

# Using Big Data to Evaluate the Association between Periodontal Disease and Rheumatoid Arthritis

Michael A. Grasso, MD, PhD<sup>1</sup>, Angela C. Comer, MPH<sup>1</sup>, Dana D. DiRenzo, MD<sup>1</sup>,  
Yelena Yesha, PhD<sup>2</sup>, Naphtali D. Rische, PhD<sup>3</sup>

<sup>1</sup>University of Maryland School of Medicine, Baltimore, Maryland

<sup>2</sup>University of Maryland Baltimore County, Baltimore, Maryland

<sup>3</sup>Florida International University and NOA Data Science Institute Inc, Miami, Florida

## Abstract

An association between periodontal disease and rheumatoid arthritis is believed to exist. Most investigations into a possible relationship have been case-control studies with relatively low sample sizes. The advent of very large clinical repositories has created new opportunities for data-driven research. We conducted a retrospective cohort study to measure the association between periodontal disease and rheumatoid arthritis in a population of 25 million patients. We demonstrated that subjects with periodontal disease were roughly 1.4 times more likely to have rheumatoid arthritis. These results compare favorably with those of previous studies on smaller cohorts. Additional work is needed to identify the mechanisms behind this association and to determine if aggressive treatment of periodontal disease can alter the course of rheumatoid arthritis.

## Introduction

Rheumatoid arthritis is an autoimmune disease characterized by chronic and systemic inflammatory changes. The specific disease mechanism is not fully understood, but it is believed to involve a combination of genetic and environment factors. It can involve different parts of the body but principally attacks the joints. It is a disabling and painful condition that can lead to substantial loss of function and mobility. Rheumatoid arthritis has a significant impact on society and affects roughly 1% of adults in the developed world<sup>1</sup>. The prevalence of rheumatoid arthritis increases with age, and the disease is roughly three times more common among women. The treatment includes both medical and non-pharmacologic therapies to control inflammation and prevent joint damage.

Periodontal disease is a chronic inflammatory condition of the periodontal tissues. It is one of the most common oral diseases, affecting roughly half of all adults over age 30. The disease is the result of an opportunistic infection by specific microorganisms in the oral cavity. This leads to a destructive inflammatory process, and ultimately to bone and tooth loss, as well as other sequelae<sup>2</sup>. Good oral hygiene and regular professional cleanings are used to prevent and treat periodontal disease. More extensive interventions can be used to treat refractory disease.

The creation of very large clinical repositories has opened new opportunities for data-driven research<sup>3</sup>. One such repository is the national patient database from the Veterans Health Administration (VA), which is available through the VA Informatics and Computing Infrastructure<sup>4</sup>. For each patient, structured information is available, including demographics, diagnostic codes, outpatient visits, hospital admissions, patient orders, vital signs, laboratory testing, inpatient and outpatient pharmacy data, clinical consults, immunizations, mental health screening, associated physicians, and payment information. The repository also includes unstructured and semi-structured information in the form of progress notes, radiology reports, procedure reports, images, and other clinical narratives.

The VA dataset is one of the largest clinical repositories available, providing detailed patient information for approximately 25 million patients who received care at 152 medical centers and more than 800 outpatient clinics across the United States over the past 15 years. The repository includes more than 4 billion progress notes, 2 billion procedure and imaging reports, 1.6 billion medication fills, and 1.5 billion diagnoses. These data will eventually be combined with genomic data from the Million Veteran Program<sup>5</sup>, which when complete, will be one of the largest genomic databases in the world. The VA data set is one of the best, biggest, and most detailed data repositories for studying disease with unprecedented statistical power.

A number of research projects have explored the relationship between periodontal disease and rheumatoid arthritis<sup>6,7,8,9,10</sup>. The exact nature of this relationship has yet to be defined, and most studies to date have been performed on relatively small patient cohorts. The two conditions might have a non-causal relationship through a common mechanism, due to shared genetic and environmental risk factors. Alternatively, they could have a causal

relationship, with one condition influencing the course of the other. In this paper, we present the results of our inquiries using the VA repository to study the association between periodontal disease and rheumatoid arthritis.

**Methods**

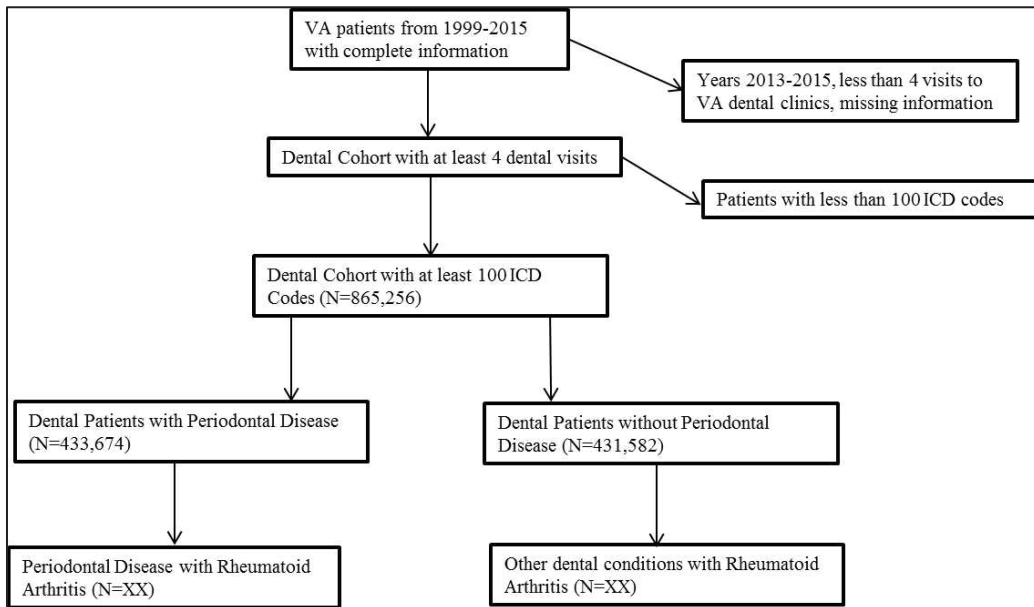
We conducted a retrospective cohort study to measure the association between periodontal disease and rheumatoid arthritis using a very large clinical repository. This work was supported through a grant from The Arthritis Foundation. Our research was also supported with resources and facilities from the Baltimore VA Medical Center, the Veterans Affairs Informatics and Computing Infrastructure, and the Department of Emergency Medicine at the University of Maryland School of Medicine. Regulatory approval for this work was obtained from the University of Maryland School of Medicine, the Baltimore VA Medical Center, and the Veterans Affairs Informatics and Computing Infrastructure.

The VA repository is housed by the Veterans Affairs Informatics and Computing Infrastructure. A secure workspace is provided through a Microsoft Windows-based remote desktop connection within the national VA network. The data are store in a Microsoft SQL Server database. Other applications available in the secure workspace include Excel, MATLAB, Protégé Knowtator, SAS, SPSS, and R.

The VA repository contains data for all patients who received care from 1999 through the present day. As outlined in Figure 1, we limited our search to the years 1999 through 2012. Our dental cohort consisted of all people in the 25-million-patient VA repository who had at least 4 visits in one of the VA dental clinics for any dental issue. We also limited our cohort to people who had enough clinical encounters to have accumulated at least 100 International Classification of Disease (ICD) codes. This focus ensured that enough dental and other clinical information was present to minimize the possibility of bias due to incomplete data. Within this cohort, we identified patients with periodontal disease as those who had at least 4 dental encounters for the condition, indicated as ICD code 523.x.

Also within the dental cohort, we identified people with rheumatoid arthritis based on those who had the appropriate ICD code (714.0, 714.1, 714.2, 714.4, 714.8, or 714.9) and who had at least one positive serologic test (rheumatoid factor or anti-CCP). Our approach to the diagnosis of rheumatoid arthritis was taken from previous work to verify the accuracy of ICD codes<sup>11</sup>. That earlier work concluded that, within the VA population, ICD code 714 is a sensitive screening tool for identifying patients with rheumatoid arthritis. Combined with a positive serology result, it achieved 88% sensitivity and 91% specificity, and had a 93% positive predictive value.

**Figure 1.** Identification of the dental cohort.



We excluded people with missing or incomplete information regarding date of birth, gender, location, or clinical encounters. We calculated odds ratios using 2x2 tables to measure the association between periodontal disease and rheumatoid arthritis. We also calculated 95% confidence intervals to estimate the precision of each odds ratio.

## Results

Half of the patients in the dental cohort had periodontal disease (433,674 out of 865,256). A total of 21,442 patients in the dental cohort (2.5%) had rheumatoid arthritis. Patient demographics are presented in Table 1.

**Table 1.** Patient demographics of the dental cohort.

Number of dental patients	865,256
Number with periodontal disease	433,674 (50.1%)
Number with rheumatoid arthritis	21,442 (2.5%)
Age (years)	
Median	44
25 <sup>th</sup> percentile	36
75 <sup>th</sup> percentile	56
Male	806,944 (93.3%)
Race/Ethnicity	
Caucasian	509,394 (58.9%)
African American	194,546 (22.5%)
Hispanic	49,926 (5.8%)
Asian	5,955 (0.7%)
Other/Unknown	105,435 (12.2%)

As shown in Table 2, the odds ratio between periodontal disease and rheumatoid arthritis is 1.42 (95% CI, 1.37–1.46). This finding holds for men and women, and for those above and below the median age of the cohort. It also holds for Caucasians, African Americans, and Hispanics. It did not hold for Asians, who represented only 0.7% of the cohort.

**Table 2.** Odds ratios for periodontal disease (PD) and rheumatoid arthritis (RA).

Cohort	N	RA+PD+	RA+PD-	RA-PD+	RA-PD-	Odds Ratio	95% CI
All patients	865,256	12,576	8,866	421,098	422,716	1.42	1.37 - 1.46
Age < 44 years	455,070	6,844	4,753	229,472	214,001	1.34	1.29 - 1.39
Age ≥ 44 years	410,186	5,732	4,113	191,626	208,715	1.52	1.46 - 1.58
Male	806,944	10,862	7,754	393,124	395,204	1.41	1.37 - 1.45
Female	58,312	1,714	1,112	27,974	27,512	1.52	1.40 - 1.64
Caucasian	509,394	7,422	5,425	248,078	248,469	1.37	1.32 - 1.42
African American	194,926	2,876	2,100	89,692	99,878	1.53	1.44 - 1.62
Hispanic	49,926	864	432	27,686	20,944	1.51	1.35 - 1.70
Asian	5,955	97	52	3,415	2,391	1.31	0.93 - 1.84

## Discussion

A growing corpus of evidence points to an association between periodontal disease and rheumatoid arthritis<sup>6,12</sup>. Most previous reports were based on case-control studies and meta-analyses, all of which had relatively low sample sizes. Our approach was the first large-population study to confirm this association, demonstrating that subjects in our cohort with periodontal disease were roughly 1.4 times more likely to have rheumatoid arthritis compared to other dental patients.

Several hypotheses have been proposed to explain this association. One suggests that chronic inflammation from periodontal disease can trigger systemic factors, which can induce rheumatoid arthritis<sup>13</sup>. Another hypothesis suggests that rheumatoid arthritis is triggered through an autoimmune response to periodontal bacterial byproducts<sup>14</sup>.

A third hypothesis is based on the existence of a shared pathway in the development of both periodontal disease and rheumatoid arthritis<sup>15</sup>.

Our approach has a number of notable limitations. We identified patients with periodontal disease by ICD codes, and not by chart review or by a direct observation of clinical findings. We attempted to mitigate ICD code accuracy<sup>11,16</sup> by requiring at least 4 dental clinical visits for periodontal disease. Other studies were based on patient-reported information<sup>17</sup>, physician interviews, or review of medical records<sup>18</sup>, which would be very expensive to perform with the very large VA repository. In future work, we plan to refine our patient cohort by applying text analytics for information extraction from electronic health records, as we and other investigators have done on past efforts<sup>19,20,21</sup>.

Our definition of rheumatoid arthritis, based on previously reported work<sup>11</sup>, was shown to be very sensitive in the VA population. However, it was not based on the gold standard for the diagnosis of rheumatoid arthritis diagnosis using clinical and radiographic findings. Another limitation was the exclusion of individuals with negative serologic test results, frequently associated with early rheumatoid arthritis or diagnosis at advanced age<sup>22</sup>. A recent meta-analysis of reports related to the diagnosis of rheumatoid arthritis documented that anti-CCP antibody or rheumatoid factor positivity had only 78% sensitivity (CI, 76%–80%)<sup>23,24</sup>. If we had relied on that definition in our study, we would have missed a substantial number of patients who actually met our criteria.

Finally, our research was limited to the national VA repository. About 90% of the patients whose records are stored in that database are male and about 60% are Caucasian. In addition, our analysis was limited to U.S. veterans who received both medical and dental care through the Veterans Health Administration.

Additional research is needed to explore the significance of this association and of any temporal significance between the two conditions. These investigations could ask whether the treatment of periodontal disease in people at risk for rheumatoid arthritis (e.g., females, older people, those with specific human leukocyte antigen genes) can prevent rheumatoid arthritis. Research is also needed to study whether the treatment of periodontal disease in people with rheumatoid arthritis can relieve its manifestations (fatigue, joint pain, morning stiffness). Finally, research is needed to identify ways to apply big data analytics to large clinical repositories to elucidate more about this association, and to explore temporal correlations between rheumatoid arthritis and periodontal disease.

## Conclusion

We conducted a retrospective cohort study to measure the association between periodontal disease and rheumatoid arthritis. This was the first such study to employ a very large clinical repository. We demonstrated that patients with periodontal disease were roughly 1.4 times more likely to have rheumatoid arthritis compared to other dental patients. These results compare favorably with previous studies on smaller cohorts. Additional work is needed to identify the mechanisms behind this association and to determine if aggressive treatment of periodontal disease can alter the course of rheumatoid arthritis.

## References

1. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. *Lancet*. 2010;376(9746):1094-108.
2. Irfan UM, Dawson DV, Bissada NF. Epidemiology of periodontal disease: a review and clinical perspectives. *J Int Acad Periodontol*. 2001;3(1):14-21.
3. Sinha A, Hripcsak G, Markatou M. Large datasets in biomedicine: a discussion of salient analytic issues. *J Am Med Inform Assoc*. 2009;16(6):759-767.
4. Derby L, Barrett V, Schemnet JS, Trautman T, Nebeker J. VA informatics and computing infrastructure (VINCI) central. Poster Presentation, AMIA Annual Fall Symposium, Washington, DC, November 13-17, 2010.
5. D'Avolio LW, Fiore LD. New models of clinical discovery & science: progress and lessons learned the front lines of the Dept. of Veterans' Affairs transformational initiatives. *AMIA Jt Summits Transl Sci Proc*. 2013;2013:169.
6. Kaur S, White S, Bartold PM. Periodontal disease and rheumatoid arthritis: a systematic review. *J Dent Res*. 2013;92(5):399-408.
7. Detert J, Pischon N, Burmester GR, Buttgereit F. The association between rheumatoid arthritis and periodontal disease. *Arthritis Res Ther*. 2010;12(5):218.

8. Koziel J, Mydel P, Potempa J. The link between periodontal disease and rheumatoid arthritis: an updated review. *Curr Rheumatol Rep.* 2014;16(3):408.
9. Bingham CO 3rd, Moni M. Periodontal disease and rheumatoid arthritis: the evidence accumulates for complex pathobiologic interactions. *Curr Opin Rheumatol.* 2013;25(3):345-53.
10. Ortiz P1, Bissada NF, Palomo L, Han YW, Al-Zahrani MS, Panneerselvam A, Askari A. Periodontal therapy reduces the severity of active rheumatoid arthritis in patients treated with or without tumor necrosis factor inhibitors. *J Periodontol.* 2009;80(4):535-40.
11. Singh JA, Holmgren AR, Noorbaloochi S. Accuracy of Veterans Administration databases for a diagnosis of rheumatoid arthritis. *Arthritis Rheum.* 2004;51(6):952-7.
12. Persson RG. Rheumatoid arthritis and periodontitis – inflammatory and infectious connections. Review of the literature. *J Oral Microbiol.* 2012;4.
13. Golub LM1, Payne JB, Reinhardt RA, Nieman G. Can systemic diseases co-induce (not just exacerbate) periodontitis? A hypothetical "two-hit" model. *J Dent Res.* 2006;85(2):102-5.
14. Rosenstein ED1, Greenwald RA, Kushner LJ, Weissmann G. Hypothesis: the humoral immune response to oral bacteria provides a stimulus for the development of rheumatoid arthritis. *Inflammation.* 2004;28(6):311-8.
15. Bartold PM, Marshall RI, Haynes DR. Periodontitis and rheumatoid arthritis: a review. *J Periodontol* 2005;76(11 Suppl):2066S-74S.
16. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res.* 2005;40(5 Pt 2):1620-39.
17. Arkema EV, Karlson EW, Costenbader KH. A prospective study of periodontal disease and risk of rheumatoid arthritis. *J Rheumatol.* 2010;37:1800-4.
18. Demmer RT, Molitor JA, Jacobs DR Jr, Michalowicz BS. Periodontal disease, tooth loss and incident rheumatoid arthritis: results from the First National Health and Nutrition Examination Survey and its epidemiological follow-up study. *J Clin Periodontol* 2011; 38:998–1006.
19. Dhariwal D, Joshi A, Grasso MA. Text and ontology driven clinical decision support system. *AMIA Annu Symp Proc.* 2013
20. Jahiruddin Abulaish M, Dey L. A concept-driven biomedical knowledge extraction and visualization framework for conceptualization of text corpora. *J Biomed Inform.* 2010;43(6):1020-35.
21. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc.* 2010;17(5):507-13.
22. Pratt AG, Isaacs JD. Seronegative rheumatoid arthritis: pathogenetic and therapeutic aspects. *Best Pract Res Clin Rheumatol.* 2014;28(4):651-9.
23. Sun J, Zhang, Liu L. Diagnostic accuracy of combined tests of anti cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis: a meta-analysis. *Clin Exp Rheumatol,* 2014; 32:11–21.
24. Walitt B, Mackey R, Kuller L, et al. Predictive value of autoantibody testing for validating self-reported diagnoses of rheumatoid arthritis in the Women's Health Initiative. *Am J Epidemiol.* 2013 May 1;177(9):887-93.