

## What Journal Citations Teach Us About Rheumatic Diseases

Steven K. Magid, Timothy R. Roberts, and Mohsen Javaheri

The past 50-plus years have seen dramatic advances in medical discovery. Only 51 years have elapsed from the identification of the double-helix structure of DNA (1) to the publication of the first estimate of the human genome (2). Furthermore, the specialty of rheumatology is also relatively new. The name “rheumatologist” was first introduced in 1940 by Bernard Comroe (3,4), and the first subspecialty board exams took place in 1971. As a field of study, rheumatology changed radically in the early 1950s largely because of the discoveries of rheumatoid factor, the lupus erythematosus cell, and the therapeutic effect of cortisone (4).

Keeping pace with medical discovery has been an exponential increase in the volume of published literature, as well as a dramatic improvement in the technology available to search this literature. The PubMed database contains over 16 million citations published between 1951 and 2005. This is remarkable considering the fact that the National Library of Medicine’s Index-Cat (a database for citations older than those available in PubMed) lists only 2.5 million references to publications from late 17th to the early 20th centuries.

The ability to access over 50 years of publication data electronically with uniform indexing, paired with scientific developments, invited our approach to examining publication trends in the field of rheumatology. We were interested in studying how the number of articles on particular topics changed over time. We theorized that a number of patterns would be displayed that would reflect the history of our specialty.

### PubMed search

We reviewed the indices of *Arthritis & Rheumatism (A&R)* from volume 1 (1958) to the present and developed a list of diseases and concepts considered

central to the field of rheumatology. These terms were searched in PubMed. For all but two searches, we allowed PubMed’s automatic term mapping to enhance the terms entered (Table 1). This function searches the database not only for the phrase entered as text in the title or abstract of the citation, but it also maps the phrase to the appropriate Medical Subject Heading (MeSH). For example, the search for *superantigen* in PubMed mapped to the MeSH term Superantigens and performed a search that included all articles indexed with that term. Nearly 25% of the citations in the resulting set did not have the words *superantigen* or *superantigens* in the title or abstract. Additionally, the subject headings are automatically “exploded,” meaning that if a term has corresponding specific subject headings within the MeSH tree structure, these headings are included in the search. For example, the phrase *rheumatoid arthritis* automatically maps to the MeSH term Arthritis, Rheumatoid and includes citations with the more specific subject headings of Arthritis, Juvenile Rheumatoid; Caplan’s Syndrome; Felty’s Syndrome, etc.

Automatic term mapping was deemed inadequate for two of the searches: the illnesses associated with silicone breast implants and rapeseed oil poisoning. For these topics, we constructed more complex searches using subject heading combinations to retrieve a representational result set. Also, although we used the automatic term mapping for our search on fibromyalgia, we did not count citations before 1981, the year that the word was first used by an author in an article title (5). The subject heading *fibromyalgia* has been retroactively assigned to citations that were previously assigned the heading *fibrositis*. For the purpose of our study, including those results would have misrepresented the trend.

Separate searches were performed for consecutive 5-year periods, from 1951 through 2005. (We performed our searches between May 1, 2007 and September 16, 2007.) For each topic, the total number of papers was tabulated for each 5-year period and was then charted over time as: absolute number of articles, percentage of the total number of PubMed articles, and the percentage of PubMed “rheumatic disease” articles. To

Steven K. Magid, MD, Timothy R. Roberts, MLS, Mohsen Javaheri, MD: Hospital for Special Surgery, New York, New York.

Address correspondence to Steven K. Magid, MD, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021. E-mail: magids@hss.edu.

Submitted for publication September 17, 2007; accepted September 17, 2007.

**Table 1.** Searches performed with the PubMed automatic mapping translation\*

Term entered	Search strategy
Ankylosing spondylitis	spondylitis ankylosing (MeSH Terms) OR ankylosing spondylitis (Text Word)
Antiphospholipid syndrome	antiphospholipid syndrome (MeSH Terms) OR antiphospholipid syndrome (Text Word)
Dermatomyositis	dermatomyositis (MeSH Terms) OR dermatomyositis (Text Word)
Eosinophilia myalgia syndrome	eosinophilia-myalgia syndrome (MeSH Terms) OR eosinophilia myalgia syndrome (Text Word)
Etanercept	TNFR-Fc fusion protein (Substance Name) OR etanercept (Text Word)
Fibromyalgia	fibromyalgia (MeSH Terms) OR Fibromyalgia (Text Word)
Gold	gold (MeSH Terms) OR gold (Text Word)
Methotrexate	methotrexate (MeSH Terms) OR methotrexate (Text Word)
Osteoarthritis	osteoarthritis (MeSH Terms) OR osteoarthritis (Text Word)
Polymyositis	polymyositis (MeSH Terms) OR polymyositis (Text Word)
Psoriatic arthritis	arthritis psoriatic (MeSH Terms) OR psoriatic arthritis (Text Word)
Rapeseed poisoning OR rapeseed oil poisoning OR plant oils/poisoning OR toxic oil syndrome	brassica rapa (TIAB) NOT Medline (SB) OR "brassica rapa"(MeSH Terms) OR rapeseed (Text Word) OR brassica (MeSH Terms) OR Brassica (Text Word) OR rapeseed oil (Substance Name) AND Toxic-allergic (All Fields) AND syndrome (MeSH Terms) OR syndrome (Text Word) Plant Oils/poisoning (MeSH) Brassica rapa/poisoning (MeSH Terms) OR toxic (All Fields) AND oil (All Fields) AND syndrome (MeSH Terms) OR syndrome (Text Word)
Rheumatic diseases or arthritis or connective tissue diseases OR autoimmune diseases	rheumatic diseases (MeSH Terms) OR Rheumatic Diseases (Text Word) OR arthritis (MeSH Terms) OR Arthritis (Text Word) OR connective tissue diseases (MeSH Terms) OR Connective Tissue Diseases (Text Word) OR autoimmune diseases (MeSH Terms) OR Autoimmune Diseases (Text Word)
Rheumatic fever	rheumatic fever (MeSH Terms) OR rheumatic fever (Text Word)
Rheumatoid arthritis	arthritis rheumatoid (MeSH Terms) OR rheumatoid arthritis (Text Word)
Rheumatoid arthritis pathogenesis	arthritis rheumatoid (MeSH Terms) OR rheumatoid arthritis (Text Word) AND etiology (Subheading) OR pathogenesis (Text Word)
Scleroderma	systemic scleroderma (Text Word) OR scleroderma systemic (MeSH Terms) OR localized scleroderma (Text Word) OR scleroderma localized (MeSH Terms) OR scleroderma (Text Word)
Silicone gel OR silicone OR Silicones (Substance Name) AND breast implants AND rheumatic diseases or arthritis or connective tissue diseases or autoimmune diseases or disease	silicone gels (MeSH Terms) OR silicone gel (Text Word) OR silicones (TIAB) NOT Medline(SB) OR silicones (MeSH Terms) OR silicone (Text Word) OR Silicones (Substance Name) AND breast implants (MeSH Terms) OR ("breast implantation" [TIAB] NOT Medline[SB]) OR breast implantation (MeSH Terms) OR breast implants (Text Word) AND rheumatic diseases (MeSH Terms) OR Rheumatic Diseases (Text Word) OR arthritis (MeSH Terms) OR Arthritis (Text Word) OR connective tissue diseases (MeSH Terms) OR Connective Tissue Diseases (Text Word) OR autoimmune diseases (MeSH Terms) OR Autoimmune Diseases (Text Word) OR disease (MeSH Terms) OR disease (Text Word)
Steroids	Steroids (MeSH Terms) OR steroids (Text Word)
Systemic lupus erythematosus	systemic lupus erythematosus (Text Word) OR lupus erythematosus, systemic (MeSH Terms) OR systemic lupus erythematosus (Text Word)
T cell	t-lymphocytes (TIAB) NOT Medline (SB) OR t-lymphocytes (MeSH Terms) OR T cell (Text Word)
Tuberculosis arthritis	osteoarticular tuberculosis (Text Word) OR tuberculosis, osteoarticular (MeSH Terms) OR tuberculous arthritis (Text Word)

\* TIAB = title/abstract; SB = subset.

collect the Rheumatic Disease subset of PubMed, we constructed a very sensitive search to retrieve all articles that matched a subject heading within the following MeSH tree structures:

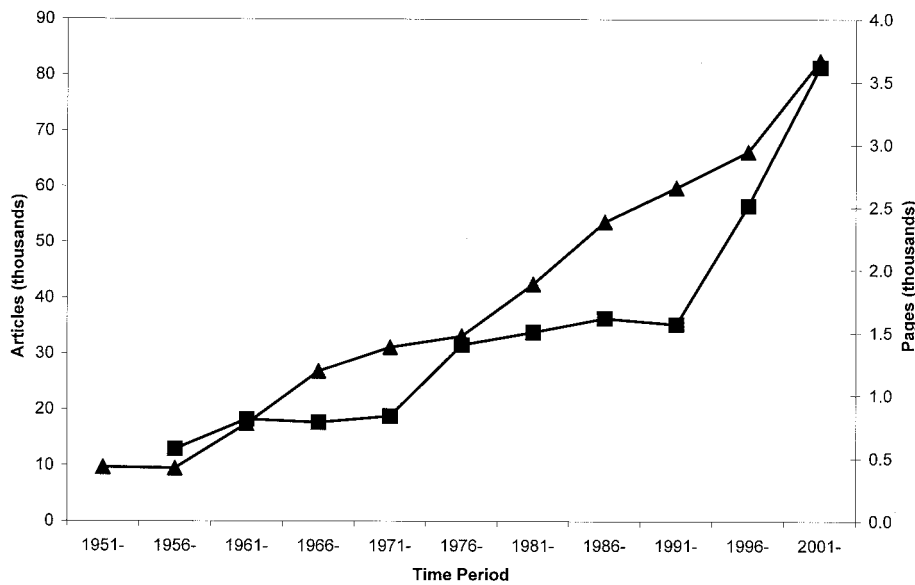
- Rheumatic Diseases [tree number C05.799]
- Arthritis (under Joint Diseases tree) [tree number C05.550.114]
- Connective Tissue Diseases [tree number C17.300]
- Autoimmune Diseases [tree number C20.111]

To construct the data set for the cause of rheumatoid arthritis (Figure 1), we searched *rheumatoid arthritis pathogenesis* and allowed PubMed to map the search. The mapping kept the text words we entered and also searched the MeSH heading Arthritis, Rheumatoid

and all more specific related terms in the MeSH tree limited with the subheading for Etiology, which is defined as follows: "Used with diseases for causative agents including microorganisms and includes environmental and social factors and personal habits as contributing factors. It includes pathogenesis" (6). Within this set we searched for the specific pathogenic factors, leveraging the benefits of PubMed's mapping (e.g., T cells mapped to a search for T-lymphocytes as a MeSH term or T cells as a text word).

### Search results

For each consecutive 5-year period, the absolute number of citations increased, from just over one half



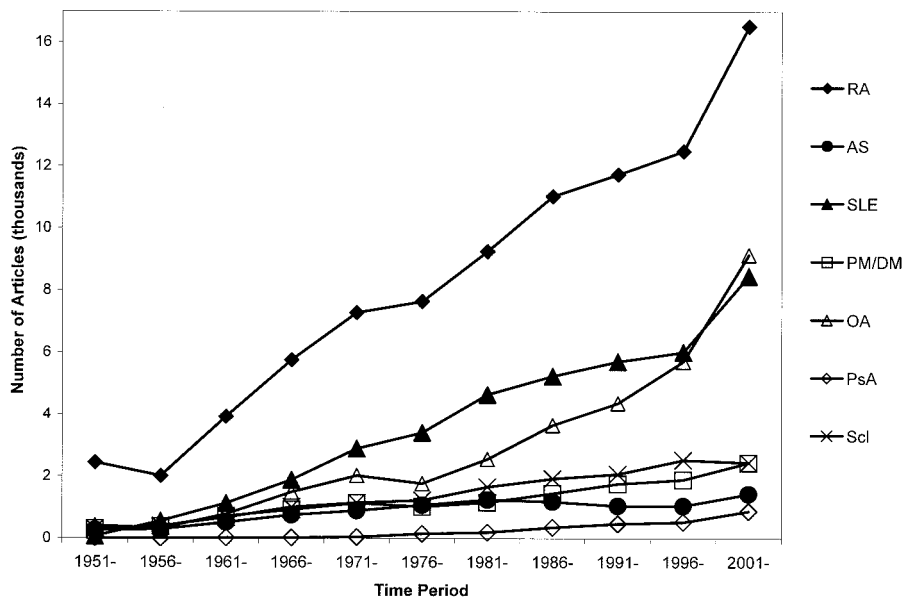
**Figure 1.** Number of citations on rheumatic diseases in PubMed (triangles) compared with number of pages in *Arthritis & Rheumatism* (squares) in the 5-year periods shown.

million published between 1951 and 1955 to over 3 million between 2001 and 2005. Similarly, the number of articles on rheumatic diseases increased from 9,643 to 82,543.

The growth in rheumatology is also illustrated by the growth of *A&R*, which comprised only 572 pages in 1958 but by 2006 had 4,027 pages. Figure 1 illustrates this growth with a parallel plot for the growth in the

number of citations on rheumatic diseases as a whole in PubMed.

**Major rheumatic diseases.** For every period studied, papers on rheumatoid arthritis (RA) far outnumber papers on the other major rheumatic diseases. The absolute number of articles on RA climbed in an almost a linear manner (Figure 2). However, articles related to



**Figure 2.** Total citations in PubMed within each 5-year segment on rheumatoid arthritis (RA), ankylosing spondylitis (AS), systemic lupus erythematosus (SLE), polymyositis/dermatomyositis (PM/DM), osteoarthritis (OA), psoriatic arthritis (PsA), and scleroderma (Scl).

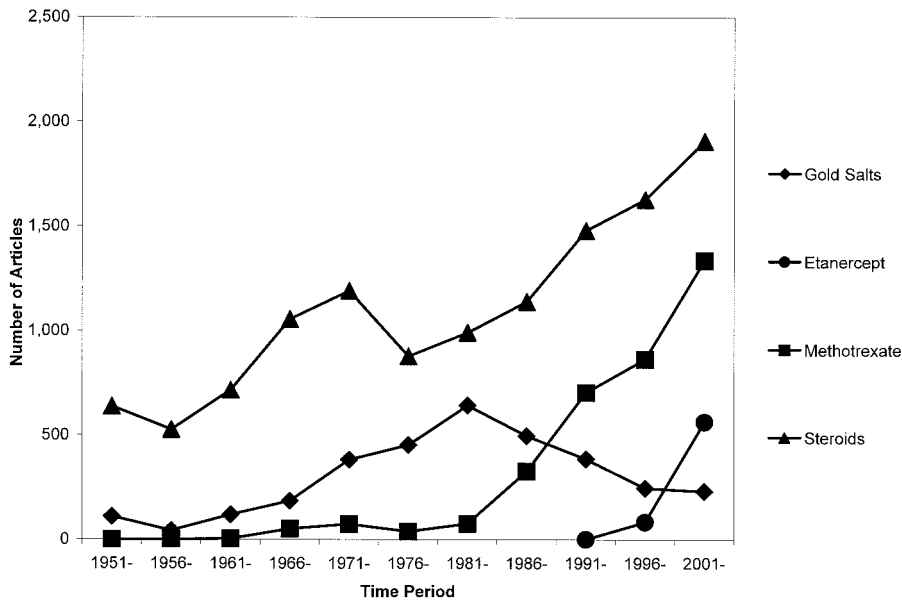


Figure 3. Number of citations on rheumatic diseases in PubMed that also discuss gold salts, etanercept, methotrexate, or corticosteroids.

RA, expressed as a percentage of the total PubMed literature, have remained stable or decreased slightly.

Papers devoted to systemic lupus erythematosus (SLE) had the second highest number until recently, when papers devoted to osteoarthritis started to outnumber those on SLE, followed by papers on scleroderma, ankylosing spondylitis, polymyositis/dermatomyositis, and psoriatic arthritis.

**Treatment.** Gold salts gained acceptance as a treatment for tuberculosis in 1925 but fell out of favor by the late 1930s (7). Later, gold salts were introduced for the treatment of RA. The popularity of this treatment in publications peaked by the early 1980s and has progressively waned. Penicillamine followed a similar pattern. Methotrexate was reported as an effective treatment for RA in 1967 (8,9) and was rarely cited when gold salts

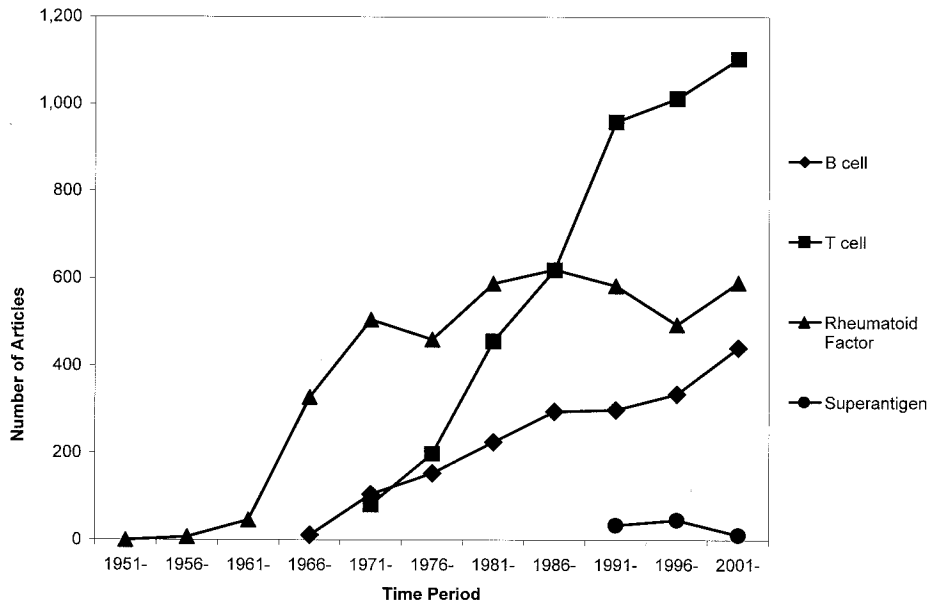
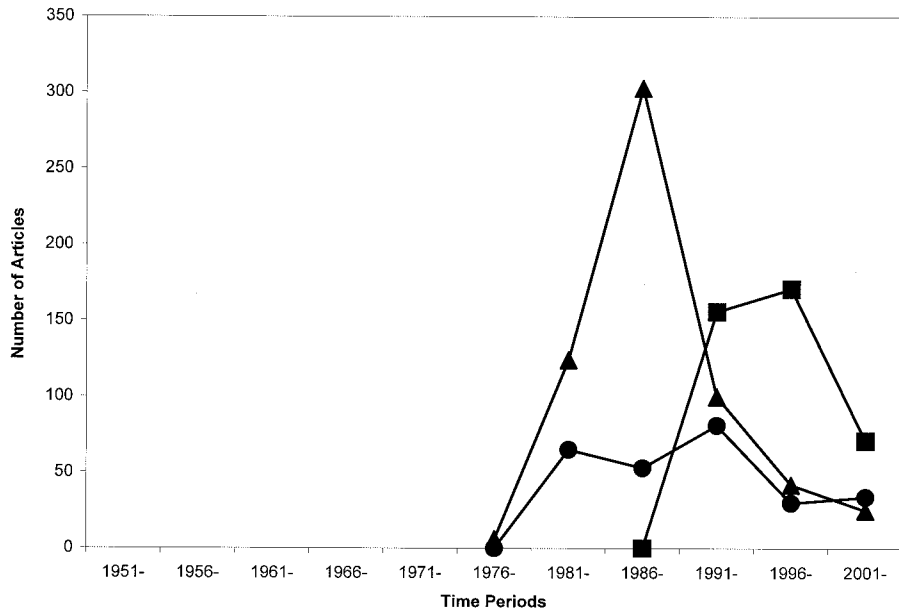


Figure 4. Trends in specific factors within articles on the pathogenesis of rheumatoid arthritis.

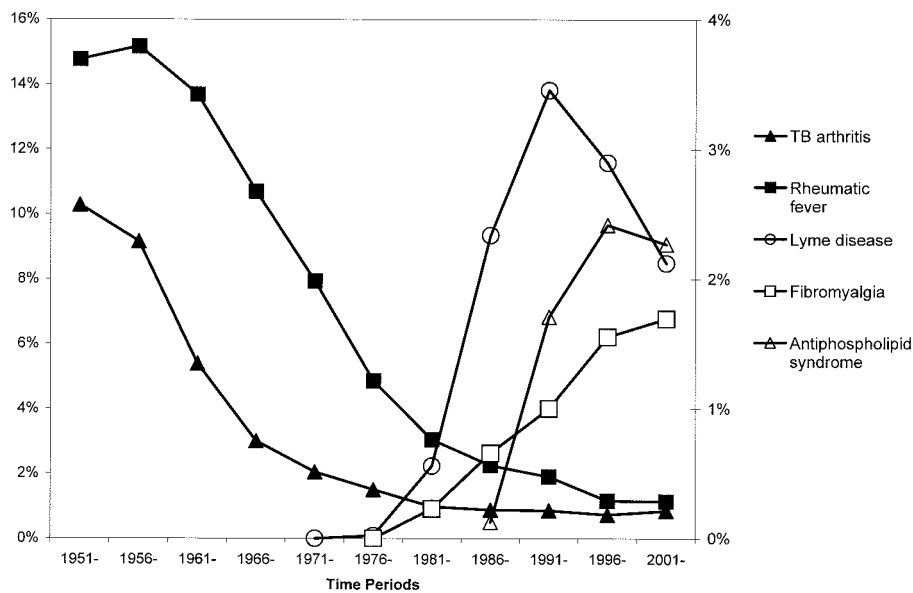


**Figure 5.** Pattern of citations on eosinophilia-myalgia syndrome (triangles), the illness associated with silicone breast implants (squares), and rapeseed oil poisoning (circles).

and penicillamine peaked, but since that time it had a rapid and continued rise, according to citations. For most (though not all) of the time periods, articles devoted to corticosteroids and nonsteroidal anti-inflammatory drugs (NSAIDs) have increased in number (Figure 3). However, those on phenylbutazone and indomethacin have all but disappeared. After

Charette's article on cyclooxygenase 2 appeared in 1995 (10), many papers have been devoted to this subject.

**Etiology of RA.** During various time periods, the predominant cause of RA has been attributed to rheumatoid factor, B cells, T cells, superantigens, infections (and others). Although publications on T cells outnumber those on B cells, those devoted to T cells have



**Figure 6.** Downward trend in tuberculosis (TB) arthritis and rheumatic fever as a percentage of publications represented in the rheumatic disease subset of PubMed, and upward trend in fibromyalgia, Lyme disease, and antiphospholipid syndrome.

leveled off over the last three 5-year periods, while there has been a slight uptick in those published on B cells in RA. Studies on rheumatoid factor have been more or less level since the late 1960s (Figure 4).

**“Come and go” diseases.** Certain diseases appeared suddenly and then waned dramatically. Two epidemic rheumatic diseases had known causes. The first, eosinophilia–myalgia syndrome, was caused by a specific manufacturing defect that occurred during the production of L-tryptophan. The initial reports on this disease were published by the Centers for Diseases Control in 1989 (11,12). The second involved the criminal adulteration of rapeseed (canola) oil, which caused an illness clinically similar to scleroderma in Spain in 1981 (13). A concern about rheumatic diseases, specifically scleroderma that developed after silicone implantation for breast enlargement, followed a similar pattern: a dramatic increase followed by a steep decline, whereas the curve for scleroderma itself has had a slow, gradual increase (Figure 5).

**“Old” diseases.** The number of articles on some diseases, such as tuberculosis arthritis and rheumatic fever, have fallen dramatically. In these cases, the publication curve likely reflects the prevalence of the disease itself (Figure 6).

**“New” diseases.** Many new diseases have appeared in the past 50 years, including toxic shock syndrome, severe acute respiratory syndrome (SARS), Ebola virus, bird flu, Legionnaires’ disease, and acquired immune deficiency syndrome (AIDS). New rheumatologic conditions also appeared during this period such as fibromyalgia, Lyme disease, and antiphospholipid syndrome. Figure 6 illustrates publications related to these entities.

### Bibliometrics of rheumatic diseases

The word “bibliometrics,” coined by Alan Pritchard, referred to the “. . . application of mathematics and statistical methods to books and other media of communication . . .” (14). The advent of electronic databases has allowed the rapid growth of this field. Bibliometrics has typically been used to evaluate the citations of individual authors, as well as the impact factor of specific journals. It has also been used to determine the scientific output of countries, and to evaluate specific research groups and university departments (15).

However, very few papers have employed these techniques to the study of musculoskeletal diseases. PubMed, for example, lists only 14 articles in response

to a query of *bibliometrics AND musculoskeletal diseases* (16).

In our research, we used a simple search strategy to examine publishing trends in the field of rheumatology. We did not endeavor to employ the methods of citation or content analysis to measure the quality of the literature in the specialty (15) or the rate of publication relative to the burden of illness (17), as other authors have. It was our goal to shed light on the history of rheumatology from 1951–2005 by charting the incidence of citations on selected topics.

We are aware that the National Library of Medicine’s PubMed has been considered incomplete when searching for citations prior to 1966. However, in the past few years the NLM has made enhancements to improve the authority and searchability of citations from 1950 through 1965. As of 2004, NLM had retrospectively added citations from the authoritative print indexes of the time period: the Current List of Medical Literature (1950–1959) and the Cumulative Index Medicus (1960–1963.) As of 2006, NLM had completed mapping all of the original subject headings from the print sources to the current MeSH, allowing PubMed’s mapping function to be extended to these older citations (18).

We observed a number of interesting trends. Certain diseases continue to be the subject of many publications. The number of articles on RA far outnumber other rheumatic diseases, and rise in an almost linear fashion. Articles on SLE also have increased each year, with a sharp increase during the last 5-year period examined. Articles on osteoarthritis increased since the mid 1970s with more articles published in the last 5-year period than those on SLE. There was a steep decline in articles on diseases such as tuberculosis arthritis and rheumatic fever, which probably parallels the decreased incidence of these diseases. As expected, the “newly” discovered conditions of Lyme disease, antiphospholipid syndrome, and fibromyalgia are receiving increased interest in the literature.

As the treatment of rheumatic diseases evolved, so did the literature about these treatments. Articles on gold therapy for RA peaked in the early 1980s and then nearly vanished. Articles on methotrexate and subsequently etanercept consistently rise. Interesting patterns also emerge when viewing the publications on the “cause” of rheumatoid arthritis, with varying focus on T cells, B cells, rheumatoid factor, and superantigens.

### REFERENCES

1. Watson JD, Crick FHC. A structure for deoxyribose nucleic acid. *Nature* 1953;171:737–8.

2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431:931–45.
3. Paget S, Lockshin M, Loebel S. *The Hospital for Special Surgery Rheumatoid Arthritis Handbook*. New York: Wiley; 2001.
4. Benedek TG. A century of American rheumatology. *Ann Intern Med* 1987;106:304–12.
5. Yunus M, Masi AT, Calabro JJ, Miller KA, Feigenbaum SL. Primary fibromyalgia (fibrositis): clinical study of 50 patients with matched normal controls. *Semin Arthritis Rheum* 1981;1:151–71.
6. Journal Subject Terms. URL: <http://www.nlm.nih.gov/bsd/journals/subjects.html>.
7. Keers RY. The gold rush 1925–35. *Thorax* 1980;35:884–9.
8. Hoffmeister RT. Methotrexate in rheumatoid arthritis [abstract]. *Arthritis Rheum* 1972;15:114.
9. Hoffmeister RT. Methotrexate therapy in rheumatoid arthritis: 15 years experience. *Am J Med* 1983;75:69–73.
10. Charette L, Misquitta C, Guay J, Reindeau D, Jones TR. Involvement of cyclooxygenase 2 (COX-2) in intrinsic tone of isolated guinea pig trachea. *Can J Physiol Pharmacol* 1995;73:1561–7.
11. The Centers for Disease Control. Eosinophilia-myalgia syndrome and L-tryptophan-containing products—New Mexico, Minnesota, Oregon, and New York, 1989. *MMWR Morb Mortal Wkly Rep* 1989;38:785–8.
12. The Centers for Disease Control. Update: eosinophilia-myalgia syndrome associated with ingestion of L-tryptophan—United States, as of January 9, 1990. *JAMA* 1990;263:633.
13. Garcia-Dorado D, Miller DD, Garcia EJ, Delcan JL, Maroto E, Chaitman BR. An epidemic of pulmonary hypertension after toxic rapeseed oil ingestion in Spain. *J Am Coll Cardiol* 1983;5:1216–22.
14. Hertzell DH. Bibliometrics history. In: Drake MA, editor. *Encyclopedia of Library and Information Science*. 2nd ed. New York: Marcel Dekker; 2003.
15. Lewison G, Devey ME. Bibliometric methods for the evaluation of arthritis research. *Rheumatology (Oxford)* 1999;38:13–20.
16. PubMed Home. URL <http://www.ncbi.nlm.nih.gov/sites/entrez/>.
17. Glazier R, Fry J, Badley E. Arthritis and rheumatism are neglected health priorities: a bibliometric study. *J Rheumatol* 2001;28: 706–11.
18. OLDMEDLINE Data. URL: [http://www.nlm.nih.gov/databases/databases\\_oldmedline.html](http://www.nlm.nih.gov/databases/databases_oldmedline.html).