Assessment of Research Topic Prevalence by Journal Impact Quartile in Oral Health Sciences Using Bayesian Methods

Maria Carolina Cortes1, Luis Vicente Gonzalez2, Laura H. Gunn3,4, Enrique ter Horst5, German Molina6, Silvia Restrepo5, and Juan Diego Zambrano5

Abstract
The relationship between research topics and academic prestige of journals is of relevance to assess venues for current research as well as trending areas of new research. This is of special relevance for those developing a research agenda or with defined productivity outcome expectations. This manuscript extracts prevalent topics using titles and abstracts from more than 10,000 manuscripts, constituting all published research in International Scientific Indexing (ISI) journals within the oral health specialties of oral surgery, orthodontics, and periodontics during 2018. Journals are clustered across four quartile categories according to their impact factors. The novelty of our work includes (a) an examination of a neglected unit of analysis (bigram) in oral health sciences which is of higher relevance than single-word topic definitions and (b) the use of an efficient Bayesian hierarchical approach to extract and rank topics across quartiles with information borrowing. Some topics persisted across quartile groups, while others show higher prevalence in specific quartiles, indicating that topics may find some journal quartiles a more appropriate venue for publication. All quartile groups show a prevalence of empirical research. The approach described in this manuscript offers the possibility to adjust/generate research agendas based on research topic prevalence and dynamics. This methodology is relevant for researchers looking to define their research agendas with potential outcomes aligned with the expectations of quantity and quartile set by their home institutions. It also serves researchers to assess most likely quartiles for publication of their work.

Keywords
oral surgery, orthodontics, periodontics, text analysis, MNIR, research topics

Introduction
Dentistry encompasses multiple clinical specialties. These can be surgical (oral surgery and periodontics) or non-surgical (orthodontics), with novel research topics constantly stemming from them. Quantity and quality of research productivity have become a mainstream expectation across increasing numbers of institutions, further spreading the “publish or perish” concept in academia (McGrail et al., 2006; McKiernan et al., 2019). Knowledge of current trends in research topics becomes relevant to support the definition of research agendas that meet productivity expectations of the researcher’s home institution. While expectations are institution-based, research agendas must meet globally defined, peer-review standards of validity and relevance.

The academic literature is a key distribution channel for new research findings and hypotheses. As scientific venues for publication continue to grow, the number of published manuscripts across disciplines becomes larger by the millions each year (Jinha, 2010). Obtaining relevant scientific information from reputable sources and identifying connections between different scientific topics have become academic challenges (Pletscher-Frankild et al., 2015). Text analysis of discipline-specific relevant topics in specialized journals requires novel statistical tools that can adapt to the big data nature of the problem.
Text mining tools have been successful across multiple disciplines, including computer vision fields (Fei-Fei & Perona, 2005; Luo et al., 2015), statistical sciences (Butt et al., 2021), and social networks (Jiang et al., 2015). Within the health disciplines, the study of cancer (Zhu et al., 2013) or gene-disease associations (Pletscher-Frankild et al., 2015) are among the increasing areas of application where text analysis has been extensively utilized.

One developing area of text analysis is quotation-based bibliometric analysis, which provides information regarding the predominant areas within a research field (Park et al., 2017). The academic interest in scientific content analysis dates at least back to 1987. Eugene Garfield, a pioneer in the field of Scientometry, cataloged the classic references of studies cited in more than 100 scientific articles published in the Journal of the American Medical Association (Garfield, 1987). Since then, this type of analysis has been used in many dentistry disciplines such as dental education (Ullah, Adnan, & Afzal, 2019), oral health (Ullah, Zafar, et al., 2019), dental caries (Gansky, 2003; Workie & Belay, 2019), and orthodontics (Tarazona et al., 2018). These studies are mostly descriptive in nature, focusing on enumeration of terms in highly cited manuscripts or journals. They often-times ignore the differing relevance of topics based on their venue of publication, an issue that is relevant for researchers who may not be able to aim at highest impact journals, but still need to define a competitive research agenda. Also looking at each journal or group of journals independently ignores cross-information (information borrowing) that can enhance the analysis. Citation-based analyses, on the contrary, are biased against novel yet prevalent topics, which may not have been in the literature long enough to achieve high citation numbers. For example, the medical literature on Covid-19 was non-existent prior to 2020, and highly prevalent in the highest impact (medical and otherwise) journals since then, but citations of those works remained low during early 2020. Yet very impactful articles published around that topic in the early stages were hosted in top journals.

In recent years, many academic tools have emerged, all with different academic purposes, such as the altimeter analysis, which is complemented by bibliometrics and provides a better view of the impact of a research topic (Melero, 2015). The automatic literature analysis, and its integration with biomedical data resources, has reached new levels of sophistication with artificial intelligence tools (Feng et al., 2019; Rebholz-Schuhmann et al., 2012). Bayesian methods introduce new tools for information borrowing across clusters (Taddy, 2013b), which is more relevant to jointly assess topic frequency and relevance. This manuscript builds on the methodology of the latter reference and explores current trends within the oral health sciences literature across its disciplines. Some of the tools used for bibliographic analysis have morphed into pay-per-use products, such as inCites (Clarivite, 2021). This has increased the access gap for younger researchers in developing countries. While researchers in wealthier countries/institutions may have access to these tools to shape and gear their research agendas, others in developing countries face an informational gap when defining their areas of research focus. This manuscript provides a novel, freely available, self-contained approach, allowing for both information-borrowing and non-citation dependent analysis of topic prevalence within each discipline and provides younger researchers in low-income countries with a tool to extract relevant information to develop and support their research agendas. Self-containment is key for usability beyond the study period of any manuscript and to help address the issue of equity of access to information (American Library Association, 2021).

Simple word counts per category (e.g., journal quartile) in a descriptive fashion, as is common in the existing literature, assumes independence among the different research categories (e.g., journal quartiles). A more promising approach builds on text-specific dimension reduction methods, which are based on the multinomial form and exchangeability of token counts (e.g., words or combinations of words forming a topic) to borrow information efficiently among categories (Taddy, 2013b). A topic model treats document contents as drawn from multinomial distributions with topic probabilities arising as weighted combinations of “topic” factors. These probabilities are further modeled through a hierarchical approach, which allows for information borrowing among quartiles, thus extracting and utilizing for inference the dependence structure among them. We build on this approach to construct a quartile-based representation of the current state of the oral health literature, and to provide a framework for future analyses.

This manuscript has a dual contribution to the oral health sciences literature, which has focused historically on enumeration or identification of most-cited articles: (a) It provides a novel approach to study multi-word topic extraction beyond keyword analysis, which is demonstrated with a bigram example (two word combination), and (b) it builds on the Bayesian text analysis literature to extract information across journals with different impact characteristics (quartiles) and provides a ranking of topics with information borrowing across quartiles.

Knowledge of quartile-specific topic distributions can provide researchers in the oral health sciences and its disciplines with a tool to both statically and, when performed over time, dynamically assess the (relative) prevalence of topics across quartiles. This assessment can be relevant to avoid mis-alignment between researchers’ agendas and research productivity expectations within their home or prospective institutions. It can also serve doctoral students exploring a research career to assess whether their preferred research topics align with existing and future trends.

**Method**

**Data**

This study identifies the topics of highest prevalence in the oral health sciences’ academic literature during the year
Abstracts (and titles) are considered exchangeable sets of tokens (which can be defined as uni-grams), or combinations of tokens (n-grams; Jurafsky & Martin, 2008). When dealing with text documents, tokens can be understood as regular stemmed words, where, for example, the words “periodontological,” “periodontology,” and “periodontal” are all mapped to a common stem “periodont.” This mapping is performed to reduce the risk of splitting equivalent topics into the multiple possible forms in which they can be expressed, with each such form ranking lower in prevalence than the combined form would rank. Lemmatization would require an up-to-date dictionary, as concepts may evolve and lemmas may develop or change over time. The proposed approach, which only uses stemming, does not require a dictionary to implement. Stop words are removed prior to stemming, following the aforementioned approach.

Following the notation in Taddy (2013b) and Casarin et al. (2019), \(x_i = (x_{i1}, \ldots, x_{ip})\)' is the vector of counts for \(p\) possible bigrams in each document \(i\) (combination of title and abstract). The empirical frequencies can be defined as \(f_i = x_i / m_i\), where \(m_i = \sum_{i=1}^{p} x_{ij}\). Each of the possible \(i = 1, \ldots, n\) documents are linked to a sentiment variable \(y_i\). This sentiment variable represents the quartile of the journal where the manuscript was published, and consists of four ordered categories.

Modeling the conditional distribution of \(y_i | x_i\), can be computationally taxing. Taddy (2013b) proposes to collapse token counts as \(x_i = \sum_{y_j} x_{ij}\), for every sentiment category \(y \in \mathcal{Y}\), leading to

\[
x_i \sim MN[q, m_i]\text{ with } q_{ij} = \frac{\exp(\alpha_j + y_m_j)}{\sum_{j=1}^{p} \exp(\alpha_j + y_m_j)} \text{ for } j = 1, \ldots, p, y \in \mathcal{Y}
\]

where \(x_i\) is a \(p\)-dimensional multinomial distribution with size parameter vector \(m_i = \sum_{y_j} m_{ij}\) and probabilities \(q_{ij} = [q_{ij}, \ldots, q_{ij}]\).

A latent Sufficient Reduction Score \(z_i = \psi f_i\) is defined, and the model is completed with independent gamma-Laplace non-informative priors. These priors allow for efficient calculation of the posterior mode through a simple optimization, avoiding traditional Markov Chain Monte Carlo approaches. This combines the efficiency of optimization with the interpretability of the Bayesian paradigm.

One additional advantage of the method is its tractability. As the information content is effectively added, the computational time increases linearly with the number of documents explored. This makes the technique scalable as the literature in the oral health sciences expands. For more details regarding the algorithm, which is freely available, and implementation of MNIR in R software, see Taddy (2013b).
Results

Bigrams were extracted and analyzed across manuscripts, journal quartiles, and specialty areas (oral surgery, orthodontics, and periodontics). The top 40 ranked topics are listed by specialty and journal quartile in Tables 1 and 2 (oral surgery), Tables 3 and 4 (orthodontics), and Tables 5 and 6 (periodontics). This is an arbitrary number of the highest ranked topics selected to fit the tables, although a significantly larger number of topics is available from the algorithm’s outcome. This range of topics of high prevalence was deemed sufficiently broad to provide readers with an overall picture of the most significant topics within each of the disciplines.

Oral Surgery

Most bigrams were prevalent among all journal quartile categories, with relative rankings within each quartile also relatively stable across most topics. This indicates that, in most circumstances, the quality of the manuscript will likely be the factor determining the journal quartile of publication, rather than current prevalence of the topic in the literature.

Table 1. Forty Most Prevalent Research Topics in Oral Surgery Among Journals in The Top Two Quartiles.

| Rank | Topic                                      | Q1                          | Q2                          |
|------|--------------------------------------------|-----------------------------|-----------------------------|
| 1    | oral maxillofacial                        | oral maxillofacial          |                             |
| 2    | maxillofacial surgeons                    | maxillofacial surgeons      |                             |
| 3    | soft tissue                               | significant differences     |                             |
| 4    | bone loss                                 | control group               |                             |
| 5    | implant placement                         | significant difference      |                             |
| 6    | dental implants                           | head neck                   |                             |
| 7    | craniomaxillofacial surgery               | soft tissue I               |                             |
| 8    | prosthetic dentistry                      | computed tomography         |                             |
| 9    | implants placed                           | bone loss                   |                             |
| 10   | marginal bone                             | implant placement           |                             |
| 11   | cleft lip                                 | dental implants             |                             |
| 12   | bone formation                            | squamous cell               |                             |
| 13   | orthognathic surgery                      | cell carcinoma              |                             |
| 14   | temporomandibular joint                   | craniomaxillofacial surgery |                             |
| 15   | dental implant                            | prosthetic dentistry        |                             |
| 16   | cleft palate                              | implants placed             |                             |
| 17   | oral cavity                              | marginal bone               |                             |
| 18   | alveolar bone                             | cleft lip                   |                             |
| 19   | implant survival                          | bone formation              |                             |
| 20   | bone regeneration                         | orthognathic surgery        |                             |
| 21   | maxillary sinus                           | temporomandibular joint     |                             |
| 22   | third molar                               | dental implant              |                             |
| 23   | bone graft                                | oral cavity                 |                             |
| 24   | class iii                                 | new bone                    |                             |
| 25   | class ii                                  | alveolar bone               |                             |
| 26   | clinical radiographic                     | implant survival            |                             |
| 27   | periimplant bone                          | bone regeneration           |                             |
| 28   | implant stability                         | maxillary sinus             |                             |
| 29   | maxillofacial surgery                     | oral squamous               |                             |
| 30   | le fort                                   | third molar                 |                             |
| 31   | third molars                              | lip palate                  |                             |
| 32   | clinical dentistry                        | cleft palate                |                             |
| 33   | bone level                                | bone graft                  |                             |
| 34   | oral cancer                               | class iii                   |                             |
| 35   | orthodontic treatment                     | class ii                    |                             |
| 36   | sinus floor                               | clinical radiographic       |                             |
| 37   | tooth extraction                          | periimplant bone            |                             |
| 38   | surgeons aim                              | implant stability           |                             |
| 39   | electron microscopy                       | maxillofacial surgery       |                             |
| 40   | first molar                               | le fort                     |                             |

Table 2. Forty Most Prevalent Research Topics in Oral Surgery Among Journals in the Bottom Two Quartiles.

| Rank | Topic                                      | Q3                          | Q4                          |
|------|--------------------------------------------|-----------------------------|-----------------------------|
| 1    | oral maxillofacial                        | oral maxillofacial          |                             |
| 2    | maxillofacial surgeons                    | maxillofacial surgeons      |                             |
| 3    | control group                             | soft tissue                 |                             |
| 4    | head neck                                 | bone loss                   |                             |
| 5    | soft tissue                               | implant placement           |                             |
| 6    | bone loss                                 | dental implants             |                             |
| 7    | implant placement                         | squamous cell               |                             |
| 8    | dental implants                           | cell carcinoma              |                             |
| 9    | squamous cell                             | craniomaxillofacial surgery |                             |
| 10   | cell carcinoma                            | prosthetic dentistry        |                             |
| 11   | craniomaxillofacial surgery               | implants placed             |                             |
| 12   | prosthetic dentistry                      | marginal bone               |                             |
| 13   | implants placed                           | cleft lip                   |                             |
| 14   | marginal bone                             | bone formation              |                             |
| 15   | cleft lip                                 | orthognathic surgery        |                             |
| 16   | bone formation                            | temporomandibular joint     |                             |
| 17   | orthognathic surgery                      | dental implant              |                             |
| 18   | temporomandibular joint                   | oral cavity                 |                             |
| 19   | dental implant                            | alveolar bone               |                             |
| 20   | oral cavity                              | cone beam                   |                             |
| 21   | new bone                                  | implant survival            |                             |
| 22   | alveolar bone                             | bone regeneration           |                             |
| 23   | bone graft                                | maxillary sinus             |                             |
| 24   | implant survival                          | oral squamous               |                             |
| 25   | bone regeneration                         | lip palate                  |                             |
| 26   | maxillary sinus                           | bone graft                  |                             |
| 27   | oral squamous                             | class iii                   |                             |
| 28   | third molar                               | class ii                    |                             |
| 29   | lip palate                                | mandibular canal            |                             |
| 30   | bone graft                                | clinical radiographic       |                             |
| 31   | class iii                                 | periimplant bone            |                             |
| 32   | class ii                                  | implant stability           |                             |
| 33   | clinical radiographic                     | maxillofacial surgery       |                             |
| 34   | case report                               | le fort                     |                             |
| 35   | report describes                          | third molars                |                             |
| 36   | periimplant bone                          | clinical dentistry          |                             |
| 37   | prostodontic society                      | neck cancer                 |                             |
| 38   | implant stability                         | bone level                  |                             |
| 39   | materials total                           | oral cancer                 |                             |
| 40   | maxillofacial surgery                     | orthodontic treatment       |                             |
For example, oral maxillofacial and maxillofacial surgeons were the most frequent bigrams across all journal quartiles. There are some cases, however, where bigrams were more prevalent in specific journal quartiles. Bigrams such as cleft palate were more prevalent among the top two quartiles, indicating that they are topics where researchers succeeded more at publishing in higher quartile journals. This may also be a reflection of the complexity to produce novel results in those areas of study. The low number of manuscripts addressing those topics in lower journal quartiles is also an indication that researchers were successful at publishing them in higher impact journals. In contrast, bigrams such as cone beam were more prevalent in lower quartile journals, indicating the opposite effect. Researchers pursuing such topics were mostly successful in publishing their manuscripts in lower quartile journals, potentially making it harder for them to meet institutional expectations of publication in higher quartile journals.

When looking at the overall trends, dental implant and soft tissue terms were prioritized over temporomandibular joint, facial trauma, craniomaxillofacial surgery, orthognathic surgery, and oral cancer issues.
Prevalence of trends was also similar across quartiles among specialized orthodontics journals. Bigrams like orthodontic treatment, class ii, and class iii had sustained interest across quartiles. The same applies for computed tomography, a topic of great importance for diagnosis based on volumetric three-dimensional imaging, also associated with the emerging virtual planning technology. The converse was found in the lateral cephalograms bigram, which ranks lower.

A prevalent bigram was cleft lip, which was present across quartiles and represents a constant topic in research involving patients with growth and development alterations. This topic poses a clinical management challenge for orthodontic specialists, positioning it among the most prevalent topics of 2018.

Topics such as dental casts did not make it to the top 40 in the highest journal quartile, but they were more prevalent in the lower quartiles. On the contrary, topics such as bond strength were highly prevalent in the top quartile journals (ranked 6th), while almost negligible among other quartiles. This, again, demonstrates that some topics find their homes more easily within journals in specific quartiles.

Table 5. Forty Most Prevalent Research Topics in Periodontics Among Journals in the Top Two Quartiles.

| Rank | Q1                                      | Q2                                      |
|------|-----------------------------------------|-----------------------------------------|
| 1    | periodont disease                       | periodont disease                       |
| 2    | chronic periodont                       | chronic periodont                       |
| 3    | bone loss                               | bone loss                               |
| 4    | alveolar bone                           | alveolar bone                           |
| 5    | soft tissue                             | clinic dentistry                        |
| 6    | clinic attach                           | soft tissue                             |
| 7    | attach level                            | clinic attach                           |
| 8    | dental implant                          | attach level                            |
| 9    | squamou cell                            | dental implant                          |
| 10   | cell carcinoma                          | squamou cell                            |
| 11   | implant placement                       | cell carcinoma                          |
| 12   | connect tissue                          | implant placement                       |
| 13   | periodont therapi                       | connect tissue                          |
| 14   | crevicular fluid                        | periodont therapi                       |
| 15   | periodont tissue                        | crevicular fluid                        |
| 16   | sever periodont                         | periodont tissue                        |
| 17   | gingiv recess                           | gingiv recess                           |
| 18   | bleed probe                             | pocket depth                            |
| 19   | tooth loss                              | bleed probe                             |
| 20   | periodont treatment                     | tooth loss                              |
| 21   | gingiv crevicular                       | periodont treatment                     |
| 22   | plaqu index                             | gingiv crevicular                       |
| 23   | oral health                             | plaqu index                             |
| 24   | porphyromona gingivali                 | oral health                             |
| 25   | stem cell                               | porphyromona gingivali                 |
| 26   | oral squamou                            | stem cell                               |
| 27   | implant place                           | oral squamou                            |
| 28   | oral hygien                             | implant place                           |
| 29   | periodont healthi                       | oral hygien                             |
| 30   | clinic radiograph                       | periodont healthi                       |
| 31   | gingiv tissue                           | clinic radiograph                       |
| 32   | periodont ligament                      | gingiv tissue                           |
| 33   | western blot                            | periodont ligament                      |
| 34   | periodont paramet                       | western blot                            |
| 35   | patient chronic                         | periodont paramet                       |
| 36   | class ii                                | class ii                                |
| 37   | oral caviti                             | oral caviti                             |
| 38   | gene express                            | growth factor                           |
| 39   | attach loss                             | attach loss                             |
| 40   | periodont patient                       | root canal                              |

Table 6. Forty Most Prevalent Research Topics in Periodontics Among Journals in the Bottom Two Quartiles.

| Rank | Q3                                      | Q4                                      |
|------|-----------------------------------------|-----------------------------------------|
| 1    | periodont disease                       | chronic periodont                       |
| 2    | chronic periodont                       | bone loss                               |
| 3    | bone loss                               | alveolar bone                           |
| 4    | alveolar bone                           | soft tissue                             |
| 5    | soft tissue                             | clinic attach                           |
| 6    | clinic attach                           | growth factor                           |
| 7    | attach level                            | attach level                            |
| 8    | dental implant                          | dental implant                          |
| 9    | squamou cell                            | squamou cell                            |
| 10   | cell carcinoma                          | cell carcinoma                          |
| 11   | implant placement                       | implant placement                       |
| 12   | connect tissue                          | crevicular fluid                        |
| 13   | periodont therapi                       | periodont therapi                       |
| 14   | crevicular fluid                        | sever periodont                         |
| 15   | periodont tissue                        | gingiv recess                           |
| 16   | sever periodont                         | pocket depth                            |
| 17   | gingiv recess                           | bleed probe                             |
| 18   | pocket depth                            | show signific                           |
| 19   | lead probe                              | tooth loss                              |
| 20   | tooth loss                              | periodont treatment                     |
| 21   | growth factor                           | gingiv crevicular                       |
| 22   | periodont treatment                     | plaqu index                             |
| 23   | gingiv crevicular                       | oral health                             |
| 24   | plaqu index                             | porphyromona gingivali                 |
| 25   | oral health                             | stem cell                               |
| 26   | porphyromona gingivali                 | cell death                              |
| 27   | stem cell                               | oral squamou                            |
| 28   | oral squamou                            | implant place                           |
| 29   | implant place                           | oral hygien                             |
| 30   | oral hygien                             | periodont healthi                       |
| 31   | periodont healthi                       | clinic radiograph                       |
| 32   | clinic radiograph                       | gingiv tissue                           |
| 33   | gingiv tissue                           | periodont ligament                      |
| 34   | periodont ligament                      | periodont paramet                       |
| 35   | western blot                            | probe pocket                            |
| 36   | periodont paramet                       | patient chronic                         |
| 37   | probe pocket                            | class ii                                |
| 38   | class ii                                | oral caviti                             |
| 39   | patient periodont                       | attach loss                             |
| 40   | healthi control                         | root canal                              |
Periodontics

Similar to the other specialties, most topics remained prevalent across journal quartiles, while others were more prevalent in specific quartiles. As an example, the gene express bigram is a topic that only appeared to be prevalent in top quartile journals. Similarly, topics related to western blot only appeared among the most ranked topics for the top three journal quartiles.

The most prevalent topics in periodontics were those related to clinical periodontal pathology, which is reflected in the periodontal disease, chronic periodont, bone loss, and alveolar bone bigrams. Likewise, very persistent topics were dental implants and implant placements.

Discussion

The number of scientific articles published currently exceeds 114 million, with 2.5 million new publications added each year (Jinha, 2010) across disciplines, and new journals of varying reputability are also created. As those numbers grow, the complexity of their analysis also increases. This study explores research trends across oral health science disciplines using articles published in 2018. Prevalent topics are extracted, ranked, and listed using information borrowing across journal quartiles. This is a major difference with prior studies in the dentistry literature, which focus more exclusively on enumeration with no information borrowing across journal groups (or that focus on elicitation of the most-cited articles).

This list of topics can be dynamic and depends on emerging technological trends and the interests of different research groups leading the discipline in different parts of the world. While the results provided in this manuscript represent a static view of the oral health literature, a dynamic analysis is possible by performing this analysis at multiple time points or across rolling windows.

Previous bibliometric studies have used the number of citations to define the impact of a given topic. This method suffers from several issues: (a) Self-citations, which account for up to 7% of bibliometric measurements (Kulkarni et al., 2011), can inflate specific topics; (b) citation of a manuscript on a given topic does not automatically imply that the citing paper covers the same area of research; and (c) novel topics that recently appear in the literature will have fewer citations, as those are influenced by the age of the article.

The results outlined in this manuscript demonstrate the prevalence of some topics across journal quartiles for each of the disciplines analyzed within the oral health sciences. These can be considered neutral areas of research, where quality of publication, rather than topic, will likely be the most important defining factor for the hosting journal quartile. However, some emerging research topics across oral health sciences were more prevalent in top quartile journals, while researchers focused on other topics found it more challenging for their topics to be published in those journals. This information can guide emerging oral health researchers to align their research agenda with trends in their disciplines. As academics experience increased pressure to publish (both in quantity and journal quality of their publications), identifying research topics becomes an important factor toward defining professional success as a researcher. This is also relevant for new researchers throughout their doctoral programs (and their advisors) seeking objective ways to define their doctoral research topic. Such early decisions about choice of research topics can affect where their initial work is published, as well as their associated career opportunities.

Among the limitations of this study are the constraints set to define the scope of appropriate journals. Some manuscripts in the area of oral health sciences will be published in specialized journals, while others will appear in more general health/medical journals, which were not included. In addition, some manuscripts will be written in languages other than English; however, those were not included in the study. Finally, some journals will not be ISI indexed, which were also excluded from the study. While this exclusion criteria may represent a limitation of the manuscript (a limitation shared by a large portion of the bibliometric literature), the outcomes presented remain relevant. These outcomes reflect the current interests within a large set of well-respected, specialized globally considered journals in oral health sciences. In addition, these exclusions support the provision of a picture of discipline-specific topics only, limiting the influence of areas that may be only tangentially related.

Future studies can build on the work in this study by expanding the bigrams to larger word structures, as well as exploring the dynamics of the topics over time. All bibliographic studies are static representations in nature, providing either a historical picture or, at best, a snapshot of the present. While we also provide a snapshot in time (2018) within the manuscript and demonstrate the advantages within this novel methodology, the approach can be applied over time at any point (through rolling windows of analysis) to identify not only emerging research topics but also those stalling or fading, even if still highly prevalent in the literature. This can be relevant to matriculating graduate students (and their advisors) who may prefer to focus on emerging research topics (even if still of low penetration in the literature) rather than topics already showing signs of exhaustion. For example, Covid-19 manuscripts may be prevalent in the current literature (across sciences). However, a focus on this topic solely based on current snapshots of relevance may be a risky choice for students graduating in several years.

While outside of the scope of this manuscript, the study could also have been restricted to a particular geographical area (region, group of universities, etc.) or comparative group (peer or aspirational universities). This restriction could allow researchers and department administrators to define areas
where research is lacking (and among which departmental hires may be needed) or areas of specialization and differentiation (for marketing/promoting the department).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

ORCID iDs

Laura H. Gunn https://orcid.org/0000-0003-3962-4526
Enrique ter Horst https://orcid.org/0000-0001-5153-1475

References

American Library Association. (2021). Equality and equity of access: What’s the difference? http://wwwALA.org/advocacy/intfreedom/equalityequity

Butt, N., Malik, A., & Shahbaz, M. (2021). Bibliometric analysis of statistics journals indexed in web of science under emerging source citation index. SAGE Open, 11(1), 1–8.

Casarin, R., Camargo, J., Correa, J., Dakduk, S., ter Horst, E., & Molina, G. (2019). What makes a tweet be retweeted? A Bayesian trigram analysis of tweet propagation during the 2015 Colombian political campaign. Journal of Information Science, 47, 297–305.

Clarivate. (2021). InCites benchmarking & analytics: About inCites data. https://clarivate.libguides.com/incites/aboutdata

Fei-Fei, L., & Perona, P. (2005). A Bayesian hierarchical model for learning natural scene categories. In IEEE computer society conference on computer vision and pattern recognition (CVPR’05) (Vol. 2, pp. 524–531). https://ieeexplore.ieee.org/document/1467486

Feng, X., Zhang, H., Ren, Y., Shang, P., Zhu, Y., Yanchun, L., Guan, R., & Xu, D. (2019). The deep-learning-based recommender system “pubmender” for choosing a biomedical publication venue: Development and validation study. Journal of Medical Internet Research, 21(5), e12957.

Gansky, S. (2003). Dental data mining: Potential pitfalls and practical issues. Advanced Dental Research, 17, 109–114.

Garfield, E. (1987). 100 citation classics from the Journal of the American Medical Association. Journal of the American Medical Association, 257(1), 52–59.

Jiang, S., Qian, X., Shen, J., Fu, Y., & Mei, T. (2015). Author topic model-based collaborative filtering for personalized POI recommendations. IEEE Transactions on Multimedia, 17(6), 907–918.

Jinha, A. (2010). Article 50 million: An estimate of the number of scholarly articles in existence. Learned Publishing, 23(3), 258–263.

Jurafsky, D., & Martin, J. H. (2008). Series in artificial intelligence: Speech and language processing (2nd ed.). Prentice Hall.

Kulkarni, A., Aziz, B., Shams, I., & Busse, J. (2011). Author self-citation in the general medicine literature. PLOS ONE, 6(6), Article e20885.

Luo, W., Stenger, B., Zhao, X., & Kim, T. (2015, March). Automatic topic discovery for multi-object tracking [Conference session]. Twenty-Ninth AAAI Conference on Artificial Intelligence. http://bjornstenger.github.io/papers/luo_aaai2015.pdf

McGrail, M. R., Rickard, C. M., & Jones, R. (2006). Publish or perish: A systematic review of interventions to increase academic publication rates. Higher Education Research & Development, 25(1), 19–35.

McKierman, E., Schimanski, L., Nieves, C. M., Matthias, L., Niles, M. T., & Alperin, J. P. (2019). Use of the journal impact factor in academic review, promotion, and tenure evaluations. eLife, 8, e47338.

Melero, R. (2015). Altmetrics—A complement to conventional metrics. Biochimica et Biophysica Acta, 257(2), 152–160.

Park, K., Park, B., Park, S., Yoon, D., & Bae, J. (2017). Top-100 cited articles on headache disorders: A bibliometric analysis. Clinical Neurology and Neurosurgery, 157, 40–45.

Pletcher-Frankild, S., Palleja, A., Tsafou, K., Binder, J., & Jensen, L. (2015). Text mining and data integration of disease–gene associations. Methods, 74, 83–89.

Rebholz-Schuhmann, D., Oellrich, A., & Hoehndorf, R. (2012). Text-mining solutions for biomedical research: Enabling integrative biology. Nature Reviews Genetics, 13, 829–839.

Taddy, M. (2013a). Measuring political sentiment on twitter: Factor optimal design for multinomial inverse regression. Technometrics, 55(4), 415–425.

Taddy, M. (2013b). Multinomial inverse regression for text analysis. Journal of the American Statistical Association, 108(503), 755–770.

Tarazona, B., Lucas-Dominguez, R., Paredes-Gallardo, V., Alonso-Arroyo, A., & Vidal-Infer, A. (2018). The 100 top-cited articles in orthodontics: A bibliometric study. The Angle Orthodontist, 88(6), 785–796.

Ullah, R., Adnan, S., & Afzal, A. (2019). Top cited articles from dental education journals, 2009 to 2018: A bibliometric analysis. Journal of Dental Education, 83, 1382–1391.

Ullah, R., Zafar, M., Riaz, I., & Hasan, S. (2019). Top cited publications on fluoride in relation to oral health: A bibliometric analysis. Fluoride, 52, 426–446.

Workie, M., & Belay, D. (2019). Bayesian model with application to a study of dental caries. BMC Oral Health, 19(4), Article 4.

Zhu, F., Patumcharoenpol, P., Zhang, C., Yang, Y., Chan, J., Meechai, A., Vongsangnak, W., & Shen, B. (2013). Biomedical text mining and its applications in cancer research. Journal of Biomedical Informatics, 46(2), 200–211.