been an increase in papers on gluten and IgE since 1993. It is evident that there have also been substantial recent increases in published research on gluten in combination with enzymes (+2.50,  $r^2=0.77$ , 1997) and in combination with tTG (+5.05/year,  $r^2=0.96$ , 1995) (Fig. 6). There is a moderate increase in published research on gluten in combination with HLA.

Finally, in Fig. 7, the number of publications in the biomedical literature with both a celiac disease reactive protein (gluten or

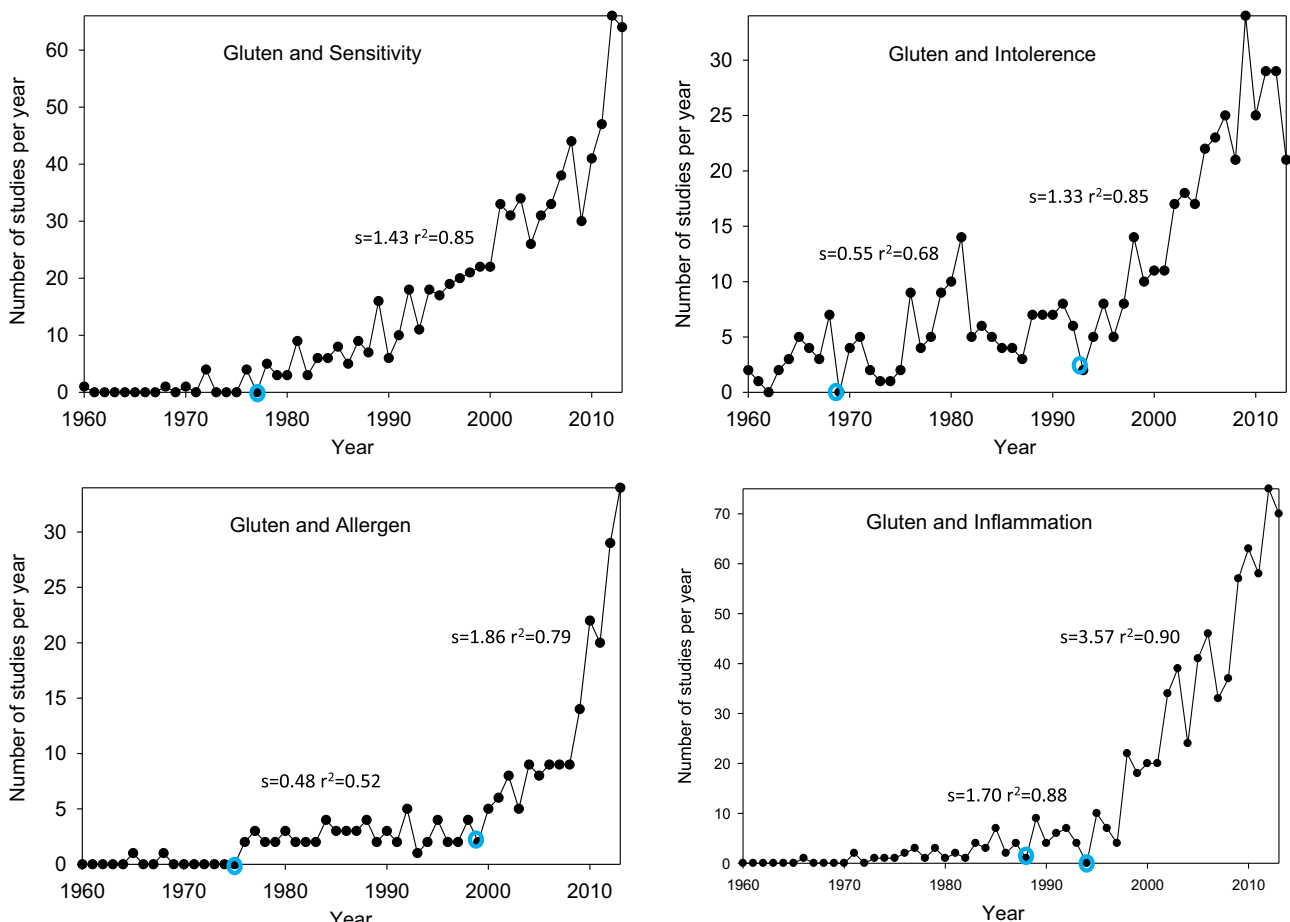
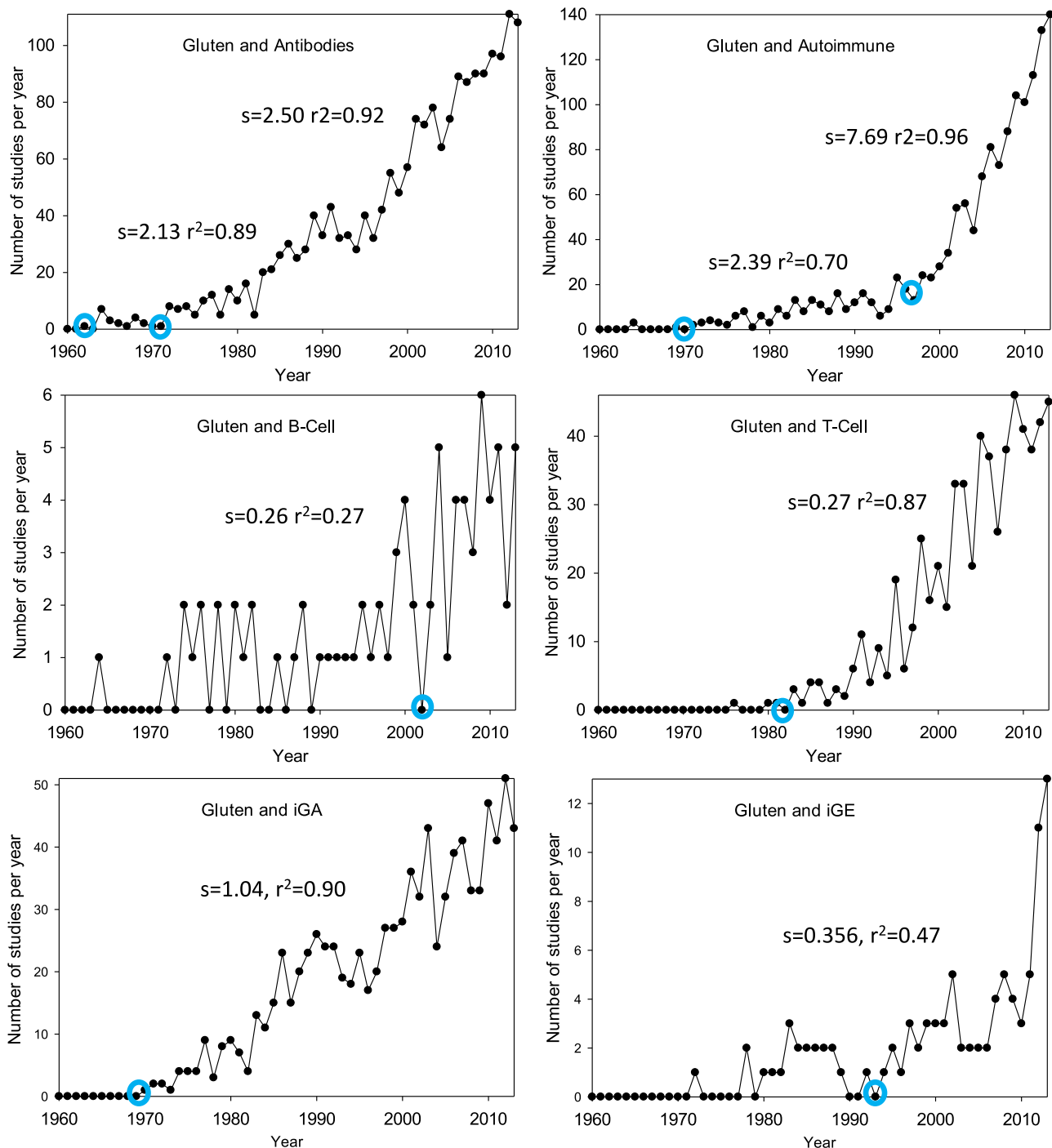


Fig. 4. The number of biomedical publications per year for research on gluten in combination with sensitivity, intolerance, allergen, and inflammation.



**Fig. 5.** Top four panels: The number of biomedical publications per year for research on gluten in combination with the immunology group of associated keywords antibodies, autoimmune, B cells, and T cells. Lower panels: The number of biomedical publications per year for research on gluten in combination with the immunology group of associated keywords IgA, IgE.

gliadin or glutenin) and an egg allergy protein (ovalbumin or lysozyme or ovomucoid) are shown. It is rather striking that in the late 1980s and early 1990s, a peak in this affiliated research occurred. However, this peak in 1988 consisted of only 10 publications in total. In contrast, there was only one article published in the biomedical literature in which both celiac disease and peanut allergy reactive proteins were mentioned (1997), while six studies in total which both egg and peanut allergy reactive proteins are mentioned appear in the published biomedical literature (occurring in 2013, 2007, 2004 (2), 1999, and 1994).

### 3.1. Summary statistics

The results of the study are summarized in Table 2, separated by topic. Shown are the keyword, year of first instance for linear regression model, the rate of change, and the  $r^2$  value. The table shows that there has been much research interest in gluten as a reactive protein in celiac disease, and that the number of research publications has been rapidly increasing in recent years. Most of the onset years for rapid linear increase in publications were in the 1990s, although interest in the association of rice with gluten, and B cells



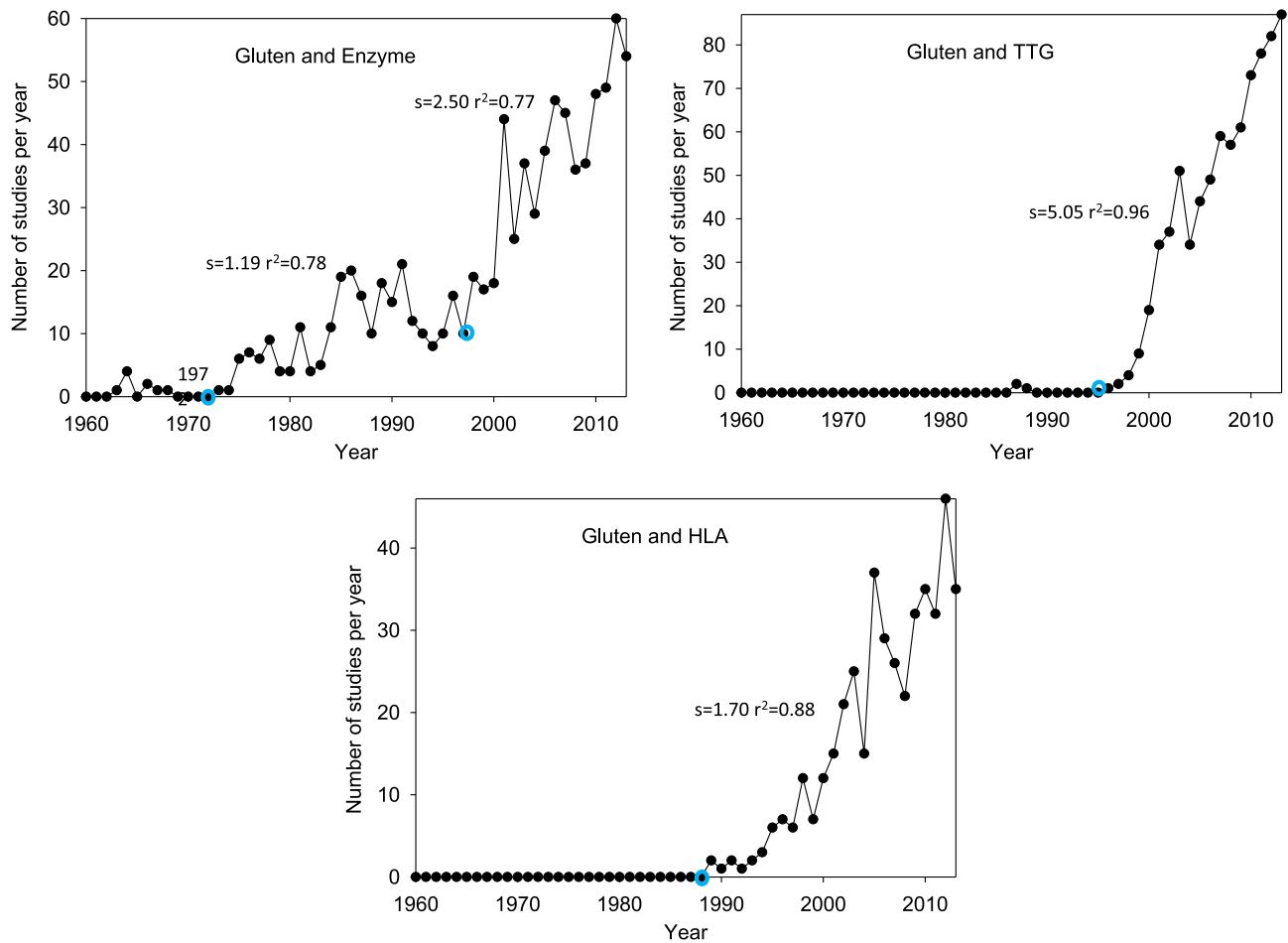


Fig. 6. The number of biomedical publications per year for research on gluten in combination with enzymes, transglutaminase, and HLA.

with gluten, made the most dramatic upswings beginning in 2000 and 2002, respectively. A few keywords have had a long association with gluten, including antibodies (1971) and sensitivity (1977).

## 4. Discussion

### 4.1. Summary

In this study, the focus and trends in published biomedical research on gluten as a reactive protein in celiac disease was investigated. The MEDLINE search tool was used to estimate the relationship. First, for perspective, proteins reactive in celiac disease were compared with proteins reactive in food allergies. Although the number of research studies for specific food allergy proteins ovalbumin and lysozyme, was greater than the celiac related proteins, the rate of increase in these published studies per year was not as much as the increase in published studies on gluten per year. Thus from Fig. 1, the rate of increase in studies involving the protein gluten increased at a rate of +20.01 per year since 1996, with a goodness of fit  $r^2=0.97$ . Although the total number of published research efforts involving ovalbumin and lysozyme in 2013 was larger, their rate of increase was less: +14.39/year for ovalbumin ( $r^2=0.91$ ) and +10.64/year for lysozyme ( $r^2=0.93$ ). Therefore, the number of published studies involving gluten is overtaking those for the food allergy proteins, i.e., there is apparently a surpassing interest in gluten and its relationship to celiac disease as a biomedical research topic, when compared to classical food allergies. Associated keywords that are

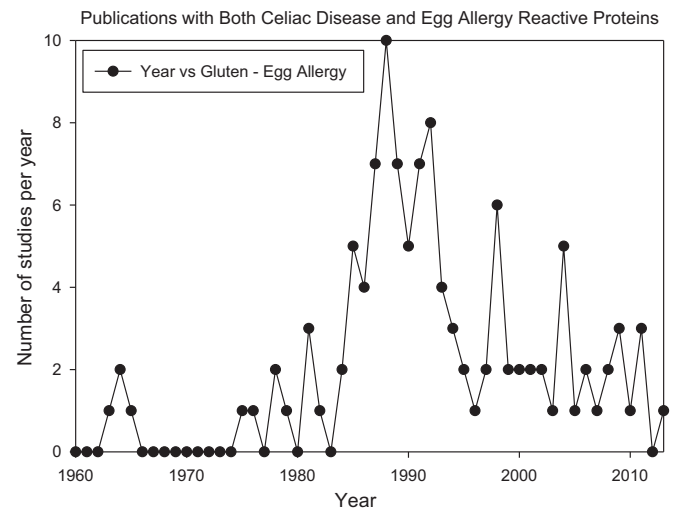


Fig. 7. The number of publications with both a celiac disease reactive protein (gluten or gliadin or glutenin) and an egg allergy reactive protein (ovalbumin or lysozyme or ovomucoid).

very much affiliated with gluten studies occurred in all the investigated areas including grains, vitamins and minerals, interaction, immunology, enzymology, and genetics (Table 2). Plots of number of publish studies in the peer reviewed literature increased at a remarkably linear rate, with many  $r^2$  values over 0.90 (Table 2). The only associated keywords having low values for

**Table 2**  
Linear regression model for each keyword.

Keyword	s	r <sup>2</sup>	Start year
Gluten	20.01	0.97	1996
Gliadin	2.83	0.92	1960
Glutenin	1.30	0.75	1981
Ovalbumin	14.39	0.91	1960
Lysozyme	10.64	0.93	1960
Ovomucoid	0.29	0.25	1983
Ara h	1.86	0.82	1993
Wheat	6.08	0.92	1999
Rye	0.95	0.72	1990
Barley	1.31	0.70	1996
Oats	0.20	0.12	1996
Rice	0.73	0.72	2000
Vitamins	0.84	0.74	1989
Calcium	0.49	0.47	1993
Iron deficiency	0.50	0.61	1986
Sensitivity	1.43	0.85	1977
Intolerance	1.33	0.85	1993
Allergen	1.86	0.79	1999
Inflammation	3.57	0.90	1994
Antibodies	2.50	0.92	1971
Autoimmune	7.69	0.96	1997
B cell	0.26	0.27	2002
T cell	0.27	0.87	1982
IgA	1.04	0.90	1970
IgE	0.36	0.47	1994
tTG	2.50	0.77	1997
Enzyme	5.05	0.96	1995
HLA	1.70	0.88	1988

goodness of fit were those with very few published studies in connection with gluten – oats and B cells. Although the associated keyword rice is not normally considered to be important to celiac disease research, there is an upward trend regarding research efforts of this affiliated keyword in combination with gluten. In fact, most keywords had upward trends in published research, and for those keywords in which a second model was used to improved linearity starting at a later time epoch, that second model had an even great upward trend.

#### 4.2. Use of the selected keywords

Although the affiliated keywords list was not exhaustive, it is comprised of commonly used keywords in celiac disease and gluten research review articles in the published literature. For example, in a review article on gluten and celiac disease [7], 18/21 keywords were present in the article or its citations, all except 'calcium', 'allergen', and 'IgE'. Similarly, in another review article, 17/21 keywords were present in the article or its citations, all except 'calcium', 'allergen', 'inflammation', and 'IgE'. These keywords reflect common themes across celiac disease research.

#### 4.3. Other studies

Data mining is important to determine trends and patterns in published biomedical research. In prior work it was determined that publications having keywords associated with celiac disease

tended to increase at a faster rate in recent years [5]. However, one keyword affiliated with celiac disease, 'malabsorption', was found to have a downward trend in this prior study, consistent with the changing clinical presentation of celiac disease. In the current study, keywords associated with gluten protein tended to have upward trends in published research since at least 2002 (B cells), and from a begin time as early as 1971 (antibodies) (Table 2). In the prior study, keywords associated with celiac disease were also used to determine the association with inflammatory bowel disease (IBD), a gastrointestinal disease with some similar properties to celiac disease. It was found that four associated keywords had sharp upswings in both celiac disease and IBD. Thus it is not surprising that some keywords affiliated with gluten protein were found in the present study to also be associated in the published literature with food allergies.

#### 4.4. Limitations and future directions

MEDLINE is a search tool with certain constraints. Although it was supposed that the list of studies published in 2013 that can be detected by MEDLINE is complete, this may not necessarily be so. In some of the graphs, there was a downward swing from 2012 to 2013 suggesting that some 2013 studies may not yet have been entered into the MEDLINE database. This can be observed for example, in all of the panels of Fig. 2. The MEDLINE database only covers a limited number of journals. Thus the findings are only valid for the journals index used in the MEDLINE database. The trends in publication rate were modeled using linear regression; however a nonlinear regression model may be more accurate in some cases. For simplicity, we did not verify that keywords associated with gluten were also associated with celiac disease, because it would have required a search using a third keyword, i.e., 'celiac disease'. In future work, more complex associations will be investigated by using several keywords for the search.

#### Acknowledgement

The authors thank Dr. Armin Alaedini of the Celiac Disease Center, Columbia University, for helpful discussions.

#### References

- [1] Green PH. The many faces of celiac disease: Clinical presentation of celiac disease in the adult population. *Gastroenterology* 2005;4(Suppl. 1):S74–8.
- [2] Green PH, Cellier C. Celiac Disease. *N Engl J Med* 2007;357:1731–43.
- [3] Cicciocioppo R, Di Sabatino A, Corazza GR. The immune recognition of gluten in coeliac disease. *Clin Exp Immunol* 2005;140:408–16.
- [4] Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2002;2:647–55.
- [5] Ciaccio EJ, Bhagat G, Lewis SK, Green PH. Trends in celiac disease research. *Comput Biol Med* 2015;65:369–78.
- [6] Greenhalgh T. How to read a paper. *The Medline database. BMJ* 1997;315:180–3.
- [7] Alaedini A, Green PH. Narrative review: celiac disease: understanding a complex autoimmune disorder. *Ann Intern Med* 2005;142:289–98.