



Leveraging Google Trends to investigate the global public interest in rheumatoid arthritis

Guo-Cui Wu¹ · Sha-Sha Tao^{2,3} · Chan-Na Zhao^{2,3} · Yan-Mei Mao^{2,3} · Qian Wu^{2,3} · Yi-Lin Dan^{2,3} · Hai-Feng Pan^{2,3}

Received: 10 March 2019 / Accepted: 1 April 2019 / Published online: 6 April 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

This study aims to investigate the global public interest in rheumatoid arthritis by evaluating search term popularity changes of the disease over a decade. Google Trends was applied to retrieve search popularity scores for the term ‘rheumatoid arthritis’ between January 2004 and December 2017, utilizing the category of “health”. Overall, relative searches volume for rheumatoid arthritis steadily decreased from January 2004 to December 2010, and then slowly rose from January 2011 to December 2017. There were significant seasonal variations in relative searches volume for the term ‘rheumatoid arthritis’ (Amplitude = 3.11; Phase: Month = 4.3; Low point: Month = 10.3; $p < 0.025$). Relative searches volume peaked in April and reached the lowest level in October. The top 11 rising topics were scleroderma, Anna Marchesini, C-reaction protein, osteoarthritis, arthritis, joint pain, autoimmune disease, rheumatoid factor, rheumatology, methotrexate, and systemic lupus erythematosus, ranking from high to low by relative growth of topic regarding rheumatoid arthritis. In conclusion, the evidence from Google Trends analysis demonstrates a significant seasonal variation in rheumatoid arthritis, with a peak in April. In addition, the top rising search queries are beneficial for physicians to search the Internet themselves for websites that provide high-quality information to recommend to their patients.

Keywords Google Trends · Global public interest · Rheumatoid arthritis · Seasonality

✉ Hai-Feng Pan
panhaifeng@ahmu.edu.cn; panhaifeng1982@sina.com

Guo-Cui Wu
gcwu82@126.com

Sha-Sha Tao
516138026@qq.com

Chan-Na Zhao
1729831835@qq.com

Yan-Mei Mao
438387736@qq.com

Qian Wu
1483118366@qq.com

Yi-Lin Dan
1282636841@qq.com

¹ School of Nursing, Anhui Medical University, 15 Feicui Road, Hefei, Anhui, China

² Department of Epidemiology and Biostatistics, School of Public Health, Anhui Medical University, 81 Meishan Road, Hefei, Anhui, China

³ Anhui Province Key Laboratory of Major Autoimmune Diseases, 81 Meishan Road, Hefei, Anhui, China

Introduction

Rheumatoid arthritis (RA) is a common systemic autoimmune disease featured by chronic inflammation in synovial joint that leads to disability and poor life quality. The incidence and prevalence of RA exhibits an increasing trend with remarkable global morbidity and mortality. The etiology and pathogenesis of RA are becoming clearer owing to a number of important research findings in the last decades [1]. Elevating numbers of inflammatory cytokines have been implicated in RA pathophysiology. Many of these cytokines are promising biomarkers for RA [2]. In recent years, biological agents are being prescribed targeting RA pathophysiology, and these treatments have shown beneficial effect in alleviating symptoms and improving disease outcome. The progress in therapeutic strategies and monitoring are helpful for the patients to achieve early and sustained clinical remission. However, both direct and indirect high cost of treatment create heavy economic burden for the patients, and reduce their health-related quality of life (HRQoL). Globally, RA causes significant years lived with disability

(YLD) [3]. Given the global health impact of RA, evaluating the global public interest in this disease is in urgent need. Patients can, nowadays, turn to the Internet for information about disorders, medication, and treatments. Those patients who have access to information about RA are prone to take a positive role in disease management [4]. However, an effective way to measure the global public interest in RA has yet to be determined.

One method to evaluate the public interest in a health-related topic is to investigate Internet searches over time. In recent years, people are increasingly referring to the Internet for information about health care [5]. In the past decade, the use of the Internet has increased sharply, and searching for health information is an important application for many users [6]. In addition, among people who seek health-related information from a source other than a physician, nearly 40% use Internet search [6]. Google is the most popular search engine and is used in 75% of the Internet searches [7]. Google Trends is a website by Google that analyzes the popularity of top search queries in Google Search across different regions and languages. The search volumes of Google Trends are usually increased for conditions with higher social influence or for periods with higher disease burden. Thus, Google Trends are gaining great importance in surveillance studies on the most epidemiologically relevant diseases. Therefore, this new approach could be serving as an effective complement to traditional data sources [8]. The previous investigations have utilized Google Trends to better understand the way that the public searched for status epilepticus to evaluate public awareness of the disease [9], assess the public interest in cancer screening [10], monitor risk of global suicide [11], and track seasonal patterns in urinary tract infections [12]. A very recent study used Google Trends to assess global public interest in osteoarthritis [13].

In the current study, we aimed to extend the work of these previous studies using Google Trends to track seasonal patterns, and investigate global public interest in RA.

Methods

The data regarding the frequency of search terms that people put in the search box were obtained from Google Trends. To perform easier comparisons between different terms, Google Trends adjusts the results to the time, and location of a search term by dividing each data point by the total searches of the geography, and time period, and then, by scaling these resulting numbers based on a given search term's proportion to all searches on a range of 0–100, the scores represent relative search volume (RSV). The data points can be downloaded in Comma-Separated Values (CSV) format. To avoid selection bias, Google Trends excludes repeated searches from the same person in a short time.

On December 06, 2018, Google Trends was searched and the CSV files were exported for the query term 'rheumatoid arthritis'. The query was searched within the global scope during the period of January 01, 2004 and December 31, 2017, using the "health" category. The downloaded CSV files had 168 (12 months per year \times 14 years) data points. Ethics approval was obtained from the Ethics Committee of Anhui Medical University. The study was carried out in accordance with ethical standards.

Statistical analysis

The cosinor analysis was conducted to explore the seasonality. The cosinor analysis and the software utilized to execute it are detailed in the previous studies by Barnett et al. [14, 15]. Briefly, the cosinor analysis is hinged on the following sinusoid:

$$St = A \cos(2\pi t/c - P), \quad t = 1, \dots, n,$$

where A is the amplitude of the sinusoid, which explains the size of the seasonal changes, P is the phase of the sinusoid, which explains where the seasonal peak occurs, c is the length of the seasonal cycle (established at 12 for monthly data), t is the time of each data point, and n is the total number of data points.

The threshold of significance in cosinor analysis was adjusted as $p < 0.025$ to decrease the false discovery rate caused by multiple testing, both sine and cosine p values were tested for statistical significance. One of the two values was presented (i.e., cosine p value). The Poisson model was used for the continuous variables, and to adjust the unequal number of days in the months, the offset was used in cosinor functions. Moreover, a time-series plot was drawn to display the consistency in the seasonal patterns. All these statistical analyses were performed by using the "season" package in R version 3.4.4.

Results

Searches for RA on Google from January 2004 to December 2017

Overall, RSV for RA steadily decreased from January 2004 to December 2010, and then slowly rose from January 2011 to December 2017 (Fig. 1).

The seasonal variation in the RSV of 'rheumatoid arthritis'

The cosinor test showed that there was statistically significant seasonal variation in RSV for the term 'rheumatoid arthritis' (amplitude = 3.11 (absolute scale); phase:

Fig. 1 Time-series plots for the relative search volume of ‘rheumatoid arthritis’

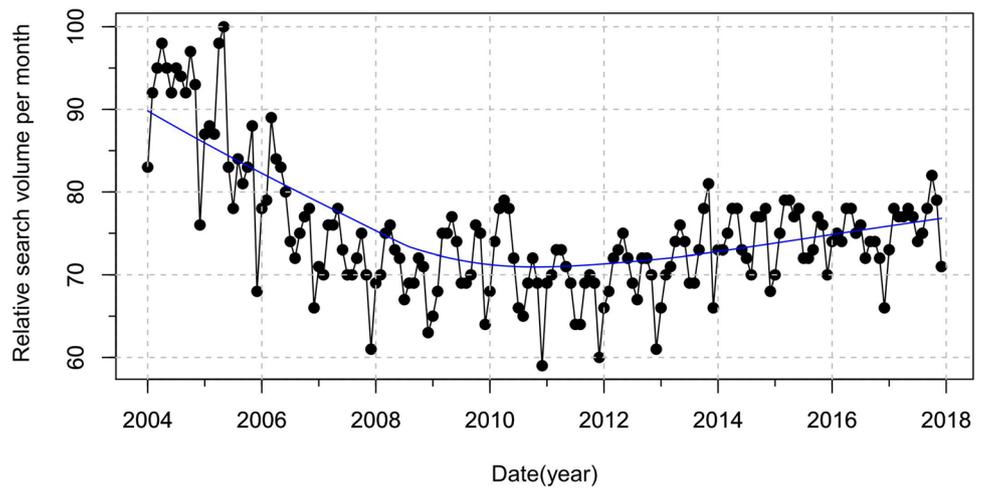


Fig. 2 The plots of cosinor models for the seasonal variation in the relative search volume of ‘rheumatoid arthritis’

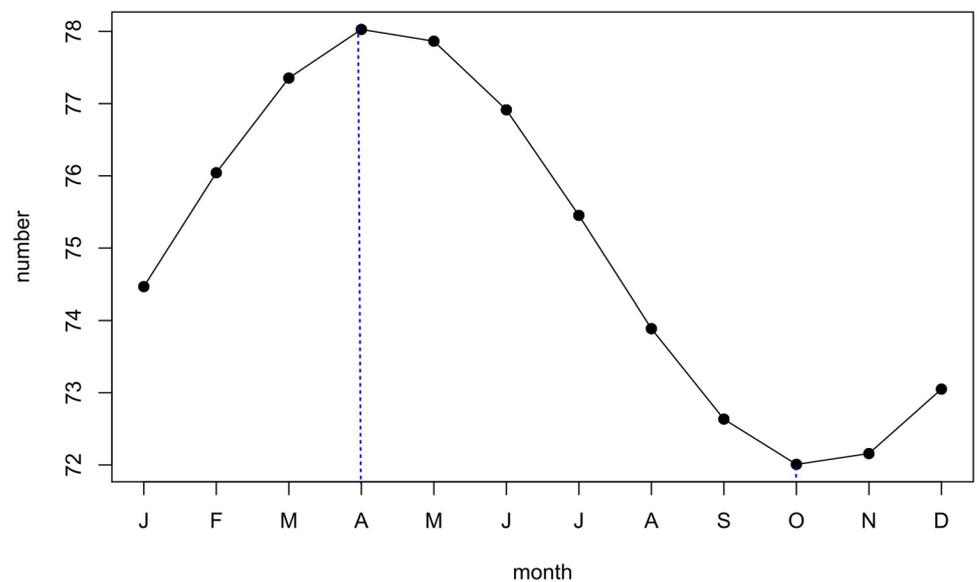


Table 1 The cosinor analysis on seasonal variation in the RSV of ‘rheumatoid arthritis’

	Estimate	Std. error	z value	p value
Intercept	4.317	0.009	484.368	$< 2e^{-16}$
Cosine	-0.007	0.013	-0.522	0.602
Sine	0.040	0.013	3.185	0.001

The z value is the Wald statistic for testing the hypothesis that the corresponding parameter (regression coefficient) is zero
RSV relative search volume

month = 4.3; low point: month = 10.3; $p < 0.025$). RSV peaked in April and reached the lowest level in October (Fig. 2) (Table 1).

Relative growth of topic regarding RA

Term’s growth is compared to the previous time period. According to the analysis on relative growth of topic regarding RA, the top rising topics were scleroderma-topic, Anna Marchesini-actress, C-reaction protein, osteoarthritis-topic, arthritis-topic, joint pain-topic, autoimmune disease-topic, rheumatoid factor-topic, rheumatology-topic, methotrexate-medicine, and systemic lupus erythematosus, ranking from high to low by relative growth of topic regarding RA (Table 2).

Discussion

Our results indicate that the number of searches for RA steadily decreased from January 2004 to December 2010, which may be due to decreasing trend in incidence and

Table 2 Relative growth of topic regarding RA

Rank	Search topic	% growth
1	Scleroderma	Breakout
2	Anna Marchesini	Breakout
3	C-reaction protein	150
4	Osteoarthritis	140
5	Arthritis	120
6	Joint pain	90
7	Autoimmune disease	90
8	Rheumatoid factor	80
9	Rheumatology	80
10	Methotrexate	70
11	Systemic lupus erythematosus	60

Term's growth is compared to previous time period. 'Breakout' is used for term search that grew by more than 5000% compared to previous period

prevalence of RA in several countries [16–18]. However, geographical and racial variations in the occurrence of RA continue to exist. The number of searches for RA has grown steadily since 2011. During the last few years, technological advancements, such as patient-centered smartphone applications, have enabled patients suffering from RA to support self-management of the disease [19]. Furthermore, the increased accessibility to health information through the Internet has grown tremendously, even though with an increased potential risk of disinformation [20]. This type of search traffic data could be a very important source for health care scientists interested in analyzing trends of different medical conditions.

It is generally accepted that the interaction between genetic and environmental factors plays an important role in the pathogenesis of RA. From cosinor analysis, Google Trends for 'rheumatoid arthritis' presented a specific seasonality, which peaked in April and reached the lowest level in October. Similarly, the previous study have shown that a peak of disease activity was observed in spring, followed by a decrease in disease activity during fall [21], which suggest the role of environmental changes occurring during varying seasons, such as vitamin D levels. The present study provided additional evidence for seasonality of RA. However, there are also different views on seasonal peculiarities of clinical course of RA. Feldthusen et al. reported fluctuations in fatigue, general, and physical fatigue being significantly greater during the winter [22]. The difference may be explained by the different environmental conditions, medical care, patient's condition, etc. In fact, data from Internet sources could serve as a real-time surveillance tool and an alert for healthcare systems, so as to allocate appropriate resources for specific moments with higher disease burden. Among the top rising queries, 'scleroderma' and 'Anna

Marchesini' were the major concerns. Scleroderma, also known as systemic sclerosis (SSc), is a chronic connective tissue disease generally classified as one of the autoimmune rheumatic diseases. In some cases of SSc, the disease may overlap with other organ-specific or systemic autoimmune diseases, such as RA [23]. Inflammatory arthritis does occur in scleroderma, and approximately 20% may have erosive arthritis [24]. Musculoskeletal (ME) pain is a common complaint of SSc patients, ranging from 40 to 80%, especially in patients with early diffuse disease [25]. The main ME symptoms are movement limitations, joint pain, and/or swelling. Arthritis may be a feature in the initial presentation of SSc, often initially misdiagnosed as RA [26]. Anna Marchesini, a popular Italian actress, has died after battling with RA for several years. A very recent study used Google Trends, Twitter, and Wikipedia to investigate the impact of Anna Marchesini's death from RA. Their results showed that, after her death, interest for RA has remarkably increased [27]. Thus, workers in the field of public health and health care delivery should be aware of the importance of mass communication. 'C reactive protein (CRP)' ranked third in the top rising queries, CRP levels directly correlate with the disease activity of various autoimmunopathies such as RA, increased CRP level can contribute to disease progression [28]. Other related queries included osteoarthritis, arthritis, joint pain-topic, autoimmune disease, rheumatoid factor, Rheumatology, methotrexate, systemic lupus erythematosus, which cover symptoms, diagnosis, treatment, and accompanying diseases.

Several limitations should be acknowledged in the current study. First, although Google is the most common search engine, it only captures the search behavior of a certain proportion of the population—those having access to Internet and those using Google rather than other search engines (such as Baidu in China). Second, Google Trends lack the detailed information on the method by which Google generates these search data and the algorithms that it employs to analyze it. Third, we also had no access to the demographic data of the users who searched for RA, and thus, subgroup analyses stratified by these parameters could not be performed. Despite these limitations, the current study also has strengths, which include the large and exhaustive amount of data, the long period of observation, and the global areas covered.

Although some privacy issues exist in using Google Web search data, tracing individuals that conduct online searches when signed into their accounts and recording and analyzing data about users' characteristics, such as gender and age, intent of web search, and "search outcomes" could improve the usefulness of this tool for public health and health education purposes [13].

In summary, this study provided another line of evidence of a seasonal variation in RA, and Google searches for RA

have steadily increased in recent years. In addition, a significant digital interest has been reported towards RA in recent years. The great majority of this interest is expressed by various concerns of people or RA patients regarding the symptoms, diagnosis, accompanying diseases, and drugs employed in the treatment of RA. Therefore, physicians should be aware of the commonly raised topics by these patients to give appropriate answers and education, which may improve their quality life and/or to prevent unnecessary worries.

Author contribution GCW, SST, CNZ, YMM, and HFP conceptualized the study, participated in the study design, and revised the manuscript. GCW and CNZ wrote the manuscript. QW and YLD collected the data, conducted the statistical analysis, and revised the manuscript. All authors read and approved the final manuscript.

Funding The Doctoral research Grant from Anhui Medical University (XJ201712), and the scientific research grant from Anhui Medical University (2017xkj010) and the National Natural Science Foundation of China (81872687).

Compliance with ethical standards

Conflicts of interest All authors declare that they have no conflict of interest.

References

- Fazal SA, Khan M, Nishi SE, Alam F, Zarin N, Bari MT, Ashraf GM (2018) A clinical update and global economic burden of rheumatoid arthritis. *Endocr Metab Immune Disord Drug Targets* 18(2):98–109. <https://doi.org/10.2174/1871530317666171114122417>
- Burska A, Boissinot M, Ponchel F (2014) Cytokines as biomarkers in rheumatoid arthritis. *Mediators Inflamm* 2014:545493. <https://doi.org/10.1155/2014/545493>
- Cross M, Smith E, Hoy D, Carmona L, Wolfe F, Vos T, Williams B, Gabriel S, Lassere M, Johns N, Buchbinder R, Woolf A, March L (2014) The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis* 73(7):1316–1322. <https://doi.org/10.1136/annrheumdis-2013-204627>
- van der Vaart R, Drossaert CH, Taal E, van de Laar MA (2011) Patient preferences for a hospital-based rheumatology Interactive Health Communication Application and factors associated with these preferences. *Rheumatology (Oxford)* 50(9):1618–1626. <https://doi.org/10.1093/rheumatology/ker161>
- Bundorf MK, Wagner TH, Singer SJ, Baker LC (2006) Who searches the Internet for health information? *Health Serv Res* 41(3 Pt 1):819–836. <https://doi.org/10.1111/j.1475-6773.2006.00510.x>
- Baker L, Wagner TH, Singer S, Bundorf MK (2003) Use of the Internet and e-mail for health care information: results from a national survey. *JAMA* 289(18):2400–2406. <https://doi.org/10.1001/jama.289.18.2400>
- NetMarketShare (2018) Search engine market share. <https://www.netmarketshare.com/search-engine-market-share>. Accessed 21 June 2018
- Cervellin G, Comelli I, Lippi G (2017) Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health* 7(3):185–189. <https://doi.org/10.1016/j.jegh.2017.06.001>
- Brigo F, Trinka E (2015) Google search behavior for status epilepticus. *Epilepsy Behav* 49:146–149. <https://doi.org/10.1016/j.yebeh.2015.02.029>
- Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, Davidson NO (2015) The utility of Google Trends data to examine interest in cancer screening. *BMJ Open* 5(6):e006678. <https://doi.org/10.1136/bmjopen-2014-006678>
- Solano P, Ustulin M, Pizzorno E, Vichi M, Pompili M, Serafini G, Amore M (2016) A Google-based approach for monitoring suicide risk. *Psychiatry Res* 246:581–586. <https://doi.org/10.1016/j.psychres.2016.10.030>
- Rosignol L, Pelat C, Lambert B, Flahault A, Chartier-Kastler E, Hanslik T (2013) A method to assess seasonality of urinary tract infections based on medication sales and google trends. *PLoS One* 8(10):e76020. <https://doi.org/10.1371/journal.pone.0076020>
- Jellison SS, Bibens M, Checketts J, Vassar M (2018) Using Google Trends to assess global public interest in osteoarthritis. *Rheumatol Int* 38(11):2133–2136. <https://doi.org/10.1007/s00296-018-4158-2>
- Barnett AG, Baker P, Dobson AJ (2012) Analysing seasonal data. *R J* 4(1):5–10. <https://doi.org/10.32614/RJ-2012-001>
- Barnett AG, Baker P (2018) Seasonal analysis of health data. R package version 0:3-5. <https://cran.r-project.org/package=season/>. Accessed 06 Dec 2018
- Abhishek A, Doherty M, Kuo CF, Mallen CD, Zhang W, Grainge MJ (2017) Rheumatoid arthritis is getting less frequent—results of a nationwide population-based cohort study. *Rheumatology (Oxford)* 56(5):736–744. <https://doi.org/10.1093/rheumatology/kew468>
- Jacobsson LT, Hanson RL, Knowler WC, Pillemer S, Pettitt DJ, McCance DR, Bennett PH (1994) Decreasing incidence and prevalence of rheumatoid arthritis in Pima Indians over a 25-year period. *Arthritis Rheum* 37(8):1158–1165. <https://doi.org/10.1002/art.1780370808>
- Doran MF, Pond GR, Crowson CS, O’Fallon WM, Gabriel SE (2002) Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a 40-year period. *Arthritis Rheum* 46(3):625–631. <https://doi.org/10.1002/art.509>
- Azevedo R, Bernardes M, Fonseca J, Lima A (2015) Smartphone application for rheumatoid arthritis self-management: cross-sectional study revealed the usefulness, willingness to use and patients’ needs. *Rheumatol Int* 35(10):1675–1685. <https://doi.org/10.1007/s00296-015-3270-9>
- Castillo-Ortiz JD, de Jesus Valdivia-Nuno J, Ramirez-Gomez A, Garagarza-Mariscal H, Gallegos-Rios C, Flores-Hernandez G, Hernandez-Sanchez L, Brambila-Barba V, Castaneda-Sanchez JJ, Barajas-Ochoa Z, Suarez-Rico A, Sanchez-Gonzalez JM, Ramos-Remus C (2016) Fifteen-year trend in information on the World Wide Web for patients with rheumatoid arthritis: evolving, but opportunities for improvement remain. *Rheumatol Int* 36(9):1281–1289. <https://doi.org/10.1007/s00296-016-3507-2>
- Iikuni N, Nakajima A, Inoue E, Tanaka E, Okamoto H, Hara M, Tomatsu T, Kamatani N, Yamanaka H (2007) What’s in season for rheumatoid arthritis patients? Seasonal fluctuations in disease activity. *Rheumatology (Oxford)* 46(5):846–848. <https://doi.org/10.1093/rheumatology/kel414>
- Feldthusen C, Grimby-Ekman A, Forsblad-d’Elia H, Jacobsson L, Mannerkorpi K (2016) Seasonal variations in fatigue in persons with rheumatoid arthritis: a longitudinal study. *BMC Musculoskelet Disord* 17:59. <https://doi.org/10.1186/s12891-016-0911-4>
- Szucs G, Szekanez Z, Zilahi E, Kapitany A, Barath S, Szamosi S, Vegvari A, Szabo Z, Szanto S, Czirjak L, Gyorgy Kiss C (2007)

- Systemic sclerosis-rheumatoid arthritis overlap syndrome: a unique combination of features suggests a distinct genetic, serological and clinical entity. *Rheumatology (Oxford)* 46(6):989–993. <https://doi.org/10.1093/rheumatology/kem021>
24. Pope JE (2002) Scleroderma overlap syndromes. *Curr Opin Rheumatol* 14(6):704–710. <https://doi.org/10.1097/00002281-20021000-00013>
 25. Clements PJ, Allanore Y, Khanna D, Singh M, Furst DE (2012) Arthritis in systemic sclerosis: systematic review of the literature and suggestions for the performance of future clinical trials in systemic sclerosis arthritis. *Semin Arthritis Rheum* 41(6):801–814. <https://doi.org/10.1016/j.semarthrit.2011.10.003>
 26. Horimoto AM, da Costa IP (2016) Overlap between systemic sclerosis and rheumatoid arthritis: a distinct clinical entity? *Rev Bras Reumatol Engl Ed* 56(4):287–298. <https://doi.org/10.1016/j.rbre.2014.12.011>
 27. Mahroum N, Bragazzi NL, Sharif K, Gianfredi V, Nucci D, Rosselli R, Brigo F, Adawi M, Amital H, Watad A (2018) Leveraging Google Trends, Twitter, and Wikipedia to investigate the impact of a celebrity's death from rheumatoid arthritis. *J Clin Rheumatol* 24(4):188–192. <https://doi.org/10.1097/RHU.0000000000000692>
 28. Vogt B, Fuhrrohr B, Muller R, Sheriff A (2007) CRP and the disposal of dying cells: consequences for systemic lupus erythematosus and rheumatoid arthritis. *Autoimmunity* 40(4):295–298. <https://doi.org/10.1080/08916930701358925>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.