Investigating Bibliometric tools for article screening in a Cochrane systematic review

by

Catherine Gracey

Submitted in partial fulfilment of the requirements

for the degree of Master of Information

at

Dalhousie University

Halifax, Nova Scotia

April 2024

Dalhousie University is located in Mi'kma'ki, the

ancestral and unceded territory of the Mi'kmaq.

We are all Treaty people

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This research applies bibliometric methods to a case study Cochrane review, to describe attributes of the groups of articles that were eligible for inclusion, included, and excluded for research integrity concerns. The variables of interest were the open access status of the articles, the type of funding received (if any), the article authors' institution types, countries, and regions, the CiteScore of each journal that articles were published in, and the citation relationships between articles. These attributes were used to conduct a descriptive analysis of the makeup of each inclusion group, and predictive analysis, comprising of a logistic regression model that predicted inclusion/exclusion on the basis of research integrity concerns. This research proposes a model for using article, journal and author-based metrics to make predictions about the relevance of articles captured in the search, and quality of articles eligible for inclusion.

LIST OF ABBREVIATIONS USED

APC	Article Processing Charge
BC	Bibliographic Coupling
CC	Co-citation
CRG	Cochrane Review Group
DC	Direct Citation
DOI	Digital Object Identifier
EBM	Evidence Based Medicine
ERD	Entity Relationship Diagram
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
ICMJE	International Committee of Medical Journal Editors
MeSH	Medical Subject Headings
OA	Open Access
ORCID	Open Researcher and Contributor ID
PICO	Patient/Population, Intervention, Comparison, Outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomized Controlled Trial

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Philippe Mongeon, for his encouragement, enthusiasm, and constant championing of my work. He is such an advocate for his students, and other early career researchers, and this work would not be possible without his support. Thank you for sparking what I hope will be a lifelong interest in research for me.

Dr. Melissa Rothfus and Dr. Jill Hayden have also both been instrumental in the process of developing and writing this work. Melissa, thank you for your invaluable comments on all things scholarly communications, and Jill, my deepest thanks to you (and your team) for sharing your data and expertise with me.

I would also like to thank my parents, who are also librarians, and encourage me professionally and academically. Your passion for this field is contagious and I'm overjoyed to be colleagues very soon. Finally, to my partner Aaron, thank you for being my rubber duck and for supporting me endlessly throughout this program. I couldn't have done it without you.

Chapter 1: Introduction

1.1 Systematic reviews in health science

Since the Industrial Revolution, researchers have produced scientific outputs at a steadily climbing rate, estimated at approximately 4% increase each year across the sciences. However, this rate is even higher in the Life Sciences, including Medicine and Biology, at 5.07% per year (Bornmann et al., 2021). This constant generation of knowledge is helpful for those who rely upon research findings to support their own evidence-based practice which includes many in the Health Sciences. However, not understanding how to navigate or evaluate information sources can result in information overload in healthcare practitioners, negating the usefulness of new research (Klerings et al., 2015). The concept of filter failure, first defined by Clay Shirky (2008), outlines the issue resulting in the amount of information alone. However, many analog systems for filtering through information and evaluating it no longer work in the digital age. According to Shirky (2008), rather than decrying the amount of scientific research produced, scholars should focus on developing adequate filters and systems to evaluate and synthesize information.

One such research study design that has emerged to filter through and evaluate health information is systematic reviews. While reviews certainly existed in various forms in the pre-digital age, online bibliographic databases, reference and systematic review management software have made large-scale systematic reviews more feasible. Adeptly conducted systematic reviews can act as information filters as authors undertake a screening process to determine which articles attempt to answer their specific research question. Then, they extract the relevant results. As a result, instead of practitioners having to read ten articles that all attempt to answer the same question, they can read a summary of the evidence found in the articles and the implications for practice all in the same document. Besides combatting information overload, systematic reviews can also improve the quality of information healthcare practitioners use. Rigorously conducted evidence synthesis projects such as systematic reviews that have more prescribed methodological expectations and high-quality studies can provide stronger evidence than an individual study, single case report, or clinician's opinion.

In many instances of the evidence pyramid, systematic reviews or other forms of evidence synthesis have been designated the top spot (Chloros et al., 2023; Murad et al., 2016). Nevertheless, the development and use of evidence is continually changing, and methodological best practices continue to evolve and vary between evidence synthesis types (systematic, scoping, rapid reviews). Additionally, reporting standards are continuing to evolve, as methodologists aim to improve these reviews' transparency and reproducibility (Page et al., 2021). Nevertheless, there is a wide range in the quality of systematic reviews, based on evaluations of their methodology and reporting, but also whether their question was relevant and sufficiently focused and whether the authors' interpretation of the results was sound (Coeytaux et al., 2014; Ioannidis, 2016). Reviews undertaken and published by the Cochrane Collaboration have a specific set of methodological requirements they must adhere to and have been assessed to be of higher quality than non-Cochrane reviews (Fleming et al., 2013; Ioannidis, 2016).

1.2 The problem

There will always be articles excluded in a screening process for a systematic review, as the process involves conducting a comprehensive search that will attempt to capture all possibly relevant evidence and then individually evaluating each based on the inclusion/exclusion criteria. Therefore, much of what reviewers exclude during a screening process is due to a lack of relevance to the specific research question rather than a lack of quality. However, plagiarism, poor reporting or study errors have been observed in the specific Cochrane review case study used in this research (Hayden et al., 2023).

These issues can result in exclusion if caught, but if not, their inclusion may dilute the strength of evidence of a review's conclusions. Identifying study characteristics reflecting research planning, conduct and reporting are critical to consider is critical, as these can impact the validity of the conclusions and may result in a high risk of bias. While a great deal of current commentary focuses on the lower quality of articles published by predatory publishers due to suspect peer-review processes, this work also explores the limitations of peer review, highlighting that articles published by reputable journals can still have quality issues.

Additionally, while observing that an article being screened for a review was published in a presumed predatory journal, may inform a reviewer’s impression of an article, exclusion by journal rather than articles is contrary to the philosophy of systematic reviews, and review authors are obligated to review articles individually. Therefore, having a better sense of problematic article-based factors may support authors in generating inclusion or exclusion criteria and expedite their review process.

1.3 Research objectives

This research is a case study of a Cochrane review, applying bibliometric methods to describe the attributes of the articles captured in the search which are then ultimately included or excluded. The purpose of this work is to delineate the bibliographic qualities of the works both included and excluded from the Cochrane Review to determine whether either the inclusion group or the exclusion group have notable ubiquitous qualities. This information could be relevant to future review groups in generating inclusion or exclusion criteria, or in maintaining an awareness of the types of articles that are repeatedly excluded. The analysis will include a descriptive aspect, as well as a predictive model for the variables of interest, to examine whether exclusion can be predicted by any attributes of the articles.

This research is a case study, using the data from a Cochrane review, *Exercise treatments for chronic low back pain: a network meta-analysis* (Hayden et al., 2023). This review is unique in that, upon completion of a relevance screening typical of a systematic review, articles eligible for inclusion were screened for reporting and data issues among other research integrity concerns, and another round of exclusions was made based on this assessment. To collect data, I used the list of article titles and DOIs to query OpenAlex for a more comprehensive list of attributes for each article, author and journal/publisher. This information was then stored in a Postgres database according to the Entity Relationship Diagram (ERD) I designed for this research. Some manual data collection was required to find the CiteScore (a measure of impact in the form of citations) for each journal, and well as the funder type (based on a funder name) as these are not available entities in OpenAlex, an open catalog of scholarly works.

The first aspect of the analysis is descriptive, in which the trends in both included and excluded articles are examined for each of the variables of interest. This section explores the body of work that is eligible for inclusion. However, the clustering analysis includes all articles captured in the review search, as clustering generates groupings based on topical focus, which is related to relevance, not quality.

The second section applies a predictive lens to these attributes to determine whether any variables can significantly and reliably predict inclusion or exclusion in the review within the group of articles eligible for inclusion. A logistic regression model was developed based on the variables of interest, excluding open access (yes/no), and authorship countries, as this data was too granular. However, open access status (type) was included, and authorship region.

The research will address the following questions:

1. What is the relationship between articles' inclusion or exclusion in the review and their open access status?
2. What is the relationship between articles' inclusion or exclusion in the review and the presence of funding, and funder types?
3. What is the relationship between articles' inclusion or exclusion in the review and the affiliation type, country and region of the authors?

4. What is the relationship between articles' inclusion or exclusion in the review and the CiteScore of the journal in which they are published?
5. How are articles included and excluded from the review distributed across publication clusters constructed with different citation-relationship types?

1.4 Assumptions

A critical note to be made here is that this research concerns the quality and relevance of articles included in the Cochrane review. However, quality is not a measure that can be assessed using bibliometric indicators. While a researcher's quality may be discussed in bibliometric research, this refers to their performance regarding research impact (Durieux & Gevenois, 2010). An evaluation of the quality of an article in terms of methodology, reporting and interpretation can be undertaken by subject experts in the form of risk of bias assessments, or assessments of heterogeneity, precision, or other criteria]. However, the authors of the case study review, who do represent subject experts, screened for relevance to their research question and quality (including trustworthiness or integrity characteristics). Therefore, this research is an independent process to describe and predict outcomes of the Cochrane review's authors' already established system, not to conduct a quality assessment of individual articles captured in the review search. Therefore, when discussing quality throughout this work, I refer to the author's assessment of quality (including identification of serious integrity concerns resulting in exclusion), not my own.

Regarding the variables of interest, I selected attributes of articles and authors that I believed to be predictors of quality concerns based on my understanding of issues with systematic reviews and the literature on low-quality articles. When approaching their screening process, the open access status, type of funding received and authorship affiliation, country and region, while known to reviewers, may have no bearing on their decision-making process or may be inappropriate indicators of the quality or reliability of articles in many instances. These variables also all have attributes that may result in spotty reporting. OpenAlex supplies open access status, but many repositories are not easily findable or articles that are more findable on a publisher (paywalled) site, called "shadowed Green" (Piwowar et al., 2018). Therefore, while an article may be available in a repository, it may be missed in my analysis. Analyzing the type of funding received requires authors to report their funding correctly. However, Khamis et al. (2018) reported that up to a third of articles published in a subsection of health policy research did not disclose their funder, and the vast majority did not disclose the role the funder played. Therefore, it is safe to assume data is missing here.

Chapter 2: Literature Review

In this chapter, a review of existing literature on the history, purpose and methodologies of systematic reviews will be explored. This chapter also aims to explain the boom of research in the digital age, and how systematic reviews combat potential information overload or conflicting results. Emphasis is placed on the current trend of open access publishing, and its conflation with predatory publishing. Finally, the concept of research quality and the role of peer review in protecting this is discussed.

2.1 Systematic reviews

2.1.1 The purpose of systematic reviews

The 18th century physician James Lind made a health science breakthrough when he conducted a controlled trial aboard a military ship to test hypothesized scurvy treatments. He has since been heralded for the forethought in controlling confounding variables when conducting his now famous research, as he specifically selected a population that was "as similar as [he] could have them" (Lind, 1753, p. 191), therefore minimizing the potential effects of age, diet, medical histories, and other variables on his results (Milne, 2012). However, Lind has a lesser-known contribution to the health science field: his subsequent systematic review on the same research question (Clarke & Chalmers, 2018). In his attempt to summarize the existing evidence to find a clear answer, he deemed it necessary to remove a great deal of rubbish science from his pool of texts (Lind, 1753).

While systematic reviews did not become widely popular until the 1980s, and experts continue to refine their methods, Lind had struck upon their importance centuries earlier (Clarke & Chalmers, 2018). Health science research has grown exponentially since

Lind's heyday, and it would be impossible for medical practitioners to keep up with the publication rate of articles (Gopalakrishnan & Ganeshkumar, 2013). Even if this were possible, clinicians would face contradictions, as different populations, intervention methodologies, and experimental conditions yield different results (DeCoursey, 2006).

Enter the systematic review, which is, simply put, a research project undertaken by a team of researchers aiming to collect all published evidence on a research question and summarize the results (Stevens, 2001). These conclusions can inform clinical practice guidelines, supporting Evidence-Based Medicine (EBM) (Hardi & Fowler, 2014; Stevens, 2001). The importance of EBM in modern medicine illustrates a shift in perceptions of best practices for decision-making in patient care. Throughout the 20th century, clinicians placed more weight on the experiences and authorities of senior practitioners, and Randomized Control Trials (RCTs) were seen as the most potent research evidence (Eddy, 2011; Kamath & Guyatt, 2016).

However, widespread inconsistencies between research-based evidence and clinicians' opinions indicated a need to refer to synthesized research summaries to inform practice (Kamath & Guyatt, 2016). Synthesized evidence uses a larger dataset than the average physician would see, and empirically measures effects and risk of bias, producing evidence that is likely stronger than other single articles or individual opinions. The development of online bibliographic databases allowed for applying a search to millions of records and export the results. Specific systematic review management software like Covidence allows for authors to review the articles and keep track of exclusion stages and reasons.

2.1.2 Accepted methodologies & best practices

When James Lind conducted an early version of a systematic review, there were no established methods for designing and reporting these projects. Over time, the importance of reproducibility has grown, necessitating more transparency in research documentation (Glasziou et al., 2014). If the authors have clearly stated the methodologies used, it is more feasible to recreate their process and verify their findings.

Therefore, there is a set of constantly evolving best practices in designing and publishing systematic reviews. In this section, I will describe the steps taken in a systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. These guidelines were designed specifically to standardize review methodologies and improve their transparency and have been adopted widely across systematic reviews (Page, McKenzie et al., 2021; Page & Moher, 2017). I will emphasize the systematic review process's screening and quality assessment steps, as these processes offer opportunities for additional quality assessment tools.

First, authors must formulate a research question (Khan et al., 2003). Researchers often use pre-existing structures, like PICO (Population/Patient, Intervention, Comparison/Control, Outcome), as this can ensure a question is clear and can be helpful when putting together the subsequent search strategy (Eriksen & Frandsen, 2018). Authors should also conduct a preliminary search to determine if there is a recent, existing systematic review on the same research question to avoid replication of research efforts

(Ioannidis, 2016). Once the researchers are satisfied that their systematic review will add a novel contribution to the field, they move on to the searching step.

Systematic reviews are intended to be comprehensive in scope. Therefore, literature searches should aim to capture all results relevant to the research question, with the subsequent screening process serving as the opportunity to weed out irrelevant studies (Tawfik et al., 2019). Authors should generate a search strategy to be used in an academic database of published literature, potentially using a framework like PICO, and include the relevant keywords and index terms for a single database (Eriksen & Frandsen, 2018). Librarians or other information professionals can be consulted or involved in designing this strategy, as they have extensive knowledge of databases and their related syntax, and librarian authorship has been linked to higher-quality search strategies (Rethlefsen et al., 2015). Researchers then translate the initial search strategy to adhere to the syntax and index terms of each other databases that will be searched (Bramer et al., 2017). The number and choice of databases used will depend on the discipline and breadth of the research question, but multiple databases should be used (Bramer et al., 2017).

Throughout the search process, documentation of the search strategies dates that reviewers searched databases, and the number of results must be maintained to meet the reporting standards dictated in PRISMA (Page, McKenzie et al., 2021). In this research, the results from all searches run for the case study review are used to generate clusters based on citation connections to generate groups that share topical focus. This is because at this stage of the review, while all results will have some type of relation to the research question (shared keywords or controlled vocabulary), they aren't necessarily relevant to the research question, and thus might be excluded based on relevance.

After the results from each database search are compiled, authors will often upload these results into a systematic review management tool like Covidence, which can aid in deduplication, and keeping track of exclusion reasons. Once results have been compiled and a system for tracking exclusions is set up, screening occurs (Kellermeyer et al., 2018). Two levels of screening must take place, first based on the titles and abstracts of each article, then in the second round, the full text of the articles are reviewed to determine whether they fit the parameters of the research question (Tawfik et al., 2019).

Two researchers complete screening independently, assigning a "yes" or "no" to an article at the first stage. When authors agree, they move an article to the next stage or exclude it. When there is a disagreement, a third researcher serves as a tiebreaker (Waffenschmidt et al., 2019). During the second round of screening, authors must assign a specific reason for exclusion when they decide not to include a study. These reasons can be based on the population, outcome, or another quality of the study design (Saif-Ur-Rahman et al., 2022). A breakdown of the number of studies excluded at this stage by reason for exclusion will be displayed in the PRISMA diagram (Page, Moher, et al., 2021).

The final steps are summarizing the evidence and then interpreting this summary (Khan et al., 2003). In the case of a meta-analysis, this will involve forest plots or other visualizations of the calculated outcome measures (Gopalakrishnan & Ganeshkumar, 2013). Discussion on the statistical implications of these findings will follow the results section.

2.1.3 Cochrane reviews

The quality of systematic reviews will vary depending on adherence to best practices, study design, and team composition. Systematic reviews published by Cochrane and stored in the Cochrane Library have been found to adhere more closely to methodological best practices than non-Cochrane reviews (Fleming et al., 2013; Ioannidis, 2016). These reviews are required to adhere to specific methodological and reporting requirements that go beyond expectations for non-Cochrane reviews and are required to improve transparency and reliability (Moseley et al., 2009). One key difference is that authors must submit a protocol to the appropriate Cochrane Review Group (CRG), a committee comprised of experts on a specific subdiscipline of health (Cumpston & Flemyng, 2023). CRGs support authors throughout the research and reporting process, and reviews cannot move forward without a connection to an existing CRG, as these reviews would go unsupported (Cumpston & Flemyng, 2023). The CRG will also evaluate the author team to ensure that the International Committee of Medical Journal Editors criteria for authorship (ICMJE 2018) is applied. While this process may change over time based on funding and new methodological developments, a collaborative approach to review development supports methodological rigour. Cochrane reviews are intended to be regularly updated to reflect new evidence in a field (Cochrane Library, 2023).

Adherence to the requirements of publishing a Cochrane review and continually updating the findings results in a heavy workload for authors, and a single review can take many years to be published (Chapman et al., 2023). However, authors continue to publish with Cochrane, potentially due to the higher quality of Cochrane reviews (Ioannidis, 2016). However, current methodological standards for Cochrane reviews do not require that authors screen for predatory journals. Individual articles published in predatory journals are generally of lower quality than those published in journals with thorough peer-review processes (Hayden, Ellis, Ogilvie, Boulos, et al., 2021), and it important to note that the authors of this review did examine the journal/publisher for each article eligible for inclusion, which informed their decision about serious research integrity concerns. In the next section, I will expand on issues with conducting systematic reviews generally but will not focus on the quality of individual articles included in reviews until section 2.2.4.

2.1.4 Current issues in systematic reviews

Systematic reviews have long been considered at the top of the evidence pyramid in the health sciences, but recent conversation has focused on whether this designation is appropriate (Murad et al., 2016). The typical figure of a pyramid, with systematic reviews occupying the top spot, has been criticized as an overly simplistic representation of the strength of evidence, particularly because the heterogeneity of studies included in a systematic review/meta-analysis can limit their applicability (Murad et al., 2016). The 6S Hierarchy of Evidence Based Hierarchy is one such instance where systematic reviews are not seen as the best source of clinical guidance. Systematic reviews are placed in the middle of the pyramid, with synopses of reviews, evidence-based clinical guidelines/textbooks, and computerized systems that use patient data and algorithms to support clinical decision making on top (Dicenso et al., 2009).

Due to their esteemed status within the evidence pyramid, systematic reviews hold significant weight in evidence-based practice and guideline creation (White & Waddington, 2012). Therefore, systematic reviews that are poorly designed or not adequately rigorous in their search or screening can cause harm to the health sciences (Ioannidis, 2016). John Ioannidis (2016) argued that the level of redundancy in the pool of biomedical systematic reviews was too high, confusing for those trying to interpret contradicting findings. Navigating these contradictions is challenging, as there is no universally accepted methodology for doing so (Ioannidis, 2016). Therefore, while systematic review authors conduct their reviews to clarify the findings on a research topic, different author groups continually generating reviews on identical research questions may do more harm than good. The quality of systematic reviews will be further explored in section 2.2.4.

Another challenge to systematic reviews was COVID-19. Systematic reviews typically take over a year to complete, depending on team size, experience and availability (Borah et al., 2017). However, during COVID-19, updated evidence was required on far shorter timelines; a year was too long to wait. Some researchers pivoted to 'living' systematic reviews, which were continually updated to incorporate new evidence in response to this (Pearson, 2021). Alternative forms of evidence synthesis were also utilized to make sense of evidence quickly, such as rapid reviews (Fretheim et al., 2020). These reviews use the same framework as systematic reviews, but reviewers employ changes that speed up the process, like having single reviewers at various screening steps and/or using filters to limit years or other factors (Ganann et al., 2010). These changes can result in a far shorter timeline, as low as three weeks, but may introduce errors or bias due to the less rigorous nature of the process (Ganann et al., 2010). The flexibility and speed of these reviews are desirable to those with limited research time or decision-makers who must come to conclusions quickly to support human health. Thus, systematic reviews no longer maintain an uncontested dominance over other evidence synthesis formats.

My research is focused on this challenge: the quality of systematic reviews can be influenced by the individual articles captured in a search (Hayden, Ellis, Ogilvie, Boulos, et al., 2021). If a search is comprehensive, the authors may have a great deal of articles to sort through, depending on the topic and scope of the question. If the articles they do decide to include have significant methodological or reporting errors, this can impact the quality and reliability of the review (Munn et al., 2021). Yet, authors must still review each individual article for inclusion. In the next section, I will describe the current academic publishing landscape to explore why there is a high quantity of articles, many of which are low quality, existing in the literature.

2.2 An overview of modern academic publishing

2.2.1 The publish or perish dynamic

Academics hoping to achieve tenure or promotion are expected to achieve a high publication level, often in specific journals (De Rond & Miller, 2005). Both teaching and research is required by most professorship appointments, but publishing is the "only activity" (Bello et al., 2023) that matters for promotion and tenure in many institutions. While notable professional service and excellence in teaching might be rewarded at the department or sometimes school level, exceptional research is often conferred with prestige at the national level, which can attract funding or desirable students to the school (De Rond & Miller, 2005). This dynamic may push academics to pursue research topics they expect to be of interest to journals, no matter what their actual interests may be, and to submit articles with 'positive' findings (Fanelli, 2012). This so-called outcome bias, or the results of the study being the determining factor in the article being published, is growing in the health sciences (Fanelli, 2012) and may be skewing the published literature to reflect only part of the story, namely, positive findings.

Academics worldwide are feeling this pressure (Grančay et al., 2017), and journals have sprung up to meet this demand for publication venues. While journals operated by traditional academic publishers are the most common, academic journals exist within numerous settings, including within academic departments or professional associations. However, these journals may be lesser known, and the articles published by them may receive fewer citations than those published in their more high-prestige counterparts (Starbuck, 2005), due to a lack of name recognition/prestige, financial resources, or not being indexed by major bibliographic databases.

The number of publications and citations often hold a great deal of weight in the tenure and promotion process and have been highlighted as an empirical way to measure success, in contrast with the influence of the 'old boys club' (Seeber et al., 2019). In the research impact game, where researchers can point to quantitative measures of their success, there will be winners and losers, as numbers of citations can separate out those who have had the most ‘impactful’ work. However, this dynamic may reduce the evaluation of the importance of research to the venue in which an author publishes, as even a run-of-the mill paper published in Nature might beat out a revolutionary article published in an institutional repository.

The journal conversation also brings into question equity in academic publishing, as journals with APCs will require payment from authors, no matter how brilliant their work is. If an author cannot afford to publish in a high-impact journal and they must find a home for their work elsewhere, their research impact metrics may suffer. This dynamic turns academia into a research impact competition in which some actors will play dirty to get ahead, where paper mills, citation mills and journals accepting APCs without doing their due diligence to ensure research integrity and quality can run rampant.

Anderson et al. (2007) found that early to mid-career researchers identified that their colleagues were hesitant to share materials and information and were reluctant to publish documents or data as they felt doing so would allow others to copy their ideas. The competitive atmosphere of their work environments, while meant to push researchers to produce high-quality research publications, was, in effect, slowing or halting the scientific process and dissemination of information (Anderson et al., 2007). Researchers trying to game the system to inflate their research impact metrics may also engage in excessive self-citation (Seeber et al., 2019). While properly attributing one's source is critical, even if that source is one's research, excessively self-citing to boost citation numbers is considered a questionable research practice. Rather than citing the best possible example or the thought leaders on a topic, authors consistently plugging their work undermines established norms in academic publishing (Seeber et al., 2019).

While it is evident that the publish-or-perish dynamic harms researchers, it is also essential to consider the impact on society more broadly. Louis et al. (1995) found that perceived competitiveness positively correlates with misconduct. Gopalakrishna et al. (2022) found a similar pattern, in which pressure to publish was the strongest predictor of questionable research practices. The disruption to the flow of information described above, in combination with potential misconduct, may compromise the quality of the health information accessible to the public.

2.2.2 Inequalities in publishing

As the nature of academic publishing is competitive, there are always winners and losers whose identities shift as metrics of success change. While research impact might seem a sound way to even the playing field and ensure that all researchers get a fair chance to be at the top of the field, there are two significant obstacles to this: the Matthew effect and biases/barriers that prevent certain demographic groups from keeping pace.

The most impactful articles, or the researchers with the most recognition, are not necessarily those with the highest quality or quantity of output. As noted, decades ago by Merton (1968), the authors who tend to get the most recognition are those who already have some recognition, also known as the Matthew effect. For instance, when reading a paper with 20 authors, the reader is more likely to remember the Nobel laureate than the junior scientist, and therefore, the laureate receives more credit in the public mind (Merton, 1968). This idea, that metaphorically, the rich get richer, applies to outcomes other than public recognition. Those who are successful in securing funding early in their career are more likely to receive mid-career funding than those who did not have funding early in their career (Bol et al., 2018). This effect may be due to the previously unsuccessful researchers self-eliminating due to a perceived waste of time and resources. However, providing information about previous grants received may likely influence selection committees to contribute to the Matthew effect (Bol et al., 2018).

The logical conclusion for early career authors would be to get published in a high-impact journal and secure funding early to establish themselves. However, academia's ‘old boys club’ is still a reality. Research from the 'scientific core,' namely Western Europe, North America and Oceania, receives more citations than similar studies out of the periphery (Gomez et al., 2022). Research conducted by non-white scientists has longer review processes and receives fewer citations (Liu et al., 2023); female principal investigators (PIs) had heavier teaching loads and lower base pay than their male counterparts (Acton et al., 2019). All of this is to say that factors outside a researcher’s control can stunt their ability to establish themselves within their field.

2.2.3 Open access publishing

Traditional, or closed academic publishing can exacerbate inequities. Tax dollars pay for research funded through grants, yet the results of the studies are often behind paywalls and, therefore, inaccessible to the public. In Canada, the Tri-Agency Open Access Policy on Publications aims to avoid this scenario by requiring that all funding from

CIHR, NSERC and SSHRC must be made open access (OA) within 12 months of publication (Government of Canada, 2023). OA's proponents presented this system as an alternative to the traditional scholarly academic publishing system, wherein readers (often through libraries) paid substantial fees to access academic content. Those without affiliation to a university, hospital or other organization with a great deal of resources may then have to pay out of pocket to access important works, or find an alternative source, potentially adding time to their research process. It is, therefore, critical to note that OA publishing was designed to support the inclusivity of research (Smits & Pells, 2022), as closed publishing disproportionally blocks readers who cannot pay associated fees (Smits & Pells, 2022) and risks insulating academic conversations and knowledge exchange from the world outside university/government settings (Rueda, 2023).

Nevertheless, the OA movement is not solely focused on reducing the financial burden for readers, as another significant component of its rationale is efficiency in the sharing of knowledge and collaboration, uninhibited by barriers to access (Ding & Li, 2021; Smits & Pells, 2022). As noted by Ding and Li (2021), scientific knowledge generation requires background knowledge and shared understanding of methodology and analysis, which is hampered in a closed publishing system. The OA system allows for a more accessible and efficient dialogue on various scientific topics.

Although OA publishing is now widespread, there are distinct subtypes of OA publishing options offered to authors by publishers. Open access status refers to categories based on where and how publishers allow a research article to be shared (Piwowar et al., 2018). Gold OA journals are entirely free and accessible to all readers.

However, publishing in gold OA journals can be inaccessible to authors with limited funding, as these journals charge Article Processing Fees (APCs) that can be thousands of United States Dollars (USD) (Singh et al., 2021). As an alternative, many publishers maintain a paywall but allow authors to post the article's content in a repository, known as Green OA (Priem et al., 2022). While OA certainly makes information more accessible to the public, it often shifts the financial responsibility for paying for research dissemination to authors, and the pressures of generating research of substantive quality to publish and financing its publishing can be challenging (Willinsky & Rusk, 2019).

2.2.4 Research quality

Perceived journal prestige is essential to authors looking to publish their work because of the associated high impact that comes with the reputation of a journal (Callaham et al., 2002). As a result, prospective authors perceived OA journals as low-quality, especially at the beginning of the OA movement, when they lacked the prestige of long-standing traditional journals (Beaubien & Eckard, 2014). However, legitimate OA journals have many of the same markers of legitimacy as traditional academic publishers, including an editorial board, a transparent publishing process for authors and a thorough peer-review process (Wicherts, 2016). Nevertheless, journal prestige and quality are not synonymous (Bray & Major, 2022).

Research quality can refer to the trustworthiness, integrity, risk of bias, or certainty of a body of evidence in the context of a systematic review. The overall trustworthiness of a systematic review may be evaluated by examining the methods of the used, i.e., whether the authors adhered to any reporting guidelines, whether their methods were transparent and repeatable, and other factors. (Garcia-Doval et al., 2017). The level of certainty of the evidence collected within a systematic review on a specific research question can be evaluated using tools like GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) (Guyatt et al., 2011). This framework guides authors through assessing the risk of bias, imprecision, inconsistency, indirectness and publication bias of the studies in the review to generate a rating on the quality or certainty of the evidence available (Siemieniuk & Guyatt, 2023).

Risk of bias evaluations are often completed, even if authors are not using GRADE (Viswanathan et al., 2017). This evaluation is conducted to identify studies in which the reported outcomes may not accurately represent a phenomenon based on methodological limitations (Siemieniuk & Guyatt, 2023) and can help explain heterogeneity in the data (Viswanathan et al., 2017). The integrity of data of articles can influence the quality of systematic reviews; if data is incorrectly reported, analyzed, manipulated, or replaced, the results will not be accurate (Zarour et al., 2021). Looking for errors, inconsistencies, or glaring omissions can help protect a systematic review from overstating the effect of an intervention based on the data in a study (Bolland et al., 2016).

The trustworthiness of a systematic review and its quality of evidence will vary based on the methodologies employed and the tools and approaches to evaluate articles. When assessing trustworthiness and quality, it is vital to take a more nuanced approach than assuming systematic reviews are inherently the best evidence, being at the top of the evidence pyramid, and that Randomized Controlled Trials (RCTs) are infallible and always of better quality than observational studies (Viswanathan et al., 2017). Theoretically, articles published in a scholarly journal should have been peer-reviewed and, therefore, should be methodologically sound, analyzed appropriately, and reported truthfully. However, peer review has some limitations, which are discussed below.

2.2.4.1 Peer-review

Peer review is a system to assure research quality, usually before publication. The rationale behind peer review is that other experts in a discipline can screen the authors' methods, analyses and conclusions to evaluate whether they were appropriate (Kelly et al., 2014). Peer reviewers read the article, provide comments and recommend that an article be published, that authors conduct major revisions, or that the article be rejected entirely (Wicherts, 2016). This process should result in a published article vetted by experts that the reader can trust (Kharasch et al., 2021). However, adequately reviewing an article can take a long time and it can be challenging to find willing and qualified reviewers, resulting in a delay in publication (Kelly et al., 2014). Additionally, it is possible that a peer reviewer may not have the level of expertise required to catch mistakes (Huutoniemi, 2012), meaning peer-reviewed articles can still be problematic.

Peer review occurs for traditional and OA journals, as peer reviewers are volunteers and do not cost money (Wicherts, 2016). However, a subpopulation of OA journals purport to have a peer-review system but publish their articles far too quickly for the process to have occurred (Hansoti et al., 2016). These journals are problematic, as articles can be interpreted as reliable when they have substantial issues with their methodology, data analysis, or reporting. The publishers who engage in this practice are sometimes called predatory publishers and are further discussed in the following section.

2.2.4.2 Predatory publishers

Predatory publishers have been identified as a significant issue in modern scholarly communications (Beall, 2012; Grudniewicz et al., 2019; Sax, 2018). The term "predatory publishers" was coined by Jeffery Beall, a librarian, who defined them as organizations "which publish counterfeit journals to exploit the open-access model in which the author pays. These predatory publishers are dishonest and lack transparency" (Beall, 2012).

Beall maintained a list of publishers he deemed predatory, which is no longer updated and has been archived (Beall, 2021). Although it is unclear exactly why Beall no longer maintains this list, publicly naming publishers or journals one deems predatory can result in repercussions like article retraction (if the naming happens in a research publication) or threats of legal action (Kincaid, 2022). Therefore, allowing authors to come to their own conclusions based on checklists has become the more popular way of protecting against predatory publishers (Cukier et al., 2020; Ross-White et al., 2019). Systematic review authors must be aware of predatory publishers and their characteristics, as including them in a review may result in a higher risk of bias due to reporting and methodological mistakes (Hayden, Ellis, Ogilvie, Boulos, et al., 2021). However, experts do not indicate that all systematic reviews should exclude articles from publishers presumed to be predatory, but instead that authors should make decisions on a case-by-case basis and ensure other analytical techniques for assessing quality, as discussed in 2.2.4, are applied (Munn et al., 2021).

2.2.5 Research quantity

A repercussion of publish or perish is that academics must aim to publish whenever possible, meaning that not every article makes a novel contribution to their field (Rawat & Meena, 2014). There has been exponential growth in scholarly article production (Fire & Guestrin, 2019). As the amount of literature on a topic grows, the more results there will be for each review update or new review (Hayden, Ellis, Ogilvie, Boulos, et al., 2021). The vast amount of information can make an evidence synthesis project overwhelming and sometimes not feasible, as the expected methodologies of a review remain rigorous, whether one is examining 20 articles or 500 (Chloros et al., 2023). This research quantity points to the need for tools that can help authors identify problematic articles to sort through them faster.

2.3 Bibliometrics

Bibliometric analyses have been applied frequently in the health sciences, as the pre-eminent method of publishing research findings in the field is in scholarly journals which are indexed in bibliographic databases. Previous scholarship in the field has focused on the growth of science (Bornmann & Mutz, 2015) and what types of study designs are related to higher citation rates (Patsopoulos et al., 2005). Bibliometric studies typically focus on data fields representing article, journal or publisher metadata, like article title, words used in the abstract, author names, and countries authors work in (Thompson & Walker, 2015).

Systematic review authors have used bibliometric methods and tools to support a descriptive analysis of the body of work they are summarizing by evaluating key authors or journals in a field, visualizing how authors are connected, or text-based factors, like frequently used keywords (Linnenluecke et al., 2019). However, very little, if any work has been done using bibliometric indicators to predict whether an article will be included in a systematic review. Applying a statistical analysis to a group of articles can involve generating regression models that predict whether an event happens based on a change in some other variable (Bornmann & Williams, 2013). Therefore, this is a gap in the literature I will explore in this research.

2.4 Variables of interest

While many characteristics could be predictive of article inclusion or exclusion, I have focused on the five following attributes: the open access status of an article, the type of funding body which provided financial resources, the authorship institution types/countries/regions, the CiteScore of the journal that published each article, and the clusters of articles based on various types of citation relationships.

Examining open access status at the journal level is becoming complicated amid the rise of alternative forms of open access publishing. For instance, hybrid journals allow authors to pay to publish their work as open access, while the rest of the articles in the journal are only available via the traditional subscription-based model (Björk, 2017). Additionally, authors can publish in a subscription-based journal (not open access) but self-archive their work in institutional or disciplinary repositories, making their work open access, regardless of whether the journal is open access. Therefore, rather than considering whether the work is published in an open access journal, this research examines the individual article's open access status, as classified in OpenAlex, the bibliographic database used in this research.

Open access status is an attribute of interest, as many predatory or low-quality journals are open access, as they capitalize on the use of Article Processing Charges (APCs) to generate profit but do not conduct adequate peer review. Therefore, articles published in these journals are often of lower quality. As noted, whether an article is open access is more complicated than which journal published the article, and systematic reviews necessitate individual article screening, meaning this attribute will be examined at the article level.

As the case study systematic review examined in this research only includes Randomized Controlled Trials (RCTs), many of the articles should have received funding. RCTs require an intervention and control group and can be costlier than observation-based studies. The type of funder is relevant as there are observed differences in the reporting and reception of articles funded by industry versus government (Flacco et al., 2015). Articles reporting sponsorship from industry are more likely to find positive efficacy in their interventions, and physicians are more critical of industry-sponsored research, which can influence their practice (Khamis et al., 2018). If articles did not receive funding, or do not report their funding source, this is still of interest.

Finally, the authors of an article may be of interest when considering article quality. Their affiliated institution type (i.e. educational), the country that their institution is based in, and the larger region are of interest in this research. The institution type may affect the expectations or requirements for authors to publish, for instance, in a university, professors often must publish regularly to be considered for tenure and promotion, which is also known as the publish or perish dynamic (Seeber et al., 2019). Therefore, those based in university settings might be more prone to publish in predatory venues, if they feel a greater level of pressure than their colleagues in government or industry.

Additionally, the country and region authors are based in is also important. Of note are academics based in the Global South, where often these requirements also involve publishing in English, and/or international journals (Grančay et al., 2017). As mentioned, certain regions might be more indicative of an author's tendency to publish in a predatory journal, Macháček & Srholec (2022) found that Asian and North African countries have a high proportion of predatory publishers. However, as previously mentioned, to mimic the article-level review process that occurs during systematic review screening, I will focus on the country where authors are based, rather than the journal or publisher location.

One attribute I am examining at the journal level is CiteScores, which is a measure of journal impact based on data regarding number of publications and citations for journals from Scopus (Elsevier) data. CiteScore is an alternative to the more recognized

Journal Impact Factor (Clarivate) but has a more extensive index of journals and is not behind a paywall (Van Noorden, 2016). While CiteScore does not represent a direct measure of journal quality, it is possible that higher impact journals may publish higher quality research works due to a more selective acceptance rate and rigorous peer review process (Bornmann et al., 2021; Severinsen et al., 2019). This perceived, or true assessment of journal quality being connected to higher impact journals may impact inclusion of individual articles.

Finally, by grouping articles based on their citation relationships, we can estimate the topical focus of these groups, if we assume papers that cite each other, or are cited together share a topical focus (Toupin et al., 2022). In this work three different methods of clustering articles are applied individually, and then in combination with each other. Rather than aiming to describe the different clusters in terms of their content and structure, this research aims to see if any clustering method groups together included or excluded articles, which might indicate a core of articles pertinent to the specific research question. While for all other variables, only the articles eligible for inclusion were included in the analysis, to attempt to analyze research quality, the clustering analysis includes all articles captured in the search. Therefore, this section is more focused on research relevance than integrity.

2.5 Conclusion

The modern academic publishing landscape for health science research has seen large changes over the past 25 years, pertaining to which research is viewed as the strongest evidence, journal attributes, and the sheer amount of work being published. Systematic reviews are widely respected for their rigour, and can aid physicians, and other practitioners in making evidence-based decisions. However, these reviews are not infallible, as issues with review reporting or methodology can result in poor quality reviews. Additionally, as systematic reviews summarize and evaluate evidence, if they include low quality articles, like those from journals without adequate peer review, their conclusions may be compromised. Bibliometric, or text-based indicators may be of use to characterize the types of articles that are included and excluded in a case study peer review, which is explored in this research.

Chapter 3: Methods

3.1 Research objectives

The general objective of this research is to investigate how the characteristics of articles screened in systematic reviews relate to their inclusion or exclusion based on a) concerns around quality and b) lack of relevance to the research question. To reiterate the research objectives included in the introduction, the following research questions will be explored:

1. What is the relationship between articles' inclusion or exclusion in the review and their open access status?
2. What is the relationship between articles' inclusion or exclusion in the review and the presence of funding, and funder types?
3. What is the relationship between articles' inclusion or exclusion in the review and the affiliation type, country and region of the authors?
4. What is the relationship between articles' inclusion or exclusion in the review and the CiteScore of the journal in which they are published?
5. How are articles included and excluded from the review distributed across publication clusters constructed with different citation-relationship types?

To answer these questions, this thesis uses a particularly well documented Cochrane review (Hayden, 2023) as a case study, uses bibliometrics to quantitatively examine the characteristics of the screened articles and model the relationship between the inclusion or exclusion of articles and their characteristics as listed in the above research questions (the rational for the use of these variables was presented in section 2.4.

3.2 Data collection

3.2.1 Cochrane review

Exercise treatments for chronic low back pain: a network meta-analysis (2023) is used here as a case study because of its status as exemplary of Cochrane reviews, and unique for the detailed inclusion/exclusion metadata recorded by the authors. As a Cochrane review, the authors had to adhere to specific methodological and reporting standards, as defined by Cochrane. Therefore, this work may be seen as representative of other Cochrane reviews, making this study a potential model or blueprint for the analysis of other Cochrane reviews. However, it could also be viewed as distinct from other Cochrane reviews, as the authors collected more data than was necessary to meet the reporting requirements, including the funders connected to each article, and measures of impact.

Cochrane reviews should all adhere to the PRISMA 2020 Checklist, which specifies that authors report their search strategy, the number of reviewers, and the ultimate reason for exclusion (if relevant) among other qualities. However, beyond reading each article that reaches the full-text stage, the authors have no responsibility to collect or report on bibliographic attributes of the article that are not relevant to methodology (study type, population), like whether an article is open access and if it received funding. However, the presence of certain attributes of articles may be correlated with problematic studies, which can be defined as having "been subject to research misconduct (plagiarism, fabrication, or falsification of data) or other serious research integrity issues" (Wilkinson et al., 2022). While methodological issues may be picked up on by existing risk of bias tools, a specific assessment of the research team and publication venue is not required in existing reporting requirements (Page et al., 2021).

Additionally, while Cochrane reviews involve a risk of bias assessment that speak to whether any studies included may have conflicts of interest of major sources of biases, but authors are not required to exclude articles based on quality concerns (Liu et al., 2023). However, the authors of the case study review did conduct a round of exclusions based on research integrity concerns after initial screening was completed. These articles represent works that would be relevant to the research question but were excluded based on research quality issues, namely not being registered, were published by a predatory journal, or had inadequate reporting (Hayden et al. 2023). These articles may have been published in a predatory journal, which are more likely than their non-predatory counterparts to be missing trial registrations and have insufficient sample sizes (Hayden, Ellis, Ogilvie, Boulos, et al., 2021). Therefore, the authors of this review's approach to data collection was conducive to bibliometric analysis, due to the additional metadata collected and the exclusions based on quality concerns.

3.2.2 OpenAlex

I used OpenAlex (Priem et al., 2022), an open catalog of scholarly records to obtain data on the open access status of the works, the institution type, country and region affiliated with authorships, as well as the citation relationships between the works. I selected OpenAlex for article metadata, due to its wide indexing (Culbert et al., 2024), and openness. The authors of the case study review provided RIS files for articles excluded in different groups, which represented different iterations and exclusion reasons. These RIS files were uploaded to Zotero, and then exported into Excel files, which were used to query the OpenAlex API. First, DOIs were used to match records in OpenAlex, and when these were unavailable, matches were made based on article titles. This information was compiled and then stored in a database that is outlined in Figure 2, the entity relationship diagram for the database.

3.2.3 Funder type

In the process of conducting the 2023 update to the review (Hayden et al., 2023), the authors collected data for the 456 eligible studies for inclusion on funding reporting. The variables were 1) funding reported (boolean), 2) funding received (boolean), and 3) funding comment (the name of the organization(s) that provided funding). Using the funding comment, I created the type of funder variable by manually searching Google for each funder. The three possible values were University, Government and Industry. These are broad categories and are generalizations. University was assigned to institutes of higher education, even if they were formally called colleges, or another term. Government was used to describe ministries and departments of national and regional governments, most notably including the European Union (EU). Industry was the term used for corporations, foundations and non-profits. While all these types of organizations may have different mandates and missions, they as a group are typically distinct from university/government grants in their application process, and research priorities (Khamis et al., 2018). The disclosure of industry provided grants can also make physicians more sceptical of the author's findings (Khamis et al., 2018). Therefore, while there might be a large variation in the funding process across these organizations, their status as non-Government/University groups give them enough in common to consider together.

3.2.4 CiteScore

Based on the OpenAlex record for each work, the journal that each work was published in was generated. Based on the list of these journals, I manually collected the CiteScore, percentile and rank within the assigned category for each journal. To do this, I manually searched each journal title in the Scopus *Source* section, with the Title field selected. When there was an exact match, I copied down the information as given. When there was not an exact match, but Scopus recommended several results, I would review each suggestion to determine if there was a match. This often occurred in the case that a journal had a short name, like Cell, as any journal with the word cell in it would be captured. To ensure I found the correct journal, I sorted results alphabetically and checked in the spot where the journal would be. Journals with non-English titles would occasionally generate a single result with an English title. In these cases, I would translate the original title to English, and check the journal website for the English title to verify that they were the same venue. For many journals, Scopus recognized the journal, but did not have CiteScores or any other information, and these were recorded as “N/A” and included in my dataset. However, if a journal was not recognized at all by Scopus, and the message that “no sources were found” appeared, I excluded the journal.

3.2.5 Exclusions

The authors of the systematic review also collected exclusion reasons for the articles that they screened for inclusion. Over various iterations of the review, the team used different software to help them screen and keep track of reasons, including DistillerSR and Cochrane's Screen4Me RCT Classifier. The exclusions during the title/abstract and full text stages are based on whether the article fits the inclusion criteria for the systematic review. Of the 20,366 citations captured by the comprehensive search, 19,842 were excluded, leaving 524 articles for inclusion. The exclusion reasons can be grouped into 3 main groups: various topical PICO reasons (n=7823), non-RCTs as determined by Screen4Me (n=6129), and no exclusion reason available (n=5890). The articles without exclusion reasons noted were screened before the team adopted the DistillerSR software, and therefore the reasons were not stored. However, they were all excluded based on topical reasons, not research quality concerns.

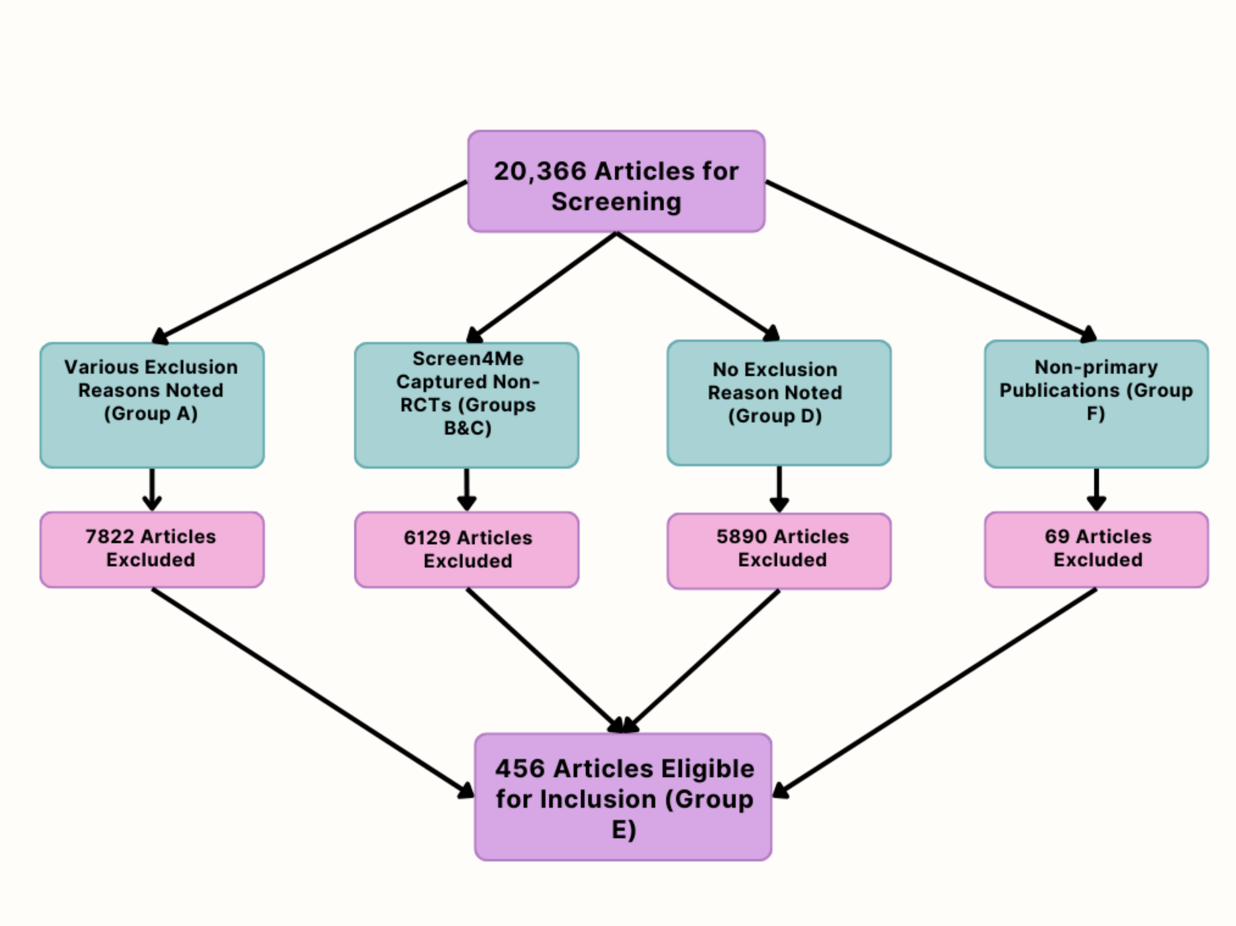

Figure 1: Breakdown of exclusion groups

Of the exclusion reasons noted for the 19,842 articles excluded in this dataset, several exclusion reasons were noted that slightly varied, and for the purposes of this research, I reclassified these to fit into a smaller list of exclusion reasons; non-RCT, study design, incorrect population, missing information, duplicate, and no reason available. Classification of this manner allows for similar attributes to be considered together that were assigned slightly different exclusion reasons that could be attributed to variations in data entry. For instance, both “not adult chronic NSLBP” and “not adults with chronic non-specific LBP” will be considered under “incorrect population”. The exclusions during the title/abstract and full text stages are based on whether the article fits the inclusion criteria for the systematic review, so the exclusion reasons for this subsection of articles has more to do with their study design or population, and in some cases missing information or data.

Reclassified Exclusion Reason	Original Exclusion Reason
Non-RCT	- Not RCT – Screen4Me - Not an RCT
Study Design	- No relevant exercise group or comparison group - No relevant outcome - Not a full text publication
Incorrect Population	- Not adult chronic NSLBP - Not adult population - Not adults with chronic non-specific LBP - Not chronic - Not chronic population - Not non-specific low back pain population
Missing Information	- Abstract not available - No PDF available
Duplicate	- Exact duplicate of previously assessed
No reason available	- Excluded before Distiller (no reason available) - Excluded on title screen

Table 1: Reclassified and original exclusion reasons for the relevance screening process

The remaining 525 articles reported results of 456 trials; 69 were publications linked to primary articles, and were therefore not included in this study, leaving 455 articles eligible for inclusion. The 455 trials were subjected to a research integrity screening in which the authors identified issues in reporting, and data integrity. While all 455 were deemed relevant to the scope of this systematic review, a subsection of articles were excluded based on concerns about research integrity. Therefore, exclusion reasons at this point don’t speak to relevance, but to research integrity/quality. Reclassifying exclusion reasons was particularly necessary for the integrity exclusion reasons, as many articles had multiple concerning attributes, and examining the specific characteristics separately were not in scope of this research.

Reclassified Exclusion Reason	Original Exclusion Reason
Plagiarism	- Article judged to have research integrity concerns: excluded in 2021 review (contains plagiarism, inadequate reporting of CONSORT items); trial not registered
Research Integrity Concerns	- Article judged to have research integrity concerns: excluded in 2021 review (published in a predatory journal, high risk of bias, inadequate reporting of CONSORT items) - Article judged to have research integrity concerns: major data inconsistencies/concerns identified - Article judged to have research integrity concerns: trial not registered, inadequate reporting, data inconsistencies/concerns identified - Article judged to have research integrity concerns: trial not registered, inadequate reporting, major data inconsistencies/concerns identified - Article judged to have research integrity concerns: trial not registered, major data inconsistencies/concerns identified - Article judged to have research integrity concerns: trial not registered, published in a predatory journal, data inconsistencies/concerns identified - Article judged to have research integrity concerns: trial not registered, published in a predatory journal, inadequate reporting - Article judged to have research integrity concerns: trial not registered, published in a predatory journal, inadequate reporting, data inconsistencies/concerns identified - Article judged to have research integrity concerns: trial not registered, published in a predatory journal, inadequate reporting, major data inconsistencies/concerns identified - Article judged to have research integrity concerns: trial not registered, published in a predatory journal, major

	data inconsistencies/concerns identified

Table 2: Reclassified and original exclusion reasons for the research integrity screening process

3.3.6 Entity relationship diagram

The collected data was stored in a relational Postgres database with tables to represent works, authorships, exclusions, sources, funders, and clusters, with associative tables linking these tables together with shared values.

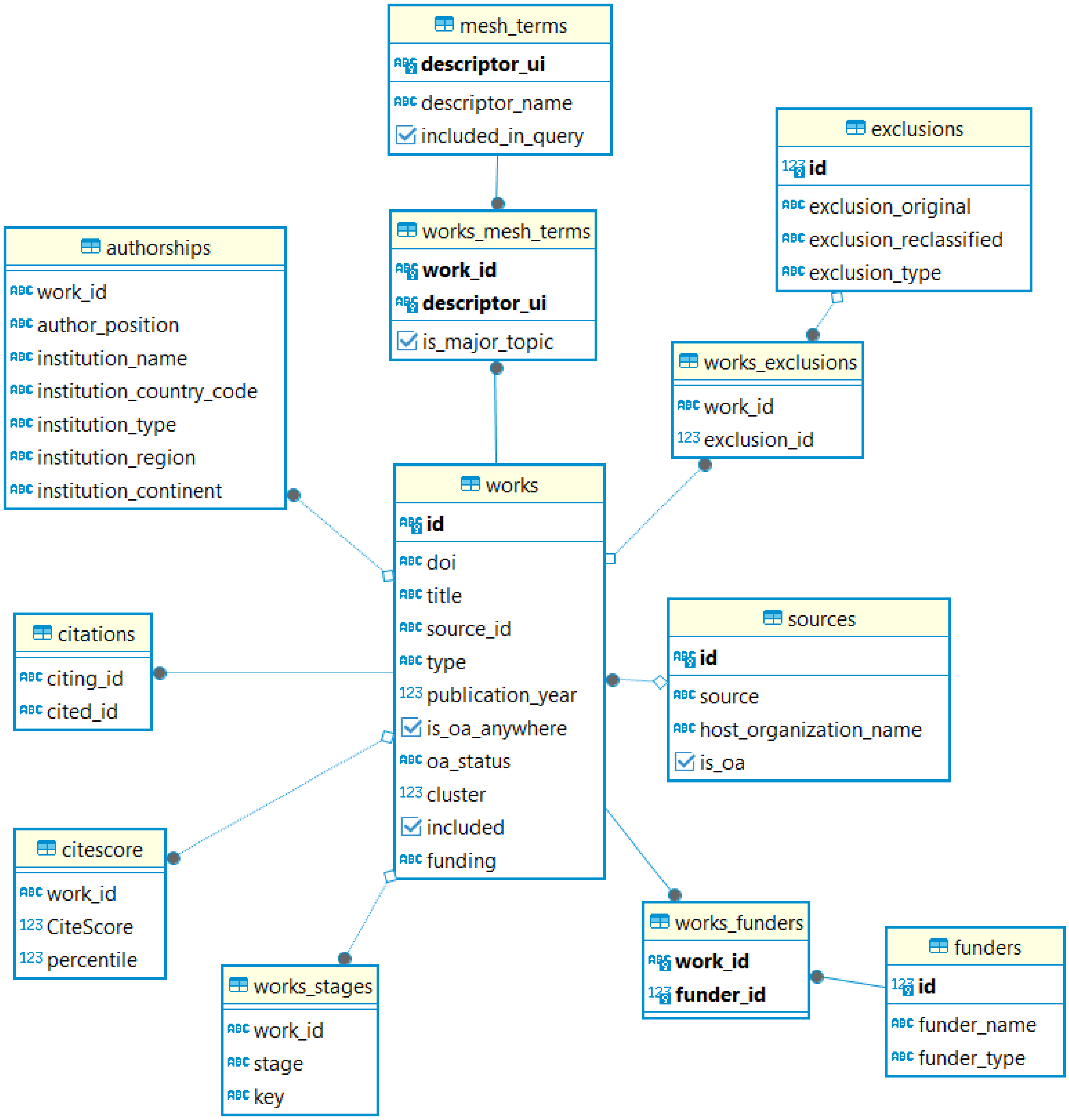

Figure 2: Entity Relationship Diagram describing the database in which values were stored for this research

As this analysis is focused on primarily text-based factors, rather than journal level, many of the tables are connected to the works table. For instance, rather than collecting information about authors (names, number of works, list of works), I focused on authorships which in this case, this means the specific relationship between an author and a work. I gathered the author_position (first, last or middle), which can speak to the authors involvement in the project. I also collected their institution/country, as I wanted to consider regions based on where the authors were based, rather than journals. Open access status was also strategically collected to focus on text rather than journal. While publishing in a gold/diamond open access journal does make an individual work OA, it is also possible for works published in journals that are closed or partially open to be OA, if authors opt to self archive or select a hybrid option respectively. Therefore, in addition to collecting information on whether a journal is open access, I also collected information on whether the article was available anywhere in an open access format, and the article's OA status.

OA status	Definition
Gold	Published in an OA journal indexed by the DOAJ
Green	Toll-access on the publisher landing page, but there is a free copy in an OA repository
Hybrid	Free under an open licence in a toll-access journal
Bronze	Free to read on the publisher landing page, but without any identifiable licence
Closed	All other articles

Table 3: Open access status definitions in the Open Alex API (Priem et al., 2022)

Clusters and funding sources are more directly tied specifically to articles, rather than journals or publishers. Clusters are created based on citation relationships between articles. Theoretically, if one article cites another, or if two articles are both cited by the same article, there would be topical overlap. Funding is typically tied to an author or project, and can be used for multiple articles, but I was interested in exploring how the type of funder, as reported within the article, might be related to the inclusion of said article. As funding should be disclosed explicitly within the reporting of each individual output for which it has been used, it is easily connectable to specific articles (if adequately disclosed). CiteScores and other measures of journal impact are more directly tied to journals, but have still been included here, as articles from high impact journals may have higher quality articles. Or even if this is not true, it might subconsciously be perceived by reviewers to be true.

3.3 Data Analysis

For the variables of open access, funders, authorships and CiteScore, only articles in groups E&F were used. These represent articles that are relevant for inclusion (group F are tied publications but are building upon a trial represented in group E). If articles were excluded from that group, it represented a major research quality concern. Examining only this group of articles was to ensure that exclusions were based on quality issues, as I was aiming to test these factors independently and together as proxies for quality measures.

For the clustering aspect of the analysis, the entire set of articles captured by the search (20,366) to see if citation relationships would predict relevance. Most articles excluded overall in the screening process, particularly in the title/abstract, full-text screening stages, were excluded based on relevance, not quality. Therefore, in this section of the analysis, I am aiming to see if the relevance to the research study can be predicted.

3.3.1 Frequency measures

To describe categorical variables, including open access (open/closed and status), funders (yes/no and funder type) and authorship characteristics (institution type, countries and region), I used frequency to describe the makeup of all articles, and then separately for included and excluded articles. To do so, I grouped articles by the variable of interest and then counted each instance of each variable. So, for open access status, each instance of green, gold, hybrid and bronze was counted, and then I formatted this into a table or figure, displaying the categories from highest to lowest frequency. Percentages were also calculated from these frequency measures, by dividing the number of instances within a group by the total number of articles, multiplied by 100. These measures were calculated for each aspect of open access, funders and authorship characteristics for all articles eligible for inclusion, and then by inclusion status to describe similarities and differences between the included and excluded groups, in addition to the larger group of articles.

When examining Open Access, I considered both the overall availability of an open access version online (Boolean), and the open access status (text) of articles. For both, I calculated their frequency at the level of all articles, just included and just excluded articles. In doing so, I calculated the expected proportion (based on the eligibile group) of articles that would fall into each category in the included and excluded group.

For the analysis on funder types, there were 346 instances of funding included in the analysis. When an article had multiple funders, each funder was counted once, so each funder type could be included. For an article with three funders, one from industry, one from government and one from an educational organization, there would be three entries in the table below. As displayed in table 7, 304 of the instances of funding were for included articles, and 42 were excluded. Instances where a funder type could not be established or there was no funding were also included in this analysis, as a lack of funding or issues with reporting is of interest to RQ2.

The variables of interest for authorships were the author's institution types, their country and region. These attributes were examined for each individual authorship, rather than for unique authors, so if the same author contributed to 5 works, they would be included 5 times. This way, they could be tied to each paper they wrote as a unique entity. There were 2008 authorships tied to the works in the eligible dataset. 1506 authorships were tied to included works, and 502 authorships were tied to excluded works. Unlike the open access analysis, many of these variables had instances N/A values. I made the decision to keep these values in the analysis, as 'unknown' as a lack of information in OpenAlex could speak to a lack of indexing in journals. N/A also made up for a significant portion of the countries and region counts, so I didn't want to misrepresent other values as making up a higher percentage of each value, which would happen if N/A values were removed.

CiteScores were collected at the journal level, and then were connected to works published in the relevant journal. There were 247 individual works included in this analysis in groups E and F, 180 of which were included and 67 which were excluded. However, of the 247 works, 59 were missing CiteScores, and were removed from the analysis, leaving

188. These works were excluded as there were instances of journals having a 0 value CiteScore, and I did not want to conflate missing, and 0 values. In the included works, 35 CiteScores were missing, while in the excluded works, 24 CiteScores were missing, leaving 145 and 43 works for analysis respectively.

3.3.2 Clustering

Citation relationships between documents are often used in bibliometrics to create article networks, to which community detection or clustering algorithms can be applied to identify groups (i.e. clusters) of articles that are more closely related to each other than other articles in the network. There are three approaches used in bibliometrics to link articles together based using citations:

1. **Direct citations (DC)**: article A cites article B. Therefore, A and B are linked to each other, and the weight of that link is 1.
2. **Bibliographic coupling (BC)**: article A and B have X references in common. Therefore, they are linked to each other, and the weight of that link is X.
3. **Co-citation (CC)**: Z documents cite both article A and article B. Therefore, they are linked to each other, and the weight of that link is Z.

In total, seven distinct networks were networks created based on these three methods used individually or jointly and each network is labelled using the methods used to create it (e.g., BC, CC, BC-CC, CC-DC, BC-CC-DC, etc.). Then each of the nodes (articles) is put into a cluster (a group of nodes) using the Louvain community detection algorithm (Blondel et al., 2008).

3.3.3 Predictive analysis

Open access status, funder types, authorship regions and institution types, and CiteScore were included in the model. Unlike in the descriptive analysis, each work was counted once, meaning only first authorship was considered. Rather than having a unique entry for each funder type, I created new variables for education, government, industry and unknown, where a 1 indicated that that funding type was present, and a 0 that it was not. So, a single article would only have one entry, and multiple funding types would be indicated with a 1 in multiple columns for funding types. Using these variables, a logistic regression model was developed, in which the dependent variable was study inclusion. The coefficients, standard error, z value, p value, and confidence intervals were calculated.

3.4 Conclusion

This research is comprised of two main components, a descriptive and predictive analysis. For each part of the analysis, other than the clustering, only articles deemed eligible for inclusion were used. This was to ensure that all articles were relevant, so exclusion reflected quality issues. However, there is no simple measure of quality. Therefore, this analysis is focused on using variables to determine if there is a proxy that can predict inclusion, whether alone or in combination with other variables.

Chapter 4: Results

4.1 Introduction

In this section, the results of the previously described methods are outlined in two primary sections, descriptive and predictive analyses. In the descriptive section, I describe the makeup and attributes of the dataset in terms of the 5 key areas of interest: open access, funders, authorships, CiteScore and clusters. The articles eligible for inclusion are described, in addition to the included and excluded groups separately, to illustrate major differences between the groups. In the predictive analysis, the results of a logistic regression model including the 5 areas of interest are described, with a focus on statistically significant predictors of inclusion/exclusion, the predictions of the model, and its success rate.

4.2 Descriptive analysis

4.2.1 Open access

In the eligible group, open access information was available for 360 articles of the 525 total articles in group E and F. The distribution of these articles is displayed in Table 4, and 261 of those articles were included, while 99 articles were excluded. Open articles (all OA options together) were slightly more prevalent than closed articles, at 51.67% closed compared to 48.33% open. A similar percentage breakdown was observed in the included articles, but a slightly larger portion of closed articles were included compared to closed articles. However, in the excluded group, open articles were notably more prevalent at

59.60%. Therefore, open articles are slightly over-represented in the excluded group, as compared to the eligible articles group.

Open access	**Eligible**		**Included**		**Excluded**	
	N	**%**	**N**	**%**	**N**	**%**
Open	186	51.67	127	48.66	59	59.60
Closed	174	48.33	134	51.34	49	40.40
Total	360	100.00	261	100.00	108	100.00

Table 4: Frequency and percentage of open and closed articles in each article group

4.2.1.1 Open access status

In table 5, a more granular way of describing open access is displayed, broken down by their status or type, namely, gold, bronze, green, hybrid and closed. While each open access type indicates that an article is accessible to all somewhere on the internet, these types vary in their point of access (journal homepage versus institutional repository), and oftentimes findability, and therefore visibility (Charles University, 2023). In the eligible group, when broken down by status, besides closed, gold was the most frequently observed type, followed by bronze, green and then hybrid. This order was consistent in the included group, with a higher percentage of closed, green, and bronze and a lower percentage of gold and hybrid. In the excluded group, a smaller proportion of articles were closed (as noted above), bronze and green, while a larger proportion of articles were gold, and hybrid. In the excluded group, gold was the most frequently observed article type, followed by closed, hybrid, green then bronze, a different descending order from the eligible and included groups.

Open access	Frequency	Eligible %	Frequency	Included %	Frequency	Excluded %
Closed	174	48.33	134	51.34	40	40.40
Gold	110	30.56	63	24.14	47	47.48
Green	34	9.44	29	11.11	5	5.05
Bronze	23	6.39	22	8.43	1	1.01
Hybrid	19	5.28	13	4.98	6	6.06
Total	360	100	261	100	99	100

Table 5: Frequency and percentage of open access types in each article group

The percentage values of each type of open access status for the eligible group represents the expected values for the included and excluded groups in a case where OA status does not impact the inclusion status. Therefore, by dividing each included eligible percentage by itself, we get 1, which represents an exact match between the expected and actual values. Then, by dividing each percentage for the included and excluded groups by the eligible percentage, a value helpful for comparison between variables is calculated. A number over 1 indicates an overrepresentation, and a value under 1 indicates an underrepresentation.

As indicated by Table 6, bronze is the most overrepresented status in the included group, and gold is the most underrepresented. Closed and hybrid were the closest to the expected value of 1. Among the excluded group, the values were further from the expected values, most notably an overrepresentation of gold and underrepresentation of bronze and green. Closed articles were also underrepresented, and hybrid articles were slightly overrepresented.

Open access	Expected Value	Actual Included Value	Actual Excluded Value
Closed	1.00	1.06	0.84
Gold	1.00	0.79	1.55
Green	1.00	1.18	0.53
Bronze	1.00	1.32	0.16
Hybrid	1.00	0.94	1.15

Table 6: Expected values for each open access type proportion

4.2.2 Funders

In the eligible and included groups, the 3 most common funders were government, unknown, and no funding. In the eligible group, the proportion of government and unknown groups were very similar, within a percentage point of one another, which represented 2 funders. However, this difference was starker in the included group, in which 30.26% of article funders were government agencies, and only 22.70% were unknown. The proportion of no funding was similar across these groups, in addition to university (more prevalent in eligible group) and industry (more prevalent in included group). In the excluded group, a far smaller percentage of funders were from government agencies, at 9.52%, and a larger percentage of articles received funding from funder types with no discernable type. Notably, there were no excluded articles which were funded by an industry group.

Funder Type	Eligible		Included		Excluded	
	Frequency	%	Frequency	%	Frequency	%
Government	96	27.75	92	30.26	4	9.52
Unknown	94	27.17	69	22.70	25	59.52
No funding	60	17.34	54	17.76	6	14.29
University	51	14.74	44	14.47	7	16.67
Industry	45	13.00	45	14.80	0	0
Total	346	100	304	100	42	100

Table 7: Frequency and percentage of funder types

When comparing actual to expected values in table 8, the included group is not far off, with the most notable discrepancies being an underrepresentation of unknown funders, and an overrepresentation of industry. Differences between the expected and actual excluded values are more significant. Government funders, no funding, and industry are underrepresented, while unknown and university funders are overrepresented.

Funder Type	Expected Value	Actual Included Value	Actual Excluded Value
Government	1.00	1.09	0.34
Unknown	1.00	0.84	2.19
No funding	1.00	1.02	0.82
University	1.00	0.98	1.13
Industry	1.00	1.14	0

Table 8: Expected values for each funder type proportion

4.2.3 Authorships

4.2.3.1 Authorship institutions

As shown in table 9, the authorships of works in the eligible dataset were most frequently tied to educational institutions. A significant portion of the authorships were also linked to unknown and healthcare institutions. A similar proportion of authorship institution types were observed in the eligible and included groups. However, in the excluded group, a higher proportion of authorships were tied to educational institutions, and a lower proportion were tied to healthcare institutions.

Institution		Eligible		Included		Excluded
Type	Frequency	%	Frequency	%	Frequency	%
Education	1,251	62.30	907	60.22	344	68.52
Unknown	379	18.87	284	18.85	95	18.92
Healthcare	264	13.15	217	14.41	47	9.36
Facility	42	2.09	35	2.32	7	1.39
Non-profit	31	1.54	29	1.93	2	0.40

Company	18	0.90	17	1.13	1	0.20
Government	13	0.65	9	0.60	4	0.80
Other	10	0.50	8	0.53	2	0.40
Total	2,008	100	1,506	100	502	100

Table 9: Frequency and percentage of author institution types

In the excluded group, authorships tied to educational institutions were overrepresented, as shown in table 10. Authorships tied to non-profits and companies were significantly underrepresented in the excluded group as well as healthcare and facilities to a lesser extent. These groups, particularly non-profits and companies were therefore overrepresented in the included group. Unknown author institutions were consistently represented across all three groups.

Institution Type	**Expected Value**	**Actual Included Value**	**Actual Excluded Value**
Education	1.00	0.97	1.02
Unknown	1.00	1.00	1.00
Healthcare	1.00	1.10	0.72
Facility	1.00	1.11	0.67
Non-profit	1.00	1.25	0.26
Company	1.00	1.26	0.22
Government	1.00	0.92	1.23
Other	1.00	1.06	0.26

Table 10: Expected values for Author institution type proportion

4.2.3.2 Authorship countries

Due to the granularity of the data when examining the individual countries, I chose to focus on the 10 most frequent authorship countries and compare these visually, rather than breaking down the proportion of each. All authorships are included in this analysis (i.e., each individual authorship is counted as a separate entity), but only articles from groups E and F (n = 525) are included. As displayed in figure 3, the most frequent value for institution country was unknown, at 18.87% in the eligible articles group. However, with NA values removed, the top 10 countries were Iran, the USA, Brazil, South Korea, Australia, the United Kingdom, Spain, Canada, Germany and the Netherlands.

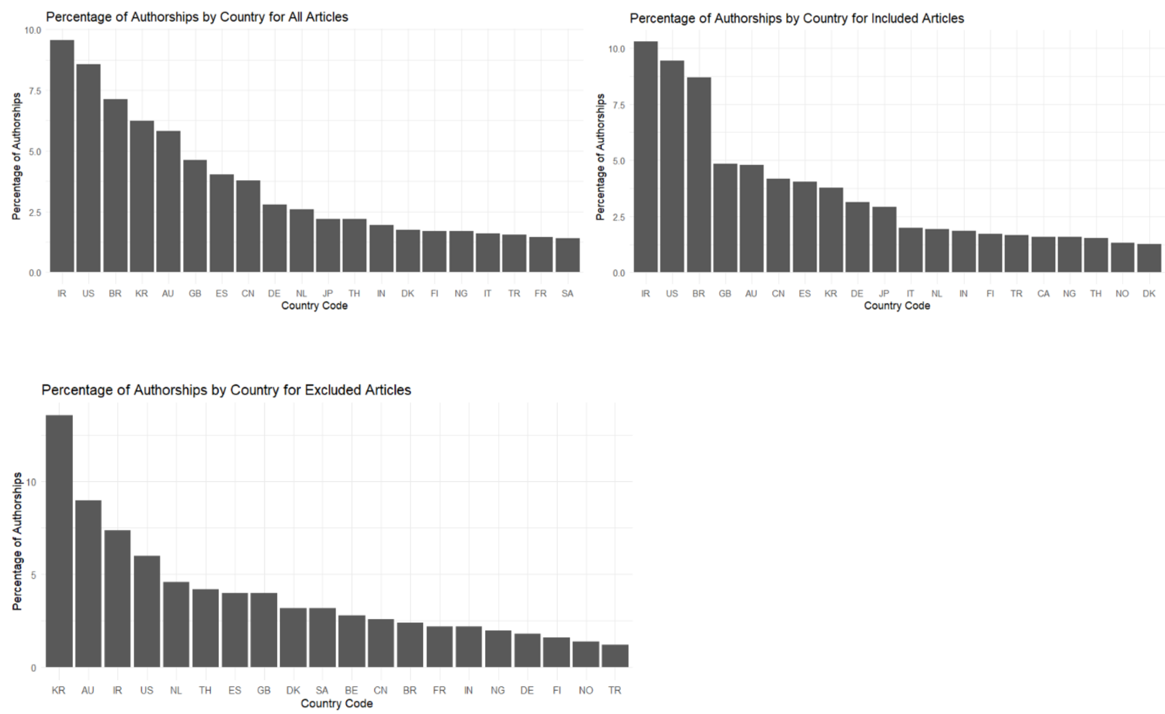

Figure 3: The top 20 authorship countries for each inclusion status group

When only considering included articles, the United Kingdom and Canada move up the rankings to claim the 4th, and 6th place rankings, compared to 6th and 8th in the eligible group. South Korea moves down to the 8th most frequent spot, and the Netherlands drops out of the top 10 countries, and Japan enters the top 10. There isn't a large change in the percentage makeup though, as unknown articles make up 18.86%, and only Iran passes the 10% mark among all the countries.

In the excluded group, NA is still the most frequent value. South Korea has the highest percentage of all exclusions for known countries at 13.55%, followed by Australia, Iran and the USA. Iran and the USA shifted down the list, from the 1st and 2nd most frequent, to the 3rd and 4th spots. Thailand, Denmark and South Africa entered the top 10 countries.

4.2.3.3 Authorship regions

These country affiliations are grouped into regions by OpenAlex. In the eligible article group, 35.96% of all articles came from Western Europe and North America, and 35.41% were from unknown regions. The smallest percentage of articles came from Arab, Africa and Eastern Europe, which also holds true for included articles (although in different proportions, and ordered Africa, Arab and Eastern Europe in terms of decreasing proportion). Unknown articles made up a higher percentage in the excluded group, and Latin America and the Caribbean made up a smaller percentage.

Region		**Eligible**		**Included**		**Excluded**
	Frequency	**%**	**Frequency**	**%**	**Frequency**	**%**
Western Europe and North America	722	35.96	552	36.65	170	33.86
Unknown	711	35.41	509	33.80	202	40.24
Asia and the Pacific	336	16.73	246	16.33	90	17.93
Latin America and the Caribbean	145	7.22	133	8.83	12	2.39
Arab	43	2.14	25	1.66	18	3.59
Africa	38	1.89	28	1.86	10	1.99
Eastern Europe	13	0.65	13	0.86	0	0

Table 11: Frequency and percentage of authorship regions

Unknown regions, Asia and the Pacific, and Arab and Africa were underrepresented among included articles, with Arab being the most significantly underrepresented of these regions, with a value of 0.78. Western Europe, Latin America and the Pacific and Eastern Europe regions were all overrepresented with positive values. Eastern Europe was the most significantly overrepresented, with a value of 1.32 and no authors related to excluded articles. However, a very small portion of articles came from authors in Eastern Europe, Arab and Africa overall, and a very small number were linked with included articles.

Region	**Expected Value**	**Actual Included Value**	**Actual Excluded Value**
Western Europe and North America	1.00	1.02	0.94
Unknown	1.00	0.95	1.14
Asia and the Pacific	1.00	0.98	1.07
Latin America and the Caribbean	1.00	1.22	0.33
Arab	1.00	0.78	1.68
Africa	1.00	0.98	1.05
Eastern Europe	1.00	1.32	0

Table 12: Expected values for authorship region proportion

In figure 4, each value for institution region and type are displayed, with the pink bar representing the proportion of excluded articles, and blue indicating included. While for each variable, most articles were included (because only a small proportion of eligible articles were excluded based on research integrity issues), the pink bars highlight areas where a higher proportion of articles were excluded.

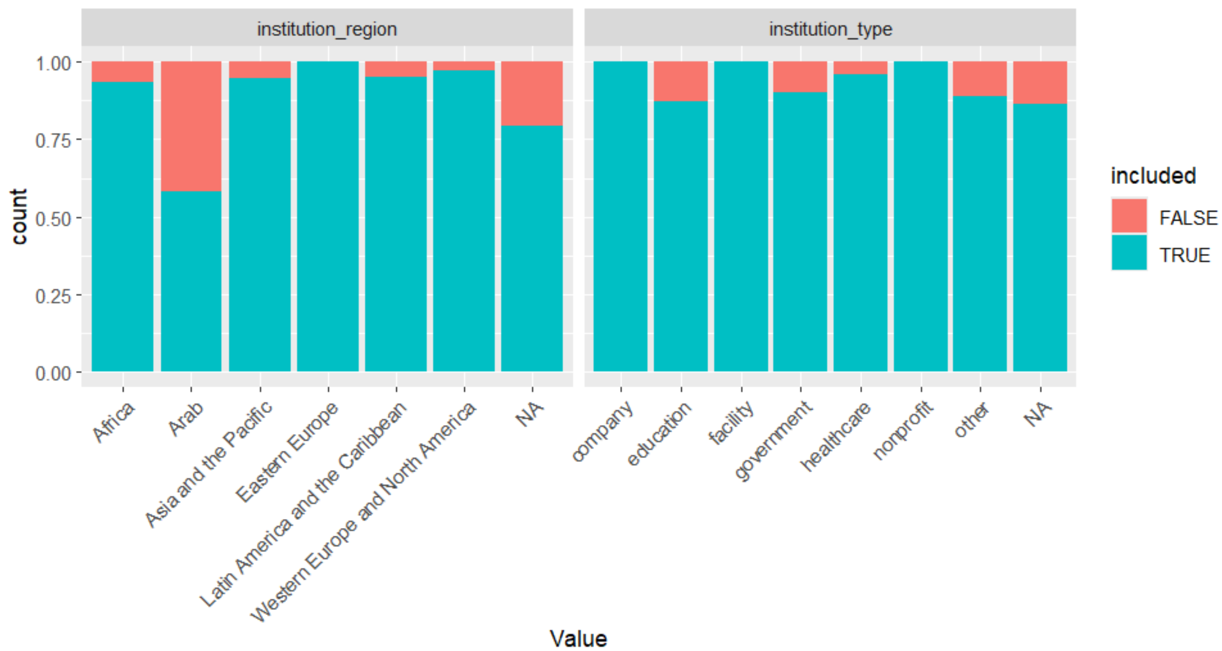

Figure 4: Inclusion compared to exclusion for each institution region and type

4.2.4 CiteScore

In table 13, the summary statistics for each group are displayed. Excluded papers had a lower mean CiteScore, at 4.85, compared with the eligible articles mean (6.22) and the included mean (6.63). However, the standard deviation and variance in these groups were far larger than the excluded group, indicating a wider spread from the mean in eligible and included papers. The median was 4 and q3 was 5.7 in each group. The maximum value was 134.4 in the eligible and included groups, but only 15.3 in the excluded group. This indicates a small number of very high CiteScores in the included/eligible groups that skew the average up to be higher than the excluded, which are all in a smaller range of possible values.

Group	n	mean	sd	var	q1	median	q3	min	max
Eligible	188	6.22	13.84	191.67	2.5	4	5.7	0.1	134.4
Included	145	6.63	15.66	245.11	2.5	4	5.7	0.2	134.4
Excluded	43	4.85	3.24	10.52	2.7	4	5.7	0.1	15.3

Table 13: Descriptive statistics summary for CiteScores in each inclusion status group

4.2.5 Clustering

In Figure 5, the 7 networks, based on different types of citation relationships are displayed separated into facets, with the number of clusters displayed on the x-axis, and the number of papers per cluster on the y axis. Figure 7 represents all papers captured by the search that were available in OpenAlex. Clusters based on direct citation alone were separated into the highest number of clusters, with smaller groups of papers making up each cluster, while bibliographic coupling and co-citation methods produced smaller number of clusters, with a larger number of papers per cluster. Combinations of methodologies were also used to generate clusters, and the combination of all 3 methodologies generated a smaller number of clusters than direct citation or co-citation alone.

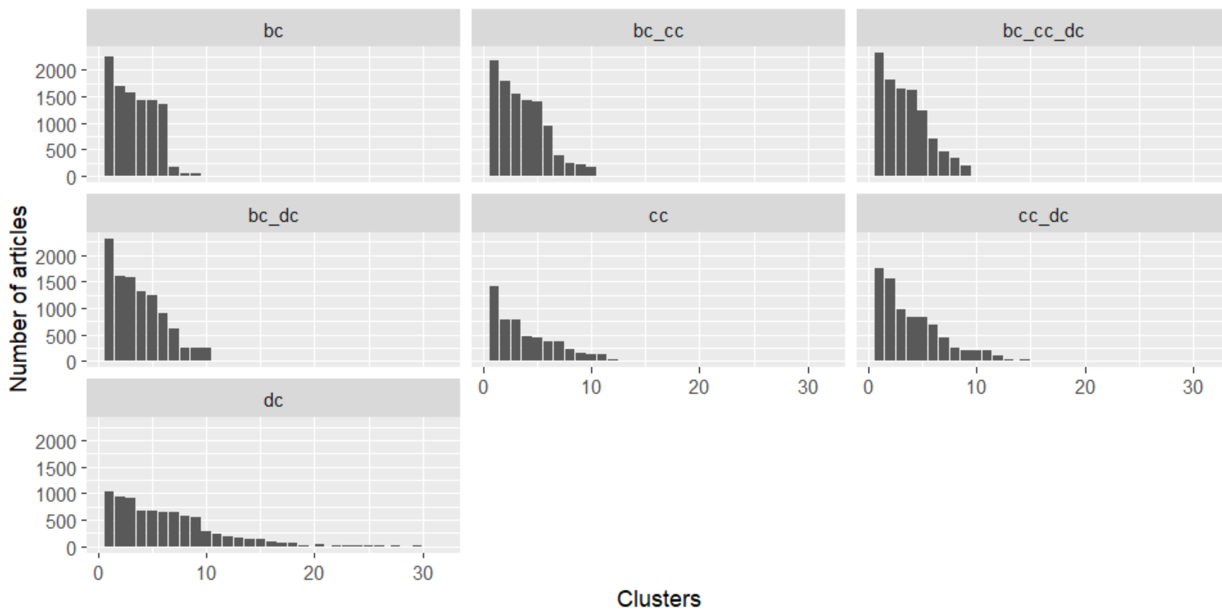

Figure 5: Distribution of articles across clusters for each citation-based network.

When only highlighting included papers, as shown in figure 6, it becomes easier to see if clustering methods group together included papers. Bibliographic coupling does not appear to be successful at grouping included papers together into a single group, as there are 2-3 groups that have a close to equal number of papers in each methodological group that includes bibliographic coupling. However, direct citations and co-citations do appear to be more successful at grouping included papers together. In both the direct citation and co-citations groups where the methodologies were used alone, there is a single group that contains most included papers, with a few outliers dispersed across clusters. Direct citation and co-citation also work in conjunction with each other, both methodologies combined generate clusters that capture most included papers in a single cluster.

Figure 6: Distribution of articles across clusters for each citation-based network

4.3 Predictive Analysis

4.3.1 Logistic regression model

While open access, funder types, authorships and CiteScore were all examined in this analysis, for authorships, only region and institution type were included in this model, as the countries were too granular, with too large a spread of values to provide meaningful predictions on inclusion. Only the CiteScore of the journal that a work was published in was found to be a statistically significant predictor of inclusion ($p < 0.05$). Despite this significance, the effect appears to be small, based on the coefficient, making papers with higher CiteScores only slightly more likely to be included. The coefficients that imply a larger effect include Green OA, authorships affiliated with institutions based in Latin America and the Caribbean and Eastern Europe, and Industry funders. However, none of these effects are statistically significant.

Variable		Coeff	Std.Err.	Z value	Pr	Odds Ratio	2.5%	97.5%
Intercept	---	1.540	2.069	0.744	0.457	4.666591e+00	0.107	4.909205e+02
Open access status	Closed	-0.770	1.610	-0.478	0.632	4.629008e-01	0.010	7190626e+00
	Gold	-1.949	1.597	-1.220	0.222	1.424622e-01	0.003	2.146350e+00
	Green	15.815	2937.565	0.008	0.994	7.386635e+06	0.000	2.394414e+291
	Hybrid	-0.680	2.007	-0.339	0.735	5.063771e-01	0.006	2.784868e+01
Institution Region	Arab	-0.116	1.917	-0.582	0.560	3.275275e-01	0.005	1.309411e+01
	Asia	2.397	1.712	1.400	0.161	1.098732e+01	0.298	4.547017e+02
	Eastern Europe	17.953	6176.925	0.003	0.998	6.261433e+00	0.000	Inf
	Latin America	19.215	2683.094	0.007	0.994	2.213867e+08	0.000	NA
	Unknown	0.537	1.361	0.395	0.693	1.711280e+00	0.069	2.480252e+01
	Western Europe and North America	0.964	1.486	0.649	0.517	2.621433e+00	0.091	4.978990e+01
Funder Type	Education	0.130	0.839	0.154	0.877	1.138401e+00	0.230	6.744898e+00
	Government	-0.221	0.976	-0.226	0.821	8.019487e-01	0.128	6.856373e+00
	Industry	16.183	1968.528	0.008	0.993	1.067442e+07	0.000	1.546935e+274
	Unknown	-0.103	0.591	-1.866	0.062	3.320300e-01	0.099	1.027141e+00
CiteScore		0.469	0.163	2.871	**0.004**	1.598274e+00	1.192	2.269278e+00

Table 14: Coefficients of the logistic regression model

When the model predicts inclusion (based on the probability of inclusion from the model), it correctly predicts 160 inclusions, incorrectly predicts 16. For exclusions, it correctly predicts 13, and incorrectly predicts 3. The model has an overall accuracy rate of 90.1%.

Actual data	**Model prediction**	
	Excluded	**Included**
Excluded	13	16
Included	3	160

Table 15: Model Predictions

4.4 Conclusion

In this chapter, OA status, funder types, authorships and CiteScores were described for the set of eligible articles and broken down by included and excluded articles. Frequency and percentage of these attributes were determined and contrasted between each inclusion status group. Clusters were formed through direct citations, bibliographic coupling and co-citations and were examined for all articles captured by the search. By filtering for included articles, methodologies were evaluated for their success in grouping them. Finally, using the attributes from the descriptive analysis section, a logistic regression model that predicted inclusion of articles among the eligible articles was generated.

Chapter 5: Discussion

5.1 Introduction

This section details the interpretation of the results presented in chapter 4 in terms of descriptive and predictive analysis. The characteristics of the overall group of eligible articles, included and excluded that were outlined in results chapter of this research are compared to the literature on attributes of articles that may speak to lower article quality. The implications of the predictive model are explored, as well as limitations of this research, and future research opportunities and questions.

5.2 Descriptive analysis

5.2.1 Open access

RQ1: What is the relationship between articles' inclusion or exclusion in the review and their open access status?

According to Piwowar et al (2018) approximately 25% of scholarly literature is open access. However, this was calculated based on years up until 2015, a year in which OA papers made up 45% of all articles, and the authors noted that newer works were more likely to be OA. Therefore, while the percentage of OA articles making up more than half of the eligible articles may initially seem high, it is in line with increasing OA trend being observed (Piwowar et al., 2018). This trend may be driven by funding requirements, especially in countries like Canada where the Tri-Agency policy requires funding recipients to make their works OA (Government of Canada, 2023).

Piwowar et al., (2018) observed that bronze was the most common type of OA article, and that gold and hybrid had been increasingly contributing to the growth of OA. However, in this case study, bronze articles were the second to least common type of OA in the eligible and included groups, and the least common type in the excluded group. Notably, bronze OA is a fairly new term, coined in the 2018 Piwowar et al. article, and it refers to cases when the article appears on the journal website, and is free to read, but there is no accompanying license that indicates whether reuse or redistribution is allowed, unlike in gold/hybrid journals. Therefore, without close examination for license information, a bronze journal may be accidently mischaracterized as gold or hybrid. Alternatively, (Jeong & Huh, 2016) found that articles published in journals that used Creative Commons licenses were more likely to have identifiers (ORCID, DOI) than their free access (with no license) counterparts. Therefore, it is also possible that a greater portion of the articles generated by the search were bronze, but these articles are not included in OpenAlex and as a result, this analysis, due to a lack of identifiers.

Another dynamic which may contribute to the proportions of OA types is predatory publishing. In hybrid/gold/bronze OA journals, authors pay a fee to make their work open. Predatory publishers accept articles regardless of quality and pocket this money without conducting the editing and peer review process typical of academic journals. Therefore, in this model, articles that are green OA are the least likely to be predatory, as the APC model is not a part of this infrastructure, as authors can self-archive for free. This is not to say that the presence of any APC is indicative of a predatory publisher, only that these publishers wouldn't necessarily gain from a closed model which allows authors to self-archive. The authors of this study used publication in a presumed predatory journal as one criterion that could lead to exclusion in a trial, which may have contributed to the overrepresentation of gold and hybrid, and the underrepresentation of green articles in the excluded group.

5.2.2 Funders

RQ2: What is the relationship between articles' inclusion or exclusion in the review and the presence of funding, and funder types?

In this analysis, funders were grouped into government, educational and industry, where no information was provided or no information could be found about a stated funder, an unknown type was assigned, and when authors explicitly stated that they did not receive funding, no funding was assigned. Therefore, the unknown category represents both issues with funding reporting and funding from entities that have little to no current web presence/searchability. However, having no funding at all made up a smaller proportion of articles, and was underrepresented in the excluded group. Therefore, one can conclude that issues with reporting funding might be more indicative of an article that will be excluded in this review, than an article not having any funding at all.

Based on the issues outlined in the literature on research funding from industry, including being more likely to report positive findings, increased physician weariness (Khamis et al., 2018), and a focus on developing more tangible products or services rather than knowledge production (Thelwall et al., 2023), the lack of exclusion of any industry sponsored research was surprising. However, this could also be based on the decision to group together corporations, foundations and non-profits, as the quality issues and bias against industry funded research may not extend to non-profits and foundations. Similarly, the overrepresentation of educational funders in the excluded group was not anticipated, as most universities have policies on research integrity and plagiarism. This may be because of a lack of education regarding journal quality or a lack of enforcement on these policies. However, several of the articles that were in the excluded group were funded by the same educational institution, suggesting that a single bad actor may be skewing the results. Additionally, it is possible that authors are reporting their employer (a university) as a funder as they would have paid researchers a salary, when in reality this would represent no specific trial funding. These results indicate that funding types should not be defined with attributes linked to research quality across disciplines and countries, i.e.: not all research funded by industry has a positive skew and not all university funded work is high quality. Rather, transparency in reporting seems to be the more important indicator.

5.2.3 Authorships

RQ3: What is the relationship between articles' inclusion or exclusion in the review and the affiliation type, country and region of the authors?

5.2.3.1 Institution types

Regarding authorship institutions, the spread of authorships from educational institutions between the included and excluded group is the most notable feature of the data. Educational institutions made up most authorship affiliations, while organizations like companies, government and other each had less than 20 authorships total. A higher research output coming from educational institutions is in line with the publish or perish dynamic in academia, and expectations for professors or scholars to be conducting research. In a field focused on a human health research question, one might expect to see a higher number of authors affiliated with healthcare institutes. However, it is possible that some of the authors attributed here as having educational affiliations also have relationships with healthcare providers. OpenAlex authorship research institution data comes from authors self-reporting on publications. Therefore, if they have only recorded their academic affiliation and not their healthcare affiliation, that information is missing. This affiliation may also be reported more frequently, as universities may provide more research time and resources than hospitals or other health institutions. This case might be especially likely in cases where authors can claim APC discounts based on their affiliation with a university that holds a transformative agreement with a publisher.

Non-profits and companies were overrepresented in the included group, and underrepresented in the excluded group, which was surprising, because of the factors addressed in the funders section. One might expect articles coming from these sources to be more likely to be excluded based on patterns in publication and reporting biases, but they were not. Again, this reinforces the notion that quality can’t be tied to only educational or healthcare institutes. However, the values for the companies, government and other groups are so small as to make comparison difficult. As in this analysis, each authorship was considered, there are a variable number of authorships per paper. So, twenty authors may represent five papers, or two. Therefore, it is not necessarily helpful to compare differences of one authorship.

5.2.3.2 Countries

There were a large portion of unknown author institution countries (~18% across each inclusion group), and due to the large number of potential values (each country being examined separately), comparing the relative proportions for each value was not feasible. Therefore, rather than calculating the relative value for each country, the top 20 countries in each group were displayed in a graph. Of note were a high proportion of authorships connected to institutions in Iran and the USA. While the USA is unsurprising, due to the large population and number of research/educational institutions, Iran is a much smaller country. It is possible that this field is an area of particular interest to institutions within that country, or that there is government or industry funding allocated to support low-back pain research, as these can contribute to public health costs. Also notable was the high proportion of excluded articles that came from authorships connected with South Korea, a total of 68 authorships, affiliated with 26 universities and one hospital. Since this problem of research integrity does not appear to be localized to a particular university, further examination of the specific articles and the associated issues is warranted.

5.2.3.3 Regions

Western Europe and North America was the most frequent region to which authorships were connected, followed by unknown. In this case, unknown values represented a lack of information in OpenAlex, suggesting spotty metadata, or no affiliation listed for authors in the publishing journal. Unknown regions were overrepresented in the excluded group, again suggesting that there may be a connection between reporting issues and exclusion. Latin America and the Caribbean was overrepresented in the included group, and underrepresented in the excluded group, making up 7.22% of overall eligible articles authorships. Latin America has long been noted for the unique OA culture cultivated in the region, with a mostly diamond (free for authors to publish and for the public to read), setup (Berger, 2021). Without the APC model, predatory publishing is of low concern in Latin America (Jones, 2015). The region has long been hailed as a leader, not just for the Global South, but for the rest of the academic community, achieving platinum journals at a fraction of the cost seen elsewhere (Morrison, 2019). Therefore, it follows that many Latin American authors would publish locally, and their articles would be subject to an adequate peer review process, and a small portion of articles would be excluded based on research integrity issues.

Eastern Europe is a less talked about region in the realm of OA, so the fact that none of the authorships from this region was excluded is notable. On the other hand, Arab authorships were tied to papers that were overrepresented in exclusions. In an opinion piece (2023), Emily Choynowski detailed the resistance she encountered when promoting OA initiatives in the Arab region, where she notes that OA and predatoriness are often conflated. If authors are not well versed in identifying predatory journals among OA, it is possible authors from this region might be more likely to publish in predatory journals – however, more research would have to be done on author perceptions and understandings of predatory journals to make any type of conclusion.

5.2.4 CiteScore

RQ4: What is the relationship between articles' inclusion or exclusion in the review and the CiteScore of the journal in which they are published?

The mean CiteScore of included articles was the higher than the excluded, and eligible groups. However, the minimum values of all 3 groups are within 0.1 of each other, the excluded q1 value is higher than the included, and the q3 and median are the same across all groups. While this does initially seem odd, there are multiple articles published in the same journals, meaning multiple instances of the same CiteScore value for each instance of a repeated journal. It is possible that these values are due to popular or frequent journals in the article group. The main difference comes from the maximum value, which is 15.3 in the excluded group, but the included/eligible group have a max value of 134.4.

While journal impact measures are not direct measures of quality, measures like JIF and CiteScore have long been thought to be markers (Saha et al., 2003). Severin et al. (2023) found that higher impact biomedical journals have more thorough peer review processes, and rejection rates of many notable high impact journals are so high that most articles don't even make it to peer review. This is not to say that high impact journals never publish articles with research integrity issues – oftentimes, higher impact journals have high retraction rates (Lu et al., 2013). Nevertheless, reputable journals would aim to ensure that they aren't knowingly publishing, or failing to retract articles that contain plagiarism, errors, or other research integrity issues.

5.2.5 Clusters

RQ5: How are articles included and excluded from the review distributed across publication clusters constructed with different citation-relationship types?

The assumption of bibliometric clustering is that papers that have citation relationships will have similar topical foci (Waltman et al., 2010). The hypothesis of this research was that if the screening process of a systematic review is to remove or filter out papers that are not relevant to the research question, these papers would be grouped into clusters based on their topical foci (whether relevant to the research question or not). Papers without any type of citation relationship with other papers would not be likely to be included, due to the nature of scientific citation behavior, in which an author is expected to acknowledge and describe previous work that explored similar questions.

Of the 3 methodologies applied to clustering, bibliographic coupling was the least successful at grouping together included papers. This methodology involves grouping papers that cite the same papers together into a group. So, if papers A and B both cite paper C, they'll be put in the same group together. Co-citations, which are a similar type of grouping were far more successful, and they are characterized by grouping together papers that are cited by the same papers. So, if paper C cites both papers A and B they are grouped together. The success of co-citations compared to bibliographic coupling implies that papers being cited by a shared paper is more indicative of their relevance to this research question than two papers both citing another paper. Direct citations were also effective at grouping together included papers. These papers are grouped when one paper cites another. Therefore, papers that are relevant and high-quality cite other relevant high-quality papers.

These results do support the notion of core relevant papers, that cite each other, and are cited by the same papers. In contrast with other sections of this analysis, which aimed to predict inclusions/exclusions for eligible papers, this analysis involves all papers captured in the search, so rather than quality, this section aims to predict relevance. Therefore, clustering should occur at the early screening stages and not be used to predict articles with quality concerns.

5.3 Predictive analysis

When predicting inclusion of articles from a list of eligible studies, the model was 90.1% accurate, but only CiteScore was independently significant as a predictor of inclusion. This is notable, as this was the only metric considered at a journal level, rather than article level. While in combination, article-based factors, like funding, attributes of authorships and open access status fairly successfully predict inclusion, they are not significant in and of themselves. This illustrates the complexity of screening, which only involves reading the articles and making decisions based on their attributes and topical focus (Munn et al. 2021); however, venue attributes do seem to matter in this case.

It is key to note that this analysis was conducted on papers that were relevant, and were only excluded based on quality concerns, so a model of this type would not be applicable in the case where authors were aiming to predict inclusion of all articles collected for screening. Additionally, due to the methodological expectations of authors to individually screen each article before inclusion in a systematic review, this paper does not advocate for a replacement of this process with a similar model but does suggest that consideration of a combination of article and journal-based factors, may be helpful for authors when reviewing their list of eligible studies.

5.4 Limitations

5.4.1 Missing data

While OpenAlex was selected for its wide coverage, there were still a number of articles that could not be matched based on DOI or title. Also, records that were found still had missing data. For tens of thousands of records, it was not feasible to manually search for each of these missing values, meaning there were instances where records were not included in the analysis. Therefore, this analysis cannot be described as accurately describing the entire collection of records captured in the search.

5.4.2 Data skew

When considering the entire group of articles captured in the search (n=20,366), only a very small percentage of articles were eligible for inclusion (n=455). Among the eligible articles, only a small percentage of articles were excluded based on research integrity concerns. Therefore, I was often contrasting a very large group of articles, with a very small group. To make these comparisons meaningful, in addition to frequency, I used percentages, and expected values. Nevertheless, comparisons between groups were not between two equal groups, and sample sizes could be quite small.

5.4.3 Selection of variables of interest

At the time of writing, no predictive models using bibliometric analysis have been used to predict article inclusion in a review. Therefore, the variables of interest chosen were selected based only on the existing research and commentary on issues contributing to low quality articles, which is mostly qualitative in nature. From systematic review to systematic review, there is no singular, agreed upon checklist that can be used to identify problematic articles, subject experts are required to conduct such an analysis, based on their understanding of methodologies and reporting responsibilities within their field. Therefore, a quantitative approach in which attributes are reduced to numbers and applying them to a systematic review risk losing some of the important context. As there is no existing research with a similar objective or questions to this research, it is difficult to interpret and contextualize these findings within the literature.

5.4.4 Limited applicability

As noted in chapter 3, this case study review was chosen, both because it is characteristic of all Cochrane reviews, and an exemplary contribution, based on the additional metadata collection. However, this analysis and level of detail are not required, and therefore, do not uniformly exist within other review groups. While future research could attempt to apply these methods to another systematic review, the lack of detail about research integrity exclusions might prohibit an analysis of this kind. Additionally, the findings of this research, and the qualities of articles/journals that are related in this case to article inclusion or exclusion cannot be extrapolated to apply to any case study. The publishing norms, venues or methodologies of a discipline vary greatly between different fields. Additionally, a different search would have gathered a different list of articles, which may have been included or excluded differently.

Chapter 6: Conclusion

6.1 Summary of findings

In this research, I used the metadata from a Cochrane systematic review to describe the characteristics of articles eligible for inclusion. Open access status, funder types, author affiliations, CiteScores were considered. For all articles included in the search for the systematic review, the Louvain clustering approach was applied, for 3 different citation relationships to group together articles on similar topics.

6.1.1 Open access

Approximately half of articles eligible for inclusion were open access, which is higher than would be expected, as a smaller percentage of total articles are OA. However, measuring the proportion of articles across disciplines and publishing venues that are OA would be challenging, and likely estimates would quickly become outdated. Therefore, estimates within various disciplines could be more helpful for comparison. Narratives on the lower quality of OA articles are not supported by this data, as the proportion of OA versus closed articles does not vary significantly between the included and excluded article groups. While this conversation was more prevalent decades ago, as OA became more common place, this information is still important evidence for those who are doubtful of the quality of OA articles.

Additionally, when broken down by OA status, gold OA was the most frequently observed type, in both included and excluded groups. However, it was underrepresented in the group of included articles, while green was overrepresented. The authors did explicitly identify journals thought to be predatory, and flagged these as an item considered when screening articles for research integrity, so it is possible predatory journals benefiting from APCs were most likely to use a gold or hybrid model. Among all OA articles, Piwowar et al. (2018) found bronze OA to be the most common. However, a small number of articles eligible for inclusion were bronze, suggesting that either 1) more journals are adding licenses to the works they publish, 2) OpenAlex identifies some bronze journals incorrectly, potentially due to mistaking them as gold, or another type, or 3) disciplinary norms or patterns in this field result in authors infrequently publishing in bronze journals. Despite over or underrepresentation in the included/excluded groups, no OA type was an independent, significant predictor of inclusion in the review. This information reiterates the idea that while predatory journals may be more likely to be a certain OA type, it is still important to examine articles individually to assess their quality, as OA type alone cannot identify questionable research practices.

6.1.2 Funder type

Funder types and reporting of funding received was important to consider in this analysis, as identifying financial contributions to a work is a key component of transparency best practices in science. Many instances of no reporting or listing funders who I could not find were noted, but also many articles that explicitly stated that they did not receive funding. Of identifiable funders, government organizations were the most common, followed by educational institutions, and then industry groups. I expected most articles eligible for inclusion to have some type of funding, as RCTs can be expensive to run, so I did expect frequent government funding, which can be quite large, or span multiple years. However, many of the trials in this field are small with short-term follow up, so a lack of funding was more common than initially anticipated. Educational institutions were more frequently observed in the eligible group than industry, which was an interesting observation. I expected industry to be a frequent funder, to support Research & Development endeavors, however, it is possible that the research funded by industries is not as frequently published in academic articles, as another format may be more conducive to product/service design.

Additionally, for profit industry sponsoring in health is not evenly distributed across all subdisciplines, and typically focuses more on product specific research, like specific pharmacological or nutrient interventions, rather than behaviour, such as exercise (which is the intervention in this review) (Fabbri et al., 2018). Also surprising was the lack of exclusion of any articles funded by industry groups from the review, as practitioners can be more critical of industry-funded research. This may again speak to the diversity of quality, and limitations of funder types alone predicting inclusion, or industry may be too broad of a category, as it encompasses for profit, non-profit and charity groups.

6.1.3 Authorship characteristics

To study specific articles rather than the journals that publish them, I focused on authorships of articles, what institution types they were affiliated with, and where they were geographically located. There is more than one author associated with each work, so we considered each authorship in the descriptive section, to get a sense of overall contributions. However, for the predictive aspect, only the first author was considered. Most authors were affiliated with educational institutions, followed by unknown institution types, and then healthcare. The prevalence of educational institution affiliated authors does line up with the publish or perish dynamic, which is particularly felt by researchers seeking tenure and promotion at post-secondary institutions. However, it is important to mention that there are many researchers in medicine who may be affiliated with both a hospital/health service and an educational institution due to the professional teacher emphasis in medicine. So, I cannot say whether an author only has a single affiliation, or if this data speaks instead to the affiliation authors are more likely to report. However, in allied health, many healthcare appointments are in private primary care, which doesn't always support research, so their university affiliation may be the organization providing resources for their research work.

Author countries were most frequently unknown, with significant numbers coming from Iran, the USA, Brazil, South Korea, Australia, the United Kingdom, Spain, Canada and Germany. Among the included group, the most frequent countries were the same, with only slight differences in the percentage of each. In the excluded group, Thailand, Denmark and South Africa entered the top 10, replacing Canada, Brazil and Germany. By regions, authors were most frequently located in Western Europe and North America, unknown regions, Asia and the Pacific, and Latin America and the Caribbean. Notably, works associated with authors based in Latin America and Eastern Europe were very overrepresented in the included group and Arab was underrepresented.

Latin America has long been considered a leader in OA publishing, due to a primarily diamond OA system, wherein predatory publishers would not thrive, due to a lack of APCs. However, Eastern Europe and Arab regions have been understudied for their publishing landscape. It has been noted that authors in the Arab region might be resistant to OA (Choynowski, 2023). However, there may also be issues with how countries were considered, as I counted each authorship, so large author teams, or a geographic skew in OpenAlex reporting could have resulted in an inaccurate representation of the countries. The review authors found only 10% of studies in the 2021 review were conducted in the Middle East (Hayden, Ellis, Ogilvie, Malmivaara, et al., 2021). Considering regions was more conducive to comparison than countries, due to small numbers of articles belonging to some countries, and a large number of unknown author locations.

6.1.4 CiteScore

Articles in the included group came from journals with a higher average CiteScore than those that were excluded. The maximum value for CiteScores in the included group was 134.4, and 15.3 in the excluded group. Both the included and excluded groups had similar minimum values and medians. Therefore, a small number of articles in journals with high CiteScores may be bringing up the mean. However, CiteScores were the only characteristic that were statistically significant independent of other characteristics in predicting inclusion. Therefore, while articles still must be screened individually, CiteScores may be an effective predictor of inclusion, meaning articles that are more highly cited may be more likely to be included in this review. While research quality and impact are separate concepts, these results do suggest a relationship between the two in this review.

6.1.5 Clustering

The clustering analysis was conducted on all articles captured in the database searches, and clusters were aiming to capture groups of articles with the same topical focus. Direct citations, co-citations and bibliographic coupling-based clusters were generated, and direct citations, co-citations and combinations of the two were more effective at grouping together included articles. The success rate was not 100%, meaning that there were instances of included articles being grouped into a different cluster than the majority of articles. Direct citations resulted in the largest number of clusters, with smaller numbers of articles per group. Therefore, direct citations and co-citations are more effective at grouping together articles by relevance and inclusion status for the topic of this review.

6.1.6 Logistic regression model

Using open access status, funder types, author affiliation types, and CiteScores, a logistic regression model was generated, which predicted inclusion in the Cochrane review. The model was able to successfully predict inclusion 90.1% of the time. It predicted exclusion for 3 articles that were ultimately included, and predicted inclusion for 16 articles that were ultimately excluded. It is important to note that the model was created with a dataset in which the vast majority of studies were included, so this finding must be interpreted with this in mind.

6.2. Theoretical contributions

While this research represents a case study of a single Cochrane review, and the findings cannot be broadly applied to all systematic reviews, or even Cochrane reviews, the descriptive aspect of this analysis does provide insight into characteristics that are over and underrepresented in included/excluded groups. One risk of extrapolating these findings is that generalizations like 'gold open access journals are predatory' or 'research funded by educational institutions is low quality' are untrue. Nevertheless, these findings can illustrate areas of future research, to determine if they are specific to this review/discipline, or if they are observed in other cases. If so, collecting information on OA status might be a suggested best practice for authors involved in systematic reviews. Collecting large amounts of metadata on articles for systematic reviews is challenging, as tens of thousands of articles must be screened. However, the work of the authors of this review, and the use of OpenAlex combined were able to support this data collection, and an illustration of the makeup of articles eligible for the review, a novel contribution to the field.

6.3. Methodological contributions

In the process of screening for articles, authors aim to sort articles based on relevance to the research question, and inclusion criteria. In this case study, authors additionally screened based on research integrity concerns. While describing articles and inclusion groups does not replicate a screening process, using article metadata and bibliometric methods to describe a group of works can support authors in understanding the field their work is situated in. The use of OpenAlex to collect metadata to support in the data collection for a systematic review is becoming more widespread, with EPPI-Reviewer, a systematic review management software, recently integrating with OpenAlex to give reviewers an option to keep their reviews up to date. Therefore, a working knowledge of OpenAlex will be an asset to reviewers moving forward – and rather than just focusing on keeping reviews updated, reviewers could use the methods outlined in this paper to describe the characteristics of their dataset and use this information to alter their search or screening process.

6.4. Practical contributions

The objective of this research is not to replace individual article screening in systematic reviews with bibliometric tools, but to apply these tools to describe the dataset. A process like this could be conducted after independent screening to help illustrate patterns noticed by reviewers. In this case, by separating included and excluded articles, bibliometric tools can be used to generate a descriptive analysis, which can illustrate differences between groups. Identifying a particular pattern or difference (like articles from authors based in a certain geographic region), can alert authors to potential biases in their screening, or potentially identifying problematic actors who consistently contribute to works with research integrity issues.

Clustering could be used to outline key concepts or journals in the field, to support in generating future inclusion criteria, or search terms. Direct and co-citation relationships were most successful at grouping together articles, so examining keywords of the groups that contained included articles, or conversely, avoiding (if possible) terms used in groups with only excluded articles may support in search generation and efficiency.

6.5. Concluding remarks and future research

In this case study, I aimed to describe the attributes of articles captured in the search (through clustering), and articles eligible for inclusion (all other variables) and predict inclusion/exclusion on the basis of research integrity. Only CiteScore was a significant predictor of inclusion, but the model was 90.1% accurate. However, there were a number of missing values for multiple variables in OpenAlex, meaning that the results may not entirely capsulate the articles in the Cochrane review.

Besides screening, another key aspect of the systematic review process is the search strategy. OpenAlex contains a list of MeSH terms associated with each article, which could support reviewers in identifying the most common MeSH terms associated with eligible articles and included/excluded groups. This information could be compared to MeSH terms used in the search strategy, to determine if keywords or MeSH are contributing to bringing in most eligible articles. This research could support reviewers in updating their search and potentially selecting additional MeSH terms to include in a search.

Other potential predictors of inclusion could be added to the logistic regression model, to see if the accuracy could be improved. The characteristics that were used in this model were based on the information available from OpenAlex and the authors of the review, but another review in a different discipline may have different key characteristics that would be of interest to explore.

References

Acton, S. E., Bell, A. J., Toseland, C. P., & Twelvetrees, A. (2019). A survey of new PIs in the UK. *eLife*, *8*, e46827. https://doi.org/10.7554/eLife.46827

Anderson, M. S., Ronning, E. A., De Vries, R., & Martinson, B. C. (2007). The Perverse Effects of Competition on Scientists' Work and Relationships. *Science and Engineering Ethics*, *13*(4), 437–461. https://doi.org/10.1007/s11948-007-9042-5

Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, *489*(7415), Article 7415. https://doi.org/10.1038/489179a

Beall, J. (2021). *Beall's List – of Potential Predatory Journals and Publishers*. https://beallslist.net/

Beaubien, S., & Eckard, M. (2014). Addressing Faculty Publishing Concerns with Open Access Journal Quality Indicators. *Journal of Librarianship and Scholarly Communication*, *2*(2). https://doi.org/10.7710/2162-3309.1133

Bello, S. A., Azubuike, F. C., & Akande, O. A. (2023). Reputation disparity in teaching and research productivity and rewards in the context of consequences of institutionalization of Publish or Perish culture in academia. *Higher Education Quarterly*, *77*(3), 574–584. https://doi.org/10.1111/hequ.12417

Berger, M. (2021). Bibliodiversity at the Centre: Decolonizing Open Access. *Development and Change*, *52*(2), 383–404. https://doi.org/10.1111/dech.12634

Björk, B.-C. (2017). Growth of hybrid open access, 2009–2016. *PeerJ*, *5*, e3878. https://doi.org/10.7717/peerj.3878

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, *115*(19), 4887–4890. https://doi.org/10.1073/pnas.1719557115

Bolland, M. J., Avenell, A., Gamble, G. D., & Grey, A. (2016). Systematic review and statistical analysis of the integrity of 33 randomized controlled trials. *Neurology*, *87*(23), 2391–2402. https://doi.org/10.1212/WNL.0000000000003387

Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*, *7*(2), e012545. https://doi.org/10.1136/bmjopen-2016-012545

Bornmann, L., Haunschild, R., & Mutz, R. (2021). Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanities and Social Sciences Communications*, *8*(1), Article 1. https://doi.org/10.1057/s41599-021-00903-w

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222. https://doi.org/10.1002/asi.23329

Bornmann, L., & Williams, R. (2013). How to calculate the practical significance of citation impact differences? An empirical example from evaluative institutional bibliometrics using adjusted predictions and marginal effects. *Journal of Informetrics*, *7*(2), 562–574. https://doi.org/10.1016/j.joi.2013.02.005

Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Systematic Reviews*, *6*(1), 245. https://doi.org/10.1186/s13643-017-0644-y

Bray, N. J., & Major, C. H. (2022). Impact Factors, Altmetrics, and Prestige, Oh My: The Relationship Between Perceived Prestige and Objective Measures of Journal Quality. *Innovative Higher Education*, *47*(6), 947–966. https://doi.org/10.1007/s10755-022-09635-4

Callaham, M., Wears, R. L., & Weber, E. (2002). Journal Prestige, Publication Bias, and Other Characteristics Associated With Citation of Published Studies in Peer-Reviewed Journals. *JAMA*, *287*(21), 2847–2850. https://doi.org/10.1001/jama.287.21.2847

Chapman, A., Rankin, N. M., Jongebloed, H., Yoong, S. L., White, V., Livingston, P. M., Hutchinson, A. M., & Ugalde, A. (2023). Overcoming challenges in conducting systematic reviews in implementation science: A methods commentary. *Systematic Reviews*, *12*(1), 116. https://doi.org/10.1186/s13643-023-02285-3

Charles University. (2023, January 17). *How to publish OA?* https://openscience.cuni.cz/OSCIEN-28.html

Chloros, G. D., Prodromidis, A. D., & Giannoudis, P. V. (2023). Has anything changed in Evidence-Based Medicine? *Injury*, *54*, S20–S25. https://doi.org/10.1016/j.injury.2022.04.012

Choynowski, E. (2023, August 24). Supporting open science in the Arab world. *Research Information*. https://www.researchinformation.info/analysis-opinion/supporting-open-science-arab-world

Clarke, M., & Chalmers, I. (2018). Reflections on the history of systematic reviews. *BMJ Evidence-Based Medicine*, *23*(4), 121–122. https://doi.org/10.1136/bmjebm-2018-110968

Cochrane Library. (2023). *About the Cochrane Database of Systematic Reviews | Cochrane Library*. https://www.cochranelibrary.com/cdsr/about-cdsr

Coeytaux, R. R., McDuffie, J., Goode, A., Cassel, S., Porter, W. D., Sharma, P., Meleth, S., Minnella, H., Nagi, A., & John W Williams, J. (2014). Criteria Used in Quality Assessment of Systematic Reviews. In *Evidence Map of Yoga for High-Impact Conditions Affecting Veterans [Internet]*. Department of Veterans Affairs (US). https://www.ncbi.nlm.nih.gov/books/NBK242394/

Cukier, S., Helal, L., Rice, D. B., Pupkaite, J., Ahmadzai, N., Wilson, M., Skidmore, B., Lalu, M. M., & Moher, D. (2020). Checklists to detect potential predatory biomedical journals: A systematic review. *BMC Medicine*, *18*(1), 104. https://doi.org/10.1186/s12916-020-01566-1

Culbert, J., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2024). *Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus*. https://doi.org/10.48550/ARXIV.2401.16359

Cumpston, M., & Flemyng, E. (2023). Chapter II: Planning a Cochrane Review. In *Cochrane Handbook for Systematic Reviews of Interventions* (6.4). https://training.cochrane.org/handbook/current/chapter-ii

De Rond, M., & Miller, A. N. (2005). *Publish or Perish: Bane or Boon of Academic Life? 14*(4). https://doi.org/10.1177/105649260527685

DeCoursey, T. E. (2006). It's difficult to publish contradictory findings. *Nature*, *439*(7078), Article 7078. https://doi.org/10.1038/439784b

Dicenso, A., Bayley, L., & Haynes, R. B. (2009). Accessing pre-appraised evidence: Fine-tuning the 5S model into a 6S model. *Evidence-Based Nursing*, *12*(4), 99–101. https://doi.org/10.1136/ebn.12.4.99-b

Ding, D., & Li, Z. (2021). The theoretical origin of the knowledge-sharing mode of open access: From knowledge communism to academic capitalism. *Cultures of Science*, *4*(4), 199–207. https://doi.org/10.1177/20966083221075424

Durieux, V., & Gevenois, P. A. (2010). Bibliometric indicators: Quality measurements of scientific publication. *Radiology*, *255*(2), 342–351. https://doi.org/10.1148/radiol.09090626

Eddy, D. M. (2011). The Origins of Evidence-Based Medicine: A Personal Perspective. *AMA Journal of Ethics*, *13*(1), 55–60. https://doi.org/10.1001/virtualmentor.2011.13.1.mhst1-1101

Eriksen, M. B., & Frandsen, T. F. (2018). The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: A systematic review. *Journal of the Medical Library Association : JMLA*, *106*(4), 420–431. https://doi.org/10.5195/jmla.2018.345

Fabbri, A., Lai, A., Grundy, Q., & Bero, L. A. (2018). The Influence of Industry Sponsorship on the Research Agenda: A Scoping Review. *American Journal of Public Health*, *108*(11), e9–e16. https://doi.org/10.2105/AJPH.2018.304677

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7

Fire, M., & Guestrin, C. (2019). Over-optimization of academic publishing metrics: Observing Goodhart's Law in action. *GigaScience*, *8*(6), giz053. https://doi.org/10.1093/gigascience/giz053

Flacco, M. E., Manzoli, L., Boccia, S., Capasso, L., Aleksovska, K., Rosso, A., Scaioli, G., De Vito, C., Siliquini, R., Villari, P., & Ioannidis, J. P. A. (2015). Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor. *Journal of Clinical Epidemiology*, *68*(7), 811–820. https://doi.org/10.1016/j.jclinepi.2014.12.016

Fleming, P. S., Seehra, J., Polychronopoulou, A., Fedorowicz, Z., & Pandis, N. (2013). Cochrane and non-Cochrane systematic reviews in leading orthodontic journals: A quality paradigm? *European Journal of Orthodontics*, *35*(2), 244–248. https://doi.org/10.1093/ejo/cjs016

Fretheim, A., Brurberg, K. G., & Forland, F. (2020). Rapid reviews for rapid decision-making during the coronavirus disease (COVID-19) pandemic, Norway, 2020. *Eurosurveillance*, *25*(19), 2000687. https://doi.org/10.2807/1560-7917.ES.2020.25.19.2000687

Ganann, R., Ciliska, D., & Thomas, H. (2010). Expediting systematic reviews: Methods and implications of rapid reviews. *Implementation Science*, *5*(1), 56. https://doi.org/10.1186/1748-5908-5-56

Garcia-Doval, I., van Zuuren, E. j., Bath-Hextall, F., & Ingram, J. r. (2017). Systematic reviews: Let's keep them trustworthy. *British Journal of Dermatology*, *177*(4), 888–889. https://doi.org/10.1111/bjd.15826

Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., Michie, S., Moher, D., & Wager, E. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, *383*(9913), 267–276. https://doi.org/10.1016/S0140-6736(13)62228-X

Gomez, C. J., Herman, A. C., & Parigi, P. (2022). Leading countries in global science increasingly receive more citations than other countries doing similar research. *Nature Human Behaviour*, *6*(7), Article 7. https://doi.org/10.1038/s41562-022-01351-5

Gopalakrishna, G., Ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The

Netherlands. *PLOS ONE*, *17*(2), e0263023. https://doi.org/10.1371/journal.pone.0263023

Gopalakrishnan, S., & Ganeshkumar, P. (2013). Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. *Journal of Family Medicine and Primary Care*, *2*(1), 9–14. https://doi.org/10.4103/2249-4863.109934

Government of Canada, I. (2023, March 10). *Tri-Agency Open Access Policy on Publications*. Innovation, Science and Economic Development Canada. https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/open-access/tri-agency-open-access-policy-publications

Grančay, M., Vveinhardt, J., & Šumilo, Ē. (2017). Publish or perish: How Central and Eastern European economists have dealt with the ever-increasing academic publishing requirements 2000–2015. *Scientometrics*, *111*(3), 1813–1837. https://doi.org/10.1007/s11192-017-2332-z

Grudniewicz, A., Moher, D., Cobey, K. D., Bryson, G. L., Cukier, S., Allen, K., Ardern, C., Balcom, L., Barros, T., Berger, M., Ciro, J. B., Cugusi, L., Donaldson, M. R., Egger, M., Graham, I. D., Hodgkinson, M., Khan, K. M., Mabizela, M., Manca, A., … Lalu, M. M. (2019). Predatory journals: No definition, no defence. *Nature*, *576*(7786), 210–212. https://doi.org/10.1038/d41586-019-03759-y

Guyatt, G. H., Oxman, A. D., Kunz, R., Atkins, D., Brozek, J., Vist, G., Alderson, P., Glasziou, P., Falck-Ytter, Y., & Schünemann, H. J. (2011). GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of Clinical Epidemiology*, *64*(4), 395–400. https://doi.org/10.1016/j.jclinepi.2010.09.012

Hansoti, B., Langdorf, M. I., & Murphy, L. S. (2016). Discriminating Between Legitimate and Predatory Open Access Journals: Report from the International Federation for Emergency Medicine Research Committee. *The Western Journal of Emergency Medicine*, *17*(5), 497–507. https://doi.org/10.5811/westjem.2016.7.30328

Hardi, A. C., & Fowler, S. A. (2014). Evidence-Based Medicine and Systematic Review Services at Becker Medical Library. *Missouri Medicine*, *111*(5), 416–418.

Hayden, J. A., Ellis, J., Ogilvie, R., Boulos, L., & Stanojevic, S. (2021). Meta-epidemiological study of publication integrity, and quality of conduct and reporting of randomized trials included in a systematic review of low back pain. *Journal of Clinical Epidemiology*, *134*, 65–78. https://doi.org/10.1016/j.jclinepi.2021.01.020

Hayden, J. A., Ellis, J., Ogilvie, R., Malmivaara, A., & van Tulder, M. W. (2021). Exercise therapy for chronic low back pain. *Cochrane Database of Systematic Reviews*, *2021*(10). https://doi.org/10.1002/14651858.CD009790.pub2

Hayden, J. A., Ogilvie, R., Kashif, S., Singh, S., Boulos, L., Stewart, S. A., Wieland, L. S., Jesus-Moraleida, F. R., Saragiotto, B. T., Yamato, T. P., Zoete, A. de, Bülow, K., Oliveira, L. A. de, Bejarano, G., & Cancelliere, C. (2023). Exercise treatments for chronic low back pain: A network meta-analysis. *Cochrane Database of Systematic Reviews*, *6*. https://doi.org/10.1002/14651858.CD015608

Huutoniemi, K. (2012). Communicating and compromising on disciplinary expertise in the peer review of research proposals. *Social Studies of Science*, *42*(6), 897–921. https://doi.org/10.1177/0306312712458478

Ioannidis, J. P. a. (2016). The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, *94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210

Jeong, G. H., & Huh, S. (2016). Status of digital standards in Korean medical journals in 2016. *Science Editing*, *3*(2), 100–104. https://doi.org/10.6087/kcse.73

Jones, P. (2015, August 10). Defending Regional Excellence in Research or Why Beall is Wrong About SciELO. *The Scholarly Kitchen*. https://scholarlykitchen.sspnet.org/2015/08/10/defending-regional-excellence-in-research-or-why-beall-is-wrong-about-scielo/

Kamath, S., & Guyatt, G. (2016). Importance of evidence-based medicine on research and practice. *Indian Journal of Anaesthesia*, *60*(9), 622–625. https://doi.org/10.4103/0019-5049.190615

Kellermeyer, L., Harnke, B., & Knight, S. (2018). Covidence and Rayyan. *Journal of the Medical Library Association : JMLA*, *106*(4), 580–583. https://doi.org/10.5195/jmla.2018.513

Kelly, J., Sadeghieh, T., & Adeli, K. (2014). Peer Review in Scientific Publications: Benefits, Critiques, & A Survival Guide. *EJIFCC*, *25*(3), 227–243.

Khamis, A. M., Bou-Karroum, L., Hakoum, M. B., Al-Gibbawi, M., Habib, J. R., El-Jardali, F., & Akl, E. A. (2018). The reporting of funding in health policy and systems research: A cross-sectional study. *Health Research Policy and Systems*, *16*(1), 83. https://doi.org/10.1186/s12961-018-0356-3

Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, *96*(3), 118–121.

Kharasch, E. D., Avram, M. J., Clark, J. D., Davidson, A. J., Houle, T. T., Levy, J. H., London, M. J., Sessler, D. I., & Vutskits, L. (2021). Peer Review Matters: Research Quality and the Public Trust. *Anesthesiology*, *134*(1), 1–6. https://doi.org/10.1097/ALN.0000000000003608

Kincaid, A. E. (2022, December 5). Board members decry their own journal's retraction of paper on predatory publishers. *Retraction Watch*. https://retractionwatch.com/2022/12/05/board-members-decry-their-own-journals-retraction-of-paper-on-predatory-publishers/

Klerings, I., Weinhandl, A. S., & Thaler, K. J. (2015). Information overload in healthcare: Too much of a good thing? *Zeitschrift Fur Evidenz, Fortbildung Und Qualitat Im Gesundheitswesen*, *109*(4–5), 285–290. https://doi.org/10.1016/j.zefq.2015.06.005

Lind, J. (1753). *A treatise of the scurvy. In three parts. Containing an inquiry into the nature, causes, and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. By James Lind, M.D. Fellow of the Royal College of Physicians in Edinburgh*. printed by Sands, Murray, and Cochran. For A. Kincaid & A. Donaldson; Eighteenth Century Collections Online. https://link.gale.com/apps/doc/CW0107813863/ECCO?sid=bookmark-ECCO&xid=f3654712&pg=204

Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2019). Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*. https://doi.org/10.1177/0312896219877678

Liu, F., Rahwan, T., & AlShebli, B. (2023). Non-White scientists appear on fewer editorial boards, spend more time under review, and receive fewer citations. *Proceedings of the National Academy of Sciences*, *120*(13), e2215324120. https://doi.org/10.1073/pnas.2215324120

Louis, K. S., Anderson, M. S., & Rosenberg, L. (1995). Academic Misconduct and Values: The Department's Influence. *The Review of Higher Education*, *18*(4), 393–422. https://doi.org/10.1353/rhe.1995.0007

Lu, S. F., Jin, G. Z., Uzzi, B., & Jones, B. (2013). The Retraction Penalty: Evidence from the Web of Science. *Scientific Reports*, *3*(1), Article 1. https://doi.org/10.1038/srep03146

Macháček, V., & Srholec, M. (2022). Predatory publishing in Scopus: Evidence on cross-country differences. *Quantitative Science Studies*, *3*(3), 859–887. https://doi.org/10.1162/qss_a_00213

Merton, R. K. (1968). The Matthew Effect in Science. *Science*, *159*(3810), 56–63.

Milne, I. (2012). Who was James Lind, and what exactly did he achieve. *Journal of the Royal Society of Medicine*, *105*(12), 503. https://doi.org/10.1258/jrsm.2012.12k090

Morrison, H. (2019, July 15). National open access journal subsidy. *Sustaining the Knowledge Commons*.

https://sustainingknowledgecommons.org/2019/07/15/national-open-access-journal-subsidy/

Moseley, A. M., Elkins, M. R., Herbert, R. D., Maher, C. G., & Sherrington, C. (2009). Cochrane reviews used more rigorous methods than non-Cochrane reviews: Survey of systematic reviews in physiotherapy. *Journal of Clinical Epidemiology*, *62*(10), 1021–1030. https://doi.org/10.1016/j.jclinepi.2008.09.018

Munn, Z., Barker, T., Stern, C., Pollock, D., Ross-White, A., Klugar, M., Wiechula, R., Aromataris, E., & Shamseer, L. (2021). Should I include studies from "predatory" journals in a systematic review? Interim guidance for systematic reviewers. *JBI Evidence Synthesis*, *19*(8), 1915–1923. https://doi.org/10.11124/JBIES-21-00138

Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *BMJ Evidence-Based Medicine*, *21*(4), 125–127. https://doi.org/10.1136/ebmed-2016-110401

O'Reilly (Director). (2008, September 19). *Web 2.0 Expo NY: Clay Shirky (shirky.com) It's Not Information Overload. It's Filter Failure*. https://www.youtube.com/watch?v=LabqeJEOQyI

Page, M. J., & Moher, D. (2017). Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: A scoping review. *Systematic Reviews*, *6*, 263. https://doi.org/10.1186/s13643-017-0663-8

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *The BMJ*, *372*, n160. https://doi.org/10.1136/bmj.n160

Patsopoulos, N. A., Analatos, A. A., & Ioannidis, J. P. A. (2005). Relative citation impact of various study designs in the health sciences. *JAMA*, *293*(19), 2362–2366. Scopus. https://doi.org/10.1001/jama.293.19.2362

Pearson, H. (2021). How COVID broke the evidence pipeline. *Nature*, *593*(7858), 182–185. https://doi.org/10.1038/d41586-021-01246-x

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, *6*, e4375. https://doi.org/10.7717/peerj.4375

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. https://doi.org/10.48550/ARXIV.2205.01833

Rawat, S., & Meena, S. (2014). Publish or perish: Where are we heading? *Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences*, *19*(2), 87–89.

Rethlefsen, M. L., Farrell, A. M., Osterhaus Trzasko, L. C., & Brigham, T. J. (2015). Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *Journal of Clinical Epidemiology*, *68*(6), 617–626. https://doi.org/10.1016/j.jclinepi.2014.11.025

Ross-White, A., Godfrey, C. M., Sears, K. A., & Wilson, R. (2019). Predatory publications in evidence syntheses. *Journal of the Medical Library Association*, *107*(1), Article 1. https://doi.org/10.5195/jmla.2019.491

Rueda, A. (2023, October 10). Q&A: Science is a public good, not a business. *SciDev.Net*. https://www.scidev.net/global/role-models/qa-science-is-a-public-good-not-a-business/

Saha, S., Saint, S., & Christakis, D. A. (2003). Impact factor: A valid measure of journal quality? *Journal of the Medical Library Association: JMLA*, *91*(1), 42–46.

Saif-Ur-Rahman, K. M., Hasan, Md., Hossain, S., Anwar, I., Hirakawa, Y., & Yatsuya, H. (2022). Prioritization and sequential exclusion of articles in systematic reviews. *Campbell Systematic Reviews*, *18*(2), e1229. https://doi.org/10.1002/cl2.1229

Sax, P. (2018, May 28). *Predatory Journals Are Such a Big Problem It's Not Even Funny*. HIV and ID Observations. https://blogs.jwatch.org/hiv-id-observations/index.php/predatory-journals-big-problem-not-even-funny/2018/05/28/

Seeber, M., Cattaneo, M., Meoli, M., & Malighetti, P. (2019). Self-citations as strategic response to the use of metrics for career decisions. *Research Policy*, *48*(2), 478–491. https://doi.org/10.1016/j.respol.2017.12.004

Severin, A., Strinzel, M., Egger, M., Barros, T., Sokolov, A., Mouatt, J. V., & Müller, S. (2023). Relationship between journal impact factor and the thoroughness and helpfulness of peer reviews. *PLOS Biology*, *21*(8), e3002238. https://doi.org/10.1371/journal.pbio.3002238

Severinsen, A., Midtgaard, J., Backhausen, M. G., Broberg, L., & Hegaard, H. K. (2019). Pregnant women's experiences with sick leave caused by low back pain. A qualitative study. *Work*, *64*(2), 271–281. c8h. https://doi.org/10.3233/WOR-192991

Siemieniuk, R., & Guyatt, G. (2023). *What is GRADE? | BMJ Best Practice*. https://bestpractice.bmj.com/info/toolkit/learn-ebm/what-is-grade/

Singh, M., Prasad, C. P., & Shankar, A. (2021). Publication Charges Associated with Quality Open Access (OA) Publishing and Its Impact on Low Middle Income Countries (LMICs), Time to Reframe Research Policies. *Asian Pacific Journal of Cancer Prevention*, *22*(9), 2743–2747. https://doi.org/10.31557/APJCP.2021.22.9.2743

Smits, R.-J., & Pells, R. (2022). *Plan S for Shock: Science. Shock. Solution. Speed.* Ubiquity Press. https://doi.org/10.5334/bcq

Starbuck, W. H. (2005). How Much Better Are the Most-Prestigious Journals? The Statistics of Academic Publication. *Organization Science*, *16*(2), 180–200. https://doi.org/10.1287/orsc.1040.0107

Stevens, K. R. (2001). Systematic reviews: The heart of evidence-based practice. *AACN Clinical Issues*, *12*(4), 529–538. https://doi.org/10.1097/00044067-200111000-00009

Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical Medicine and Health*, *47*(1), 46. https://doi.org/10.1186/s41182-019-0165-6

Thelwall, M., Kousha, K., Abdoli, M., Stuart, E., Makita, M., Wilson, P., & Levitt, J. (2023). Why are co-authored academic articles more cited: Higher quality or larger audience? *Journal of the Association for Information Science and Technology*, *74*(7), 791–810. https://doi.org/10.1002/asi.24755

Thompson, D. F., & Walker, C. K. (2015). A Descriptive and Historical Review of Bibliometrics with Applications to Medical Sciences. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, *35*(6), 551–559. https://doi.org/10.1002/phar.1586

Toupin, R., Millerand, F., & Larivière, V. (2022). Who tweets climate change papers? Investigating publics of research through users' descriptions. *PLOS ONE*, *17*(6), e0268999. https://doi.org/10.1371/journal.pone.0268999

Van Noorden, R. (2016). Controversial impact factor gets a heavyweight rival. *Nature*, *540*(7633), Article 7633. https://doi.org/10.1038/nature.2016.21131

Viswanathan, M., Patnode, C. D., Berkman, N. D., Bass, E. B., Chang, S., Hartling, L., Murad, M. H., Treadwell, J. R., & Kane, R. L. (2017). Assessing the Risk of Bias in Systematic Reviews of Health Care Interventions. In *Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]*. Agency for Healthcare Research and Quality (US). https://www.ncbi.nlm.nih.gov/sites/books/NBK519366/

Waffenschmidt, S., Knelangen, M., Sieben, W., Bühn, S., & Pieper, D. (2019). Single screening versus conventional double screening for study selection in systematic reviews: A methodological systematic review. *BMC Medical Research Methodology*, *19*(1), 132. https://doi.org/10.1186/s12874-019-0782-0

Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). *A unified approach to mapping and clustering of bibliometric networks*. https://doi.org/10.48550/ARXIV.1006.1032

White, H., & Waddington, H. (2012). Why do we care about evidence synthesis? An introduction to the special issue on systematic reviews. *Journal of Development Effectiveness*, *4*(3), 351–358. https://doi.org/10.1080/19439342.2012.711343

Wicherts, J. M. (2016). Peer Review Quality and Transparency of the Peer-Review Process in Open Access and Subscription Journals. *PLOS ONE*, *11*(1), e0147913. https://doi.org/10.1371/journal.pone.0147913

Wilkinson, J., Antoniou, G., Boughton, S., Hilton, J., Bero, L., & Kirkham, J. (2022). *Protocol for a survey to identify methods to detect problematic studies*. OSF. https://osf.io/https://osf.io/s34hx

Willinsky, J., & Rusk, M. (2019). *If Research Libraries and Funders Finance Open Access: Moving Beyond Subscriptions and APCs | Willinsky | College & Research Libraries*. https://doi.org/10.5860/crl.80.3.340

Zarour, M., Alenezi, M., Ansari, M. T. J., Pandey, A. K., Ahmad, M., Agrawal, A., Kumar, R., & Khan, R. A. (2021). Ensuring data integrity of healthcare information in the era of digital health. *Healthcare Technology Letters*, *8*(3), 66–77. https://doi.org/10.1049/htl2.12008