Practice of Epidemiology

Regional Epidemiologic Assessment of Prevalent Periodontitis Using an Electronic Health Record System

Amit Acharya*, Jeffrey J. VanWormer, Stephen C. Waring, Aaron W. Miller, Jay T. Fuehrer, and Gregory R. Nycz

* Correspondence to Dr. Amit Acharya, Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, 1000 North Oak Avenue, Marshfield, WI 54449 (e-mail: acharya.amit@mcrf.mfldclin.edu).

Initially submitted March 28, 2012; accepted for publication June 15, 2012.

An oral health surveillance platform that queries a clinical/administrative data warehouse was applied to estimate regional prevalence of periodontitis. Cross-sectional analysis of electronic health record data collected between January 1, 2006, and December 31, 2010, was undertaken in a population sample residing in Ladysmith, Wisconsin. Eligibility criteria included: 1) residence in defined zip codes, 2) age 25–64 years, and 3) ≥1 Marshfield dental clinic comprehensive examination. Prevalence was established using 2 independent methods: 1) via an algorithm that considered clinical attachment loss and probe depth and 2) via standardized Current Dental Terminology (CDT) codes related to periodontal treatment. Prevalence estimates were age-standardized to 2000 US Census estimates. Inclusion criteria were met by 2,056 persons. On the basis of the American Academy of Periodontology/Centers for Disease Control and Prevention method, the age-standardized prevalence of moderate or severe periodontitis (combined) was 407 per 1,000 males and 308 per 1,000 females (348/1,000 males and 269/1,000 females using the CDT code method). Increased prevalence and severity of periodontitis was noted with increasing age. Local prevalence of periodontitis was consistent with national estimates. The need to address potential sample selection bias in future electronic health record–based periodontitis research was identified by this approach. Methods outlined herein may be applied to refine oral health surveillance systems, inform dental epidemiologic methods, and evaluate interventional outcomes.

electronic health records; oral disease; periodontal disease; periodontitis

Abbreviations: AAP, American Academy of Periodontology; CDC, Centers for Disease Control and Prevention; CDT, Current Dental Terminology; CI, confidence interval; EHR, electronic health record; MESA, Marshfield Epidemiologic Study Area; NIDCR, National Institute of Dental and Craniofacial Research; PTP, periodontal treatment procedures.

Periodontitis is a common dental health condition caused by bacterial infection that leads to chronic inflammation of gingival tissue, degeneration of periodontal ligament and bone, and eventual tooth loss. Periodontitis affects up to half of US adults in some form (1–3), resulting in decreased quality of life (4) and increased health-care costs (5). Left untreated, periodontitis also promotes cardiovascular disease and type 2 diabetes by supporting a chronic inflammatory state (1, 6–8).

Despite the serious nature of periodontitis in the context of oral and systemic health, surprisingly little is known about the national prevalence of periodontitis and variability with respect to disease severity. The National Health and Nutrition Examination Survey has been the primary epidemiologic research tool for studying periodontitis (3), with few studies to date utilizing administrative records to conduct periodontitis surveillance at more local levels (where interventions are most likely to occur). This is an important and modifiable gap in the scientific literature, since information captured in electronic health records (EHRs) in dental and medical settings is useful for both patient care and population-based research. Indeed, the current National Institute of Dental and Craniofacial Research (NIDCR) strategic plan states, “Documenting the...
nation’s prevalence of the full range of oral, dental, and craniofacial diseases is an important element of a strategic investment in basic and clinical research. . . . The NIDCR will seek and validate new methods to measure and document oral, dental, and craniofacial diseases, disorders and conditions” (9, p. 51). There has been a strong push by the federal government for health-care systems to adopt and use the EHR, including direct incentives for dentists. In addition, there has been widespread adoption of EHRs in the majority of US dental schools (10).

The generally positive impact of EHRs in patient care has been well documented (11–15), but relatively few dental studies have been completed with secondary use of EHR data. The increasing adoption of EHRs in dentistry highlights the opportunity to “reuse” such clinical data for epidemiologic, outcomes-based, and comparative effectiveness studies in dental/oral/craniofacial research. However, there are few models of oral health surveillance described in the dental science literature, which is a barrier to conducting even basic EHR-based dental research.

In the current article, we describe a regional oral health surveillance platform that queries an electronic clinical/administrative data warehouse, as well as findings from an application of this system in estimating the prevalence of moderate-to-severe periodontitis in an economically disadvantaged area of northwestern Wisconsin. Implications for future EHR-based oral health research in this region and in other health-care systems are discussed.

MATERIALS AND METHODS

Setting

This oral health study was conducted using data from the Marshfield Clinic health-care system (headquartered in Marshfield, Wisconsin), specifically the dental services provided in partnership with the Family Health Center of Marshfield Clinic. The Family Health Center is a Federally Qualified Health Center that has partnered with the Marshfield Clinic since it was first established in 1974. The Family Health Center has a mission to improve access to quality health-care services for low-income and underserved communities in the Marshfield Clinic service area. In 2010, an integrated medical-dental EHR was launched that enabled dentists and physicians to seamlessly access medical and dental health information on their patients across the system (16).

Participants

A cross-sectional analysis was performed using EHR data collected between January 1, 2006, and December 31, 2010, from part of the Marshfield Epidemiologic Study Area (MESA), where Marshfield Clinic dental clinics serve as the primary provider of dental services. As is described in more detail elsewhere (17), MESA is a population-based health surveillance/research resource developed and maintained by the Marshfield Clinic. The sample in this analysis was drawn from residents of 6 zip codes surrounding the city of Ladysmith, Wisconsin (see Figure 1), a rural region of approximately 13,000 people primarily served by Marshfield Clinic medical and dental care systems. Eligibility criteria for this analysis, as of December 31, 2010, were: 1) inclusion in the MESA registry with a home address in the 54848, 54526, 54731, 54530, 54819, or 54563 zip code; 2) age 25–64 years; and 3) completion of 1 or more Marshfield dental clinic comprehensive examinations. Over 90% of adult residents in this geographic study area are included in the MESA registry (18). The age range was selected on the basis of the low likelihood of periodontitis before age 25 years and the low likelihood of dental care coverage after age 64 years (typically due to the transition away from private insurance and the lack of an oral health benefit in Medicare). Because this was a retrospective analysis of health-care data that was part of an ongoing epidemiologic surveillance system, informed consent procedures were waived, and the study was reviewed and approved by the Marshfield Clinic Institutional Review Board.

Measures

All measures were extracted from the Marshfield Clinic’s enterprise data warehouse, which stores data from EHRs and is linked to the MESA registry file. At their source, data on dental variables are collected from comprehensive, full-mouth dental examinations (by licensed dentists and hygienists) and recorded in the EHR. Several sociodemographic and medical covariates were also reported in this study, including age, sex, race/ethnicity, smoking status, personal history of diabetes (i.e., type 1, type 2), personal history of cardiovascular disease (i.e., myocardial infarction, angina, stroke), medical insurance status, and dental insurance status. These data were collected for each individual as part of routine clinical care and documented in the EHR during medical and dental office visits. For each measure, the most recent recorded value as of December 31, 2010, was used for analytical purposes.

The outcome of interest was the prevalence of periodontitis. Periodontal status was established using 2 separate methods. The first method used a rule-based algorithm based on operational definitions of moderate and severe periodontitis outlined by the American Academy of Periodontology (AAP) and the Centers for Disease Control and Prevention (CDC) (19). Specifically, a person was categorized as having moderate periodontitis if 2 or more interproximal sites had a clinical attachment loss of ≥4 mm (not on the same tooth) or if 2 or more sites had a probing depth of ≥5 mm (not on the same tooth). Severe periodontitis was recorded if 2 or more sites had a clinical attachment loss of ≥6 mm (not on the same tooth) and if 1 or more sites had a probing depth of ≥5 mm. The periodontal chart data on interproximal sites came from comprehensive dental examinations performed at Marshfield Clinic’s dental centers. These examinations are part of routine care provided during dental office visits and typically include a formal periodontal disease assessment. Notably, comprehensive dental examinations do not typically occur during emergency oral care visits (e.g., tooth extraction) or orthodontic surgeries.

The second method used for establishing periodontal condition was based on standardized Current Dental Terminology...
(CDT) codes (20) documented in the EHR as they specifically related to periodontal treatment procedures (PTPs). Participants were categorized as having moderate-to-severe periodontitis if 1 or more PTP codes/claims were present in their EHRs. This approach was based on previous methodology developed by Spangler et al. (21), which screened for up to 16 CDT codes indicative of periodontitis.

Statistical analyses

Summary statistics were calculated separately for each method of periodontitis case-finding and stratified by sex. Prevalence was calculated by dividing the number of periodontitis cases by the number of eligible persons present in the analytical sample. Prevalence estimates were then directly age-standardized (22) to the year 2000 US Census estimates of all residents in the targeted zip codes and age ranges. Using $\chi^2$ tests, sociodemographic and medical characteristics were compared between the analytical sample and the remaining MESA residents who had no dental data available between January 1, 2006, and December 31, 2010. This latter group made up the analytical sample. The remaining 5,620 persons were excluded from the analytical sample. As outlined in Table 1, relative to the analytical sample, the excluded group contained significantly more males, persons of unknown race/ethnicity, uninsured persons (both medical and dental insurance), and persons with unknown smoking status, as well as lower rates of known diabetes and cardiovascular disease. When unknown responses were excluded from the race/ethnicity variable, the difference between those included and excluded from the analyses was not significant; however, this was not the case for other sociodemographic variables.

Numbers of periodontitis cases derived using both the AAP/CDC and PTP methods are given in Table 2 by age to ascertain periodontal status based on recent dental records). All analytical procedures were conducted using SAS software (version 9.2; SAS Institute Inc., Cary, North Carolina).

RESULTS

As of December 31, 2010, there were 7,676 persons aged 25–64 years with a known residence in one of the target MESA zip codes. Of these persons, 2,056 (27%) had undergone 1 or more Marshfield Clinic dental clinic comprehensive examination(s) between January 1, 2006, and December 31, 2010. This latter group made up the analytical sample. The remaining 5,620 persons were excluded from the analytical sample. As outlined in Table 1, relative to the analytical sample, the excluded group contained significantly more males, persons of unknown race/ethnicity, uninsured persons (both medical and dental insurance), and persons with unknown smoking status, as well as lower rates of known diabetes and cardiovascular disease. When unknown responses were excluded from the race/ethnicity variable, the difference between those included and excluded from the analyses was not significant; however, this was not the case for other sociodemographic variables.
and sex. Age-standardized prevalence estimates, by sex and case-finding method, are shown in Table 3. Based on the AAP/CDC method, the age-standardized prevalence of moderate periodontitis was 373 (95% confidence interval (CI): 342, 404) cases per 1,000 males and 285 (95% CI: 258, 312) cases per 1,000 females. The age-standardized prevalence of severe periodontitis was 33 (95% CI: 22, 45) cases per 1,000 males and 24 (95% CI: 15, 33) cases per 1,000 females. The age-standardized prevalence of moderate or severe periodontitis (combined) was 407 (95% CI: 375, 438) cases per 1,000 males and 308 (95% CI: 281, 336) cases per 1,000 females. For both males and females, there was a general pattern of increased prevalence and severity of periodontitis with increasing age.

Table 1. Sociodemographic and Health History Characteristics of Residents of the Marshfield Epidemiologic Study Area Aged 25–64 Years With and Without at Least 1 Dental Clinic Comprehensive Examinationa During the Period 2006–2010b

<table>
<thead>
<tr>
<th>Measure</th>
<th>≥1 Comprehensive Dental Examinations (n = 2,056)</th>
<th>No Comprehensive Dental Examination (n = 5,620)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
</tr>
<tr>
<td>Age group, years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–34</td>
<td>534</td>
<td>26</td>
<td>1,459</td>
</tr>
<tr>
<td>35–44</td>
<td>459</td>
<td>22</td>
<td>1,187</td>
</tr>
<tr>
<td>45–54</td>
<td>592</td>
<td>29</td>
<td>1,550</td>
</tr>
<tr>
<td>55–64</td>
<td>471</td>
<td>23</td>
<td>1,424</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1,109</td>
<td>54</td>
<td>2,607</td>
</tr>
<tr>
<td>Male</td>
<td>947</td>
<td>46</td>
<td>3,013</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>1,469</td>
<td>71</td>
<td>2,522</td>
</tr>
<tr>
<td>Nonwhite, non-Hispanic</td>
<td>45</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>Unknown</td>
<td>542</td>
<td>26</td>
<td>3,038</td>
</tr>
<tr>
<td>Medical insurance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>601</td>
<td>29</td>
<td>1,910</td>
</tr>
<tr>
<td>Public</td>
<td>999</td>
<td>49</td>
<td>1,105</td>
</tr>
<tr>
<td>None</td>
<td>362</td>
<td>18</td>
<td>2,112</td>
</tr>
<tr>
<td>Unknown</td>
<td>94</td>
<td>5</td>
<td>493</td>
</tr>
<tr>
<td>Dental insurance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1,461</td>
<td>71</td>
<td>1,143</td>
</tr>
<tr>
<td>No or unknown</td>
<td>595</td>
<td>29</td>
<td>4,477</td>
</tr>
<tr>
<td>Smoker</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>638</td>
<td>31</td>
<td>947</td>
</tr>
<tr>
<td>Former smoker</td>
<td>353</td>
<td>17</td>
<td>625</td>
</tr>
<tr>
<td>Never smoker</td>
<td>726</td>
<td>35</td>
<td>1,621</td>
</tr>
<tr>
<td>Unknown</td>
<td>339</td>
<td>16</td>
<td>2,427</td>
</tr>
<tr>
<td>History of diabetes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>178</td>
<td>9</td>
<td>318</td>
</tr>
<tr>
<td>No</td>
<td>1,878</td>
<td>91</td>
<td>5,302</td>
</tr>
<tr>
<td>History of cardiovascular disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>149</td>
<td>7</td>
<td>256</td>
</tr>
<tr>
<td>No</td>
<td>1,907</td>
<td>93</td>
<td>5,364</td>
</tr>
</tbody>
</table>

a Persons with 1 or more Marshfield dental clinic comprehensive examinations in 2006–2010 were included in the analytical data set, whereas those without 1 or more examinations were excluded because periodontitis status could not be ascertained.

b For each measure, the most recent known value recorded between January 1, 2006, and December 31, 2010, was reported.

Prevalence estimates were somewhat lower using the PTP code method, where the age-standardized prevalence of moderate or severe periodontitis was 348 (95% CI: 318, 379) cases per 1,000 males and 269 (95% CI: 242, 295) cases per 1,000 females. Note that the specific subcategories of periodontitis (moderate vs. severe) were not possible to ascertain using the PTP code method, because such treatment codes do not correspond to periodontitis severity level. Also note that only 4 PTP codes were observed among the periodontitis cases: codes D4341, D4910, D4342, and D4355. Consistent with previous research (21), codes D4910 and D4341 (i.e., manual scaling, root planing, and other maintenance procedures secondary to periodontal disease) were clearly the most common codes, appearing at least once in 77% and 76%, respectively, of all periodontitis cases identified using the PTP method.

A sensitivity analysis combining the AAP/CDC method with the PTP code method was performed to identify all possible moderate-to-severe periodontitis cases and provide a more robust estimate of periodontitis prevalence in the population. Any person with an indication of moderate-to-severe periodontitis, whether via the AAP/CDC algorithm or via 1 or more PTP codes, was considered a case. Under this combined method, the age-adjusted prevalence of moderate-to-severe periodontitis was 496 (95% CI: 464, 528) cases per 1,000 males and 408 (95% CI: 379, 438) cases per 1,000 females. The age- and sex-standardized prevalence of moderate or severe periodontitis overall was 453 (95% CI: 431, 475) cases per 1,000 persons. As outlined in Figure 1, there was also noticeable regional variation in that the overall periodontitis prevalence was highest in the zip codes furthest from Ladysmith. Further sensitivity analyses that restricted the analytical sample to only the 7,156 subjects who resided in the study area for ≥1 year during the data collection timeframe yielded nearly identical results (data not shown).

**DISCUSSION**

To our knowledge, this was the first attempt to characterize periodontitis prevalence using EHRs in a population-based sample. Under a combination of 2 methods of
identifying periodontitis cases (the AAP/CDC algorithm and CDT codes), almost half of the target population was estimated to have moderate-to-severe periodontitis. This is consistent with previous national prevalence estimates of periodontitis in all US adults aged 30 years or older (3). Also consistent with previous national research (23), the burden of periodontitis was greater in males and in older age groups. In terms of regional variation, periodontitis prevalence seemed to be higher in zip codes outside the municipality of Ladysmith, particularly in the eastern areas of the study region, where the population density is very low (approximately 4 residents per square km) and driving distances to the dental clinic can be as great as approximately 50 km (31 miles) (maximum distance in the western study region is less than 40 km (25 miles)). More research is needed to determine the impact of distance from care on periodontitis prevalence, which may reflect a lower willingness to seek any form of dental care (i.e., preventive check-ups, periodontal treatment) because of transportation and/or convenience barriers. In addition, the analysis was generally limited by the lack of comparison with an established “gold standard” assessment of periodontitis, as well as potential selection biases given the observed differences between persons included in the analytical sample and those excluded from the sample.

Both methodological approaches studied seem appropriate for use in other health-care systems depending on availability of data, but the relative advantages and disadvantages of each method should be considered. The CDT code method seems more likely to underestimate prevalence, because it fails to capture some persons with periodontitis who do not seek treatment. Notably, there were no osseous surgeries or bone graft procedures observed in this population, despite the fact that up to 5% of commercially insured adults are known to have received at least 1 of these treatments (21). This is likely related to regional factors in that only general dentists practice at this clinic and most patients receive public health-care assistance. As such, there is very limited infrastructure or financial incentive for providing more advanced periodontal treatments in this area. The CDT code method may also misclassify some cases, because it is possible that some patients receive treatment that is consistent with periodontal disease therapy, yet it is actually being used for a condition other than periodontitis. In addition, this method cannot reliably categorize periodontitis severity levels; thus, some periodontitis cases identified with this approach could have been very mild.

As for the AAP/CDC algorithm, periodontitis status was unknown for 20% of persons using the AAP/CDC algorithm, despite receiving a recent comprehensive dental examination. A periodontal assessment was essentially skipped for these persons, which may have been due to the patient’s appearing edentulous or in perfect oral health, or to other unknown technical barriers to fully completing the examination. The combined approach of using the AAP/ CDC algorithm and CDT codes was sensitive in that it identified the most cases; however, in future research, investigators should conduct validation/adjudication of both methods to determine the extent to which they agree with research “gold standard” periodontal disease assessments (e.g., radiographic examinations).

Approximately one-fourth of the target population had comprehensive oral examination data available during the 5-year data collection time frame. Reasons for the unavailability of data are only speculative, but it may be related to the very low baseline rate of dental care utilization in this area relative to the rest of Wisconsin (University of Wisconsin Population Health Institute (Madison, Wisconsin), unpublished data, 2011). In addition, data on some persons were probably unavailable because they received their dental care outside of the Marshfield dental clinic system (note that of the 9 dentists located within the study area, only 2 were not part of the Marshfield Clinic’s dental centers), although this may be a less influential factor given the area’s low economic status and the Marshfield Clinic’s being the sole provider of subsidized dental care in that area.

A potentially major limitation of the methodology used in this study is related to the sample representativeness of the target region. Missing data is of particular concern if the group with available data substantially differed from the group without it (i.e., data not missing at random), because prevalence estimates may be biased. For example, persons in the analytical sample were much more likely to have known dental insurance than the excluded group. Among the two-thirds of area adults who did not have known dental insurance, only 12% had recent comprehensive oral examination data available. This raises suspicions of residual confounding and suggests that because of the low rate of dental insurance, the observed prevalence of periodontitis in the sample may be a conservative estimate of the true, unobserved prevalence of periodontitis in the population. In general, differences between persons included in and excluded from the analysis in terms of sociodemographic (or other) variables can over- or underestimate prevalence estimates.

Such issues have practical implications for other periodontitis surveillance systems that rely on EHR data. Where possible, it may be useful or necessary to combine data from multiple care systems (particularly in large populations) to construct analytical cohorts that are as representative as possible of the region under study (24, 25). Note that in this and other underserved areas of northern Wisconsin, there are several initiatives currently under way to improve access to dental care (26). If proven successful, such initiatives may also improve the validity of our future informatics-based periodontitis surveillance efforts, because a higher and presumably more representative fraction of the target population will have recent periodontal disease assessment data available. Such data could then be used for more advanced epidemiologic comparisons across finer geographic strata and subpopulations (e.g., incidence ratios, municipalities, and children), thereby assisting program planners in adjusting, refining, and otherwise more precisely identifying regional oral care gaps to address over time.

There remain barriers, however, to the broader translation of the methods outlined in this study. Many dentists do not use EHRs, and those that do may not have standard data
collection and reporting practices. Recent studies suggest that approximately one-third of general dentists in the United States use computers chairside and/or electronically maintain clinical information, with only about 2% of practices being completely paperless (27–29). Although this rate of EHR utilization is below that of medical practices, a California survey found that approximately one-third of dentists who currently do not use an EHR are in the process of implementing one in the near future (29). Such findings suggest that opportunities to conduct regional EHR-based oral health research will continue to increase (30, 31).

In conclusion, consistent with national estimates, approximately 45% of adults in this northwestern Wisconsin study region had evidence of moderate-to-severe periodontitis, with increased prevalence and severity being noted among older males in particular. As in other areas of the country, the target population in this study is of priority interest to the NIDCR, the main dental research institute that promotes Federally Qualified Health Centers as a venue for reaching populations in which major oral health disparities exist. Patients served by Federally Qualified Health Centers tend to be low-income, to be uninsured, and to otherwise live in geographic areas with very limited access to dental health care. Unfortunately, this problem is hastened by a lack of reliable systems for precisely characterizing the “person, place, and time” in which oral health disparities exist and thus where more intense programs, services, and outreach are needed most. As EHR use increases in dental healthcare systems around the country, it will become increasingly important and (likely) cost-efficient to capitalize on the secondary use of such data to improve the public’s oral health. The methods outlined in this study can be built upon to scale and refine current oral health surveillance systems, thereby informing dental epidemiology methods in general and evaluations of dental health-care interventions in particular.

ACKNOWLEDGMENTS

Author affiliations: Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, Wisconsin (A. Acharya, A. W. Miller, J. T. Fuehrer); Epidemiology Research Center, Marshfield Clinic Research Foundation, Marshfield, Wisconsin (J. J. VanWormer); Essentia Institute of Rural Health, Essentia Health, Duluth, Minnesota (S. C. Waring); and Family Health Center of Marshfield, Marshfield Clinic, Marshfield, Wisconsin (G. R. Nycz).

This study was supported in part by funds from the Marshfield Clinic, a grant from Delta Dental of Wisconsin, and Clinical and Translational Science Award 1UL1RR025011 from the National Center for Research Resources, National Institutes of Health.

We thank Marie Fleisner of the Marshfield Clinic Research Foundation for editorial assistance in the preparation and submission of the manuscript.

Conflict of interest: none declared.

REFERENCES


