Using software engineering principles to improve the completeness and efficiency of the systematic review ecosystem

RABIA BASHIR

BS (Computer Science), MS (Computer Software Engineering)

Centre for Health Informatics

Australian Institute of Health Innovation

Faculty of Medicine and Health Sciences

March 2019

This thesis is submitted in total fulfillment of the requirements for the degree of

Doctor of Philosophy

Table of Contents

List of Figures

List of Tables

List of Abbreviations

WHO	World Health Organization
ICTRP	International Clinical Trial Registry Platform
DOI	Digital Object Identifiers
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ICMJE	International Committee of Medical Journal Editors
IBM	International Business Machines
FDA	Food and Drug Administration
IQR	Interquartile Range
SR	Systematic Reviews
MA	Meta-analysis
PICO	Population, Interventions, Comparators, and Outcomes
ANZCTR	Australian New Zealand Clinical Trials Registry
ReBec	Brazilian Clinical Trials Registry
ChiCTR	Chinese Clinical Trial Registry
CRiS	Clinical Research Information Service
CTRI	Clinical Trials Registry- India
RPCEC	Cuban Public Registry of Clinical Trials
EU-CTR	European Clinical Trials Register
DRKS	German Clinical Trials Register
IRCT	Iranian Registry of Clinical Trials
NTR	Netherlands Trial Register
ISRCTN	International Standard Randomised Controlled Trials Number
PACTR	Pan African Clinical Trials Registry
SLCTR	Sri Lanka Clinical Trials Registry
JPRN	Japan Primary Registries Network
IQR	Interquartile Range
AUC	Area Under Characteristics Curve
CDSR	Cochrane Database of Systematic Reviews

Abstract

Systematic reviews are a critical component of evidence-based medicine because of their importance in clinical practice guidelines and decision making in practice, but it can be challenging to keep them up to date because of the rate at which new evidence is produced. Despite the number of systematic reviews published each year, new studies are incorporated into systematic reviews relatively slowly, which can delay the recognition of important safety issues. Guidelines on how and when systematic reviews should be updated appear to have little influence over how systematic reviews are updated in practice, suggesting that there may be benefit in developing tools to help systematic reviewers decide which reviews to update, avoiding redundancy and better targeting efforts where they are most needed.

In this thesis I use software engineering principles to examine inefficiencies across the systematic review ecosystem, with a particular focus on the role of systematic review updates. First, I highlight the importance of improving data interoperability between trial registries and bibliographic databases. In a literature review I show that trial registries are often disconnected from the articles reporting their results. Second, I show that there is no clear evidence that systematic review updates are undertaken earlier following a signal of new evidence. Rather, systematic reviews often add no new evidence and rarely produce a change in conclusion. Third, I propose a new approach to help identify systematic reviews for which an update is warranted by modelling the risk of conclusion change in a curated set of published systematic review updates. To support the creation of a larger database for use in models of this type, I use a rule-based approach to automatically extract relevant information from published reviews and their updates. Finally, I make several recommendations about the need for an interoperable repository of structured systematic review information.

Declaration

I certify that contents of this work have not previously been submitted for a higher degree to any university. I also declare that this submission is my own work and thesis contains no material previously published or written by another person except where due reference is made in the thesis itself. Any contribution made to the research by others, is explicitly acknowledged in the thesis.

(Signed)_____ Date: March 20, 2019

Acknowledgements

Undertaking PhD has been a truly life-changing experience for me, and it would not have been possible without the support, motivation and guidance that I received from a bunch of people.

Special profound gratitude goes to my primary supervisor, Dr Adam Dunn, for his empathic guidance, kindness and support. Without his critical insight and encouragement, I may have never been able to accomplish this degree. He always inspired me by his incisive intellect and wisdom, and I got an excellent opportunity of learning critical, analytical and research writing skills under his supervision. His guidance and timely feedback helped me throughout my research and writing this thesis. I could not have imagined having a best supervisor for my PhD.

I am also hugely appreciative to Dr Didi Surian, especially for his generous support in performing statistical analyses and visualisations, and for being so dedicated to his role as my co-supervisor.

My sincere appreciation also goes to Dr Xujuan Zhou, who introduced me to Dr Adam Dunn. Without her support it was not possible to find a brilliant supervisor.

I gratefully acknowledge the support of Macquarie University for providing International Macquarie Research Excellence scholarship and Centre for Health Informatics for providing the opportunity to carry out this research.

Finally, but by no means least, I am exceedingly grateful for continuous support of my whole family, especially my mother for her prayers and encouragement to follow my dreams. And to my husband Dr Sajjad Akbar, for being supportive and to motivate me during this long journey of doctoral studies. These are the most important people in my life, and I dedicate this thesis to them.

Research outputs

Articles included in this thesis:

- **R Bashir**, D Surian, AG Dunn (2019) The risk of conclusion change in systematic review updates can be estimated by learning from a database of published examples, Journal of Clinical Epidemiology, 110:42-49, doi:10.1016/j.jclinepi.2019.02.015.
- **R Bashir**, AG Dunn (2019) Software engineering principles address current problems in the systematic review ecosystem, Journal of Clinical Epidemiology, 109:136-141, doi:10.1016/j.jclinepi.2018.12.014.
- **R Bashir**, D Surian, AG Dunn (2018) Time-to-update of systematic reviews relative to the availability of new evidence, Systematic Reviews, 7(1):195, doi:10.1186/s13643-018-0856-9.
- **R Bashir**, FT Bourgeois, AG Dunn (2017) A systematic review of the processes used to link clinical trial registrations to their published results, Systematic Reviews, 6(1):123, doi:10.1186/s13643-017-0518-3.
- **R Bashir**, AG Dunn (2016) Systematic review protocol assessing the processes for linking clinical trial registries and their published results, BMJ Open, 6(10):e013048, doi:10.1136/bmjopen-2016-013048.

Conference abstracts related to this thesis:

- **R Bashir**, D Surian, AG Dunn (2018) An empirically defined decision tree to predict systematic reviews at risk of change in conclusion, 25th Cochrane Colloquium, September 16-18, 2018, Edinburgh, Scotland.
- **R Bashir**, AG Dunn (2017) Do systematic review updates target questions where evidence accumulates faster? Evidence Live Conference, June 22-23, 2017, Oxford, UK.

Other relevant articles not included in this thesis:

- P Martin, D Surian, **R Bashir**, FT Bourgeois, AG Dunn (2019) Trial2rev: Combining machine learning and crowd-sourcing to create a shared space for updating systematic reviews, JAMIA Open, ooy062, doi:10.1093/jamiaopen/ooy062.
- D Surian, AG Dunn, L Orenstein, **R Bashir**, E Coiera, FT Bourgeois (2018) A shared latent space matrix factorisation method for recommending new trial evidence for systematic review updates, Journal of Biomedical Informatics, 79:32-40, doi:10.1016/j.jbi.2018.01.008.

Chapter 1: Introduction

1.1 Background

Systematic reviews aim to identify, evaluate and then synthesise the findings of all relevant studies to make the available evidence accessible to decision makers [1]. They are the standard way to evaluate the benefits and harms of clinical interventions. Systematic review evidence has become increasingly important as it provides a basis for healthcare policies, clinical practice guidelines and a comprehensive synthesis for clinical decision making and patient decision aids [2-4]. They also help to identify and mitigate publication and reporting biases [5-8].

An inability to maintain an up-to-date synthesis of clinical evidence can have a negative effect on human health by delaying necessary changes to policy and practice [9]. Ideally, systematic reviews should be updated as soon as new evidence is available to ensure that clinical practice has access to the most up-to-date evidence, as well as to reveal any emerging safety issues that are identified only after synthesising evidence from across relevant trials [10]. However, the rate at which the clinical evidence is produced [11, 12], and the time and resources needed to produce systematic reviews [4] both hinder the ability to keep systematic reviews current.

Systematic reviews are updated for a range of reasons but primarily to consider the impact that new evidence should have on policy and practice. Studies examining the update timing of systematic reviews have revealed that one third of systematic reviews are updated within two years and the median time to update is more than five years [7-11, 13]. However, a substantial proportion of systematic review updates are redundant, misleading, or do not target the clinical questions where evidence accumulates faster [14, 15]. Given the complexities involved in quickly updating systematic reviews, the policies and guidelines about how and when systematic reviews should be updated also vary across different organisations [2]. For example, the Cochrane Collaboration's recommendations about systematic review updates have considered updating reviews whenever new evidence is made available, updating reviews every two years, and updating based on priority and need [16-18].

Many of the existing methods and tools proposed to help decide when a systematic review update is needed are based on availability of new evidence or the likelihood of conclusion change [19-21].

Historically, decisions about updating systematic reviews have been made at the individual systematic review level but with a growing recognition of problems related to redundancy and the inability to keep up with the production of new evidence, new approaches might instead aim to allocate resources to updating reviews where the risk of conclusion change is greatest [22].

Machine learning and information retrieval technologies have the potential to improve the efficiency of systematic review updates, and methods and tools for automating the underlying processes of systematic reviews—searching, screening, information extraction, and synthesis—have already been proposed or developed [23-30]. However, there are relatively few data-driven tools available for deciding when an update may be warranted, and to date have been restrictive in terms of where they can be used [31].

Enabling smarter approaches to systematic review updates will also require improving how metadata describing and linking all sources of clinical evidence are made available. To improve the interoperability and transparency of clinical evidence, Sim et al. [32] proposed the structured and computable reporting of trial results, but progress in the space has taken many years [33]. To check the methodological quality of reporting trials and ensure the transparency, The Consolidated Standards of Reporting Trials (CONSORT) Statement [34] was introduced as a standard, and recently revised [35]. Zarin et al. [36] suggested the use of ClinicalTrials.gov as a central location that links each trial via unique registry identifier to its publication. Despite these efforts, systematic reviews are typically not reported in ways that make it easy to link with trial registries and bibliographic databases. This is due to lack of standardisation and interoperability between trial registries, bibliographic databases and systematic reviews to enable cross-study analyses [37].

1.2 Thesis aim and objectives

The overarching aim of my thesis was to examine and address inefficiencies in the systematic review ecosystem, taking a software engineering perspective to identify the gaps and opportunities for improvement. To achieve this aim, my research was focused on two objectives.

- Improving the completeness of evidence synthesis in systematic reviews by linking all sources of information related to a clinical trial.

- Improving the efficiency of systematic review updates by making better decisions about when to update systematic reviews.

1.3 Thesis structure

This is a thesis by publication and is structured around a set of ten research outputs. These include five published research articles, two conference presentations, one manuscript in preparation, and two additional journal articles to which I contributed but do not reproduce here as part of the thesis.

In Chapter 1, I characterise the current issues in the systematic review ecosystem from the perspective of software engineering, looking specifically at completeness and efficiency of evidence synthesis. In Chapter 2, I highlight the importance of links between trial registries and published results in completeness of trial reporting; and present a systematic review of the set of studies that examine links between trial registries and published trial reports. In Chapter 3, I present an analysis of the update timing of systematic reviews relative to the availability of new evidence, testing whether the availability of new trial evidence is associated with faster updates. In Chapter 4, I propose a possible solution for this problem, where I develop models to learn which systematic review characteristics might be used to quickly or automatically estimate the risk of conclusion change in systematic review updates. Chapter 5 is a discussion and summary of the current problems affecting the efficiency of systematic review practice. I use quality of service attributes from software engineering as a lens to examine the entire systematic review ecosystem and discuss where I think future efforts could be made to address the current critical challenges associated with systematic reviews. In Chapter 5, I also provide an overarching discussion of the contributions of the thesis and implications of my research for systematic review practice and further research in the area.

Chapter 2: Interoperability between trial registries and published results

2.1 Chapter background

This chapter is based on a protocol and systematic review of processes used to link clinical trial registrations to their published results. The aim of this systematic review was to quantify the processes used by studies to evaluate the completeness and accuracy of trial reporting and to examine the use and utility of automatic linkage over time. To do this, I sought to identify all studies examining a cohort of clinical trials to identify links from clinical trial registries to bibliographic databases and from bibliographic databases to clinical trial registries. The protocol and systematic review address the first objective of thesis, but also lay the foundation for the other objectives.

1. **R Bashir**, AG Dunn (2016) Systematic review protocol assessing the processes for linking clinical trial registries and their published results, BMJ Open, 6(10): e013048, doi:10.1136/bmjopen-2016-013048.
2. **R Bashir**, FT Bourgeois, AG Dunn (2017) A systematic review of the processes used to link clinical trial registrations to their published results, Systematic Reviews, 6(1):123, doi:10.1186/s13643-017-0518-3.

Author contributions: For each of the above manuscripts, I developed the search strategies, designed the protocol, performed the data collection and screening, drafted and revised the manuscripts. Adam Dunn contributed in each of the above steps and critically revised the manuscript. Florence Bourgeois critically revised the systematic review and approved the final version of the manuscript.

2.2 Systematic review protocol assessing the processes for linking clinical trial registries and their published results

Abstract

Introduction: Clinical trial registries are an important source of information for tracking clinical trials from their inception through to their reporting, and have been used to measure publication bias and outcome reporting bias. Our aim is to survey and quantify the processes that have been used to identify links between clinical trial registries and published trial reports in studies that rely on these links to evaluate the completeness and accuracy of trial reporting.

Methods and analysis: We will identify studies that describe a process for identifying the links between a trial registry included in the WHO International Clinical Trial Registry Platform and published trial results, and use those links to evaluate the completeness and accuracy of trial reporting. Information extracted from the studies will include the purpose and application domain of the study, registries used or searched, processes by which the links were identified, the study period, and proportions for which links were found. We will summarise what is known about the number and availability of links between clinical trial registries and published results, and examine how automatic linking, inference, and inquiry processes have been used to identify links since the introduction of trial registries.

Ethics and dissemination: The systematic review is focused on the analysis of secondary data and does not require ethics approval. The results of the systematic review will be used to inform standard processes used to identify links to and from clinical trial registries in studies that evaluate the completeness and accuracy of clinical trial reports, as well as systematic reviews. Our findings will be disseminated by publishing the systematic review in a peer-reviewed journal, and by engaging with stakeholders from clinical trial registries and bibliographic databases. These include investigators, funders and sponsors of trials, and authors of journal articles.

2.2.1 Strengths and limitations of this study

- This systematic review will quantify the processes used to link clinical trial registries to clinical trial results and determine how these may have changed since the introduction of clinical trial registries.
- The processes used to link clinical trial registries to published reports of clinical trials vary across studies that rely on those links to evaluate the accuracy and completeness of trial reports, and this systematic review will quantify these differences to inform the way this is performed in the future.
- By producing a baseline measurement of the availability of automatic links and the number of other links that must be identified through inquiry or inference, the systematic review will help determine the potential value of using clinical trial registries to augment current methods used to identify trials for systematic reviews of clinical interventions.
- Because studies linking clinical trial registry data to published results are designed for a range of different purposes, the processes used to identify links are not always reported completely, making information extraction difficult.

2.2.2 Introduction

Clinical trial registries were designed to provide information to researchers, clinicians, and the public about trials that are underway or for which the results have not been reported [38, 39]. Since their introduction, their use has increased substantially following changes in requirements for journal publication and changes to the law in several countries [40-42], and a number of studies have examined publication bias [43-54], and outcome reporting bias [55-64], using one or more of the registries.

Studies that use clinical trial registries to examine the completeness and accuracy of clinical trial reporting rely on being able to establish links between registries and reports of clinical trials. A proportion of those links can be accessed automatically [65, 66], but the remainder must be determined by inference or inquiry. The manner in which these processes are used vary from study to study and are known to be time consuming [43, 60, 64]. It is not yet known whether differences in the way links are established by these processes have influenced the results of studies examining

publication bias or outcome reporting bias. There is a current need to survey the studies that have used these processes to identify links between clinical trial registries and their published results.

Clinical trial registries are sometimes used to identify trials for inclusion in systematic reviews [67-69]. Systematic reviews benefit from clinical trial registries not only because they can be used to quantify reporting bias for an intervention or condition, but may also be used to assist in scheduling updates [2, 31, 70, 71], could be used as an external corpus in machine learning methods that automate or assist in searching and screening methods [28], and in some cases as a source of trial results that have not been published in peer-reviewed literature [68, 69]. By understanding the processes that have been used to establish links between trial registries and published trial reports, we may be able to provide guidance on how each of these processes can be used to identify a complete set of trials, supporting new methods that use clinical trial registries in systematic reviews.

The objective of this systematic review is to quantify the processes that have been used to link clinical trial registrations to their published results in studies that examined the completeness and accuracy of clinical trial reporting.

2.2.3 Methods

Inclusion and exclusion criteria

We will include all English-language studies that use one or more of the clinical trial registries included in the World Health Organization International Clinical Trial Platform (WHO ICTRP) [72], to compare what was registered with what was published, determine the proportion of published trial reports that have been registered or the proportion of registered trials that have been published. Studies will be excluded if they do not report the number of clinical trials for which they identified links or if the study is describing a trial or reviewing clinical evidence.

Search strategy

Relevant articles will be identified by searching PubMed and EMBASE for studies that meet the inclusion criteria. These databases were selected because they are known to have good coverage of

clinical research [73, 74], and other databases typically used in systematic reviews of clinical evidence (such as the Cochrane Central Register of Controlled Trials) were irrelevant to the topic of the review. We designed the search strategy with the support of a medical research librarian, and it was designed to balance the number of articles returned by the broad terms covering the clinical trial registries by constraining the search using terms that were common to the set of relevant studies. We considered a set of 50 articles we knew met the inclusion criteria and used their titles, abstracts, and keywords to define a search strategy that returned all 50 articles without dramatically increasing the number of articles that needed to be screened (Appendix, Table 2.1). We will additionally hand-search the reference lists of all included studies to identify any other articles that may have been missed by our searches. The complete search strategies for both databases are included in Table 2.2 and Table 2.3 in the Appendix.

Two reviewers will evaluate the articles returned by the searches against the inclusion and exclusion criteria. Duplicate studies will be removed by automatically comparing digital object identifiers (DOIs) across the search results where possible, and by manually evaluating titles, authors, and journal names for the remainder. In each of the two phases of screening for eligibility (title/abstract and then full text review), disagreements about inclusion will be resolved by a third author and by discussion, as needed.

Data extraction process

Data from studies will be extracted independently by both reviewers and then compared, reporting the level of agreement for each information element. The information to be extracted includes: (a) the number of trial registry entries examined or identified in the study; (b) the number of published trial reports examined or identified in the study (c) the trial registry or registries used; (d) the purpose of the study (such as measuring publication bias, outcome reporting bias, or the number of published trials that were registered); (e) the application domain; (f) the processes used to identify the links; and (g) the proportions of the links found for each method if available.

In relation to the method for identifying the links, we categorise links as one of three types — automatic, inferred, or inquired. Automatic links are those for which the unique identifier from the

trial registry entry is used to identify links to published results without requiring any inference or manual work. Inferred links are those for which the investigators used the characteristics of the trials to search and reconcile links to or from published trial reports. Inquired links are those confirmed by contacting trial investigators or authors to identify the location of published results.

Data synthesis

Using the information extracted from the articles, we will pool the overall proportions of trials for which links were identified. Because they represent different types of links between trial registries and published reports, studies that start from a cohort of trial registry entries and identify published results will be pooled separately from studies that start from a cohort of published trial reports and identify trial registry entries.

Heterogeneity in the overall number and proportion of links identified in these studies is expected to come partially from differences in the processes being used to identify links, the period in which the trials were completed and published (reflecting temporal changes in the policy and practice of trial registration and reporting), and the specific application domains (some conditions or interventions may be more likely to have registered trials published or published trials registered). To account for these differences in the overall pooled estimates, we will estimate the contributions of each of the three categories of linking processes to the overall estimates wherever the information is available. The result will include estimates of the proportions of links that can be automatically captured, the larger proportion that can be reliably identified when investigators search for and infer links, and the larger proportion that can be identified when investigators contact trial investigators for more information.

We are also interested in examining whether the processes for identifying links between registry entries and trial results have changed over time. To measure the differences over time, we will examine the trend in the proportion of links identified overall—as well as using each of the three categories of linking processes—by applying linear regression relative to the mid-points of the data collection periods specified in each of the studies. All statistical analyses will be performed using SPSS statistical software version 24.0 (IBM, Armonk, NY).

2.2.4 Discussion and dissemination

To our knowledge**,** this is first systematic review surveying the processes used to identify links between clinical trial registries and published clinical trial results. By aggregating the results of many trials in the area to estimate the proportions of links that can be identified through automatic linking, inference, and inquiry, this systematic review is expected to advance the field in several ways. First, the systematic review will be used to determine whether differences in the processes for identifying links between registry entries and trial reports can partially explain differences in the results of existing studies of publication bias and outcome reporting bias, which in turn may be used to help standardise the way these studies are undertaken in the future. Second, by determining the proportions of trial registry entries that can be automatically linked to their results, we can help guide new systematic review technologies that rely on links to improve methods used in the identification of trials.

One limitation of the systematic review process is the exclusion of studies that are not published in English, which may mean that we miss some articles describing registries based in non-English speaking countries. Because a substantial proportion of the studies that will be included are designed for purposes other than simply identifying the links between registries and published results, the description of the processes used to link the two may be limited, and this may limit our ability to determine the proportions of links captured automatically, by inference, or by inquiry.

Our findings will be reported on the basis of guidelines from Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement[1]. The results of this review will be submitted for publication in a peer-reviewed medical journal. Other forms of dissemination will include direct engagement with clinical trial registry developers.

[1] Available online: doi: 10.1136/bmjopen-2016-013048

2.3 A systematic review of the processes used to link clinical trial registrations to their published results

Abstract

Background: Studies measuring the completeness and consistency of trial registration and reporting rely on linking registries with bibliographic databases. In this systematic review, we quantified the processes used to identify these links.

Methods: PubMed and Embase databases were searched from inception to May 2016 for studies linking trial registries with bibliographic databases. The processes used to establish these links were categorised as automatic when the registration identifier was available in the bibliographic database or publication, or manual when linkage required inference or contacting of trial investigators. The number of links identified by each process was extracted where available. Linear regression was used to determine whether the proportions of links available via automatic processes had increased over time.

Results: In 43 studies that examined cohorts of registry entries, 24 used automatic and manual processes to find articles; 3 only automatic; and 11 only manual (5 did not specify). Twelve studies reported results for both manual and automatic processes and showed that a median of 23% (range from 13 to 42%) included automatic links to articles, while 17% (range from 5 to 42%) of registry entries required manual processes to find articles. There was no evidence that the proportion of registry entries with automatic links had increased ($R^2 = 0.02$, $p = 0.36$). In 39 studies that examined cohorts of articles, 21 used automatic and manual processes; 9 only automatic; and 2 only manual (7 did not specify). Sixteen studies reported numbers for automatic and manual processes and indicated that a median of 49% (range from 8 to 97%) of articles had automatic links to registry entries, and 10% (range from 0 to 28%) required manual processes to find registry entries. There was no evidence that the proportion of articles with automatic links to registry entries had increased ($R^2 = 0.01$, $p = 0.73$).

Conclusions: The linkage of trial registries to their corresponding publications continues to require extensive manual processes. We did not find that the use of automatic linkage has increased over time. Further investigation is needed to inform approaches that will ensure publications are properly linked to trial registrations, thus enabling efficient monitoring of trial reporting.

Keywords: Clinical trials as topic; Trial registration; Publication bias; Reporting bias; Systematic reviews as topic.

2.3.1 Background

Clinical trial registries were established to improve transparency and completeness in the reporting of clinical trials [38, 39, 75-78]. Since they were established, a number of policies have been implemented to encourage or mandate their use, and this has led to substantial growth in the number of trials that have been registered [79-83]. For example, since 2005, prospective trial registration has been a condition for publication in member journals of the International Committee of Medical Journal Editors (ICMJE) [75, 84]. The European Union and USA have also passed legislation requiring prospective registration of clinical trials involving drugs or devices [85].

Clinical trial registries provide the ability to measure biases in the reporting of clinical trials that arise due to non-publication, delayed publication, or incomplete publication of results [86]. Studies examining these issues rely on the ability to establish a link between the original trial registration and subsequent published article. These links can be established in an automatic fashion if the publication abstract or metadata includes the registry identifier [66, 87]. However, if this identifier is not included by trial investigators or added by journals, manual processes are needed to create these links, either through searches and inference or through direct contact with investigators. Despite the number of studies that have examined reporting biases by linking trial registry entries and publications, the processes for linking are variable and poorly described.

Clinical trial registries are a critical source of information for systematic reviewers who use these registries to augment bibliographic database searches when compiling relevant evidence from clinical trials [67-69]. Systematic reviewers may seek to identify links from published trial reports to their

respective registry entries to fill in gaps for information that is missing or incompletely reported. They may also independently search trial registries to identify additional trials [88, 89] and follow links from the registry to reports of the trials.

Our aim was to quantify the processes that have been used to link clinical trial registries with published results and to examine the use and utility of automatic linkage over time. To do this, we conducted a systematic review of all studies examining a cohort of clinical trials to identify links from clinical trial registries to bibliographic databases and from bibliographic databases to clinical trial registries, following a published systematic review protocol [90].

2.3.2 Methods

Inclusion criteria and search strategy

We identified all primary studies that examined links between any of the registries in the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) and published articles in bibliographic databases. Studies were excluded if there was no English-language version, if they did not unambiguously report the total number of clinical trials for which links were identified, if they were reporting on a specific clinical trial, or if the identification of links was not the primary focus of the study. Studies that did not unambiguously report the processes used to identify links were included in the review but excluded from the analyses.

PubMed and Embase were searched from inception to May 27, 2016, [73, 74]. The search strategy was developed with the assistance of a medical research librarian with details described in a previously published protocol [90]. The full version of the search strategy for both databases is provided in the Appendix. This strategy included searching of all study references to identify any other relevant articles not captured in the original search. Duplicate studies were removed using digital object identifiers and manually comparing titles, authors, publication dates, and article metadata. All identified studies were screened individually by two reviewers for inclusion, and disagreement was resolved through discussion.

Data extraction

Two reviewers evaluated all the included studies to extract relevant information from the studies and resolved ambiguities by discussion. For each study, the following information was extracted: (a) number of reported clinical trials, (b) number of published articles, (c) trial registries used, (d) the study purpose (such as publication bias, outcome reporting bias, or assessing the publication rate of registered trials), (e) application domain (any constraints such as journal lists, conditions, or specialties), (f) processes for identifying links, and (g) proportions of links found using each process.

The processes used to identify links were categorised as one of three types: automatic, inferred, and inquired. Automatic links were defined by any process that used the unique registry identifier to reconcile the link into or from a bibliographic database without the need for a search or inquiry. This included searching PubMed for registry identifiers to find published articles in cohorts of registry entries or using identifiers in the metadata, abstract, or full text of published articles to find registry entries in cohorts of published articles. Inferred links were defined by any manual processes in which investigators searched for matches across databases using characteristics of the trial such as the names of the investigators, titles, and acronyms associated with the trial, location, sample size, or the population, intervention, or measurable outcome information to find a match in a bibliographic database or trial registry. Inquired links were defined by any manual process where the study authors attempted to contact the investigators or authors of a trial to request or confirm the presence or absence of a registry entry or a published article for each included trial.

Data synthesis and analysis

We examined the proportions of links that were identified through each of these three processes. Using the publication year of the studies that used both automatic and manual processes, we applied linear regression to determine whether the utility of the automatic processes—the proportion that were found automatically compared to the proportion that required manual processes—had increased over time. We did not undertake a pooled analysis of the utility of automatic links because many studies did not specify proportions found by each process used and because of the heterogeneity in the study designs.

All statistical analyses were conducted using SPSS statistical software version 24.0 (IBM, Armonk, NY).

The protocol for this systematic review was published in 2016 [90]. We did not register the systematic review with PROSPERO because it does not directly examine at least one outcome of direct patient or clinical relevance. This systematic review is reported in accordance with the reporting guidance provided in the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement[2].

2.3.3 Results

The initial search returned 11,986 results (after non-English articles were excluded), which produced 9486 articles after de-duplication (Figure 2.3.1) [91]. A set of 348 studies remained after screening titles and abstracts, and of these, 81 studies were included in the review. One study considered links from both cohorts of registry entries and published articles [66, 92], for a total of 82 analyses. Excluded studies included conference abstracts, studies for which information about the proportions of registry entries or published articles that were identified was ambiguous [93-95] and studies that considered reporting biases but could not be included because the linking was atypical or there was no linking performed [96-99]. Some studies were excluded because they did not measure links between trial registries and bibliographic databases and, instead, considered links to or from other source of clinical trial information. These included links to or from protocols [100-103], conference or meeting abstracts [104-108], internal company documents [67], Food and Drug Administration (FDA) documents or new drug approvals [109-113], or other databases of published articles [114, 115].

Studies identifying published articles from cohorts of registry entries

We identified 43 studies that examined links to published articles from registries, typically with the aim of examining publication bias or outcome reporting bias (Appendix, Table 2.4). The application domains varied by types of studies (e.g., terminated and withdrawn trials [52, 62], trials funded by

[2] Available online: doi: 10.1186/s13643-017-0518-3

specific organisations or from certain countries [53, 116], and by specialty and condition (e.g., paediatric or surgical trials [117, 118]).

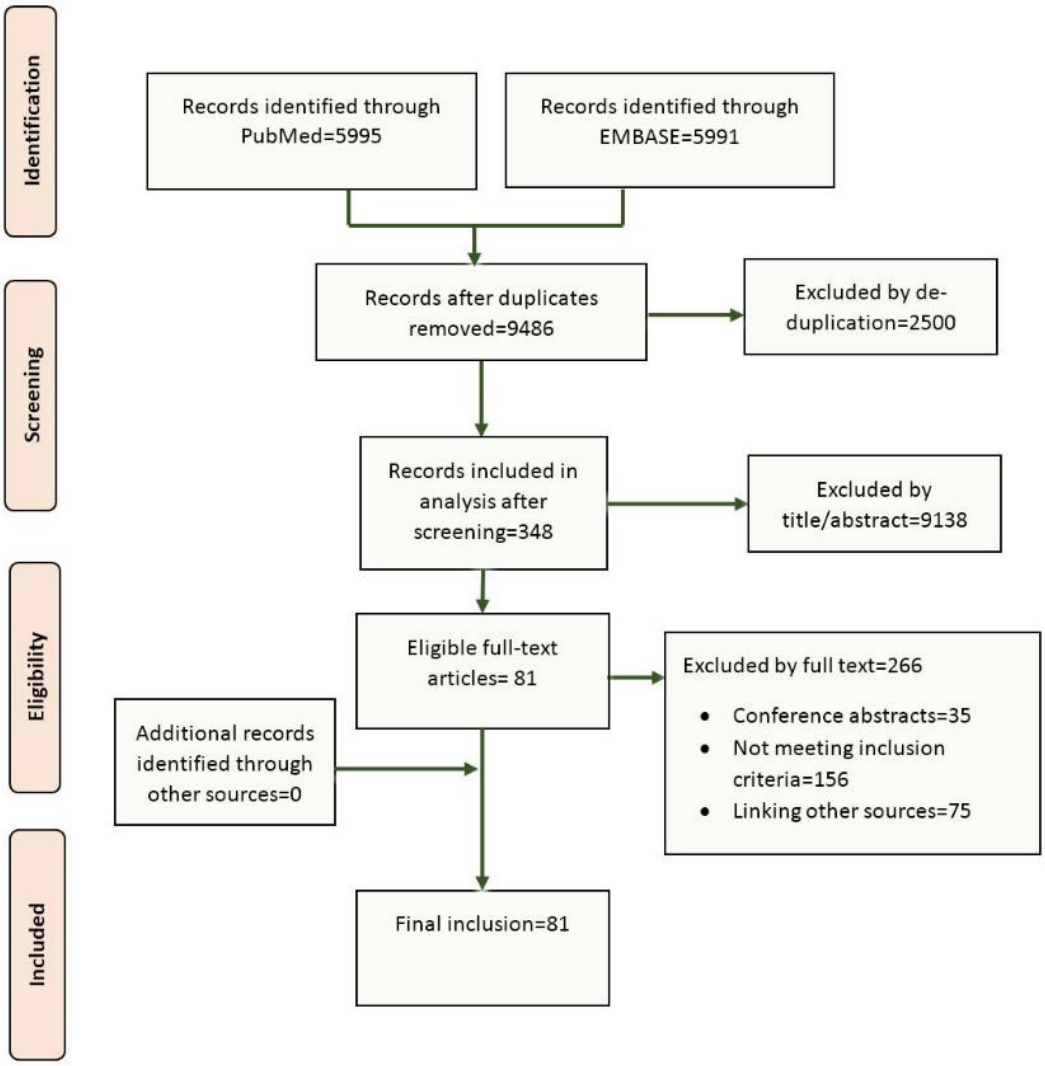

Figure 2.3.1: PRISMA flow diagram of study selection for a search and screening process that resulted in the inclusion of 81 studies.

The most commonly studied registry was ClinicalTrials.gov only (35 studies), followed by some or all the registries of the WHO ICTRP (8 studies). The most commonly examined bibliographic databases were PubMed alone (22 studies), or Embase in combination with PubMed or other bibliographic databases (20 studies). The studies included cohorts of registry entries that ranged in size from 34 to 8907 (median 305) entries. The median proportion of registry entries for which published articles were found was 47%, and these proportions ranged from 4% (2 published articles in a cohort of 46 registry entries) to 76% (47 published articles in a cohort of 62 registry entries).

The processes used to identify links between clinical trial registries and published articles varied across the set of studies (Figure 2.3.2 and 2.3.3). The most common process was to use a combination of automatic and manual processes (24/43, 56%), followed by manual processes only (11/43, 26%), and automatic processes only (3/43, 7%). There were five studies for which the process for identifying published articles was not clear or not provided.

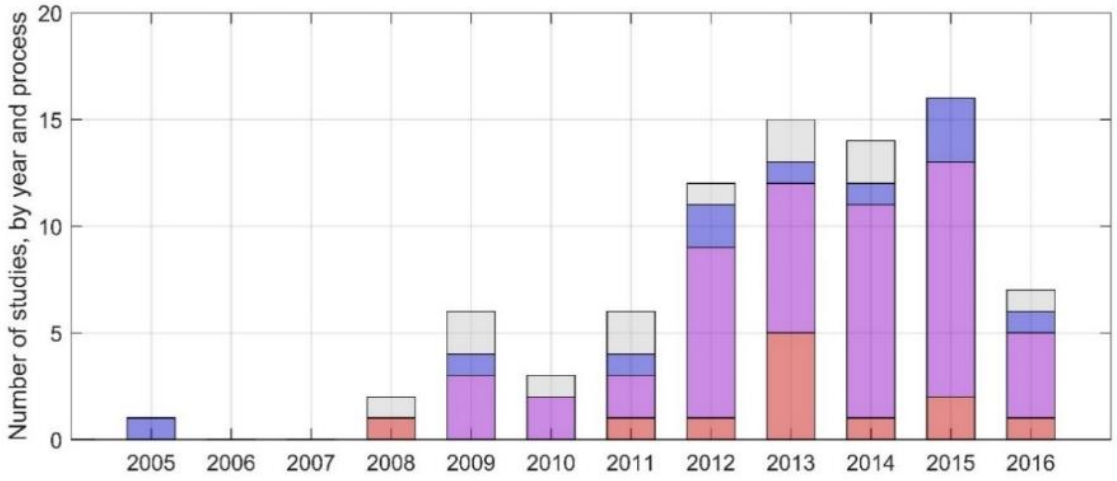

Figure 2.3.2: The processes used to identify links in 81 included studies, including studies that examined automatic links only (red), both automatic and manual processes (purple), manual processes only (blue), and studies that did not report the processes used (grey).

Of the 24 studies that looked for published articles among a cohort of registry entries and used both manual and automatic processes, 12 studies specified the number of published articles identified via each process (Figure 2.4). Among these studies, automatic links were used to identify between 13 and 42% (median 23%) of the published articles, and manual processes were used to find a further 5–42% (median 17%) articles that were not available via automatic links.

We found no evidence of a change in the overall proportion of publications that could be found via automatic links. A linear regression over the 12 studies—using the publication year as the independent variable—indicated no significant trend in the proportion of available links that can be identified by automatic processes ($R^2 = 0.02$, $p = 0.36$, $\beta = 1.28\%$ increase per year).

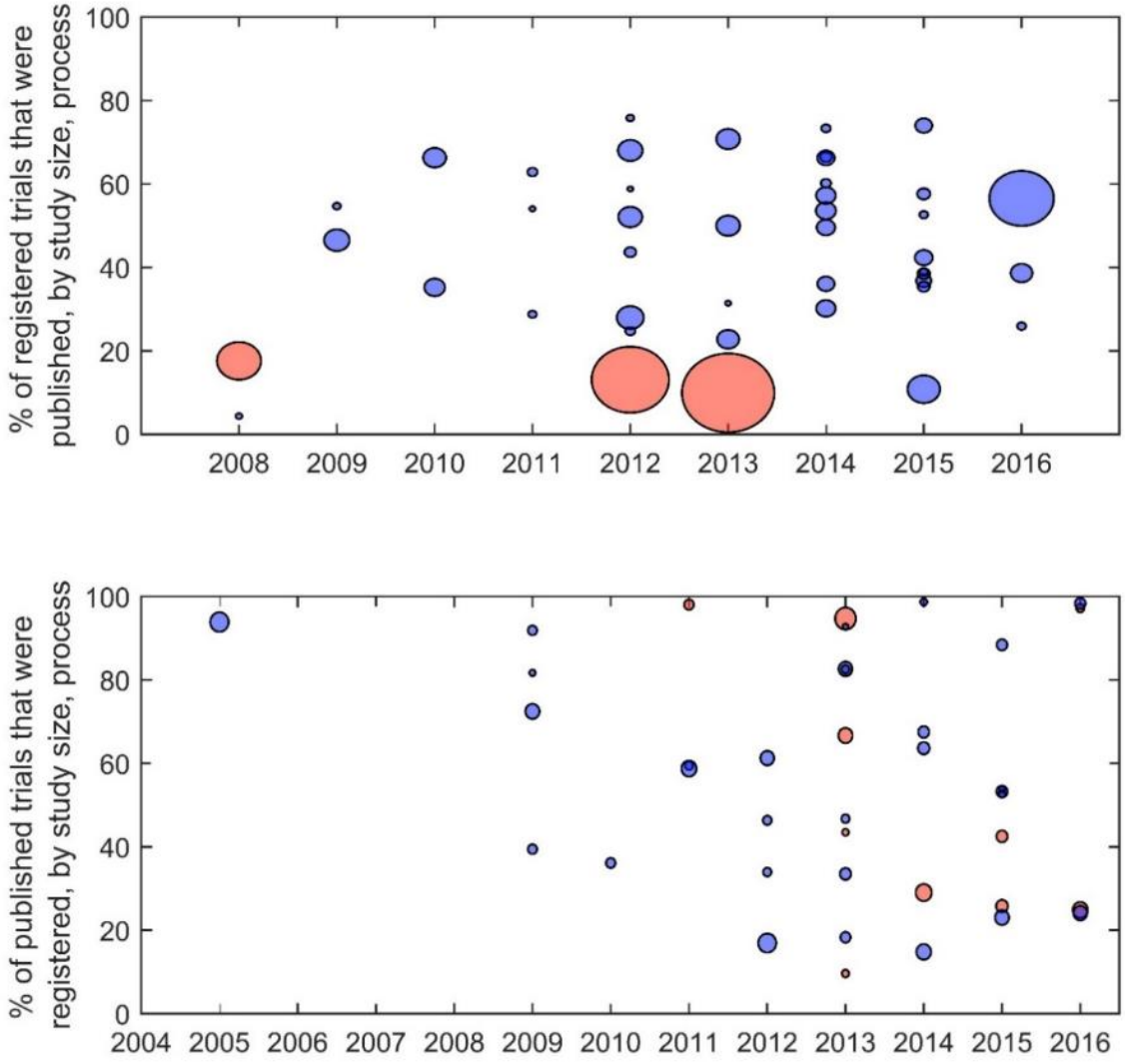

Figure 2.3.3: The proportions of published articles identified in cohorts of registry entries (top, 43 studies, ranging from 34 to 8907 registry entries) and the proportions of registry entries found in cohorts of published articles (bottom, 39 studies, ranging from 54 to 698 articles), with studies that only considered automatic links (red) and all other studies (blue). The circle areas are proportional to the study size.

Studies identifying registry entries from cohorts of publications

There were 39 studies that considered cohorts of publications and identified associated registry entries in one or more of the WHO ICTRP clinical trial registries (Appendix, Table 2.5). These studies included a range of 51–698 (median 181) published articles. These studies also covered a range of application domains, varying by the selection of journal, discipline, or study design [57, 60, 119-123]. The most commonly used bibliographic database was PubMed alone (19 studies), followed by

PubMed in combination with other bibliographic databases (7 studies). To identify registrations, the studies most commonly searched ClinicalTrials.gov in combination with other registries (25 studies), followed by all trial registries included in the WHO ICTRP (9 studies). The median proportion of registry entries that were identified from cohorts of published articles was 54%, ranging from 10% (8 registrations from a cohort of 83 published articles) to 99% (75 registrations from a cohort of 76 published articles).

The processes used to identify links between clinical trial registries and published articles varied across the set of studies (Figures 2.3.2 and 2.3.3). The most common process was to use a combination of automatic and manual processes (21/39, 54%), followed by automatic processes only (9/39, 23%), and manual processes only (2/39, 5%). There were 7 studies for which the processes used to identify registry entries were not clear or not provided.

Of the 21 studies that looked for registry entries among a cohort of published articles and used both manual and automatic processes, 16 reported the number of registry entries found using each process (Figure 2.3.4). Among these studies, automatic links identified between 8 and 97% (median 49%) of registry entries and the manual processes identified between 0 and 28% (median 10%) additional entries.

We found no evidence of a change in the overall proportion of published articles for which registry entries could be found via automatic links. A linear regression over the 16 studies—using the publication year as the independent variable—indicated no significant trend in the proportion of links that can be identified via automatic processes ($R^2 = 0.01$, $p = 0.73$, $\beta = 1.40\%$ increase per year).

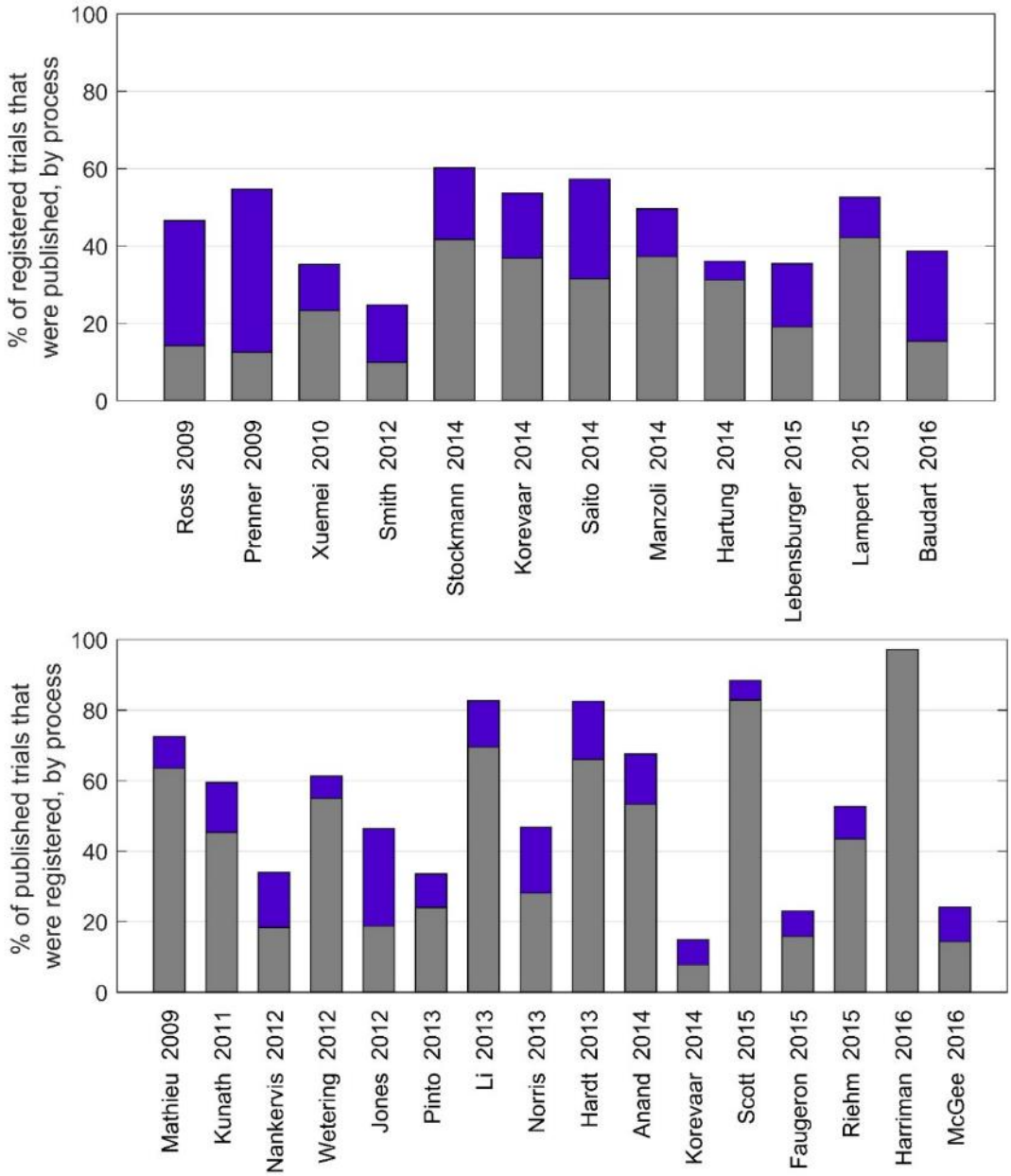

Figure 2.3.4: The proportions of published articles found in cohorts of registry entries (12 studies, top) and the proportions of registry entries found in cohorts of published articles (16 studies, bottom), by automatic links (grey) and manual processes (blue).

2.3.4 Discussion

In this systematic review, we found that investigators use both automatic and manual processes to link registry entries and publications and that automatic links could be used to identify some but not all links between registry entries and published articles. We found no evidence that the utility of automatic processes had increased over time.

To the best of our knowledge, no other systematic review has examined the utility of automatic links between trial registries and bibliographic databases. Previous studies that examined the availability of automatic links provided a broad analysis of automatic links made available through ClinicalTrials.gov and PubMed but did not systematically evaluate the proportion of links that could additionally be resolved using manual processes [65, 66, 87]. Other systematic reviews have examined reporting biases as a topic and included subsets of the studies we included [44, 86], but focused on publication rates and the completeness and consistency of outcome reporting, which we did not evaluate here. Our review adds to this area of research by compiling information about a broader group of studies and synthesising what is known about the utility of automatic links, and the need for supplementing automatic processes with manual processes, in studies that rely on links between trial registries and bibliographic databases.

Implications

Our results indicate that automatic links alone are a useful but not sufficient process for measuring rates of registration and publication or associated biases. Relying on automatic links to draw conclusions about the rate of non-publication will likely over-estimate the rate of non-publication. When aiming to monitor compliance with prospective registration of clinical trials, or monitoring publication practices and patterns, the limits of automatic links should be considered.

In general, the proportion of links identified by automatic processes was lower in studies that started with a cohort of registry entries and aimed to identify published articles, compared to studies that started with a cohort of published articles, and aimed to identify registrations. This may be a consequence of journals that have not yet established standards for registration [65] or have not

implemented standards for incorporating registry identifiers in the information they pass to bibliographic databases.

The results also have implications for systematic reviews. Systematic review technologies for automating or supporting reviewers rarely consider information from clinical trial registries to improve the searching or screening processes [30] or the prioritisation or scheduling of systematic review updates. Because systematic reviews are already time-consuming [12, 13], the need for additional manual effort in the linking of trial registry entries with their published results may have hindered the development of tools based on this linkage. Areas for development include processes where systematic reviewers compare published reports with information in a registry or use trial registries to identify trials not found in bibliographic databases. By removing these barriers, machine-readable information linking all published studies with all registry entries may provide the catalyst for the increased use of registries in the searching, screening, and prioritising of systematic reviews.

Recommendations

We recommend continued pressure to ensure that journals and publishers adhere to standards of reporting that require unique trial identifiers to be specified in the abstract of the article and reported as part of the metadata provided to bibliographic databases. Trial investigators should also be encouraged to update registry entries with links to published results when journals do not provide the information to bibliographic databases. As we move into an era where the structured reporting of clinical trial results and individual participant data become the standard for responsible clinical trial reporting [36], the inability to automatically identify all sources of information about a clinical trial hinders our ability to reuse and synthesise results across trials. Given the number of extra links that could be identified by examining the full text of articles, we also recommend that journals ensure that clinical trial identifiers are included in the abstract or metadata provided to bibliographic databases.

We additionally recommend a standardised method for identifying links between registry entries and published articles that, for the time being, includes manual validation and checking and avoids drawing conclusions based only on automatic links. A standardised method should include details

about what elements of a registry entry should be used to search for published articles and a standard definition for what constitutes published results. Standard reporting for these studies should include the number of registry entries for which searches were performed, the proportion that were identified by automatic links, by inference or by inquiry, and the full details of the dates of trial completion and the length of follow-up. Presenting studies in terms of the time to publication rather than the presence or absence of publication would make a greater proportion of the studies comparable and amenable to meta-analysis.

Limitations

There are three limitations to this review. First, the exclusion of studies for which there was no English language version available meant that we may have missed some studies examining WHO ICTRP registries from countries where English is not the primary language. Second, we included one meta-research article that examined links between articles and registrations of cohort studies. Because cohort studies are registered less often than trials, we could have excluded this study or examined it separately. Third, we used the publication year of the studies as a proxy for estimating changes in the proportions of links identified by each process without considering the period of study that each of the studies covered. This was necessary because a substantial proportion of studies did not report the range and distribution of publication and registration dates in the cohorts they examined, and this may have influenced our analysis of the trends in the utility of the automatic processes.

2.3.5 Conclusions

In this systematic review, we have quantified the use and utility of the processes that are used to link trial registries to bibliographic databases. The results indicate that manual processes are still used extensively and that the gap between what can be identified via automatic processes and what must be identified via manual processes persists. Future improvements in the quality of automatic linking between clinical trial registries and bibliographic databases should come from continued pressure on journals to enforce policies and practices to consistently include registry identifiers in published reports.

Chapter 3: Availability of clinical evidence and systematic review updates

3.1 Chapter background

This chapter is based on work presented as part of one conference abstract and one published article. The conference abstract aimed to determine whether systematic reviews were targeted for updating after relevant trials were published. This abstract was presented at *Evidence Live Conference* in Oxford, UK. The conference abstract was then extended as a journal article to examine the update timing of systematic reviews relative to the availability of new clinical trial evidence and determines whether the availability of new trial evidence was associated with shorter update times. To observe associations between the availability of new evidence and update timing, I compared the update time for systematic reviews with a publication signal of new evidence to systematic reviews without a publication signal of new evidence. This chapter addresses the second objective of the thesis.

1. **R Bashir**, AG Dunn (2017) Do systematic review updates target questions where evidence accumulates faster? Evidence Live Conference, June 22-23, 2017, Oxford, UK.
2. **R Bashir**, D Surian, AG Dunn (2018) Time-to-update of systematic reviews relative to the availability of new evidence, Systematic Reviews, 7(1):195, doi:10.1186/s13643-018-0856-9.

Author contributions: For the above conference abstract and manuscript, I designed the methodology, performed the data collection, undertook the analysis, and drafted and critically revised the abstract and manuscript. Adam Dunn contributed to the methodology and data analysis, and critically revised both the abstract and manuscript. Didi Surian supported with the visualisation and statistical analysis, and critically revised the manuscript.

3.2 Do systematic review updates target questions where evidence accumulates faster?

Objective: There are several methods available for determining if a systematic review needs to be updated, but little is known about whether reviewers prioritise clinical questions with new evidence. Our aim was to determine whether systematic reviews were targeted for updating after relevant trials were published.

Methods: Systematic reviews published in the Cochrane Database of Systematic Reviews in 2010 were selected if they included at least one clinical trial; updated before December 1 2016; performed a new search; and did not change populations, interventions, outcomes, or comparators. Using the updated set of trials to define the most recent evidence base, we retrospectively quantified the accumulation of new evidence between the search dates of the reviews and their updates. Recording trial publication dates (and trial completion dates where registration information was available), the ongoing completeness of a review was determined by the number of participants included in the review as a proportion of the total number of participants available as new evidence was published. We determined whether reviews with a signal of new evidence (≤ 90% completeness within a year of the search date) were updated faster (using time between search dates) than reviews without a signal.

Results: From 773 articles published in 2010, 53 systematic reviews were sampled for analysis. The median update time was 41 months (IQR 35-60). For 55% (29/53) of the reviews, no new trials were added in the update. Within a year of search date, the reviews covered between 72.4% and 100% of the published trial participants. Clinical questions (with ≤ 90% completeness) were not targeted for update (N=12, median 49 months) earlier than those without (N=41, median 40 months); p=0.017 in a Wilcoxon rank sum test. The 14 reviews with complete registration information covered a median of 93.9% (IQR 83.8%-100%) of completed trial participants at the search date, and 85.8% (IQR 73.8%-96.0%) within a year of search.

Conclusion: Updates to Cochrane systematic reviews mostly found no new evidence, and updates were not targeted at questions where new evidence was published. Methods for automatically monitoring registrations and publications may help to prioritise systematic review updates.

3.3 Time-to-update of systematic reviews relative to the availability of new evidence

Abstract

Background: A number of methods for deciding when a systematic review should be updated have been proposed, yet little is known about whether systematic reviews are updated more quickly when new evidence becomes available. Our aim was to examine the timing of systematic review updates relative to the availability of new evidence.

Methods: We performed a retrospective analysis of the update timing of systematic reviews published in the Cochrane Database of Systematic Reviews in 2010 relative to the availability of new trial evidence. We compared the update timing of systematic reviews with and without signals defined by the completion or publication of studies that were included in the updates.

Results: We found 43% (293/682) systematic reviews were updated before June 2017, of which 204 included an updated primary outcome meta-analysis (median update time 35.4 months; IQR 25.5-54.0); 38% (77/204) added new trials and 4% (8/204) reported a change in conclusion. In the 171 systematic reviews with reconcilable trial reporting information, we did not find a clear difference in update timing (p=0.05) between the 15 systematic reviews with a publication signal (median 25.3 months; IQR 15.3-43.5) and the 156 systematic reviews without a publication signal (median 34.4 months; IQR 25.1-52.2). In the 145 systematic reviews with reconcilable trial completion information, we did not find a difference in update timing (p=0.33) between the 15 systematic reviews with a trial completion signal (median 26.0 months; IQR 19.3-49.5) and the 130 systematic reviews without a trial completion signal (median 32.4 months; IQR 24.1 to 46.0).

Conclusion: A minority of 2010 Cochrane reviews were updated before June 2017 to incorporate evidence from new primary studies, and very few updates led to a change in conclusion. We did not find clear evidence that updates were undertaken faster when new evidence was made available. New approaches for finding early signals that a systematic review conclusion is at risk of change may be useful in allocated resources to the updating of systematic reviews.

Keywords: Systematic reviews; Updating systematic reviews; Clinical trial registries; Evidence synthesis

3.3.1 Background

Systematic reviews provide an important source of clinical evidence for informing policy and health decision-making [2, 124-126], but need to be kept up to date to avoid the potential for unnecessary waste or harm in clinical decision-making [127, 128]. However, updating a systematic review is a resource-intensive process and ensuring that they are kept up to date is a challenge. Studies examining the timing of systematic review updates have found that around a third are updated within two years and that the median update time is more than five years [13, 128-131].

Updating a systematic review is not simply a matter of mechanistically repeating the processes used for the previous version but involves consideration of changes in methods, new standards, and the broader context in which clinical decisions are made [17]. Reflecting the complexity of the process for updating systematic reviews, policies and guidelines about how and when systematic reviews should be updated vary from organisation to organisation [2]. For example, the recommendations produced by the Cochrane Collaboration have changed over time to match the availability of resources and adapt to new methods and technologies designed to support the process [16-18].

The current methods and tools available for deciding whether a systematic review needs to be updated are primarily based on estimating the likelihood that new evidence is available or that the conclusions are likely to change for an individual systematic review [19-21]. Given the challenges posed by the increasing rate at which evidence is being produced [11], the focus appears to be shifting away from making decisions about updating individual systematic reviews and towards more pragmatic approaches for prioritising clinical questions most at risk of being an incorrect reflection of current evidence [132, 133]. Prioritisation is particularly important in relation to safety, where the rapid detection of post-approval safety issues could be improved [10].

A 2007 study examined the availability of new and relevant trial evidence after a systematic review was published, looking for signals that a systematic review may be out of date [12]. Other studies have focused on the time between systematic review updates [13, 130]. Our aim was to examine the update

timing of systematic reviews relative to the availability of new trial evidence, and determine whether the availability of new trial evidence was associated with earlier decisions to update.

3.3.2 Methods

The study was a retrospective analysis of the update timing of systematic reviews. We analysed updated systematic reviews published in the Cochrane Database of Systematic Reviews in 2010 and extracted information about the availability of results for the trials that were added in the systematic review updates.

Inclusion criteria

We performed a search on June 1, 2017 and identified all articles published between January 1, 2010 and December 31, 2010 in Cochrane Database of Systematic Reviews using PubMed. Articles from this set were included in the study if they were systematic reviews based on interventional studies and had an update published before June 2017. Articles were excluded if they were editorials, systematic review protocols, or if they were withdrawn.

To be included in the study, the updated systematic review must have included a new search date, indicating that the systematic reviewers performed a search to identify new studies for inclusion. Systematic reviews that only corrected errors in the text, or made minor changes without conducting a new search were excluded from the analysis.

Data extraction

Two investigators (RB and AD) evaluated all systematic reviews and resolved ambiguities in the extraction of information by discussion. This included extracting information available in the systematic review and its update, including publication dates and the final search dates, the set of primary outcomes for which a meta-analysis was performed, the set of trials that were included in primary outcome meta-analyses, and the number of participants from those trials. The primary outcomes were used to reconcile the consistency of the primary outcomes between the systematic reviews and their updates. For those systematic reviews where the primary outcomes were not explicitly mentioned, the first outcome was considered as a primary outcome. We additionally

recorded the results of the primary outcome meta-analyses, typically a relative risk or an odds ratio with its 95% confidence interval.

We examined the set of trials included in the first primary outcome meta-analysis that was consistent in the systematic review and its update and that had added new trials included in the update. To do this, we compared the first primary outcome meta-analyses of both systematic reviews and their updates. After identifying the first consistent primary outcome meta-analysis from the original and updated systematic reviews (systematic reviews with inconsistent meta-analyses were excluded), we compared the set of trials included in the original to the set of trials included in the update and considered any trials that were not included in the original systematic review as newly added trials. For each of the included trials we used references to published articles and trial registry information to reconcile when the study results were first published in full, when the study was completed, and the number of participants in the study. To identify registrations for the trials that were not provided in the systematic review or the published articles reporting the trials, we searched ClinicalTrials.gov and the International Clinical Trial Registry Portal (ICTRP) using a standard process [90]. The process included checking for metadata links available in PubMed, then searching for the intervention (including its synonyms), trial acronyms, and reconciling information about the investigators and authors, the study design, and the number of participants.

Where this information was not available, we attempted to estimate the completion date using information about recruitment and follow-up presented in the published results. This information was then used to define the accumulation of new evidence relevant to the meta-analysis prior to it being updated.

Outcome measures

We defined the update time of a systematic review by the number of months between the publication date of the systematic review and the search date of the subsequent update. We defined the completeness of a systematic review as the proportion of study participants from relevant and published studies covered by the systematic review. This means that the completeness is a value that decreases over time as new and relevant evidence is publicly reported and the systematic review

covers a decreasing proportion of all of the relevant studies. Our completeness of a systematic review is the inverse of the participant ratio used by Takwoingi et al. [31].

We then defined two kinds of update signals based on the retrospective analysis of the registrations and publications of the trials included in the systematic review updates. A publication signal was defined by the publication of a study that was included in the update of primary outcome meta-analysis within a year of the systematic review being published (i.e. completeness by published article is less than 100% within a year). A trial completion signal was defined by the completion date of a study that was included in the update of primary outcome meta-analysis within a year of the systematic review publication date (i.e. completeness by trial completion is less than 100% within a year). In another study [12], authors also used a signal to examine new evidence but they used change in statistical significance of results or new information about efficacy and safety to define signal.

Analysis

We compared the update time for systematic reviews across a number of groups to establish any baseline differences across systematic reviews that varied by type or conclusion. These included a comparison between systematic reviews that included meta-analyses and systematic reviews that did not include a meta-analysis. We also compared systematic reviews that added new trials to at least one primary outcome meta-analysis in an update to systematic reviews that did not add any new trials to primary outcome meta-analyses. Finally, we compared the update timing of systematic reviews in which the conclusions changed to the systematic reviews where conclusions did not change.

To examine associations between the availability of new evidence and update timing, we compared the update time for systematic reviews with a publication signal of new evidence in the first twelve months after a systematic review was published to systematic reviews without a publication signal of new evidence in the first twelve months. To compare systematic reviews that had an early signal of new evidence to those that did not have an early signal of new evidence, we used a Wilcoxon rank sum test to compare the time to update across the two groups, and considered a p-value of less than 0.05 to be significant. We then repeated the same analysis using the trial completion signal rather than the publication signal. We additionally performed a sensitivity analysis by varying the time threshold

used to define an update signal. All statistical analyses were performed using Python version 2.7 and the Lifelines library was used to visualise update timing comparisons.

3.3.3 Results

There were 773 articles published in the Cochrane Database of Systematic Reviews in 2010, of which 682 systematic reviews were included in the study. We excluded 91 articles from the analysis, including 37 editorials and protocols of systematic reviews, 53 withdrawn systematic reviews, and 1 systematic review that was published twice in 2010.

Characteristics of systematic review updates

In the remaining set of systematic reviews, we found 43.0% (293 of 682) had an update that included a new search date and was published before June 2017 (Figure 3.3.1). Of the 293 systematic reviews that were updated, 204 (69.6%) included a primary outcome meta-analysis.

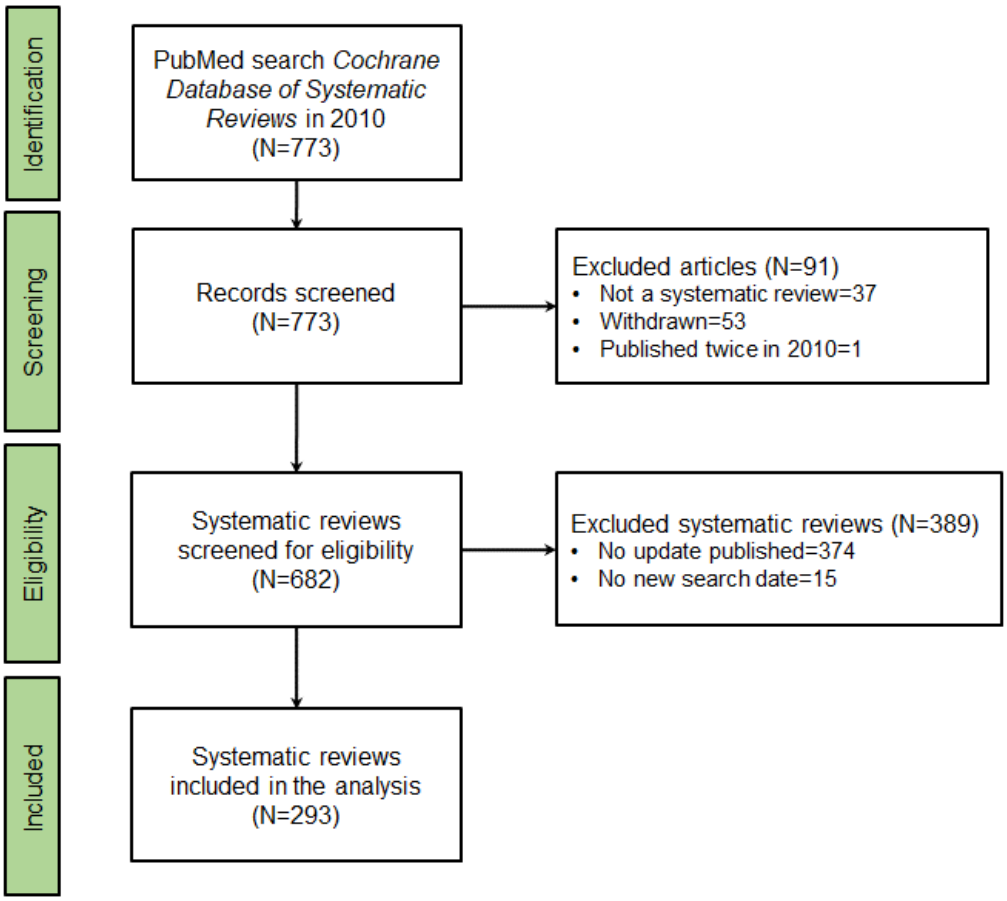

Figure 3.3.1: From 773 articles published in 2010 in the Cochrane Database of Systematic Reviews, 293 were included in the analysis.

This included 60 systematic reviews that had new trials added to a primary outcome meta-analysis in the update, 111 that included no new trials in a primary outcome meta-analysis, 17 that added new non-English trials, and 16 that had outcomes or populations that were substantially different from the systematic review that was updated (Figure 3.3.2; [3]Supplementary File 1). In 8 of the 204 updated systematic reviews with primary outcome meta-analyses, we identified a change in conclusion.

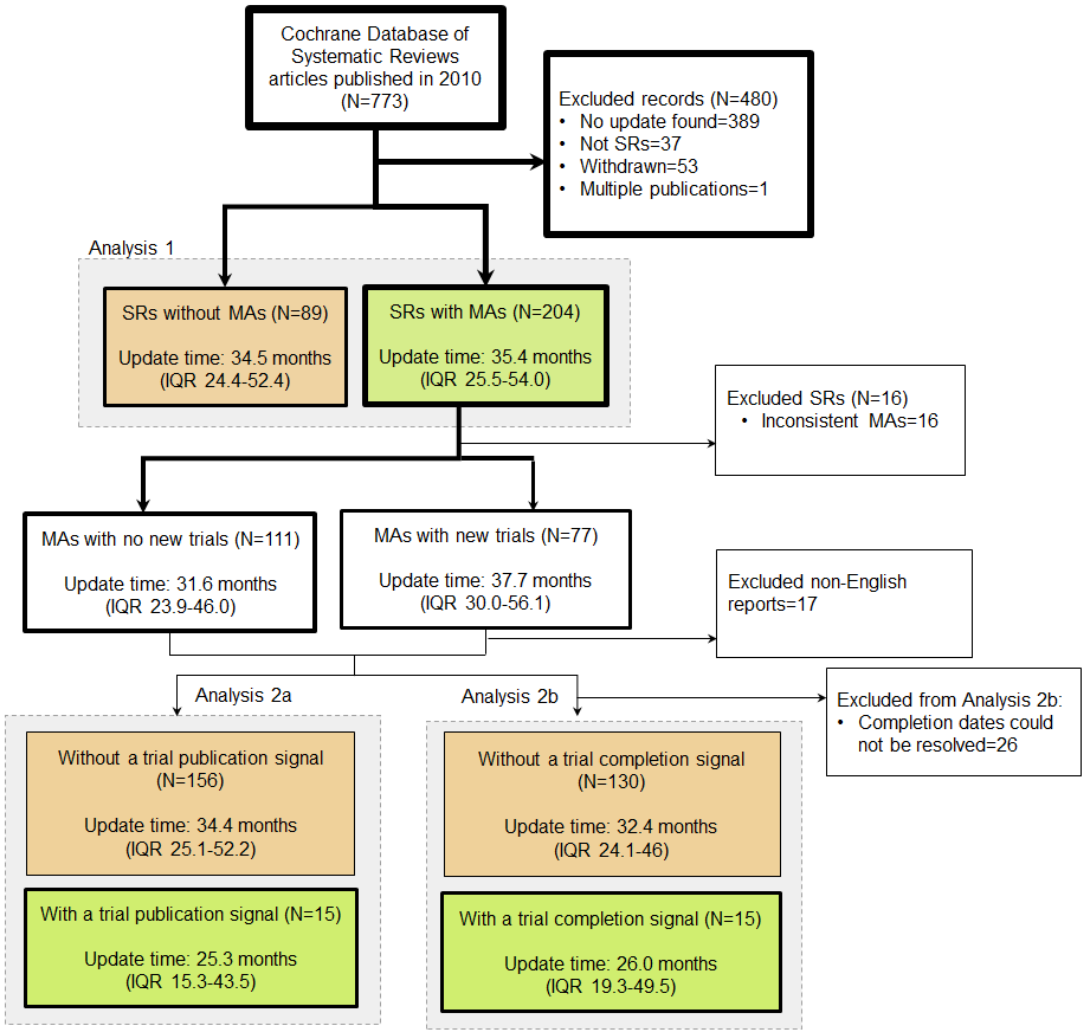

Figure 3.3.2: Update timing in systematic reviews published in Cochrane Database of Systematic Reviews in 2010 (SRs=systematic reviews, MAs=meta-analyses)

[3] Available online. doi: 10.1186/s13643-018-0856-9

Completeness of systematic reviews over time

Among 60 systematic reviews that included new trials in a primary outcome meta-analysis, we determined the completeness by publication of results for the time period between the publication date of the systematic review and the search date of its subsequent update (Figure 3.3.3). Published systematic reviews in this group covered a median of 90.1% (IQR 73.4% to 100%) of the available participants after 12 months. At the search date of the systematic review update, the median completeness of the systematic reviews in this group was 73.6% (IQR 60.0% to 87.2%). In the analyses reported below, we chose to use 12 months and completeness scores of less than 100% to represent as a signal of new evidence. A sensitivity analysis did not change the results of the statistical tests.

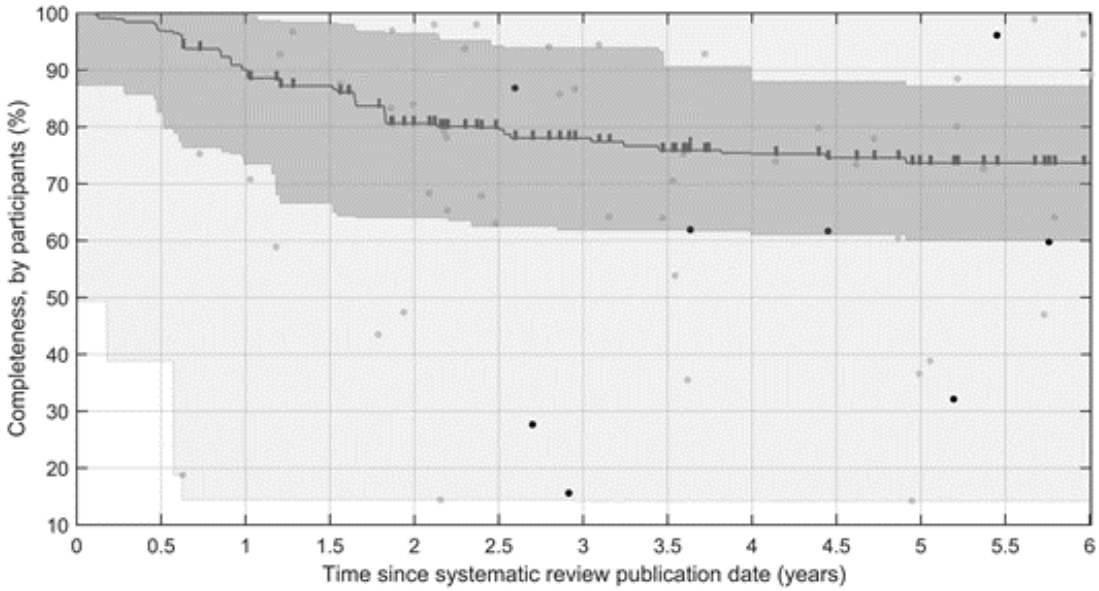

Figure 3.3.3: The completeness of 60 systematic reviews that had new trials added to a primary outcome meta-analysis in an update. The median completeness is represented for each systematic review from its publication date to the search date of its update (length of follow-up is marked), interquartile range (dark grey), and range (light grey). Individual completeness values at the search date of an update are illustrated for those with changes in conclusion (black dots) and no change in conclusion (grey dots).

Associations between new evidence availability and update timing

The median update time among the 204 systematic reviews with primary outcome meta-analyses was 35.4 months (IQR 25.5 to 54.0). The median time to update for the 89 systematic reviews without a primary outcome meta-analysis was 34.5 months (IQR 24.4 to 52.4). We found no evidence of a difference between the two groups in a Wilcoxon rank sum test (p=0.86) (Figure 3.3.4a).

We were able to reconcile enough information about the publication timing of the included trials for 83.8% (171 of the 204) systematic reviews with primary outcome meta-analyses, and used these as the basis for analysing update timing relative to trial publication signals. Among the 171 systematic reviews, 60 had new trials and 15 had new trials published soon after the systematic review was published (within 12 months). In these 15 systematic reviews, the median update time was 25.3 months (IQR 15.3 to 43.5). In the 156 systematic reviews without an early signal of new published evidence, the median update time was 34.4 months (IQR 25.1 to 52.2). The median update time was 9.2 months shorter when there was a trial publication signal, but this difference was not statistically significant (p=0.05) (Figure 3.3.4b).

In the 8 systematic reviews that reported a change in conclusion, the median update time was 48.4 months (IQR 33.6 to 63.7). In the 163 systematic reviews that reported no change in conclusion, the median update time was 32.7 months (IQR 23.9 to 49.5). The difference between two groups indicate that systematic reviews updated faster where the conclusions were not changed than systematic reviews where the conclusions changed (p=0.04).

We were able to reconcile enough information about the completion dates to examine trial completion signals for 71.1% (145 of the 204) systematic reviews, and used these as the basis for examining update timing relative to trial completion signals. In the 15 systematic reviews with a trial completion signal, the median update time was 26.0 months (IQR 19.3 to 49.5). In the 130 systematic reviews without a trial completion signal, the median update time was 32.4 months (IQR 24.1 to 46.0). The median update time was 6.4 months shorter when there was a trial completion signal, but this difference was not statistically significant (p=0.33) (Figure 3.3.4c).

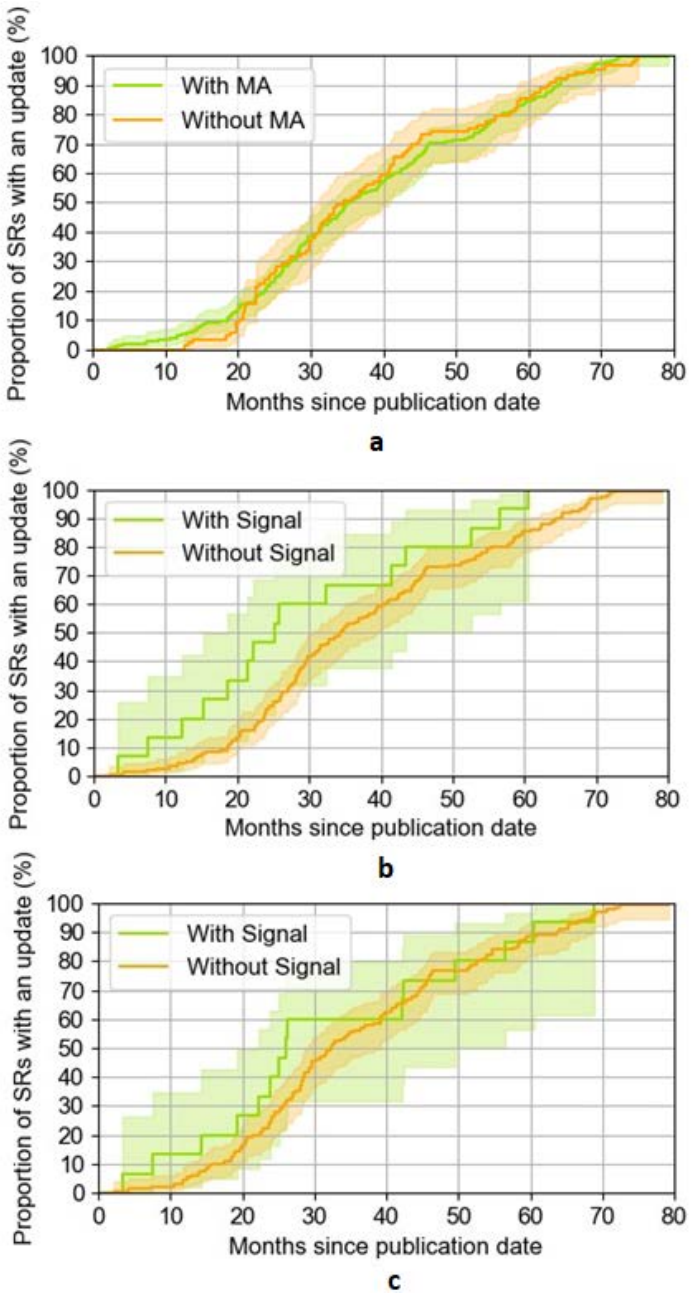

Figure 3.3.4: Time to update for (a) 204 systematic reviews with a primary outcome meta-analysis (green) compared to 89 systematic reviews without a primary outcome meta-analysis (orange); (b) 15 systematic reviews with a publication signal (green) compared to 156 systematic reviews without a publication signal (orange); and (c) 15 systematic reviews with a trial completion signal (green) compared to 130 systematic reviews without a trial completion signal (orange). Shaded regions indicate the 95% confidence interval.

3.3.4 Discussion

Among systematic reviews that were published in Cochrane Database of Systematic Reviews in 2010, fewer than half were updated. Of those that included a meta-analysis for a primary outcome in an update, fewer than half added new trials, and just 8 of 52 that were updated exhibited a change in conclusion. We found no clear evidence that a publication signal or a trial completion signal was associated with a difference in update timing but this may be because relatively few systematic reviews with updates exhibited signals that new evidence was available, and there were relatively few systematic reviews for which we could reconcile the completion dates of the studies included in the update.

In 2007, Shojania et al. [12] examined signals of new evidence for 100 systematic reviews published between 1995 and 2005, defining a signal using information about changes in statistical significance and new information about efficacy and safety. They found that 15% had a signal that new evidence was available within one year and 23% within two years. While our results are not directly comparable because we did not define a signal in the same way, we found comparable proportions. In other studies, examining the timing of systematic reviews, the time between updates has varied between a median of 14 months and 40 months [13, 130].

There are a number of methods that have been developed for deciding if and when to update a systematic review [17, 21, 31, 134-137]. Despite being important as additional sources of trial results [67, 89], and important for identifying biases that can affect systematic review conclusions [86, 138, 139], clinical trial registries are not yet routinely used to support the signalling of updates. Given their potential to provide an early signal that a relevant trial has been completed, clinical trial registries could play an important role in helping to determine which systematic reviews should be prioritised for updating.

We found that only a small proportion of systematic review updates produced a change in conclusion for a primary outcome, and that it was much more common for the conclusions to remain unaffected by new evidence, or for no new evidence to be found when a search was repeated. These results may appear to suggest that there is little value in monitoring trial registries and bibliographic databases to support the allocation of resources to evidence synthesis. However, it is precisely these systematic

reviews that should be detected as early as possible because they represent the clinical questions where current conclusions are at risk of missing important harms, or claiming benefits that are not real. To address this gap in research, future studies in this area could be aimed at developing and testing early signals of conclusion change risks that are simple to compute and precise enough to support the targeting of systematic reviews.

Several limitations should be considered when interpreting the results of this study. First, we applied our method only on Cochrane systematic reviews, which means that other non-Cochrane systematic reviews might have been published between updates and these could have influenced the perceived need for updating. Second, we were unable to include in our analysis the systematic reviews for which publication or registration details were incomplete or inaccessible in English, which may have introduced a language or geographical bias in the set we analysed. Third, there may be sampling bias from using only systematic reviews with updates since 2010 because other systematic reviews published in 2010 may be updated later. However, to analyse all systematic reviews (with or without updates) would require searching and screening new trials relevant to a primary meta-analysis and this would not be feasible. Fourth, we included the first-listed primary outcome meta-analysis that was consistent between the systematic reviews and their updates and included new trials, rather than individually assessing all primary outcome meta-analyses. Fifth, we defined the trial completion and publication signals at a year after the publication date of the systematic review, but this choice was arbitrary. Sixth, we considered only the availability of new evidence for update timing of systematic reviews, but there are other factors such as funding source, conflict of interest, geographical locations and disease area that could also affect the update timing. Also, using the publication date of the systematic review and the search date of the subsequent update to define the update time of systematic review might introduce bias because there are number of factors that affect the publication date. The alternative approach could be to use the search dates of systematic reviews and their updates. Therefore, in unpublished analysis we calculated the update time with search dates. However, using alternative approach didn't make a significant difference in results. Finally, we considered systematic reviews published in 2010 to allow enough time to check for the publication of updates. Methods for

deciding if and when systematic reviews should be updated may have changed and influenced both the update timing as well as the factors that influence the decision to update.

3.3.5 Conclusion

Among systematic reviews published in the Cochrane Database of Systematic Reviews in 2010, less than half had updates published by June 2017, a relatively small proportion had consistent primary outcome meta-analyses with new trials added, and very few reported a change in conclusion for a primary outcome. We found no clear evidence that systematic review updates were undertaken earlier when a relevant study was completed or published within a year of the systematic review publication date. The results suggest that update prioritisation could be improved by developing tools that can use trial registries and bibliographic databases to quickly estimate or predict when a systematic review is at risk of a change in conclusion.

Chapter 4: Modelling the risk of conclusion change in systematic review updates

4.1 Chapter background

The chapter is based on the work presented as part of one conference abstract, one published article, and a manuscript that is currently being prepared for submission. At the *Cochrane Colloquium* in Edinburgh, Scotland, I presented a method for examining the characteristics of a large set of published systematic review updates to determine which are likely to be useful for estimating the risk of conclusion change in future systematic review updates. Extending this work for an article, I trained a set of classification trees to model the risk of change in conclusion. Given that, this study was based on a small, manually-curated dataset, I developed a rule-based approach to extract relevant characteristics from published systematic reviews.

1. **R Bashir**, D Surian, AG Dunn (2018) An empirically-defined decision tree to predict systematic reviews at risk of change in conclusion. Cochrane Colloquium, September 16-18, Edinburgh, Scotland.

2. **R Bashir**, D Surian, AG Dunn (2019) The risk of conclusion change in systematic review updates can be estimated by learning from a database of published examples, Journal of Clinical Epidemiology, 110:42-49, doi:10.1016/j.jclinepi.2019.02.015.

3. **R Bashir**, P Martin, D Surian, AG Dunn (In preparation) A rule-based approach for automatically extracting data from systematic reviews and their updates.

Author contributions: For each of the above manuscripts, I designed the methodology, performed the data extraction, undertook the analysis, and drafted and critically revised the manuscripts. Didi Surian supported with the visualisation and statistical analysis, and critically revised the manuscripts. Adam Dunn contributed to the methodology and data analysis, and critically revised the manuscripts.

4.2 An empirically-defined decision tree to predict systematic reviews at risk of change in conclusion

Background: Systematic reviews are resource-intensive so it is important to focus on reviewing interventions for which new evidence might warrant a change in practice.

Objectives: To determine whether basic information about new relevant trials can be used to estimate the risk of a change in conclusion in published systematic reviews.

Methods: We identified systematic reviews that had updates published between October 2016 and December 2017, including pairs with consistent search strategies, inclusion criteria, outcomes, and where most included studies were trials. We analysed reviews that added new trials and reported the numbers of participants. We extracted: the total number of trials and participants in the original review; the time between the two search dates; and the completeness—the number of participants in the original review as a proportion of the number of participants in the update. A change in conclusion was defined by a change in significance of a primary safety or efficacy outcome (evaluated independently by two investigators; disagreements resolved by discussion). We trained a Classification and Regression Tree to predict (5-fold cross validation) a change in conclusion using some or all of the factors; reporting average precision and recall.

Results: We analysed 63 pairs of reviews, of which 20 reported a change in conclusion in the update. Using the number of trials/participants in the original review and time elapsed to the new search date, the decision tree produced an average precision of 40% and a recall of 70%. After adding completeness to the decision tree, this increased to an average precision of 60% and a recall of 90%. The decision tree (Figure 4.2.1) showed that reviews were most at risk of a change in conclusion when completeness was low (≤ 13.5%), the original review had fewer trials (<23), and more time had elapsed (>53 months).

Conclusions: An empirically-defined decision tree using simple information extracted from a published systematic review and basic information about trials that may be relevant can estimate the risk of a change in conclusion. The results can be used to better target resources for updating systematic reviews and would benefit patients by identifying evidence reversals earlier.

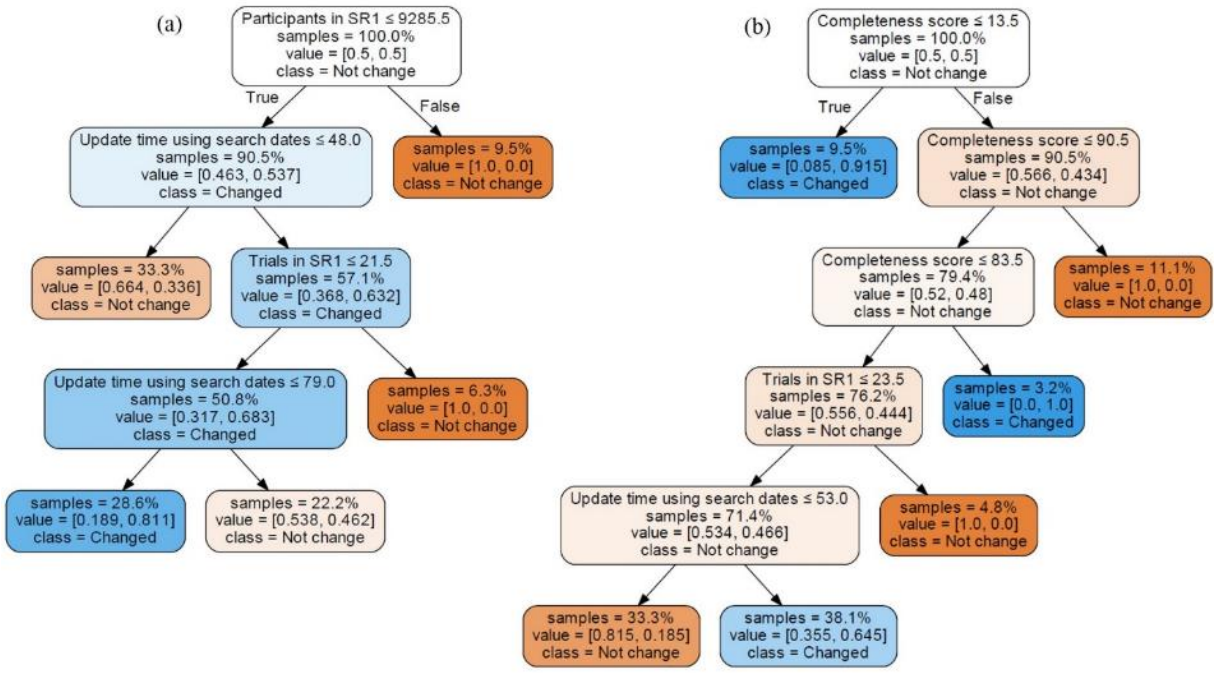

Figure 4.2.1: The classification and regression tree (CART) using (a) only information from the original review and the time to the new search date; a high risk in reviews with fewer participants (<=9285.5) and fewer trials (<22); and (b) adding basic information about new relevant trials; high risk with lowest completeness (<=13.5%) or fewer trials (<=24) and longer times (>=53 months). Changed: Change in conclusion; Not Changed: No change in conclusion

4.3 The risk of conclusion change in systematic review updates can be estimated by learning from a database of published examples

Abstract

Objectives: To determine which systematic review characteristics are needed to estimate the risk of conclusion change in systematic review updates.

Design and Setting: We applied classification trees (a machine learning method) to model the risk of conclusion change in systematic review updates, using pairs of systematic reviews and their updates as samples. The classifiers were constructed using a set of features extracted from systematic reviews and the relevant trials added in published updates. Model performance was measured by recall, precision, F_1-score, and area under the receiver operating characteristic curve (AUC).

Results: We identified 63 pairs of systematic reviews and updates, of which 20 (32%) exhibited a change in conclusion in their updates. A classifier using information about new trials exhibited the highest performance (AUC: 0.71; F_1-score 0.54; recall: 0.75; precision: 0.43) compared to a classifier that used fewer features (AUC: 0.65; F_1-score 0.52; recall: 0.75; precision: 0.39).

Conclusion: When estimating the risk of conclusion change in systematic review updates, information about the sizes of trials that will be added in an update are most useful. Future tools aimed at signalling conclusion change risks would benefit from complementary tools that automate screening of relevant trials.

Keywords: Machine learning; Classification trees; Automation of systematic reviews; Systematic reviews as topic; Clinical trial registries; Updating systematic reviews

4.3.1 What is new?

Key findings

- When modelling the risk of a conclusion change in systematic review updates, information about the sizes of new relevant trials was more useful than information extracted from the original review.
- Identifying and extracting data from systematic review updates for use as training data is challenging.

What this adds to what was known?

- Existing methods for predicting whether a systematic review conclusion would change were based on measures extracted from primary meta-analyses.
- The risk of conclusion change in a more general set of systematic review updates was modelled in a database of examples of systematic reviews paired with their updates, using features that were relatively simple to extract.

What is the implication, what should change now?

- Future tools aimed at estimating when a systematic review is at risk of a change in conclusion would benefit from being coupled with tools that automate trial screening because information about new relevant trials was found to be most useful for estimating risk.

4.3.2 Introduction

Systematic reviews are used to provide a comprehensive synthesis of clinical evidence to guide clinical decision making, form the basis for clinical practice guidelines, and suggest directions for new research [22, 70]. To fulfil that purpose, systematic reviews need to be kept up to date [140]. As a consequence of the resource-intensive processes involved in producing a systematic review, it can be a challenge to keep up with the rate at which evidence from new trials is made available [12]. At the same time, a substantial proportion of the systematic reviews being published are redundant, unnecessary, or focused away from the clinical questions where accumulating evidence could influence the conclusions in ways that would influence clinical practice [14].

An approach for improving the efficiency of systematic reviews is to make them easier to do, by individually automating the underlying processes—searching, screening, information extraction, and synthesis [23, 25-29, 141-144]. However, these tools alone are unlikely to help avoid unnecessary or redundant systematic reviews. According to a recent study [145], screening accounts for 25% of the effort required to produce a systematic review, which means that tools for avoiding unnecessary systematic reviews could save a majority of the time costs of undertaking systematic reviews. General guidance on when to update a systematic review considers not only the accumulation of new evidence but contextual factors like the importance of the topic of the review and the potential impact on guidelines and clinical decision making. However, there is no evidence that review updates are undertaken faster when new evidence is made available [15]. Similarly, relatively little work has been done to develop statistical methods to quantify the potential for new evidence to change the results of a review, and these methods are mostly confined to examining meta-analyses [12, 137, 146-148]. Some tools and checklists that are intended to support the decision to update a systematic review use the availability of new evidence as an input [133, 149].

One tool used to support the decision to update a systematic review was developed to take advantage of information about what happened in a database of previously updated systematic reviews to predict whether a primary meta-analysis would change given the accumulation of new evidence [31]. This form of empirically-derived tool has the potential to improve the efficiency of systematic review efforts but there are several practical limitations. A focus on meta-analyses means that the tool would

be less useful in systematic reviews with no meta-analyses or where new outcomes were added, which may be of particular importance for interventions where new safety issues arise. The approach makes use of continual surveillance of new and relevant studies to determine how much of the currently available evidence is covered by the original meta-analysis. This can be time-consuming, so it would be useful to know if this information is necessary for estimating whether a systematic review conclusion is likely to change. Since systematic reviews are rarely made available in structured and machine-readable formats that would make them amenable to data mining, the tool was limited to learning from examples in one journal where data extraction could be standardised.

Our aim was to determine which systematic review characteristics are useful for estimating the risk that a systematic review would change its conclusion if updated to include new studies. To do this, we extracted information from a set of systematic reviews paired with their published updates to model the risk of a change in conclusion, and examined how those features might be operationalised to create a risk-signalling tool.

4.3.3 Methods

Study data

We searched PubMed for systematic reviews published prior to December 2017 with the aim of identifying pairs of systematic reviews; starting with the most recent systematic review updates and finding their most recent previous version. To identify updates, we limited the search to articles that included the terms "systematic review" and "update" in the titles or abstracts. We then read the abstract to exclude any article that was not a systematic review and used information in the abstract or background to determine whether the systematic review referred to a previous version.

From pairs of identified systematic reviews and updates, we excluded any that were not written in English, and any pairs for which either review had a published erratum or were withdrawn. We then excluded pairs where there was a major change in the clinical question answered or where we were unable to extract a minimum set of features. This included updates that had substantially changed the inclusion criteria, included only observational studies, added no new evidence, or did not clearly state the number of participants in the set of included studies. We additionally excluded systematic reviews

that added no new studies in the update because they do not exhibit changes in conclusion and are not useful in the models.

Main outcomes and measures

Classification trees are used in machine learning to model a categorical outcome feature from several input features, producing a decision tree. We chose to use classification trees to model the risk of a change in conclusion because of their simplicity and interpretability—the contribution of the input features to the decision is clear and the tree can be implemented for use in practice more easily than non-interpretable models. For input, we characterised each pair of systematic reviews and updates by four features. The time elapsed since the search date was defined by the number of days between the search date of the systematic review and the search date in the update. We extracted the number of trials and participants from information available in the systematic review. For already published systematic review updates, a relevant trial is one that has already been evaluated by the authors of the systematic review update and included. The coverage score was defined by the total number of participants in the trials included in the systematic review as a proportion of the total number of participants in the trials included in the update. The coverage score is similar to the inverse of the participant ratio used by Takwoingi et al. [31].

The primary outcome used in the construction of the classification trees was the presence of a change in conclusion. To determine which of the updates exhibited a conclusion change, two authors (RB and AD) read the systematic reviews and their updates to determine whether there was a change in conclusion, using information extracted from the results and conclusion statements. We first used the abstract to look for a clear indication that the conclusion had changed in the update, and then compared the primary outcomes between the systematic review and the update. Where the structure of the abstract or the terms used to describe outcomes changed, or where there was a substantial shift in focus highlighting a finding that was not included in the original review, we labelled the review as having a changed conclusion. After independently labelling the updated reviews, we discussed any systematic review where we disagreed and read the full text of both the original and updated reviews in detail to make a decision. The level of agreement between the evaluations was measured using Cohen's kappa [150], and disagreements were resolved by discussion.

To ensure feasibility and to capture a sample of systematic reviews that was reasonably balanced in terms of the main outcome, we only included systematic reviews if their updates added new studies. Systematic reviews that do not include new studies can change conclusions, but we expected that this would be rare. The exclusion of systematic reviews without updates and the under-representation of reviews for which no new evidence was found may introduce biases into the models. A fairer sampling approach would have been to sample from across a general set of systematic reviews (regardless of whether they have an update) and manually search and screen for the availability of new evidence. However, this would severely limit the number of examples that could be used to construct the classifiers.

Classification tree construction, analysis, and evaluation

We built three classification trees using sets of features extracted from systematic reviews and their updates, to model the risk of conclusion change in systematic review updates. Our rationale for each of the three classifiers was based on how we expected to use the decision tree as a tool, where the classifiers represented differences in how quickly we could estimate the risk of a conclusion change in updates given the amount of time and effort involved to extract the information needed to apply the tool. The first classifier uses all four features: the time elapsed since the search date, the number of trials in the systematic review, the number of participants in the systematic review, and the coverage score based on information from new and relevant trials. This is resource intensive—users need to extract multiple types of information from the review and identify and extract information about the new and relevant trials not included in the review. While there are a range of methods used to support searching and screening of published articles and trial registrations [28, 144, 151, 152], this still requires manual effort by experts. The second classifier excludes the coverage score and uses the time elapsed since the search date, and the number of trials and number of participants in the original systematic review. This reduces the amount of manual effort required to predict the risk of a conclusion change by limiting what is needed to only include information available in the systematic review. The third classifier uses only the time elapsed since the search date and excludes information about the number of trials and participants as well as the coverage score. Early guidelines about how often a systematic review should be updated were based on this feature (often two years was given as a

reasonable time), and this third classifier represents an empirically-derived version of these guidelines. The information is trivial to extract from systematic reviews making the classifier amenable to automation.

For each of the four sets of features, we used the classification tree method to model the change in conclusion using all pairs of systematic reviews and their updates. The approach produces an interpretable set of rules that splits the set of systematic reviews into increasingly smaller groups comprising mostly changed or mostly unchanged conclusions. The ability to discern systematic reviews with changed conclusions from those with unchanged conclusions might then be useful for signalling when a systematic review has features that are most similar to others that previously exhibited a conclusion change when they were updated. The assumption is that we can learn the features that predict conclusion changes in systematic reviews from a large general database of already-published systematic review updates.

To compare the performance of three models, we calculated the precision, recall, and F_1-score of the three classifiers. The precision was defined as the number of correctly identified systematic reviews with conclusion change divided by the total number of reviews, and recall was the number of correctly identified systematic reviews with conclusion change divided by the total number of reviews with conclusion change or true positive divided by true positive and false negative (sensitivity). The F_1-score is the harmonic mean of precision and recall, which is used as a more robust measure of accuracy in cases where data are unbalanced in terms of positive and negative results. The model produces a risk estimate between 0 and 1 for each systematic review, so to classify the systematic reviews as high-risk or low-risk, we selected the threshold that maximises the F_1-score across the set of reviews. We also produced receiver operating characteristic (ROC) curves and calculated the area under the ROC (AUC) to compare the performance of the three classifiers. We conducted all experiments using Python 3.6.

Practical demonstration

To demonstrate how the classification trees might work in practice to signal when a systematic review is at risk of a conclusion change, we selected two additional systematic reviews that included

randomised controlled trials and had updates published in 2018. We then applied the tool retrospectively to examine how the estimated risk of a conclusion change varied as new evidence was published over time. To do this, we extracted information about the accumulation of new evidence by identifying the set of trials added to the updates and reconciled information about their timing and relevant characteristics. For each trial, we extracted the number of participants and the date when the results were first publicly reported in a published article, on ClinicalTrials.gov, or on company websites. We used the earliest date if the results of the trial were reported in more than one location and used the number of participants from the published article if the number of participants varied from the registration to the trial report.

To determine how the estimated risk of a conclusion change varied over time, we applied the three classification trees to calculate the risk for each day in the period between the publication date of the systematic review and the search date of the update. The number of trials and number of participants in the review were constant, the time elapsed since the search date increased each day, and the coverage score decreased to align with the public reporting of results for each of the new and relevant trials.

4.3.4 Results

We screened 1,047 records returned by the search. Of these, we excluded 207 by screening the titles and abstracts to remove articles that were not systematic reviews. We excluded 656 because they were not updates of previously published systematic reviews. Of the 184 that were updates of systematic reviews 63 met our inclusion criteria (Figure 4.3.1).

Of the 63 systematic review and update pairs included in the models, 40 (63%) were published in the Cochrane Database of Systematic Reviews. After two investigators independently reviewed the conclusions (Cohen's kappa 0.40) of 63, we found that 20 (32%) exhibited a conclusion change. Across the 63 systematic review updates, the two investigators agreed that there was a conclusion change on 13, agreed that there was no conclusion change on 37, and disagreements in each direction were 6 and 7 respectively. Our level of agreement was low because it is often difficult to identify changed conclusions as these are not always clearly mentioned in systematic reviews.

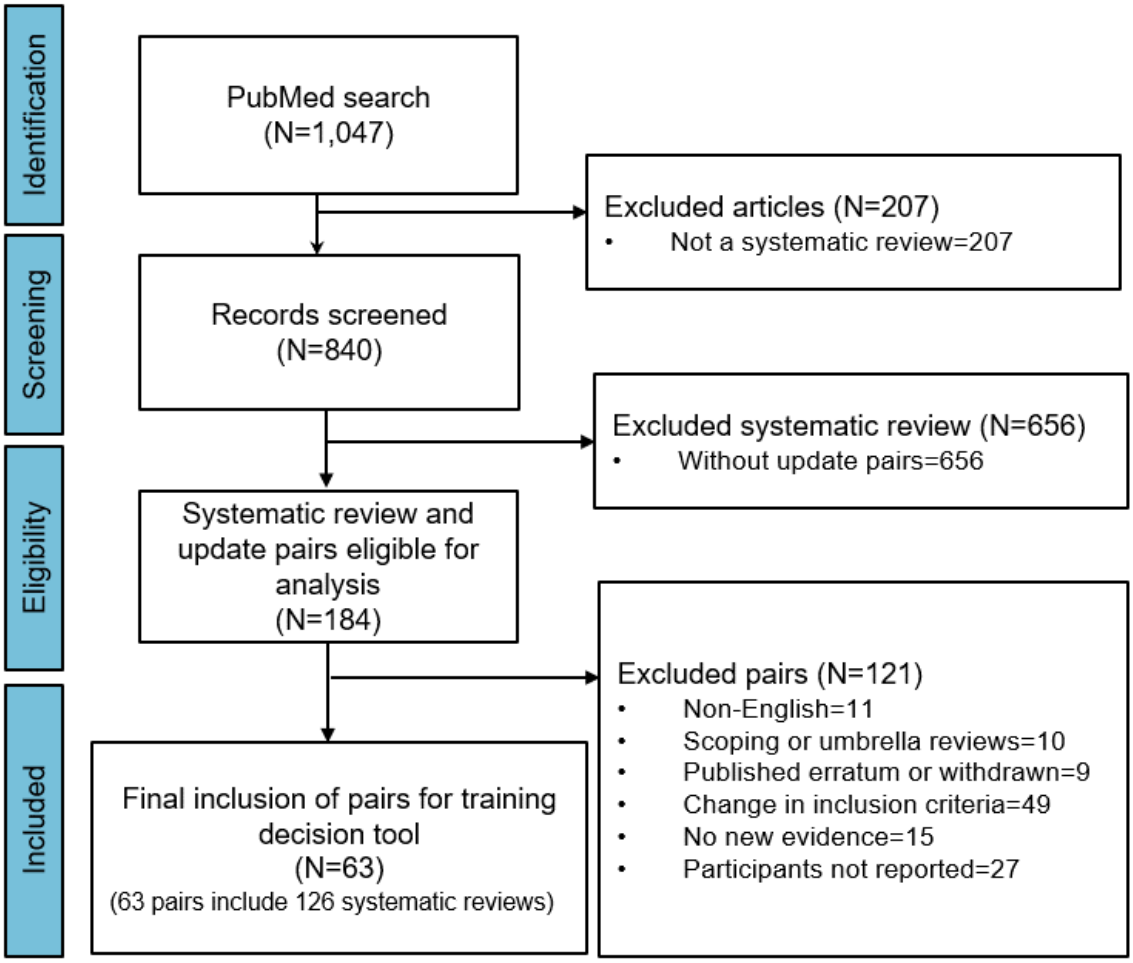

Figure 4.3.1: PRISMA flow diagram of study selection for a search and screening process that resulted in the inclusion of 63 systematic review and update pairs for constructing the classification tool.

Classification tree performance

The first classifier that used all features including the coverage score produced the highest performance: recall 0.75, precision 0.43 and F_1-score 0.54 at a threshold of 0.45, the value that maximised the F_1-score (Table 4.1).

Table 4.1: Performance of the three classifiers by precision, recall, F_1-score and area under the receiver operating characteristic curve (AUC)

Classifier	Precision	Recall	F_1-score	AUC
1: Full, including coverage score	0.43	0.75	0.54	0.71
2: Partial, including included trial details	0.39	0.75	0.52	0.64
3: Partial, including only search date details	0.39	0.75	0.52	0.61

The CART based on all features illustrates that changes in conclusion were more common in scenarios where the coverage score was lower, where the systematic review included a smaller number of participants and fewer trials, and where more time had elapsed since the search date of the review (Figure 4.3.2).

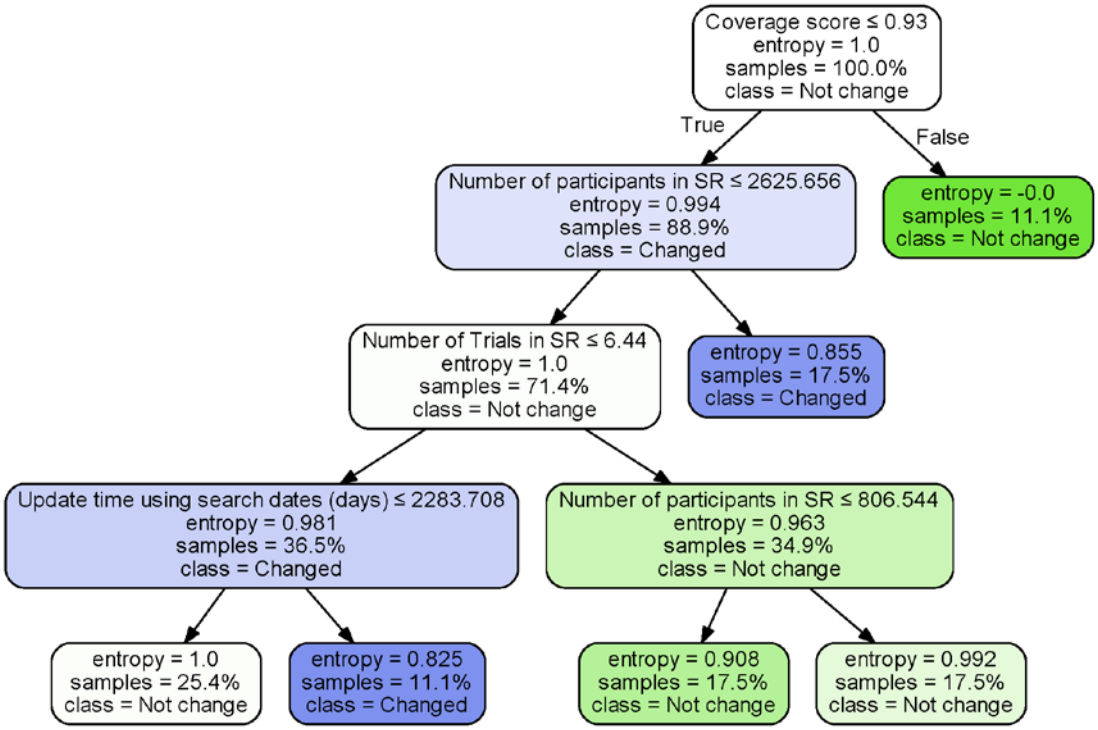

Figure 4.3.2: The classification tree produced by using all features to estimate change in conclusion. Update time is given in days; entropy shows how uniform are all samples of a node; and samples represent the proportion of 63 systematic review and update pairs that appear within the node of the tree. The tree shows that conclusion changes were more common in systematic reviews that had more participants in the original review (>2625) and a lower coverage score (≤0.93); or where there were fewer participants in the original review (≤2,625), fewer trials (≤6), and more time had elapsed since the search date (>2,283 days).

The second classifier used only features extracted from the systematic review, which reduced the performance from the full classifier: recall 0.75, precision 0.39 and F_1-score 0.52 at a threshold of 0.45. The third classifier used only the time elapsed since the search date and exhibited the lowest performance: recall 0.75, precision 0.39 and F_1-score 0.52 at a threshold of 0.45. The area under receiver operating characteristic curve (AUC) is determined independent of the choice of threshold and was highest for the first classifier (AUC: 0.71) compared to the second classifier (AUC: 0.64) and the third classifier (AUC: 0.61) (Figure 4.3.3). The results show that information about the sizes of new and potentially relevant trials produced the largest positive impact on the ability to estimate the risk of a conclusion change.

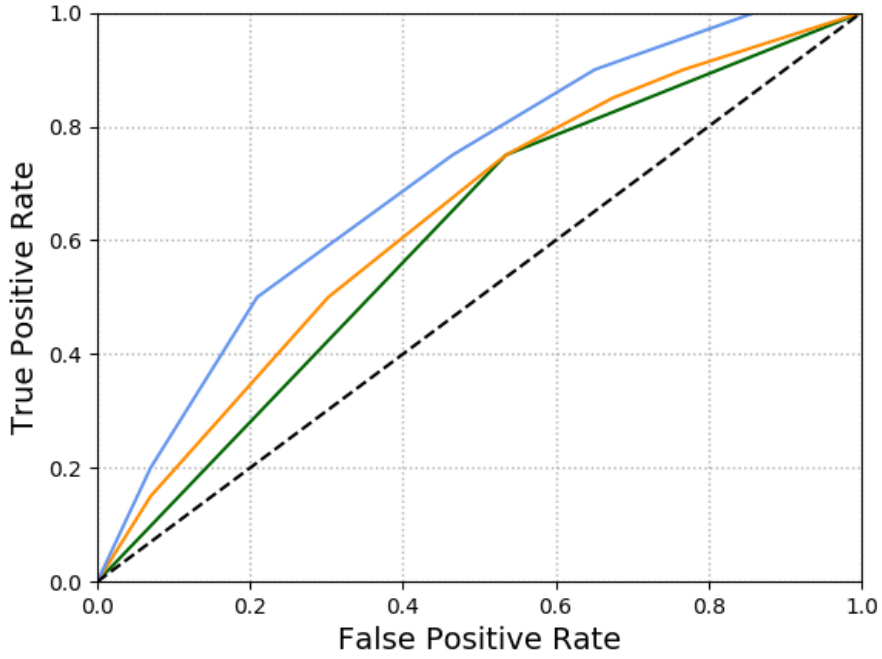

Figure 4.3.3: Area under the receiver operating characteristic curve for classifiers using all features including the coverage score (blue, AUC: 0.71); only features extracted from the systematic review (orange, AUC: 0.64) and only using time elapsed since the search date (green, AUC: 0.61).

Practical demonstration

We demonstrated how the models might be used to produce signals of conclusion change risks over time by retrospectively applying them to 2 systematic reviews. The systematic reviews used in the demonstration were new and were not included in the training data. The threshold value that

maximised the F_1-score in the best-performing model was 0.45, so we used this as the value to signal when a systematic review was at higher risk of a conclusion change.

The first systematic review and its update examined evidence for physical therapy to reduce patient length of stay and did not exhibit a change in conclusion [153, 154]. The original search date was May 2010 (it was published in September 2011), and the search date for the update was June 2017 (it was published in April 2018). When we used the first classifier (all features), the estimated risk increased to 0.44 after 1.8 years (Figure 4.3.4a). The second classifier produced a signal 6.3 years after the systematic review publication date, estimating the risk at 0.52. The third classifier produced a signal after 4.2 years. A tool based on the first classifier would not have produced a strong signal of risk in conclusion change and systematic reviewers using the tool may have chosen to delay the update of the systematic review.

The second systematic review was examining intravenous thrombolysis for acute ischemic stroke and exhibited a conclusion change [155, 156]. The original search date for the second systematic review was June 2010 (it was published in November 2011), and the search date for the update was August 2016 (it was published in March 2018) [155, 156]. The first classifier (all features) produced a signal 1.8 years after the systematic review was published, estimating the risk at 0.72 (Figure 4.3.4b). The second classifier never produced a signal of a risk of conclusion change and reached a maximum estimated risk of 0.42 during the period of analysis. The third classifier produced the second signal after 4.2 years. If systematic reviewers were using the tool, they would have seen a signal of a risk in conclusion change 4.8 years earlier than the search was performed for the update.

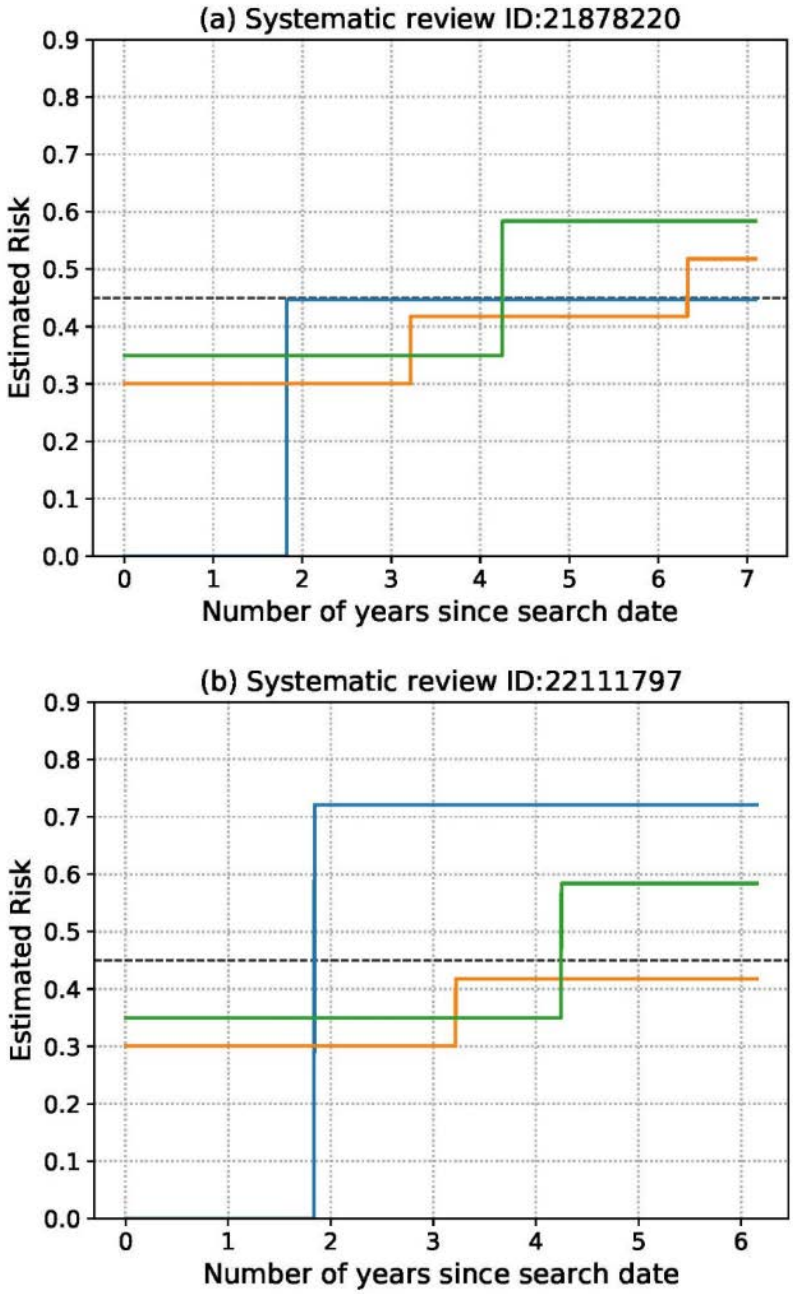

Figure 4.3.4: Estimated risk for (a) a systematic review exhibiting no change in conclusion; and (b) a systematic review exhibiting a change in conclusion. We compare three classifiers: using all features (blue) using only features extracted from the systematic review (orange); and using only the time elapsed since the search date (green). A threshold at 0.45 (dotted line) is used to determine the presence of a signal.

4.3.5 Discussion

We found that accessing information about the presence and size of new and potentially relevant trials made the greatest improvement to our ability to estimate the risk of a conclusion change. The results

showed that access to information that was more time-consuming to collect was the most useful in improving the classification tree's performance in estimating risk. As part of a broader consideration of factors that might influence the decision to update a systematic review, tools like the one we propose here could help systematic reviewers, journals, and funders avoid potentially unnecessary updates and focus on systematic reviews with conclusions that do not reflect currently available evidence. A tool that learns to estimate in advance whether a systematic review update would produce a change in conclusion could be constructed by learning from a large, general set of already-published systematic review updates.

Previous studies have proposed or used different types of information to support the decision to update a systematic review [149, 157]. Prospective evaluation of different approaches in this space is challenging because of the resource-intensive nature of undertaking systematic reviews. Takwoingi et al. [31] constructed a multicomponent tool for deciding whether to update a systematic review, using a set of 9 features extracted from the primary meta-analyses of systematic reviews published in the Cochrane Database of Systematic Reviews. This tool was used as part of a process for prioritising systematic review updates [133], but its application is limited to systematic reviews with meta-analyses of at least two trials and may only apply to systematic reviews published in that journal. Our approach differed in that we examined which features could be extracted to potentially predict the risk of conclusion change from a broader set of systematic reviews. The assumption is that with a broad and large set of published examples of systematic review updates, we would be able to determine which general characteristics would be most useful in a more general tool.

There are several implications to the work we presented here. The results suggest that it is feasible to develop tools that can applied quickly to signal when a systematic review conclusion is no longer an accurate reflection of currently available evidence, and focus resources on their update. Conversely, we might also be able to use these types of tools to avoid allocating funding and resources to systematic reviews that are unnecessary [14]. The results of the research here also indicate the value of knowing which trials are likely to be relevant to a systematic review in advance, beyond avoiding the time taken to search and screen for trials when undertaking a systematic review. The results suggest that a tool for quickly estimating the risk of conclusion change might rely on knowing in advance the

set of trials that are most likely to be included in an update. This means that existing tools for automating or supporting trial screening could be coupled with tools for estimating risks of conclusion change. Automated surveillance of ongoing and completed trials relevant to a published systematic review are likely to improve with time [152, 158-160], and these could be used to reduce or eliminate the need for screening and may eventually be used to automate signals for prioritising systematic review updates.

Future work in this area would benefit from deeper integration with trial registries. Given that trial registrations represent an early indication that new trial evidence will become available, surveillance of relevant clinical trials could help us to estimate when a systematic review may be at risk of a conclusion change. Surveillance of ongoing and completed trials may make it possible to prepare in advance for systematic reviews that are likely to be at risk of a conclusion change as soon as new results are made available to the public. While the structured, machine-readable, and connected reporting of trial results is improving [36], less effort has been spent on establishing public access to information connecting systematic reviews to their updates and the sets of studies they include. Improvements in structured reporting and transparency in this space would make it easier to track how evidence coverage degrades over time for published systematic reviews and may also help to reduce the number of redundant and unnecessary systematic reviews [161]. There are likely to be a range of other characteristics of systematic reviews that can be extracted and may be useful as features in predictive models of conclusion change. For example, information about the specialty or class of interventions might be indicative of the fragility of the conclusions in a particular area and could be integrated. In the future, it may also be useful to consider different types of conclusion changes. For example, a change in conclusion from a lack of evidence to clearer evidence of safety and efficacy is quite difference from a change in conclusion that identifies a new safety risk.

There were limitations to this study. First, we considered only English language systematic reviews that included clinical trials for our analysis and the models may not generalise to other languages or systematic reviews of other study designs. Second, identifying systematic review updates that ask and answer the same clinical questions is a challenge, and our dataset included a substantial number of systematic reviews from the Cochrane Database of Systematic Reviews, where updates are clearly

demarcated. Third, we did not include systematic reviews that did not have updates because this would have required manual screening of new and relevant trials, which was not feasible. Excluding systematic reviews without published updates could introduce biases because the training data are a special subset of the systematic reviews for which we used the classifiers. Fifth, the dataset we used to construct the classification tree models was relatively small and we did not test the resulting models on unseen systematic review updates. Once a tool based on a larger dataset has been constructed, a prospective evaluation of its ability to predict conclusion changes in advance of a systematic review update would be needed. Future work in the area would benefit from a structured database of systematic reviews with information about included trials [158].

4.3.6 Conclusion

We built three classifiers to determine which characteristics of systematic reviews are useful for estimating the risk of conclusion change in systematic review updates. The aim is to improve decisions about when to update systematic reviews, and we suggest that it may be useful to systematic reviewers who want to know in advance whether the conclusion of a review is likely to change if they incorporate newly-available evidence. The tool is different from previous approaches because it uses a set of existing systematic review updates to learn how characteristics of systematic reviews and the trials that meet their inclusion criteria correspond to the risk of a conclusion change. The results show that access to information about the presence and size of new and potentially relevant trials is most useful for estimating risk. Given the potential value that these tools may have in improving the efficiency of systematic reviews, we think that further work building a database of systematic review updates from which to learn is warranted.

4.4 A rule-based approach for automatically extracting data from systematic review articles and their updates

4.4.1 Introduction

Systematic reviews are essential in evidence-based medicine because of their role in healthcare policy and practice. They provide a comprehensive synthesis for clinical decision making and basis for clinical practice guidelines [162, 163], but safety issues identified after approval for up to a third of drugs highlight the need for the ongoing evaluation of available evidence [10]. Because systematic reviews and safety meta-analyses can play an important role in revealing safety issues after an intervention is approved for use in practice, improved alignment between systematic review updates and the accumulation of new evidence could help to improve regulatory decision making.

A number of factors influence the decisions made about when to update a systematic reviews [71]. Data-driven methods for supporting the decision to update a systematic review are rare [19-21]. However, due to the challenges imposed by the rate at which new evidence is produced, there is a need to consider how to allocate resources to systematic reviews where updates are most needed [22].

One data-driven approach uses information extracted from the primary meta-analyses of systematic reviews to estimate the risk of a conclusion change [31]. Its reliance on the presence of a primary meta-analysis limits the generalisability of the tool. Tools that can be applied to a broader set of systematic reviews are therefore likely to be useful. In a previous study we proposed a tool of this type, using a manually-curated set of systematic review updates from which to train a model of conclusion change [164].

A range of machine learning and text mining approaches are being developed to improve or automate individual processes in systematic reviews. They have the potential to reduce the manual workload associated with searching, screening, extraction, and synthesis [30]. Some methods and techniques were developed to help in searching by identifying the set of potentially relevant trials for systematic reviews, while others support screening [28, 144, 160, 165-170]. Similarly, the extraction of data is a time-consuming task in systematic reviews, and attempts have been made to automate these steps [171, 172].

Other studies have made use of the semi-structured nature of the articles published in the Cochrane Database of Systematic Reviews. These have included extracting systematic review inclusion criteria and sentences describing population, intervention, outcomes and controls (PICO) from included studies [171-173]. We know of no other studies that aim to consolidate information across systematic reviews and their updates with the aim of predicting the risk of conclusion change [164].

Our aim was to use a rule-based approach to automatically extract a set of useful features from systematic reviews and their updates, with the intention of using these features to support models for estimating the risk of conclusion change.

4.4.2 Methods

Study data

PubMed was searched from May 5, 2000 to January 31, 2019 for systematic reviews published in the Cochrane Database of Systematic Reviews. We used the original systematic reviews and their updates, which appear as the second and third linked publications assigned to a single Digital Object Identifier (DOI) published in the journal. The study data includes the full text of the articles as they were presented on webpages, from which we aimed to extract a set of features from a systematic review and their updates

Data extraction

Data extraction involved two parts: understanding the structure of systematic review articles in the journal to extract the relevant information embedded in different sections; and building the rules to extract that information. Data extraction involved identifying different sections of the systematic reviews where some features had only one value while others had more than one (e.g., search date of systematic review) and some of these are always present while others might be absent (e.g., conclusion and number of participants).

We aimed to automatically extract the following information from original systematic review and their updates: (a) search dates (b) number of included trials; (c) number of participants from included trials; (d) publication date and presence of change in conclusion.

To extract these features, we used an iterative rule-based approach that made use of the semi-structured reporting in the journal. The most challenging information to extract was the number of participants in each of the included trials in the systematic review and their updates. In each iteration we selected 50 systematic reviews at random from the pool of all systematic reviews and selected the first listed included trial from each review (Figure 4.4.1). We manually recorded information about the number of participants for evaluation. We then revised and added rules (regular expressions) until they were able to correctly extract information for at least 95% (48 of 50 trials) of the trials. Note that this included correctly identifying examples where the description of the trial did not include details of the number of participants and returning an answer of "not reported". Once the required accuracy was reached, we manually extracted information from an additional and distinct set of 50 systematic reviews and repeated the process. We stopped updating the rules when 95% accuracy was reached in an unseen set of 50 systematic reviews without adding or revising any rules from the previous round.

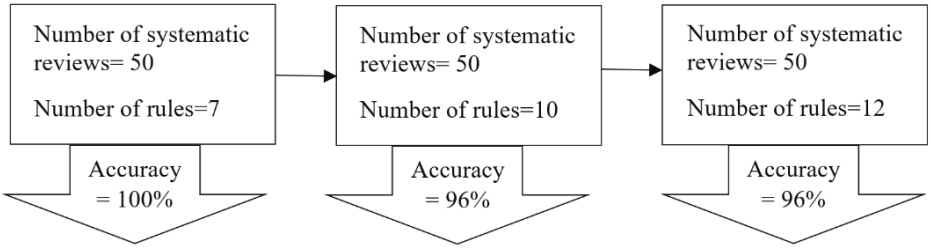

Figure 4.4.1: An iterative process of rule construction for extracting features.

The final set of rules were then applied to all remaining systematic reviews in the dataset to extract details of all included trials. To calculate the performance of our rule-based approach we measured precision for each of these features. For extracting the data from systematic reviews, we used web scraping—the process for extracting the patterned data from websites [174]. The rule development was done using Python 3.7 and Beautiful Soup was used to support the web scraping methods. The set of rules for extracting the participants' information are in the Appendix, Table 4.2.

4.4.3 Results

Our search returned 13,451 records (Figure 4.4.2). Of these, we excluded 5,060 for which the DOI was not available in PubMed. Of the 8,391 remaining records, 817 (10%) were original systematic reviews for which there was an associated update.

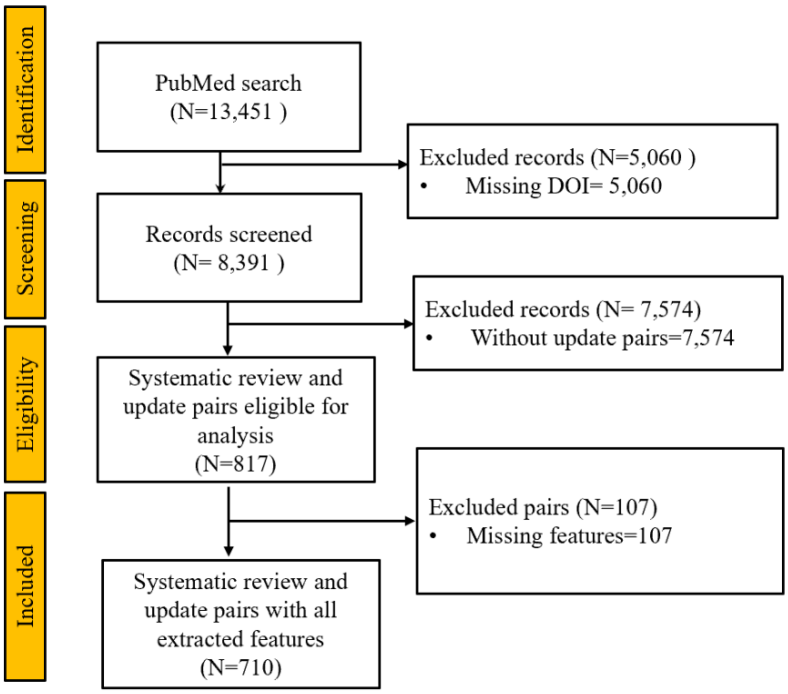

Figure 4.4.2: PRISMA flow diagram of study selection for a search and screening process that resulted in the inclusion of 817 systematic review and first update pairs for automatically extracting features through rule-based approach.

Among the 817 pairs the rules were able to extract useful features from 87% (710 of 817) of pairs. These include a complete information of search dates, number of included trials, publication dates and presence of change in conclusion, and number of participants from at least half or more than half or all of the included trials. For the remaining 107 pairs, we were unable to extract the full set of features including trial participant numbers for at least half of included trials for one or more of the following reasons: the review was withdrawn, the systematic review did not include any trials, the structure of table providing the participant information was not consistent with the set used for constructing the rules, the participant information was not reported in the table for at least half of the trials, or the

participant information was reported in an unusual way for which there was no rule available to identify the number of participants in each trial.

Among the 710 pairs for which we were able to extract the number of participants for at least half of the included trials, the precision was 100% for the number of trials, search dates, publication dates, and the presence of a conclusion change. When extracting the number of participants in included trials. The precision was 91% (648/710) in systematic review updates and 89% (634/710) in the original systematic reviews. The rule-based approach was able to extract the number of participants for all included trials in 33% (238/710) of the original systematic reviews and 63% (445/710) of the updates. Given the increase in performance from the original to the update, the results suggest that the consistency with which participant numbers are reported in the journal may be increasing over time.

4.4.4 Discussion

We found that rule-based approach was useful for extracting the relevant features from systematic reviews and their updates. We extracted complete and correct information for number of trials, search dates, publication date and presence of a change in conclusion. However, it was challenging to consistently extract the number of participants for all included trials in systematic reviews and their updates; for more than half of the systematic reviews there was at least one included trial for which the number of participants could not be extracted. This was partially because the systematic reviews did not always include the participants' information in the table of included studies, and in other cases because the information was presented in unusual ways and in combination with a range of other numbers.

Previous studies have used the Cochrane Database of Systematic Reviews and full text articles of included studies as a source of data for extracting trial and review characteristics. A well-known example, ExaCT, uses a machine learning approach to automatically extract information from published trial articles [175]. Borlawsky et al. [176] tested a set of natural language processing methods for use in extracting information from narratives in reviews published in the Cochrane Database of Systematic Reviews (based on an older format). Basu et al. [173] also used natural language processing and machine learning approach to extract sentences defining inclusion criteria of

each of the included trials in systematic reviews published in Cochrane Database of Systematic Reviews. While none of these are directly comparable to the present study because they extract different characteristics, prior work in the area suggests that automatic methods for data extraction may help to reduce the time and labour involved in systematic review processes.

Given that the Cochrane Database of Systematic Reviews uses one of the most structured formats for reporting systematic reviews and their updates, it is not clear how well the rule-based approach will work for other journals that make the full text of articles available. While full text articles from many journals are increasingly being made available for secondary use, it does not mean they are necessarily amenable to extraction. Structured and computable representations of systematic reviews would be of benefit for constructing new tools to support decisions about when a systematic review update is warranted. There are some efforts being made to support the linking of systematic reviews with their included trials [177], but this currently only includes links to trial registrations on ClinicalTrials.gov rather than published articles, and the number of participants may vary between the registration and the published article.

There were limitations to this study. First, we considered only one journal because it was known to report systematic reviews in a semi-structured format and have well-defined links between systematic reviews and their updates. It is not clear how well the approach would generalise to other journals. Second, we tested rules by taking one trial from each of the 50 systematic reviews, by assuming that pattern of reporting the participant information will be consistent for remaining included trials in each of these reviews.

4.4.5 Conclusion

We used a rule-based approach for extracting information from systematic reviews and their updates with the aim of using this information to support tools for estimating the risk of conclusion change in reviews that are yet to be updated. Preliminary results suggest that this approach can save time and effort that would be needed to manually extract information but the inconsistency with which participant numbers are reported remains a challenge. In a previous study we found that participant numbers were the most useful for estimating the risk of conclusion change, so it may be useful to

examine whether partial information about participant numbers is enough to produce a reliable estimate of conclusion change risk in systematic reviews not including the training of the model.

Chapter 5: Discussion

5.1 Chapter background

This chapter includes my editorial discussing the best places to focus efforts to meet the current challenges associated with systematic reviews. The editorial is also the part of solution that I recommend for meeting the both objectives of the thesis.

1. **R Bashir**, AG Dunn (2019) Software engineering principles address current problems in the systematic review ecosystem, Journal of Clinical Epidemiology, 109:136-141, doi:10.1016/j.jclinepi.2018.12.014.
2. P Martin, D Surian, **R Bashir**, FT Bourgeois, AG Dunn (2019) Trial2rev: Combining machine learning and crowd-sourcing to create a shared space for updating systematic reviews, JAMIA Open, ooy062, doi:10.1093/jamiaopen/ooy062.

3. D Surian, AG Dunn, L Orenstein, **R Bashir**, E Coiera, FT Bourgeois (2018) A shared latent space matrix factorisation method for recommending new trial evidence for systematic review updates, Journal of Biomedical Informatics, 79: 32-40, doi:10.1016/j.jbi.2018.01.008.

Author contributions: For the editorial above, I drafted the manuscript and Adam Dunn critically revised the manuscript. For the second manuscript, I played a minor role in the research, contributing to the data collection, advising on the use cases, and critically revising the manuscript. For the third manuscript, I helped with data collection and extraction, and critically revised the manuscript. The second and third manuscripts are not included here but are discussed in the conclusion section.

5.2 Applying software engineering principles to address current problems in the systematic review ecosystem

Abstract

Systematic reviewers are simultaneously unable to produce systematic reviews fast enough to keep up with the availability of new trial evidence while over-producing systematic reviews that are unlikely to change practice because they are redundant or biased. While the transparency and completeness of trial reporting has improved with changes in policy and new technologies, systematic reviews have not yet benefited from the same level of effort. We found that, new methods and tools used to automate aspects of systematic review processes have focused on improving the efficiency of individual systematic reviews rather than the efficiency of the entire ecosystem of systematic review production. We use software engineering principles to review challenges and opportunities for improving the interoperability, integrity, efficiency, and maintainability. We conclude by recommending ways to improve access to structured systematic review results. Major opportunities for improving systematic reviews will come from new tools and changes in policy focused on doing the right systematic reviews rather than just doing more of them faster.

5.2.1 What is new?

Key Findings

- Efforts aimed at improving systematic reviews have been focused on the quality or efficiency of performing individual reviews rather than on infrastructure to help avoid redundancy and monitor biases.

What this adds to what was known?

- There are a range of innovations aimed at improving the completeness and timeliness of trial reporting but connections across registries and bibliographic databases hinder systematic review production.
- Recent advances in the way trial study designs and results are represented in structured and machine-readable formats and stored in registries are not yet being fully utilised by systematic reviewers.

What is the implication and what should change now?

- Changes in policy and the culture of trial reporting could be expanded to cover systematic reviews, which could improve interoperability and efficiency.
- We propose establishing a centralised public repository for structured and machine-readable summaries of systematic reviews to match changes in the way clinical trials are registered and reported.

5.2.2 Current challenges in the systematic review ecosystem

Consider the systematic review ecosystem as the set of all systematic reviews in conjunction with the processes used to produce them. This includes the systems used to decide if and when a systematic review should be undertaken, the interactions with data sources used to synthesise primary research, and the processes for conducting, reporting, and publishing systematic reviews.

Systematic reviews of clinical interventions play an important role in the policy and practice of healthcare and should provide an up-to-date synthesis of available trials and other clinical studies. Ensuring that evidence synthesis is current is particularly important for recently-approved interventions, where the accumulation of new evidence might reveal safety issues and delays in their identification can cause harm [10]. Systematic reviews can also help to identify and mitigate the effect of publication and reporting biases [8], which result in delays in identifying safety issues. Despite rapid growth in the number of published systematic reviews, a substantial proportion are either redundant, conflicted, or have little clinical value [14]. Ensuring that systematic reviews do what they are meant to do is an ongoing challenge in the area.

Systematic reviews are resource-intensive and this hinders our ability to update them quickly enough to keep up with available evidence [12]. In response to this challenge, medical informatics specialists developed tools to support the automation of searching, screening, and synthesis of clinical evidence [27-30, 178]. However, these tools have typically aimed to reduce the effort required to undertake individual systematic review processes. Less effort has been used to develop informatics tools and methods for identifying which systematic reviews and clinical questions should be prioritised for review [17]. New methods and guidelines in this space may benefit the broader systematic review ecosystem by helping to improve the allocation of resources to systematic reviews that are at highest risk of a change in results of conclusions. Current approaches used to decide when to update a systematic review interpret the existing evidence and the amount of time that has elapsed since the review was undertaken [18, 146, 148]. Though there are several examples of the tools that consider the use of new evidence to predict the risk of change in conclusions, examples that evaluate the effectiveness of their tools are rare [31, 133, 157]. From a 2016 assessment of the systematic review

practices [14], it is evident that systematic reviewers still struggle with knowing in advance whether a systematic review is worthwhile.

Our aim was to examine current issues in systematic reviews, looking specifically at how well recent proposals and developments are addressing current challenges.

5.2.3 How can software engineering principles help address these problems?

In software engineering practice, systems are built to meet requirements related to a set of quality of service attributes, which describe how well a system behaves when it is implemented. In software systems, quality of service attributes describe aspects of the quality and performance of the systems and their behaviour. To define the quality of service attributes in the context of systematic review ecosystem means to them as a lens through which to examine the entire systematic review ecosystem at once, identifying gaps and opportunities for improvement.

There is no consensus or standard for list of quality of service attributes in software engineering. Therefore, the definitions of what constitutes a quality of service attribute vary [179], but four important attributes are common across most lists: interoperability, integrity, efficiency, and maintainability. We use these to frame an evaluation of current systematic review practices, encompassing the technologies, data, and resources used to produce them. For each attribute, we examine how well systematic review practices currently meet expectations and discuss the emerging initiatives and technologies that are aimed at addressing deficiencies.

Interoperability

In software engineering, interoperability relates to the capability of a system to interact with other systems. If we consider the set of all systematic reviews and the processes for producing them as our system, then interoperability is how well systematic reviews connect with trial registries, bibliographic databases, other sources of trial and study information, and the guidelines and summaries that make use of systematic reviews in practice. In software engineering, common approaches for ensuring interoperability might include using standardised data formats to ensure frictionless communication. Despite efforts to improve interoperability in the registration and reporting of clinical studies, much

less effort has been spent on connecting systematic reviews to the sources of information they use or the guidelines and policies that depend on them.

Some work has been done to improve the connectivity between systems in ways that may support interoperability with systematic reviews. There are different processes that have been used to establish interoperability between trial registries and bibliographic databases [180]. New methods have also been proposed to improve the proportion of trial registrations that include machine-readable links to bibliographic databases [181, 182]. However, systematic reviews are typically not reported in ways that make it easy to establish links to trial registries and bibliographic databases. The evidence transfer between published trials, registries, and systematic reviews is a largely ad hoc and unstructured. This is because there is a lack of standardisation and interoperability to enable cross-study analyses [37].

In 2005, Sim et al. [32] proposed the use of structured and computable reporting of trial results specifically to enhance interoperability and transparency, but progress in the space has taken many years [33]. Recently, Zarin et al. [36] suggested a further step towards interoperability by proposing the use of ClinicalTrials.gov as a central location for linking trials via their unique registry identifiers to their protocols, published results, and to any systematic reviews in which they are included. In a similar way, structured representations of systematic review registrations and results could improve interoperability through better interfacing with structured representation of trials.

Integrity

In software engineering, integrity is defined by the completeness and consistency of the data that are maintained by the system. For the systematic review ecosystem to demonstrate completeness and consistency, it would need to ensure that systematic reviews answering the same clinical question and specifying the same inclusion and exclusion criteria would include the same studies, and those studies would represent all relevant studies at the time of searching.

Incomplete representations of available evidence in systematic reviews are especially problematic when the studies that are included capture a biased subset of what is available. For example, where negative efficacy results are unpublished [183], or where safety outcomes are missing from reporting [184], systematic reviews may overestimate the efficacy and underestimate the harms of new

interventions. Statistical methods used to detect or account for publication bias are of limited value [185].

Interoperability also affects data integrity at external level where the links between registration and publication of included studies are missing. This situation makes hard to spot outcome reporting biases [60, 62, 63, 186], and selective reporting or missing outcome data [58, 59].

To address this challenge, systematic reviewers need to be able to efficiently access the complete results of trials. However, not every clinical trial gets published, which means that other sources of trial reporting become important. ClinicalTrials.gov in particular represents a very large source of structured summary results and may provide information for trials earlier and more completely. Studies examining the impact of searching for trial results in places other than in bibliographic databases conclude that trial registries are of some value [67, 89]. To be comprehensive, systematic reviewers need to consider all sources of clinical trial results information including bibliographic databases, ClinicalTrials.gov, and clinical study reports available directly from investigators, this is often challenging because of a lack of transparency in trial reporting and the effort required to search and screen multiple databases.

Efficiency

In software engineering, efficiency is defined as the degree to which a system performs without wasting resources. An efficient systematic review system is one that quickly incorporates new clinical evidence in systematic reviews without undertaking unnecessary effort or producing redundant systematic reviews. Our recent work showed that a substantial proportion of systematic review updates are not targeting the clinical questions where the evidence accumulates faster [187]. These results are aligned with the broader perspective that many systematic reviews are redundant and poorly focused where they are most needed [14]. We think that problems with efficiency in the systematic review ecosystem may contribute to the slow detection of safety issues in new interventions [10]. Wherever resources are being wasted on redundant systematic reviews, they could instead be targeted at updating systematic reviews with signals from recently reported trials. A study performed by Takwoingi et al. [31] is an example of a decision tool that estimates the risk of conclusion change in

systematic review update in advance of allocating resources to a systematic review. If we can improve tools of this type to work across a broader range of systematic reviews, we could improve the efficiency of the system by better targeting resources at clinical questions where conclusions are more likely to change.

Researchers have proposed a number of different approaches for deciding if and when a systematic review should be updated [18, 146, 148]. Novel approaches in this space use information from previous examples of systematic review updates to signal when a systematic review may be at risk of a change in results or conclusions [31, 188]. Factors used to predict which reviews are at high risk include the amount of time since the review was last updated, the number of trials and participants in the previous update, the attributes of the primary meta-analysis, or simple information about new and potentially relevant trials.

Maintainability

In software engineering, maintainability describes the capacity for a system to cope with or adapt to changes in its environment. For systematic reviews, this corresponds to the ability to adapt changes in the ways evidence is produced, synthesised, and disseminated, and how the systematic reviews are used by health providers, patients, and policy-makers. Maintainability is a challenge because it involves dealing with any changes in the resources used to produce systematic reviews, including trial registries, bibliographic databases, as well as the culture and funding of systematic reviewers. It also includes any changes to the way the results of systematic reviews are disseminated to stakeholders, such as summary reports, policy briefs, guidelines, and through news and social media.

One such change in environment has come from the perceived value of systematic reviews. Evidence about associations between conclusions with conflicts of interest and funding [189, 190] suggest that industry groups may be using systematic reviews as marketing tools. In addition, the growth in the number of available journals has made it easier to publish systematic reviews that are redundant or capture a biased subset of the available evidence. Each of these changes in practice may have introduced challenges to the credibility of systematic reviews. A number of guidelines have been made

available for managing the expected norms for systematic reviews [7, 91, 191, 192], but the use of these guidelines remains low [131].

Where are the best places to focus efforts now?

The substantial growth in the number of systematic reviews being produced suggests a recognition of their value in improving policy and practice in medicine, but this growth has also created challenges in interoperability, integrity, efficiency, and maintainability that have not yet been fully addressed. To meet these challenges, we suggest the combined efforts of the systematic reviewers and medical informatics communities.

We recommend the expanded use of standardised data formats for representing trials through their registrations and all forms of reporting including published articles, structured summary results, clinical study reports, and individual participant data [36]. The use of standardised representations of trial results data would benefit the interoperability of the system by providing a more complete representation of trial results available, which would in turn help to monitor and mitigate publication and reporting biases among prospectively registered trials. This will improve interoperability across trials that will have flow-on effects on systematic reviews—improving efficiency and integrity in the system by making it easier for systematic reviewers to account for all available trial evidence rather than just the subset of studies and outcomes reported in published articles. However, this would not be useful for systematic reviews based on old trials published before the inception of trial registration. Therefore, we can only recommend that the practice should be promoted for all newly registered trials, matching the requirements of some new policies in certain countries.

Part of the challenge of knowing when a systematic review is needed comes from keeping track of other similar systematic reviews that have been registered or published. Despite the substantial improvements that have been made to the transparency and completeness of trial reporting, systematic reviews have not seen the same level of scrutiny and development [193]. There is currently no centralised repository for identifying published systematic reviews along with information about their inclusion criteria and the outcomes they examine. Such a system would help systematic reviewers quickly ascertain if a given clinical question has been addressed by other systematic reviews and

would help manage the ecosystem by monitoring the conclusions of systematic reviews with equivalent inclusion and exclusion criteria. The availability of all key attributes in a structured format could be used to quickly evaluate overlapping systematic reviews [194], helping to avoid redundancy and improving both efficiency and maintainability. PROSPERO is a registry for systematic reviews [195]. In the same way that ClinicalTrials.gov started as a way to prospectively register trials and was expanded to include summary results data for registered trials, PROSPERO would be a logical choice for expanding to include structured summary results data for systematic reviews. Echoing the governance used to define required information in ClinicalTrials.gov, required information in a systematic review registry could be based on PRISMA and its extensions, and include links to identifiers of included trials and studies.

Biases in systematic reviews are complex. There is growing evidence of biases caused by delayed or missing publications, selective outcome reporting, differences in funding over time after new interventions are marketed, the financial conflicts of interest, and others. A registry of systematic reviews and their results would not immediately solve each of these problems, but if it were coupled with changes in policy and practice, could support improved surveillance and reduce duplication of effort.

Finally, we recommend continued pressure on systematic reviewers, funders, and journal editors to maintain expected standards for prospective registration and reporting for trials and systematic reviews. While the rates are improving, prospective trial registration is still not fully enforced across all medical journals [196], and even where systematic reviews are registered, they may not account for substantial changes in outcomes when published [197]. Similarly, investigators of trials should also understand their responsibility to update trial registries with current information after completion, and more countries and funding organisations should consider requiring the timely reporting of structured and machine-readable summary results for trials they fund.

5.3 Conclusion

In this thesis, I have examined and addressed several critical challenges facing systematic review practices. In a review of the literature I examined the processes used to link trial registries with bibliographic databases, showing that studies that use this information to evaluate biases depend critically on the availability of these links. The results showed that manual processes are still needed because of the absence of automatic links between trial registrations and their publications. These results emphasised the need to establish the links between all sources of clinical evidence.

I also sought to determine whether systematic reviews were updated earlier after a new, relevant study is first reported, finding no clear evidence of an association. While the lack of an association may be a consequence of the small proportion of reviews for which new evidence was added in an update, the results of the study nevertheless suggest the need to improve how decisions are made about updating systematic reviews in practice.

To develop a tool that could address this need, I extracted the features of systematic reviews and their updates to model the risk of conclusion change in systematic review updates. The results showed that extracting this information can be used as the basis for tools to decide about review updates. I then developed a rule-based approach for automatically extracting the features from systematic reviews and their updates. The results showed that a rule-based approach was useful for reducing labour involved in manually extracting features.

Using software engineering principles, I also examined the current challenges of the systematic review ecosystem and identified the gaps and opportunities for improvements. I found that many of the new developments from the field of clinical research informatics tend to be focused on improving the efficiency of individual systematic reviews and less often focused on broader notions of efficiency that currently lead to the over-production of redundant and unnecessary reviews.

From the research that I have undertaken as part of this thesis, I have several recommendations for improving the systematic review ecosystem. These include the adherence to standard reporting guidelines for studies, and ensuring that links to trial registration identifiers are included in the abstracts or metadata stored in bibliographic databases. My recommendations also include the

construction of a centralised repository of systematic reviews, reported in a structured and machine-readable form, and containing all key attributes, such as inclusion and exclusion criteria, outcomes of systematic review and the record of included ongoing and completed trials traceable back to trial registries and publications. This will ensure interoperability through forward and backward traceability of evidence, and will also help to identify overlapping systematic reviews to avoid redundancy.

In related work, I contributed to two additional studies [152, 177], in which the use of large databases of examples of systematic reviews linked to their included studies were used to support novel decision support tools. In the future, these tools may help to reduce or eliminate all efforts associated with searching and screening. These methods can also serve as inputs for a centralised repository by suggesting the relevant trials for systematic review updates. This will eventually help to ensure the completeness of evidence and efficiency of systematic reviews updates.

The future of systematic reviews should be based on connections between all sources of clinical evidence. This also involves the links to clinical trial registries because not all clinical trials are published [43, 45, 183]. Although the aspiration of fully automated systematic reviews might take longer to accomplish, current efforts in the area of machine learning and information retrieval can reduce systematic review workloads [144, 160, 165]. Future efforts will require a strong collaboration between systematic reviewers, computer scientists, and experts in health informatics.

5.4 Summary of contributions

In this thesis, I aimed to examine and address inefficiencies in the systematic review ecosystem, taking a software engineering perspective to identify the gaps and opportunities for improvement. In a review of the literature, I observed that there is a problem of linking between the clinical trials and their published results. Further, I found that there is no clear evidence that reviews were updated faster after new relevant studies were published, and most systematic review updates added no new studies. To deal with these challenges, I reported the need to improve the links between trial registries, bibliographic databases, and systematic reviews. I also highlighted the need to develop tools to support the prioritisation of review updates based on need. For this purpose, a set of features from systematic reviews and their updates were extracted for use in models of the risk of conclusion change in systematic review updates. Further, to improve the information extraction, rules were constructed to automatically extract information about systematic reviews and their included trials. The preliminary results showed that this rule-based approach could be used as a basis to create tools to estimate the risk of conclusion change.

The systematic review I reported in Chapter 2 was the first systematic review to examine the links between trial registries and bibliographic databases in a bidirectional way. Previous studies that examined the presence of automatic links provided broad analysis of availability of these links between ClinicalTrials.gov and PubMed, but they did not evaluate the proportion of other types of links that could be found through manual processes. Other studies used various types of links to evaluate transparency, completeness, and biases between the clinical trials and their published results. This systematic review was different in that I emphasised that links are important for interoperability between all sources of clinical evidence to improve the completeness, transparency, and the consistency of flow of evidence, while improper linking can lead to incorrect estimates of publication bias.

In Chapter 3, I reported an analysis of systematic review updates, where I examined whether systematic reviews were updated faster if they had an early signal that new trial evidence was available. In that analysis, I found that not only was there no clear evidence of an association, but also

that most systematic reviews added no new evidence when they are updated, and that only a small proportion of systematic reviews change conclusions.

My work in Chapter 3 highlighted the importance of developing tools that can be used to improve how we prioritise systematic review updates based on signals that an update may be warranted. Given what I described in Chapter 2, I decided that it would be useful to consider tools that use structured data from trial registries and bibliographic databases to learn what a systematic review at risk of conclusion change looks like.

In Chapter 4, I determined that whether the systematic review features can be useful to develop the tools that estimate the risk of conclusion change in systematic review updates. Though there are some tools available to help decide when systematic reviews are at risk of conclusion change, more general tools would be useful which consider the examples of systematic reviews and then use their features which are broadly available and relatively easy to extract. I then extracted a set of different features from systematic reviews and their updates to model the risk of conclusion change. The results showed that using the features extracted from systematic reviews and their updates could be used to estimate the risk of conclusion change in systematic review updates. In the second part of Chapter 4, I demonstrated a rule-based approach for extracting useful features directly from systematic review articles. The rule-based approach may be useful for increasing the number of systematic review updates that can be used to model the risk of conclusion change.

Looking at the development of new technologies for use with systematic reviews, the literature suggests that more effort has been spent on improving the processes of searching and screening in bibliographic databases after deciding to do a systematic review, and less effort has been spent on developing new tools to decide whether a systematic review is warranted in the first place. New tools could help make better decisions about when a systematic review is warranted, but these tools will rely on improvements in interoperability across all forms of clinical evidence and their syntheses. A structured repository of systematic reviews may be one solution, and a platform like trial2rev may be a useful starting point for building such a repository.

References

1. Gopalakrishnan S, Ganeshkumar P: **Systematic reviews and meta-analysis: understanding the best evidence in primary healthcare**. *J Family Med Prim Care* 2013, **2**(1):9.
2. Garritty C, Tsertsvadze A, Tricco AC, Sampson M, Moher D: **Updating systematic reviews: an international survey**. *PLoS ONE* 2010, **5**(4):e9914.
3. Mulrow CD, Cook DJ, Davidoff F: **Systematic reviews: critical links in the great chain of evidence**. *Ann Intern Med* 1997, **126**(5):389-391.
4. Will I: **An introduction to patient decision aids**. *BMJ* 2013, **347**:27.
5. Canestaro WJ, Hendrix N, Bansal A, Sullivan SD, Devine EB, Carlson JJ: **Favorable and publicly funded studies are more likely to be published: a systematic review and meta-analysis**. *J Clin Epidemiol* 2017, **92**:58-68.
6. Kirkham JJ, Altman DG, Chan A-W, Gamble C, Dwan KM, Williamson PR: **Outcome reporting bias in trials: a methodological approach for assessment and adjustment in systematic reviews**. *BMJ* 2018, **362**:k3802.
7. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, Davies P, Kleijnen J, Churchill R: **ROBIS: a new tool to assess risk of bias in systematic reviews was developed**. *J Clin Epidemiol* 2016, **69**:225-234.
8. Page MJ, McKenzie JE, Higgins JP: **Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review**. *BMJ Open* 2018, **8**(3):e019703.
9. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM, Woolf SH: **Validity of the Agency for Healthcare Research and Quality clinical practice guidelines: how quickly do guidelines become outdated?** *JAMA* 2001, **286**(12):1461-1467.
10. Downing NS, Shah ND, Aminawung JA, Pease AM, Zeitoun J-D, Krumholz HM, Ross JS: **Postmarket safety events among novel therapeutics approved by the US Food and Drug Administration between 2001 and 2010**. *JAMA* 2017, **317**(18):1854-1863.
11. Bastian H, Glasziou P, Chalmers I: **Seventy-five trials and eleven systematic reviews a day: how will we ever keep up?** *PLoS Med* 2010, **7**(9):e1000326.

12. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D: **How quickly do systematic reviews go out of date? A survival analysis**. *Ann Intern Med* 2007, **147**(4):224-233.

13. Sampson M, Shojania KG, Garritty C, Horsley T, Ocampo M, Moher D: **Systematic reviews can be produced and published faster**. *J Clin Epidemiol* 2008, **61**(6):531-536.

14. Ioannidis JP: **The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses**. *Milbank Q* 2016, **94**(3):485-514.

15. Bashir R, Surian D, Dunn AG: **Time-to-update of systematic reviews relative to the availability of new evidence**. *Syst Rev* 2018, **7**(1):195.

16. Chalmers I, Enkin M, Keirse MJ: **Preparing and updating systematic reviews of randomized controlled trials of health care**. *Milbank Q* 1993:411-437.

17. Garner P, Hopewell S, Chandler J, MacLehose H, Schünemann HJ, Akl EA, Beyene J, Chang S, Churchill R, Dearness K *et al*: **When and how to update systematic reviews: consensus and checklist**. *BMJ* 2016, **354**.

18. Higgins JP, Green S, Scholten R: **Maintaining reviews: updates, amendments and feedback**. *Cochrane Collaboration Cochrane handbook for systematic reviews of interventions Chichester: Wiley-Blackwell* 2008:31-46.

19. Moher D, Tsertsvadze A, Tricco A, Eccles M, Grimshaw J, Sampson M, Barrowman N: **When and how to update systematic reviews**. *The Cochrane Library* 2008.

20. Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N: **A systematic review identified few methods and strategies describing when and how to update systematic reviews**. *J Clin Epidemiol* 2007, **60**(11):1095-1095.

21. Pattanittum P, Laopaiboon M, Moher D, Lumbiganon P, Ngamjarus C: **A comparison of statistical methods for identifying out-of-date systematic reviews**. *PLoS One* 2012, **7**(11):e48894.

22. García LM, Pardo-Hernandez H, Superchi C, de Guzman EN, Ballesteros M, Roteta NI, McFarlane E, Posso M, i Figuls MR, del Campo RR: **Methodological systematic review identifies major limitations in prioritization processes for updating**. *J Clin Epidemiol* 2017, **86**:11-24.

23. Beller E, Clark J, Tsafnat G, Adams C, Diehl H, Lund H, Ouzzani M, Thayer K, Thomas J, Turner T: **Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR)**. *Syst Rev* 2018, **7**(1):77.

24. Boudin F, Nie J-Y, Bartlett JC, Grad R, Pluye P, Dawes M: **Combining classifiers for robust PICO element detection**. *BMC Med Inform Decis Mak* 2010, **10**(1):29.

25. Huang K-C, Chiang I-J, Xiao F, Liao C-C, Liu CC-H, Wong J-M: **PICO element detection in medical text without metadata: Are first sentences enough?** *J Biomed Inform* 2013, **46**(5):940-946.

26. Choong MK, Galgani F, Dunn AG, Tsafnat G: **Automatic evidence retrieval for systematic reviews**. *J Med Internet Res* 2014, **16**(10).

27. Miwa M, Thomas J, O'Mara-Eves A, Ananiadou S: **Reducing systematic review workload through certainty-based screening**. *J Biomed Inform* 2014, **51**:242-253.

28. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S: **Using text mining for study identification in systematic reviews: a systematic review of current approaches**. *Syst Rev* 2015, **4**(1):5.

29. Shemilt I, Simon A, Hollands GJ, Marteau TM, Ogilvie D, O'Mara-Eves A, Kelly MP, Thomas J: **Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews**. *Res Synth Methods* 2014, **5**(1):31-49.

30. Tsafnat G, Glasziou P, Choong M, Dunn A, Galgani F, Coiera E: **Systematic review automation technologies**. *Syst Rev* 2014, **3**(1):74.

31. Takwoingi Y, Hopewell S, Tovey D, Sutton AJ: **A multicomponent decision tool for prioritising the updating of systematic reviews**. *BMJ* 2013, **347**:f7191.

32. Sim I, Detmer DE: **Beyond trial registration: a global trial bank for clinical trial reporting**. *PLoS Med* 2005, **2**(11):e365.

33. Altman DG: **Making research articles fit for purpose: structured reporting of key methods and findings**. In.: Trials; 2015.

34. Schulz KF, Altman DG, Moher D: **CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials**. *BMC Med* 2010, **8**(1):18.

35. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P: **CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts**. *Ann Intern Med* 2017, **167**(1):40-47.

36. Zarin DA, Tse T: **Sharing individual participant data (IPD) within the context of the trial reporting system (TRS)**. *PLoS Med* 2016, **13**(1):e1001946.

37. van Valkenhoef G, Tervonen T, de Brock B, Hillege H: **Deficiencies in the transfer and availability of clinical trials evidence: a review of existing systems and standards**. *BMC Med Inform Decis Mak* 2012, **12**(1):95.

38. Dickersin K, Rennie D: **Registering clinical trials**. *JAMA* 2003, **290**(4):516-523.

39. McCray AT: **Better access to information about clinical trials**. *Ann Intern Med* 2000, **133**(8):609-614.

40. De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, Kotzin S, Laine C, Marusic A, Overbeke AJP: **Is this clinical trial fully registered?: A statement from the International Committee of Medical Journal Editors**. *N Engl J Med* 2005, **352**(23):2436-2438.

41. Manheimer E, Anderson D: **Survey of public information about ongoing clinical trials funded by industry: evaluation of completeness and accessibility**. *BMJ* 2002, **325**(7363):528-531.

42. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC: **The clinicaltrials. gov results database—update and key issues**. *N Engl J Med* 2011, **364**(9):852-860.

43. Bourgeois FT, Murthy S, Mandl KD: **Outcome reporting among drug trials registered in clinicaltrials. gov**. *Ann Intern Med* 2010, **153**(3):158-166.

44. Dwan K, Gamble C, Williamson PR, Kirkham JJ: **Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review**. *PLoS ONE* 2013, **8**(7):e66844.

45. Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF: **Non-publication of large randomized clinical trials: cross sectional analysis**. *BMJ* 2013, **347**:f6104.

46. Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P: **Timing and completeness of trial results posted at clinicaltrials. gov and published in journals**. *PLoS Med* 2013, **10**(12):e1001566.

47. Vawdrey DK, Hripcsak G: **Publication bias in clinical trials of electronic health records**. *J Biomed Inform* 2013, **46**(1):139-141.

48. De Jong JP, Ter Riet G, Willems DL: **Two prognostic indicators of the publication rate of clinical studies were available during ethical review**. *J Clin Epidemiol* 2010, **63**(12):1342-1350.

49. Gordon D, Taddei-Peters W, Mascette A, Antman M, Kaufmann PG, Lauer MS: **Publication of trials funded by the National Heart, Lung, and Blood Institute**. *NEJM* 2013, **369**(20):1926-1934.

50. Ioannidis JP: **Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials**. *JAMA* 1998, **279**(4):281-286.

51. Ross JS, Mocanu M, Lampropulos JF, Tse T, Krumholz HM: **Time to publication among completed clinical trials**. *JAMA Intern Med* 2013, **173**(9):825-828.

52. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM: **Trial publication after registration in clinicaltrials. gov: a cross-sectional analysis**. *PLoS Med* 2009, **6**(9):e1000144.

53. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM: **Publication of NIH funded trials registered in clinicaltrials. gov: cross sectional analysis**. *BMJ* 2012, **344**:d7292.

54. Thornton A, Lee P: **Publication bias in meta-analysis: its causes and consequences**. *J Clin Epidemiol* 2000, **53**(2):207-216.

55. De Oliveira GS, Jung MJ, McCarthy RJ: **Discrepancies between randomized controlled trial registry entries and content of corresponding manuscripts reported in anesthesiology journals**. *Anesth Analg* 2015, **121**(4):1030-1033.

56. Dwan K, Altman DG, Cresswell L, Blundell M, Gamble CL, Williamson PR: **Comparison of protocols and registry entries to published reports for randomised controlled trials**. *Cochrane Database Syst Rev* 2011, **1**.

57. Ewart R, Lausen H, Millian N: **Undisclosed changes in outcomes in randomized controlled trials: an observational study**. *Ann Fam Med* 2009, **7**(6):542-546.

58. Jones CW, Keil LG, Holland WC, Caughey MC, Platts-Mills TF: **Comparison of registered and published outcomes in randomized controlled trials: a systematic review**. *BMC Med* 2015, **13**(1):282.

59. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR: **The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews**. *BMJ* 2010, **340**:c365.

60. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P: **Comparison of registered and published primary outcomes in randomized controlled trials**. *JAMA* 2009, **302**(9):977-984.

61. Page MJ, McKenzie JE, Kirkham J, Dwan K, Kramer S, Green S, Forbes A: **Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions**. *Cochrane Database Syst Rev* 2014(10).

62. Pranić S, Marušić A: **Changes to registration elements and results in a cohort of Clinicaltrials. gov trials were not reflected in published articles**. *J Clin Epidemiol* 2016, **70**:26-37.

63. Tang E, Ravaud P, Riveros C, Perrodeau E, Dechartres A: **Comparison of serious adverse events posted at Clinicaltrials. gov and published in corresponding journal articles**. *BMC Med* 2015, **13**(1):189.

64. Walker KF, Stevenson G, Thornton JG: **Discrepancies between registration and publication of randomised controlled trials: an observational study**. *JRSM Open* 2014, **5**(5):2042533313517688.

65. Huser V, Cimino JJ: **Precision and negative predictive value of links between ClinicalTrials. gov and PubMed**. In: *AMIA Annu Symp Proc: 2012*: American Medical Informatics Association; 2012: 400.

66. Huser V, Cimino JJ: **Linking ClinicalTrials. gov and PubMed to track results of interventional human clinical trials**. *PLoS ONE* 2013, **8**(7):e68409.

67. Jones CW, Keil LG, Weaver MA, Platts-Mills TF: **Clinical trials registries are under-utilized in the conduct of systematic reviews: a cross-sectional analysis**. *Syst Rev* 2014, **3**(1):126.

68. Schroll JB, Bero L, Gøtzsche PC: **Searching for unpublished data for cochrane reviews: cross sectional study**. *BMJ* 2013, **346**:f2231.

69. Wolfe N, Gøtzsche PC, Bero L: **Strategies for obtaining unpublished drug trial data: a qualitative interview study**. *Syst Rev* 2013, **2**(1):31.

70. Cohen AM, Ambert K, McDonagh M: **Studying the potential impact of automated document classification on scheduling a systematic review update**. *BMC Med* 2012, **12**(1):33.

71. Garner P, Hopewell S, Chandler J, MacLehose H, Akl EA, Beyene J, Chang S, Churchill R, Dearness K, Guyatt G: **When and how to update systematic reviews: consensus and checklist**. *BMJ* 2016, **354**:i3507.

72. **Primary registries [last accessed: June 2, 2016]** [Available from: http://www.who.int/ictrp/network/primary/en/]

73. Bramer WM, Giustini D, Kramer BM: **Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study**. *Syst Rev* 2016, **5**(1):39.

74. Preston L, Carroll C, Gardois P, Paisley S, Kaltenthaler E: **Improving search efficiency for systematic reviews of diagnostic test accuracy: an exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking**. *Syst Rev* 2015, **4**(1):82.

75. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, Kotzin S, Laine C, Marusic A, Overbeke AJP: **Clinical trial registration: a statement from the International Committee of Medical Journal Editors**. *N Engl J Med* 2004, **351**(12):1250-1251.

76. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H: **Publication bias and clinical trials**. *Control Clin Trials* 1987, **8**.

77. Meinert CL: **Toward prospective registration of clinical trials**. *Control Clin Trials* 1988, **9**(1):1-5.

78. Simes RJ: **Publication bias: the case for an international registry of clinical trials**. *J Clin Oncol* 1986, **4**(10):1529-1541.

79. Califf R, Zarin D, Kramer J, Sherman R, Aberle L, Tasneem A: **Characteristics of clinical trials registered in clinicaltrials.gov, 2007-2010**. *JAMA* 2012, **307**(17):1838-1847.

80. Dwan K, Altman DG, Clarke M, Gamble C, Higgins JP, Sterne JA, Williamson PR, Kirkham JJ: **Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials**. *PLoS Med* 2014, **11**(6):e1001666.

81. Prayle AP, Hurley MN, Smyth AR: **Compliance with mandatory reporting of clinical trial results on clinicaltrials.gov: cross sectional study**. *BMJ* 2012, **344**:d7373.

82. Tse T, Williams RJ, Zarin DA: **Update on registration of clinical trials in clinicaltrials.gov**. *Chest* 2009, **136**(1):304-305.

83. Zarin DA, Tse T, Ide NC: **Trial registration at clinicaltrials.gov between May and October 2005**. *N Engl J Med* 2005, **353**(26):2779-2787.

84. Hooft L, Korevaar DA, Molenaar N, Bossuyt PM, Scholten RJ: **Endorsement of ICMJE's Clinical Trial Registration Policy: a survey among journal editors**. *The Netherlands journal of medicine* 2014, **72**(7):349-355.

85. No R: **Regulation No 536/2014 of the European Parliament and the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20**. *EC Official Journal of the European Union* 2014, **158**(27.5).

86. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C: **Systematic review of the empirical evidence of study publication bias and outcome reporting bias**. *PLoS One* 2008, **3**(8):e3081.

87. Huser V, Cimino JJ: **Evaluating adherence to the International Committee of Medical Journal Editors' policy of mandatory, timely clinical trial registration**. *Journal of the American Medical Informatics Association* 2013, **20**(e1):e169-e174.

88. Schmucker CM, Blümle A, Schell LK, Schwarzer G, Oeller P, Cabrera L, von Elm E, Briel M, Meerpohl JJ: **Systematic review finds that study data not published in full text articles have unclear impact on meta-analyses results in medical research**. *PLoS One* 2017, **12**(4):e0176210.

89. Baudard M, Yavchitz A, Ravaud P, Perrodeau E, Boutron I: **Impact of searching clinical trial registries in systematic reviews of pharmaceutical treatments: methodological systematic review and reanalysis of meta-analyses**. *BMJ* 2017, **356**:j448.

90. Bashir R, Dunn AG: **Systematic review protocol assessing the processes for linking clinical trial registries and their published results**. *BMJ Open* 2016, **6**(10):e013048.

91. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P: **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement**. *Ann Intern Med* 2010, **151**(4):264-269.

92. van de Wetering FT, Scholten RJ, Haring T, Clarke M, Hooft L: **Trial registration numbers are underreported in biomedical publications**. *PLoS One* 2012, **7**(11):e49599.

93. Carrasco M, Volkmar FR, Bloch MH: **Pharmacologic treatment of repetitive behaviors in autism spectrum disorders: evidence of publication bias**. *Pediatrics* 2012, **129**(5):e1301-e1310.

94. Nowbar AN, Mielewczik M, Karavassilis M, Dehbi H-M, Shun-Shin MJ, Jones S, Howard JP, Cole GD, Francis DP: **Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis**. *BMJ* 2014, **348**:g2688.

95. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R: **Selective publication of antidepressant trials and its influence on apparent efficacy**. *N Engl J Med* 2008, **358**(3):252-260.

96. Adewuyi TE, MacLennan G, Cook JA: **Non-compliance with randomised allocation and missing outcome data in randomised controlled trials evaluating surgical interventions: a systematic review**. *BMC Res Notes* 2015, **8**(1):1.

97. Bourgeois FT, Murthy S, Mandl KD: **Comparative effectiveness research: an empirical study of trials registered in clinicaltrials. gov**. *PLoS One* 2012, **7**(1):e28820.

98. Cole GD, Nowbar AN, Mielewczik M, Shun-Shin MJ, Francis DP: **Frequency of discrepancies in retracted clinical trial reports versus unretracted reports: blinded case-control study**. 2015.

99. Yoshimoto Y: **Publication bias in neurosurgery: lessons from series of unruptured aneurysms**. *Acta Neurochir* 2003, **145**(1):45-48.

100. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG: **Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles**. *JAMA* 2004, **291**(20):2457-2465.

101. Kasenda B, Von Elm E, You J, Blümle A, Tomonaga Y, Saccilotto R, Amstutz A, Bengough T, Meerpohl JJ, Stegert M: **Prevalence, characteristics, and publication of discontinued randomized trials**. *Jama* 2014, **311**(10):1045-1052.

102. Smyth R, Kirkham J, Jacoby A, Altman D, Gamble C, Williamson P: **Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists**. *BMJ* 2011, **342**:c7153.

103. Tharyan P, George AT, Kirubakaran R, Barnabas JP: **Reporting of methods was better in the clinical trials registry-India than in Indian journal publications**. *J Clin Epidemiol* 2013, **66**(1):10-22.

104. Harris P, Takeda A, Loveman E, Hartwell D: **Time to full publication of studies of anticancer drugs for breast cancer, and the potential for publication bias**. *International journal of technology assessment in health care* 2010, **26**(01):110-116.

105. Massey PR, Wang R, Prasad V, Bates SE, Fojo T: **Assessing the eventual publication of clinical trial abstracts submitted to a large annual oncology meeting**. *Oncologist* 2016, **21**(3):261-268.

106. Polyzos NP, Valachis A, Patavoukas E, Papanikolaou EG, Messinis IE, Tarlatzis BC, Devroey P: **Publication bias in reproductive medicine: from the European Society of Human Reproduction and Embryology annual meeting to publication**. *Hum Reprod* 2011, **26**(6):1371-1376.

107. Saeed M, Paulson K, Lambert P, Szwajcer D, Seftel M: **Publication bias in blood and marrow transplantation**. *Biol Blood Marrow Transplant* 2011, **17**(6):930-934.

108. Scherer RW, Huynh L, Ervin A-M, Dickersin K: **Using clinicaltrials. gov to supplement information in ophthalmology conference abstracts about trial outcomes: a comparison study**. *PLoS One* 2015, **10**(6):e0130619.

109. Moreno SG, Sutton AJ, Turner EH, Abrams KR, Cooper NJ, Palmer TM, Ades A: **Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications**. *BMJ* 2009, **339**:b2981.

110. Turner EH, Knoepflmacher D, Shapley L: **Publication bias in antipsychotic trials: an analysis of efficacy comparing the published literature to the US Food and Drug Administration database**. *PLoS Med* 2012, **9**(3):e1001189.

111. Withycombe B, Ovenell M, Meeker A, Ahmed SM, Hartung DM: **Timing of pivotal clinical trial results reporting for newly approved medications varied by reporting source**. *J Clin Epidemiol* 2016, **77**:78-83.

112. Chang L, Dhruva SS, Chu J, Bero LA, Redberg RF: **Selective reporting in trials of high risk cardiovascular devices: cross sectional comparison between premarket approval summaries and published reports**. *BMJ* 2015, **350**:h2613.

113. Miller JE, Korn D, Ross JS: **Clinical trial registration, reporting, publication and FDAAA compliance: a cross-sectional analysis and ranking of new drugs approved by the FDA in 2012**. *BMJ Open* 2015, **5**(11):e009758.

114. Mitchell AP, Hirsch BR, Abernethy AP: **Lack of timely accrual information in oncology clinical trials: a cross-sectional analysis**. *Trials* 2014, **15**(1):1.

115. Prasad V, Goldstein JA: **Clinical trial spots for cancer patients by tumour type: The cancer trials portfolio at clinicaltrials. gov**. *Eur J Cancer* 2015, **51**(17):2718-2723.

116. Liu J-P, Han M, Li X-X, Mu Y-J, Lewith G, Wang Y-Y, Witt CM, Yang G-Y, Manheimer E, Snellingen T: **Prospective registration, bias risk and outcome-reporting bias in randomised clinical trials of traditional chinese medicine: an empirical methodological study**. *BMJ Open* 2013, **3**(7):e002968.

117. Chapman SJ, Shelton B, Mahmood H, Fitzgerald JE, Harrison EM, Bhangu A: **Discontinuation and non-publication of surgical randomised controlled trials: observational study**. *BMJ* 2014, **349**:g6870.

118. Hill KD, Chiswell K, Califf RM, Pearson G, Li JS: **Characteristics of pediatric cardiovascular clinical trials registered on ClinicalTrials. gov**. *American heart journal* 2014, **167**(6):921-929. e922.

119. Harriman SL, Patel J: **When are clinical trials registered? An analysis of prospective versus retrospective registration**. *Trials* 2016, **17**(1):1.

120. Kunath F, Grobe HR, Keck B, Rücker G, Wullich B, Antes G, Meerpohl JJ: **Do urology journals enforce trial registration? A cross-sectional study of published trials**. *BMJ Open* 2011, **1**(2):e000430.

121. Rasmussen N, Lee K, Bero L: **Association of trial registration with the results and conclusions of published trials of new oncology drugs**. *Trials* 2009, **10**(1):1.

122. Rayhill ML, Sharon R, Burch R, Loder E: **Registration status and outcome reporting of trials published in core headache medicine journals**. *Neurology* 2015, **85**(20):1789-1794.

123. Scott A, Rucklidge JJ, Mulder RT: **Is mandatory prospective trial registration working to prevent publication of unregistered trials and selective outcome reporting? An observational study of five psychiatry journals that mandate prospective clinical trial registration**. *PLoS Med* 2015, **10**(8):e0133718.

124. Gotzsche PC: **Why we need a broad perspective on meta-analysis: it may be crucially important for patients**. *British Medical Journal* 2000, **321**(7261):585-585.

125. Jefferson T, Doshi P, Thompson M, Heneghan C: **Ensuring safe and effective drugs: who can do what it takes?** *Bmj* 2011, **342**:c7258.

126. Liberati A: **Need to realign patient-oriented and commercial and academic research**. *Lancet* 2011, **378**(9805):1777-1778.

127. Eccles M, Freemantle N, Mason J: **Using systematic reviews in clinical guideline development**. *Systematic Reviews in Health Care: Meta-Analysis in Context, Second Edition* 2001:400-409.

128. French SD, McDonald S, McKenzie JE, Green SE: **Investing in updating: how do conclusions change when Cochrane systematic reviews are updated?** *BMC Medical Research Methodology* 2005, **5**(1):33.

129. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D: **Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals**. *Jama* 1998, **280**(3):278-280.

130. Jaidee W, Moher D, Laopaiboon M: **Time to update and quantitative changes in the results of cochrane pregnancy and childbirth reviews**. *PLoS One* 2010, **5**(7):e11553.

131. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, Catalá-López F, Li L, Reid EK, Sarkis-Onofre R: **Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study**. *PLoS Med* 2016, **13**(5):e1002028.

132. Nasser M, Welch V, Tugwell P, Ueffing E, Doyle J, Waters E: **Ensuring relevance for cochrane reviews: evaluating processes and methods for prioritizing topics for cochrane reviews**. *J Clin Epidemiol* 2013, **66**(5):474-482.

133. Welsh E, Stovold E, Karner C, Cates C: **Cochrane Airways Group reviews were prioritized for updating using a pragmatic approach**. *J Clin Epidemiol* 2015, **68**(3):341-346.

134. Ahmadzai N, Newberry SJ, Maglione MA, Tsertsvadze A, Ansari MT, Hempel S, Motala A, Tsouros S, Chafen JJS, Shanman R: **A surveillance system to assess the need for updating systematic reviews**. *Syst Rev* 2013, **2**(1):104.

135. Chung M, Newberry SJ, Ansari MT, Winifred WY, Wu H, Lee J, Suttorp M, Gaylor JM, Motala A, Moher D: **Two methods provide similar signals for the need to update systematic reviews**. *J Clin Epidemiol* 2012, **65**(6):660-668.

136. Peterson K, McDonagh MS, Fu R: **Decisions to update comparative drug effectiveness reviews vary based on type of new evidence**. *J Clin Epidemiol* 2011, **64**(9):977-984.

137. Shekelle PG, Motala A, Johnsen B, Newberry SJ: **Assessment of a method to detect signals for updating systematic reviews**. *Syst Rev* 2014, **3**(1):13.

138. Dechartres A, Ravaud P, Atal I, Riveros C, Boutron I: **Association between trial registration and treatment effect estimates: a meta-epidemiological study**. *BMC Med* 2016, **14**(1):100.

139. Song SY, Koo D-H, Jung S-Y, Kang W, Kim EY: **The significance of the trial outcome was associated with publication rate and time to publication**. *J Clin Epidemiol* 2017, **84**:78-84.

140. Sutton D, Qureshi R, Martin J: **Evidence reversal—when new evidence contradicts current claims: a systematic overview review of definitions and terms**. *J Clin Epidemiol* 2018, **94**:76-84.

141. Boudin F, Nie J-Y, Bartlett JC, Grad R, Pluye P, Dawes M: **Combining classifiers for robust PICO element detection**. *BMC Med Inform Decis Mak* 2010, **10**(1):29.

142. Kim SN, Martinez D, Cavedon L, Yencken L: **Automatic classification of sentences to support evidence based medicine**. In: *BMC Bioinformatics: 2011*; 2011: S5.

143. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E: **Systematic review automation technologies**. *Syst Rev* 2014, **3**(1):74.

144. Wallace BC, Trikalinos TA, Lau J, Brodley C, Schmid CH: **Semi-automated screening of biomedical citations for systematic reviews**. *BMC Bioinformatics* 2010, **11**(1):55.

145. Bagheri E, Rios P, Pourmasoumi A, Robson RC, Hwee J, Isaranuwatchai W, Darvesh N, Page MJ, Tricco AC: **Improving the conduct of systematic reviews: a process mining perspective**. *J Clin Epidemiol* 2018, **103**:101-111.

146. Barrowman NJ, Fang M, Sampson M, Moher D: **Identifying null meta-analyses that are ripe for updating**. *BMC Med Res Methodol* 2003, **3**(1):13.

147. Sampson M, de Bruijn B, Urquhart C, Shojania K: **Complementary approaches to searching MEDLINE may be sufficient for updating systematic reviews**. *J Clin Epidemiol* 2016, **78**:108-115.

148. Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR: **Evidence-based sample size calculations based upon updated meta-analysis**. *Stat Med* 2007, **26**(12):2479-2500.

149. Sutton AJ, Donegan S, Takwoingi Y, Garner P, Gamble C, Donald A: **An encouraging assessment of methods to inform priorities for updating systematic reviews**. *J Clin Epidemiol* 2009, **62**(3):241-251.

150. McHugh ML: **Interrater reliability: the kappa statistic**. *Biochem Med* 2012, **22**(3):276-282.

151. Shao W, Adams CE, Cohen AM, Davis JM, McDonagh MS, Thakurta S, Philip SY, Smalheiser NR: **Aggregator: a machine learning approach to identifying MEDLINE articles that derive from the same underlying clinical trial**. *Methods* 2015, **74**:65-70.

152. Surian D, Dunn AG, Orenstein L, Bashir R, Coiera E, Bourgeois FT: **A shared latent space matrix factorisation method for recommending new trial evidence for systematic review updates**. *J Biomed Inform* 2018, **79**:32-40.

153. Peiris CL, Shields N, Brusco NK, Watts JJ, Taylor NF: **Additional physical therapy services reduce length of stay and improve health outcomes in oeople with acute and subacute**

conditions: an updated systematic review and meta-analysis. *Arch Phys Med Rehabil* 2018, **99**:2299-2312.

154. Peiris CL, Taylor NF, Shields N: **Extra physical therapy reduces patient length of stay and improves functional outcomes and quality of life in people with acute or subacute conditions: a systematic review**. *Arch Phys Med Rehabil* 2011, **92**(9):1490-1500.

155. Sharma VK, Ng KW, Venketasubramanian N, Saqqur M, Teoh HL, Kaul S, Srivastava PM, Sergentanis T, Suwanwela N, Nguyen TH: **Current status of intravenous thrombolysis for acute ischemic stroke in Asia**. *Int J Stroke* 2011, **6**(6):523-530.

156. Wang X, You S, Sato S, Yang J, Carcel C, Zheng D, Yoshimura S, Anderson CS, Sandset EC, Robinson T: **Current status of intravenous tissue plasminogen activator dosage for acute ischaemic stroke: an updated systematic review**. *Stroke Vasc Neurol* 2018:svn-2017-000112.

157. Langan D, Higgins JP, Gregory W, Sutton AJ: **Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis**. *J Clin Epidemiol* 2012, **65**(5):511-519.

158. P Martin DS, R Bashir, FT Bourgeois, AG Dunn: **Trial2rev: Combining machine learning and crowd-sourcing to create a shared space for updating systematic reviews**. *JAMIA Open* 2019.

159. Dalal SR, Shekelle PG, Hempel S, Newberry SJ, Motala A, Shetty KD: **A pilot study using machine learning and domain knowledge to facilitate comparative effectiveness review updating**. *Med Decis Making* 2013, **33**(3):343-355.

160. Shekelle PG, Shetty K, Newberry S, Maglione M, Motala A: **Machine learning versus standard techniques for updating searches for systematic reviews: a diagnostic accuracy study**. *Ann Intern Med* 2017, **167**(3):213-215.

161. Bashir R, Dunn AG: **Software engineering principles address current problems in the systematic review ecosystem**. *J Clin Epidemiol* 2019, **109**:136-141.

162. Tricco AC, Cardoso R, Thomas SM, Motiwala S, Sullivan S, Kealey MR, Hemmelgarn B, Ouimet M, Hillmer MP, Perrier L: **Barriers and facilitators to uptake of systematic reviews**

by policy makers and health care managers: a scoping review. *Implement Sci* 2015, **11**(1):4.

163. Murthy L, Shepperd S, Clarke MJ, Garner SE, Lavis JN, Perrier L, Roberts NW, Straus SE: **Interventions to improve the use of systematic reviews in decision-making by health system managers, policy makers and clinicians**. *Cochrane Database Syst Rev* 2012(9).

164. Bashir R, Surian D, Dunn AG: **The risk of conclusion change in systematic review updates can be estimated by learning from a database of published examples**. *J Clin Epidemiol* 2019.

165. Bannach-Brown A, Przybyła P, Thomas J, Rice AS, Ananiadou S, Liao J, Macleod MR: **Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error**. *Syst Rev* 2019, **8**(1):23.

166. Howard BE, Phillips J, Miller K, Tandon A, Mav D, Shah MR, Holmgren S, Pelch KE, Walker V, Rooney AA: **SWIFT-Review: a text-mining workbench for systematic review**. *Syst Rev* 2016, **5**(1):87.

167. Liu J, Timsina P, El-Gayar O: **A comparative analysis of semi-supervised learning: the case of article selection for medical systematic reviews**. *Inf Syst Front* 2018, **20**(2):195-207.

168. Matwin S, Kouznetsov A, Inkpen D, Frunza O, O'blenis P: **A new algorithm for reducing the workload of experts in performing systematic reviews**. *J Am Med Inform Assoc* 2010, **17**(4):446-453.

169. Cohen AM, Hersh WR, Peterson K, Yen P-Y: **Reducing workload in systematic review preparation using automated citation classification**. *Journal of the American Medical Informatics Association* 2006, **13**(2):206-219.

170. Khabsa M, Elmagarmid A, Ilyas I, Hammady H, Ouzzani M: **Learning to identify relevant studies for systematic reviews using random forest and external information**. *Machine Learning* 2016, **102**(3):465-482.

171. Wallace BC, Kuiper J, Sharma A, Zhu M, Marshall IJ: **Extracting PICO sentences from clinical trial reports using supervised distant supervision**. *The Journal of Machine Learning Research* 2016, **17**(1):4572-4596.

172. Marshall IJ, Kuiper J, Banner E, Wallace BC: **Automating biomedical evidence synthesis: RobotReviewer**. In: *Proceedings of the conference Association for Computational Linguistics Meeting: 2017*: NIH Public Access; 2017: 7.

173. Basu T, Kumar S, Kalyan A, Jayaswal P, Goyal P, Pettifer S, Jonnalagadda SR: **A novel framework to expedite systematic reviews by automatically building information extraction training corpora**. *arXiv preprint arXiv:160606424* 2016.

174. **Web scraping [last accessed: June 7, 2019]** [Available from: https://en.wikipedia.org/wiki/Web_scraping]

175. Kiritchenko S, De Bruijn B, Carini S, Martin J, Sim I: **ExaCT: automatic extraction of clinical trial characteristics from journal publications**. *BMC Med Inform Decis Mak* 2010, **10**(1):56.

176. Borlawsky T, Friedman C, Lussier YA: **Generating executable knowledge for evidence-based medicine using natural language and semantic processing**. In: *AMIA Annual Symposium proceedings: 2006*; 2006.

177. Martin P, Surian D, Bashir R, Bourgeois FT, Dunn AG: **Trial2rev: Combining machine learning and crowd-sourcing to create a shared space for updating systematic reviews**. *JAMIA Open* 2019.

178. Jonnalagadda SR, Goyal P, Huffman MD: **Automating data extraction in systematic reviews: a systematic review**. *Syst Rev* 2015, **4**(1):78.

179. A.T. Borgida VKC, P. Giorgini, E.S. Yu (Eds.): **On non-functional requirements in software engineering**: Springer, New York, NY.

180. Bashir R, Bourgeois FT, Dunn AG: **A systematic review of the processes used to link clinical trial registrations to their published results**. *Syst Rev* 2017, **6**(1):123.

181. Dunn AG, Coiera E, Bourgeois FT: **Unreported links between trial registrations and published articles were identified using document similarity measures in a cross-sectional analysis of clinicaltrials. gov**. *J Clin Epidemiol* 2018, **95**:94-101.

182. Goodwin TR, Skinner MA, Harabagiu SM: **Automatically linking registered clinical trials to their published results with deep highway networks**. *AMIA Jt Summits Transl Sci Proc* 2018, **2017**:54.

183. Schmucker C, Schell LK, Portalupi S, Oeller P, Cabrera L, Bassler D, Schwarzer G, Scherer RW, Antes G, Von Elm E: **Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries**. *PLoS One* 2014, **9**(12):e114023.

184. Golder S, Loke YK, Wright K, Norman G: **Reporting of adverse events in published and unpublished studies of health care interventions: a systematic review**. *PLoS Med* 2016, **13**(9):e1002127.

185. Jin ZC, Zhou XH, He J: **Statistical methods for dealing with publication bias in meta-analysis**. *Stat Med* 2015, **34**(2):343-360.

186. Su C-X, Han M, Ren J, Li W-Y, Yue S-J, Hao Y-F, Liu J-P: **Empirical evidence for outcome reporting bias in randomized clinical trials of acupuncture: comparison of registered records and subsequent publications**. *Trials* 2015, **16**(1):1.

187. Bashir R DA: **Do systematic review updates target questions where evidence accumulates faster?** . In: *Evidence Live.* Oxford, UK.

188. **An empirically-defined decision tree to predict systematic reviews at risk of change in conclusion**. In: *25th Cochrane Colloquium.* Edinburgh, Scotland.

189. Bes-Rastrollo M, Schulze MB, Ruiz-Canela M, Martinez-Gonzalez MA: **Financial conflicts of interest and reporting bias regarding the association between sugar-sweetened beverages and weight gain: a systematic review of systematic reviews**. *PLoS Med* 2013, **10**(12):e1001578.

190. Dunn AG, Arachi D, Hudgins J, Tsafnat G, Coiera E, Bourgeois FT: **Financial conflicts of interest and conclusions about neuraminidase inhibitors for influenza: an analysis of systematic reviews**. *Ann Intern Med* 2014, **161**(7):513-518.

191. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA: **Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation**. *BMJ* 2015, **349**:g7647.

192. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E: **AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both**. *BMJ* 2017, **358**:j4008.

193. Trinquart L, Dunn AG, Bourgeois FT: **Registration of published randomized trials: a systematic review and meta-analysis**. *BMC Med* 2018, **16**(1):173.

194. Riva N, Puljak L, Moja L, Ageno W, Schünemann H, Magrini N, Squizzato A: **Multiple overlapping systematic reviews facilitate the origin of disputes: the case of thrombolytic therapy for pulmonary embolism**. *J Clin Epidemiol* 2018, **97**:1-13.

195. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, Stewart L: **The nuts and bolts of PROSPERO: an international prospective register of systematic reviews**. *Syst Rev* 2012, **1**(1):2.

196. Dal-Ré R, Ross JS, Marušić A: **Compliance with prospective trial registration guidance remained low in high-impact journals and has implications for primary end point reporting**. *J Clin Epidemiol* 2016:100-107.

197. Page MJ, Shamseer L, Tricco AC: **Registration of systematic reviews in PROSPERO: 30,000 records and counting**. *Syst Rev* 2018, **7**(1):32.

198. Hartung DM, Zarin DA, Guise J-M, McDonagh M, Paynter R, Helfand M: **Reporting discrepancies between the ClinicalTrials. gov results database and peer-reviewed publications**. *Annals of internal medicine* 2014, **160**(7):477-483.

199. Xuemei L, Youping L, Senlin Y, Shangqi S: **Result publication of Chinese trials in World Health Organization primary registries**. *PLoS One* 2010, **5**(9):e12676.

200. Prenner JL, Driscoll SJ, Fine HF, Salz DA, Roth DB: **Publication rates of registered clinical trials in macular degeneration**. *Retina* 2011, **31**(2):401-404.

201. Wildt S, Krag A, Gluud L: **Characteristics of randomised trials on diseases in the digestive system registered in clinicaltrials. gov: a retrospective analysis**. *BMJ Open* 2011, **1**(2):e000309.

202. Gandhi R, Jan M, Smith HN, Mahomed NN, Bhandari M: **Comparison of published orthopaedic trauma trials following registration in Clinicaltrials. gov**. *BMC musculoskeletal disorders* 2011, **12**(1):1.

203. Shamliyan T, Kane RL: **Clinical research involving children: registration, completeness, and publication**. *Pediatrics* 2012, **129**(5):1291-1300.

204. Stockmann C, Sherwin CM, Koren G, Campbell SC, Constance JE, Linakis M, Balch A, Varner MW, Spigarelli MG: **Characteristics and publication patterns of obstetric studies registered in clinicaltrials. gov**. *J Clin Pharmacol* 2014, **54**(4):432-437.

205. Korevaar DA, Ochodo EA, Bossuyt PM, Hooft L: **Publication and reporting of test accuracy studies registered in Clinicaltrials. gov**. *Clin Chem* 2014, **60**(4):651-659.

206. Munch T, Dufka FL, Greene K, Smith SM, Dworkin RH, Rowbotham MC: **RReACT goes global: perils and pitfalls of constructing a global open-access database of registered analgesic clinical trials and trial results**. *Pain* 2014, **155**(7):1313-1317.

207. Khan NA, Singh M, Spencer HJ, Torralba KD: **Randomized controlled trials of rheumatoid arthritis registered at ClinicalTrials. gov: what gets published and when**. *Arthritis & Rheumatology* 2014, **66**(10):2664-2674.

208. Hakala A, Kimmelman J, Carlisle B, Freeman G, Fergusson D: **Accessibility of trial reports for drugs stalling in development: a systematic assessment of registered trials**. *BMJ* 2015, **350**:h1116.

209. Boccia S, Rothman KJ, Panic N, Flacco ME, Rosso A, Pastorino R, Manzoli L, La Vecchia C, Villari P, Boffetta P: **Registration practices for observational studies on clinicaltrials. gov indicated low adherence**. *J Clin Epidemiol* 2016, **70**:176-182.

210. Saito H, Gill CJ: **How frequently do the results from completed US clinical trials enter the public domain? A statistical analysis of the clinicaltrials. gov database**. *PLoS One* 2014, **9**(7):e101826.

211. Son C, Tavakoli S, Bartanusz V: **No publication bias in industry funded clinical trials of degenerative diseases of the spine**. *J Clin Neurosci* 2016, **25**:58-61.

212. Baudart M, Ravaud P, Baron G, Dechartres A, Haneef R, Boutron I: **Public availability of results of observational studies evaluating an intervention registered at clinicaltrials. gov**. *BMC Med* 2016, **14**(1):1.

213. Chahal J, Tomescu SS, Ravi B, Bach BR, Ogilvie-Harris D, Mohamed NN, Gandhi R: **Publication of sports medicine–related randomized controlled trials registered in clinicaltrials. gov**. *Am J Sports Med* 2012, **40**(9):1970-1977.

214. Manzoli L, Flacco ME, D'Addario M, Capasso L, De Vito C, Marzuillo C, Villari P, Ioannidis JP: **Non-publication and delayed publication of randomized trials on vaccines: survey**. *BMJ* 2014.

215. Lebensburger JD, Hilliard LM, Pair LE, Oster R, Howard TH, Cutter GR: **Systematic review of interventional sickle cell trials registered in ClinicalTrials. gov**. *Clin Trials* 2015:1740774515590811.

216. Smith HN, Bhandari M, Mahomed NN, Jan M, Gandhi R: **Comparison of arthroplasty trial publications after registration in ClinicalTrials. gov**. *J Arthroplasty* 2012, **27**(7):1283-1288.

217. Guo S-W, Evers JL: **Lack of transparency of clinical trials on endometriosis**. *Obstetrics & Gynecology* 2013, **121**(6):1281-1290.

218. Tsikkinis A, Cihoric N, Giannarini G, Hinz S, Briganti A, Wust P, Ost P, Ploussard G, Massard C, Surcel CI: **Clinical perspectives from randomized phase 3 trials on prostate cancer: An analysis of the clinicaltrials. gov database**. *Eur Urol Focus* 2015, **1**(2):173-184.

219. Chen R, Desai NR, Ross JS, Zhang W, Chau KH, Wayda B, Murugiah K, Lu DY, Mittal A, Krumholz HM: **Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers**. *BMJ* 2016, **352**:i637.

220. Ramsey S, Scoggins J: **Commentary: practicing on the tip of an information iceberg? Evidence of underpublication of registered clinical trials in oncology**. *Oncologist* 2008, **13**(9):925-929.

221. Hurley M, Prayle A, Smyth AR: **Delayed publication of clinical trials in cystic fibrosis**. *Journal of Cystic Fibrosis* 2012, **11**(1):14-17.

222. Ioannidis JP, Manzoli L, De Vito C, D'Addario M, Villari P: **Publication delay of randomized trials on 2009 influenza A (H1N1) vaccination**. *PLoS One* 2011, **6**(12):e28346.

223. Ohnmeiss DD: **The fate of prospective spine studies registered on www. ClinicalTrials. gov**. *Spine J* 2015, **15**(3):487-491.

224. Gopal RK, Yamashita TE, Prochazka AV: **Research without results: inadequate public reporting of clinical trial results**. *Contemporary clinical trials* 2012, **33**(3):486-491.

225. Lampert A, Hoffmann GF, Ries M: **Ten years after the International Committee of Medical Journal Editors' clinical trial registration initiative, one quarter of phase 3 pediatric epilepsy clinical trials still remain unpublished: A cross sectional analysis**. *PLoS One* 2016, **11**(1):e0144973.

226. Gandhi GY, Murad MH, Fujiyoshi A, Mullan RJ, Flynn DN, Elamin MB, Swiglo BA, Isley WL, Guyatt GH, Montori VM: **Patient-important outcomes in registered diabetes trials**. *JAMA* 2008, **299**(21):2543-2549.

227. Chowers MY, Gottesman BS, Leibovici L, Pielmeier U, Andreassen S, Paul M: **Reporting of adverse events in randomized controlled trials of highly active antiretroviral therapy: systematic review**. *J Antimicrob Chemothe* 2009, **64**(2):239-250.

228. You B, Gan HK, Pond G, Chen EX: **Consistency in the analysis and reporting of primary end points in oncology randomized controlled trials from registration to publication: a systematic review**. *J Clin Oncol* 2011, **30**(2):210-216.

229. Reveiz L, Bonfill X, Glujovsky D, Pinzon CE, Asenjo-Lobos C, Cortes M, Canon M, Bardach A, Comandé D, Cardona AF: **Trial registration in Latin America and the Caribbean's: study of randomized trials published in 2010**. *J Clin Epidemiol* 2012, **65**(5):482-487.

230. Nankervis H, Baibergenova A, Williams HC, Thomas KS: **Prospective registration and outcome-reporting bias in randomized controlled trials of eczema treatments: a systematic review**. *J Invest Dermatol* 2012, **132**(12):2727-2734.

231. Pinto RZ, Elkins MR, Moseley AM, Sherrington C, Herbert RD, Maher CG, Ferreira PH, Ferreira ML: **Many randomized trials of physical therapy interventions are not adequately registered: a survey of 200 published trials**. *Phys Ther* 2012, **93**(3):299-309.

232. Hannink G, Gooszen HG, Rovers MM: **Comparison of registered and published primary outcomes in randomized clinical trials of surgical interventions**. *Annals of surgery* 2013, **257**(5):818-823.

233. Rosenthal R, Dwan K: **Comparison of randomized controlled trial registry entries and content of reports in surgery journals**. *Ann Surg* 2013, **257**(6):1007-1015.

234. Hopewell S, Collins GS, Hirst A, Kirtley S, Tajar A, Gerry S, Altman DG: **Reporting characteristics of non-primary publications of results of randomized trials: a cross-sectional review**. *Trials* 2013, **14**(1):1.

235. Babu AS, Veluswamy SK, Rao PT, Maiya AG: **Clinical trial registration in physical therapy journals: a cross-sectional study**. *Phys Ther* 2013.

236. Lee S-Y, Teoh P, Camm C, Agha R: **Compliance of randomised controlled trials in trauma surgery with the consort statement: a systematic review**. *Trials* 2013, **14**(1):1.

237. Li X-Q, Yang G-L, Tao K-M, Zhang H-Q, Zhou Q-H, Ling C-Q: **Comparison of registered and published primary outcomes in randomized controlled trials of gastroenterology and hepatology**. *Scand J Gastroenterol* 2014.

238. Norris SL, Holmer HK, Fu R, Ogden LA, Viswanathan MS, Abou-Setta AM: **Clinical trial registries are of minimal use for identifying selective outcome and analysis reporting**. *Res Synth Methods* 2014, **5**(3):273-284.

239. Hardt JL, Metzendorf M-I, Meerpohl JJ: **Surgical trials and trial registers: a cross-sectional study of randomized controlled trials published in journals requiring trial registration in the author instructions**. *Trials* 2013, **14**(1):1.

240. Anand V, Scales DC, Parshuram CS, Kavanagh BP: **Registration and design alterations of clinical trials in critical care: a cross-sectional observational study**. *Intensive Care Med* 2014, **40**(5):700-722.

241. Mann E, Nguyen N, Fleischer S, Meyer G: **Compliance with trial registration in five core journals of clinical geriatrics: a survey of original publications on randomised controlled trials from 2008 to 2012**. *Age and Ageing* 2014, **43**(6):872-876.

242. Dekkers OM, Cevallos M, Bührer J, Poncet A, Rau SA, Perneger TV, Egger M: **Comparison of noninferiority margins reported in protocols and publications showed incomplete and inconsistent reporting**. *J Clin Epidemiol* 2015, **68(**5):510-517.

243. Østervig R, Sonne A, Rasmussen L: **Registration of randomized clinical trials–a challenge**. *Acta Anaesthesiol Scand* 2015, **59**(8):986-989.

244. De Oliveira Jr GS, Jung MJ, McCarthy RJ: **Discrepancies between randomized controlled trial registry entries and content of corresponding manuscripts reported in anesthesiology journals**. *Anesth Analg* 2015, **121**(4):1030-1033.

245. Reveiz L, Cortés-Jofré M, Lobos CA, Nicita G, Ciapponi A, Garcìa-Dieguez M, Tellez D, Delgado M, Solà I, Ospina E: **Influence of trial registration on reporting quality of randomized trials: study from highest ranked journals**. *J Clin Epidemiol* 2010, **63**(11):1216-1222.

246. Rongen JJ, Hannink G: **Comparison of registered and published primary outcomes in randomized controlled trials of orthopaedic surgical interventions**. *J Bone Joint Surg Am* 2016, **98**(5):403-409.

247. Vera-Badillo F, Shapiro R, Ocana A, Amir E, Tannock IF: **Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer**. *Ann Oncol* 2013, **24**(5):1238-1244.

248. McGee RG, Su M, Kelly PJ, Higgins GY, Craig JC, Webster AC: **Trial registration and declaration of registration by authors of randomized controlled trials**. *Transplantation* 2011, **92**(10):1094-1100.

249. Huić M, Marušić M, Marušić A: **Completeness and changes in registered data and reporting bias of randomized controlled trials in ICMJE journals after trial registration policy**. *PLoS One* 2011, **6**(9):e25258.

250. Chan A-W, Altman DG: **Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors**. *BMJ* 2005, **330**(7494):753.

251. Korevaar DA, Bossuyt PM, Hooft L: **Infrequent and incomplete registration of test accuracy studies: analysis of recent study reports**. *BMJ Open* 2014, **4**(1):e004596.

252. Jones CW, Platts-Mills TF: **Quality of registration for clinical trials published in emergency medicine journals**. *Annals of emergency medicine* 2012, **60**(4):458-464. e451.

253. Smaïl-Faugeron V, Fron-Chabouis H, Durieux P: **Clinical trial registration in oral health journals**. *J Dent Res* 2015, **94**(3 suppl):8S-13S.

254. Riehm KE, Azar M, Thombs BD: **Transparency of outcome reporting and trial registration of randomized controlled trials in top psychosomatic and behavioral health journals: A 5-year follow-up**. *J Psychosom Res* 2015, **79**(1):1-12.

Appendix

Appendix Table 2.1: Search strategy used for PubMed and EMBASE bibliographic databases

	Search Terms
#1	Any registry from WHO ICTRP OR “trial registry"[Title/Abstract] OR "trial register"[Title/Abstract] OR "trial registries"[Title/Abstract] OR "trials registry" [Title/Abstract] OR "registry of clinical trials" [Title/Abstract])
#2	("trial registration"[Title/Abstract] AND (discrepancy[Title/Abstract] OR discrepancies[Title/Abstract] OR consistency[Title/Abstract] OR inconsistency[Title/Abstract]))
#3	#1 OR #2
#4	(unregistered[Title/Abstract] OR non-publication[Title/Abstract] OR nonpublication[Title/Abstract] OR unpublished[Title/Abstract] OR published[Title/Abstract] OR (registered[Title/Abstract] AND (publication[Title/Abstract] OR clinical trial as topic [MeSH Terms]))) Note: for EMBASE equivalent EMTREE is "clinical trial (topic)"
#5	#3 AND #4
#6	("outcome reporting bias"[Title/Abstract] OR "selective reporting"[Title/Abstract] OR "selective outcome reporting"[Title/Abstract] OR "missing outcome data"[Title/Abstract] OR "publication bias"[MeSH Terms] OR ("reporting quality"[Title/Abstract] AND publications[Title/Abstract])) Note: for EMBASE the publication bias MeSH Term has no equivalent and is removed.
#7	#5 OR #6

Appendix Table 2.2: Search strategy for MEDLINE via PubMed

	Search Terms
#1	clinicaltrials.gov[Title/Abstract]
#2	ANZCTR[Title/Abstract]
#3	ICTRP[Title/Abstract]
#4	ReBec[Title/Abstract]
#5	ChiCTR[Title/Abstract]
#6	CRiS[Title/Abstract]
#7	CTRI[Title/Abstract]
#8	RPCEC[Title/Abstract]
#9	EU-CTR[Title/Abstract]
#10	DRKS[Title/Abstract]
#11	IRCT[Title/Abstract]
#12	JPRN[Title/Abstract]
#13	NTR[Title/Abstract]
#14	ISRCTN[Title/Abstract]
#15	PACTR[Title/Abstract]
#16	SLCTR[Title/Abstract]
#17	trial registry"[Title/Abstract]
#18	"trial register"[Title/Abstract]
#19	"trial registries"[Title/Abstract]
#20	"trials registry" [Title/Abstract]
#21	"registry of clinical trials" [Title/Abstract])]
#22	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21
#23	"trial registration"[Title/Abstract]
#24	discrepancy[Title/Abstract]
#25	discrepancies[Title/Abstract]
#26	consistency[Title/Abstract]
#27	inconsistency[Title/Abstract]
#28	#24 or #25 or #26 or #27
#29	#23 and #28
#30	#22 or #29
#31	unregistered[Title/Abstract]
#32	non-publication[Title/Abstract]
#33	nonpublication[Title/Abstract]
#34	unpublished[Title/Abstract]
#35	published[Title/Abstract]
#36	registered[Title/Abstract]
#37	#31 or #32 or #33 or #34 or #35 or #36
#38	publication[Title/Abstract]
#39	clinical trial as topic [MeSH Terms]
#40	#38 or #39
#41	#37 and #40
#42	#30 and #41
#43	"outcome reporting bias"[Title/Abstract]
#44	"selective reporting"[Title/Abstract]
#45	"selective outcome reporting"[Title/Abstract]
#46	"missing outcome data"[Title/Abstract]
#47	"publication bias"[MeSH Terms]
#48	#43 or #44 or #45 or #46 or #47
#49	"reporting quality"[Title/Abstract]
#50	publications[Title/Abstract]
#51	#49 and #50
#52	#48 or #51
#53	#42 or #52

Appendix Table 2.3: Search strategy for EMBASE via Ovid

	Search Terms
#1	"clinicaltrials.gov:".tw
#2	ANZCTR:.tw.
#3	ICTRP:.tw.
#4	ReBec:.tw.
#5	ChiCTR:.tw.
#6	CRiS:.tw.
#7	CTRI:.tw.
#8	RPCEC:.tw
#9	EU-CTR:.tw.
#10	DRKS:.tw.
#11	IRCT:.tw.
#12	NTR:.tw.
#13	ISRCTN:.tw.
#14	PACTR:.tw.
#15	SLCTR:.tw.
#16	JPRN:.tw.
#17	"trial registry:".tw.
#18	"trial register:".tw.
#19	"trial registries:".tw.
#20	"trials registry:".tw.
#21	"registry of clinical trials:".tw.
#22	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21
#23	"trial registration:".tw.
#24	discrepancy:.tw.
#25	discrepancies:.tw.
#26	consistency:.tw.
#27	inconsistency:.tw.
#28	#24 or #25 or #26 or #27
#29	#23 and #28
#30	#22 or #29
#31	unregistered:.tw.
#32	non-publication:.tw.
#33	nonpublication:.tw.
#34	unpublished:.tw.
#35	published:.tw.
#36	#31 or #32 or #33 or #34 or #35
#37	registered:.tw.
#38	publication:.tw.
#39	"clinical trial (topic)"/
#40	#38 or #39
#41	#37 and #40
#42	#36 or #41
#43	#30 and #42
#44	"outcome reporting bias:".tw.
#45	"selective reporting:".tw.
#46	"selective outcome reporting:".tw.
#47	"missing outcome data:".tw.
#48	#44 or #45 or #46 or #47
#49	"reporting quality:".tw.
#50	publications:.tw.
#51	#49 and #50
#52	#48 or #51
#53	#43 or #52

Appendix Table 2.4: Characteristics of 43 studies identifying published articles from cohorts of trial registry entries

Study	Registry entry cohort	Published articles found	Trial registries included	Study purpose	Study publication year	Application domain	Proportion of links by process
Hartung [198]	305	110	ClinicalTrials.gov	To determine consistency between registered trials and their publication.	2014	Phase III or IV trials	Automatic=95, Inferred=15
Ross [52]	677	315	ClinicalTrials.gov	To assess the publication of registered trials in clinicaltrials.gov	2009	Completed trials of phase II or higher	Automatic=96, Inferred= 215, Inquired=4 (contact=117, responded=44, published=4)
Bourgeois [43]	546	362	ClinicalTrials.gov	To determine whether funding source of these trials is associated with favourable published outcomes	2010	Anticholesteremic, Antidepressants, Antipsychotics, proton-pump inhibitors and vasodilators	Inferred=Unknown, Inquired=Unknown
Liu [199]	443	156	ANZCTR, ISRCTN, ChiCTR, IRCT, DRKS, NTR, JPRN, SLCTR, CTRI, PACTR, Clinicaltrials.gov.	Publication rate of Chinese Trials in WHO Registries	2010	Trials sponsored by China	Automatic=103, Inferred=40, Inquired=13 (contact=54, responded=all, published=1)
Prenner [200]	64	35	Clinicaltrials.gov	To evaluate the rate of publication of registered clinical trials concerning age-related macular degeneration	2009	Muscular Degeneration	Automatic=8, Inferred=27
Wildt [201]	105	66	ClinicalTrials.gov	To evaluate the adequacy of reporting of protocols for on diseases of the digestive system	2011	Gastrointestinal diseases	Inferred=66
Gandhi [202]	37	20	ClinicalTrials.gov	To compare the published orthopaedic trauma trials following registration in Clinicaltrials.gov	2011	Orthopaedic Trauma	Automatic and Inferred=Unknown
Ross [53]	635	432	ClinicalTrials.gov	To review patterns of publication of clinical trials funded by NIH in peer reviewed biomedical journals	2012	NIH funded trials in biomedical journals	Automatic and Inferred=Unknown
Shamliyan [203]	758	212	ClinicalTrials.gov	To examine registration, completeness, and publication of children studies	2012	Children studies funded by NIH	Inferred=212

Vawdrey [47]	62	47	ClinicalTrials.gov	To measure the rate of non-publication and assess the publication bias in clinical trials of electronic health records	2012	Electronic health record registered in clinicaltrials.gov	Automatic, Inferred and Inquired= Unknown
Chapman [117]	314	208	ClinicalTrials.gov	To determine the rate of early discontinuation and non-publication of RCTs	2014	Surgery	Inferred=192, Inquired=16 (contact=101, responded=25, published=16)
Liu [116]	505	115	All 14 registries in ICTRP and ClinicalTrials.gov	To estimate bias risk and outcome-reporting bias in RCTs of traditional Chinese medicine	2013	Traditional Chinese medicines	Unknown
Wetering [92]	599	312	NTR	To evaluate the reporting of trial registration numbers in biomedical publications	2012	Biomedical publications	Automatic and Inferred=Unknown, Inquired=0 (contact=42, responded=9, published=0)
Huser [66]	8907	885	ClinicalTrials.gov	Linking clinicaltrials.gov with PubMed	2013	Interventional phase II or higher clinical trials	Automatic=885
Stockmann [204]	108	65	ClinicalTrials.gov	To evaluate the publication patterns of obstetric studies registered in ClinicalTrials.gov	2014	Obstetric studies	Automatic=45, Inferred= 20
Jones [45]	585	414	ClinicalTrials.gov	To estimate the frequency with which results of large randomized clinical trials registered with ClinicalTrials.gov are not available to the public	2013	Interventional RCTs with more than one arm	Automatic and Inferred=Unknown, Inquired=4
Riveros [46]	594	297	ClinicalTrials.gov	To assess timing and completeness of trial results posted at ClinicalTrials.gov and published in journals	2013	Interventional studies of phase III and IV	Unknown
Korevaar [205]	418	224	ClinicalTrials.gov	To assess publication and reporting of test accuracy studies registered in ClinicalTrials.gov	2014	Test accuracy studies	Automatic=154, Inferred=64, Inquired=6 (contact=175, responded=119, published=6)
Munch [206]	391	118	ICTRP, ClinicalTrials.gov	To analyse the perils and pitfalls of constructing a global open-access database of registered analgesic clinical trials	2014	Analgesic clinical trials	Inferred=118

Hill [118]	90	66	ClinicalTrials.gov	To assess the characteristics of paediatric cardiovascular clinical trials registered on ClinicalTrials.gov	2014	Paediatric cardiovascular clinical trials	Unknown
Khan [207]	143	95	ClinicalTrials.gov	To examine characteristics associated with the publication and timeliness of publication of RCTs of treatment of rheumatoid arthritis	2014	Rheumatoid Arthritis	Automatic and Inferred=Unknown, Inquired=1 (contact= 58, responded=28, published=1)
Su [186]	239	88	All 14 registries in ICTRP and ClinicalTrials.gov	Outcome reporting bias	2015	Acupuncture	Automatic and Inferred=Unknown
Hakala [208]	177	102	ClinicalTrials.gov	To quantify the proportion of trials for unsuccessfully licensed drugs that are not published	2015	Stalled drugs	Automatic=Unknown, Inferred=Unknown, Inquired=0 (emails or calls=42, responded=9, published=0)
Pranic [62]	81	21	ClinicalTrials.gov	Outcome reporting bias	2016	Completed RCTs	Inferred=21
Tang [63]	300	222	ClinicalTrials.gov	Outcome reporting bias	2015	Random sample of phase II or IV trials	Automatic and Inferred=Unknown
Boccia [209]	1109	120	ClinicalTrials.gov	To assess the status of registration of observational studies	2015	Cancer	Inferred=120
Saito [210]	400	229	ClinicalTrials.gov	To determine publication rates of completed US trials	2014	Interventional studies	Automatic=126, Inferred=103
Son [211]	161	62	ClinicalTrials.gov	To assess whether there is publication bias in industry funded clinical trials of degenerative diseases of the spine	2015	Diseases of the spine	Inferred=62
Baudart [212]	489	189	ClinicalTrials.gov	To evaluate the publication rate of observational studies for intervention	2016	Observational studies with safety outcomes	Automatic=75, Inferred=99, Inquired=15 (contact= 241, responded=52, published=15)
Chahal [213]	34	20	ClinicalTrials.gov	To determine publication rates of RCTs in sports medicine	2012	Sports medicine	Automatic and Inferred=unknown
Manzoli [214]	355	176	ClinicalTrials.gov, ICTRP, ANZCTR, ChiCTR, Current Control Trails, Clinical Study Register or Indian	To evaluate the extent of non-publication or delayed publication of registered RCTs on vaccines	2014	Vaccines	Automatic=132, Inferred=44, Inquired=0, (contact=24, responded=0, published=Unknown)
Lebensburger [215]	147	52	ClinicalTrials.gov	To analyse ClinicalTrials.gov for registered sickle cell trials	2015	Sickle cells	Automatic=28, Inferred=24

Smith [216]	101	25	ClinicalTrials.gov	Outcome reporting bias	2012	Arthroplasty	Automatic=10, Inferred=15
Guo [217]	35	11	ClinicalTrials.gov	To estimate patterns of publication of clinical trials of endometriosis registered in ClinicalTrials.gov	2013	Endometriosis	Inquired= 8, Inferred=3
Tsikkinis [218]	333	141	ClinicalTrials.gov	To identify all phase III prostate cancer trials in ClinicalTrials.gov with pending results	2015	Prostate cancer	Inferred=141
Chen [219]	4347	2458	ClinicalTrials.gov	To assess publication rate and reporting of results for completed trials	2016	Interventional clinical trials	Automatic and Inferred=Unknown
Ramsey [220]	2028	357	ClinicalTrials.gov	To assess the proportion of registered trials that are published	2008	Oncology	Automatic=357
Hurley [221]	142	62	ClinicalTrials.gov	To assess the delayed publication of clinical trials	2012	Cystic fibrosis	Inferred=59, Inquired=3 (contact=83, responded=29, published=3)
Ioannidis [222]	73	21	Cochrane Controlled Clinical Trial Register, ISRCTN, ClinicalTrials.gov, ICTRP, GSK Clinical Study Register, and Indian, ANZCTR, and Chinese Clinical Trial Registries	To assess publication delay	2011	Influenza A (H1N1) vaccination	Unknown
Ohnmeiss [223]	72	28	ClinicalTrials.gov	To assess the publication of the studies registered on ClinicalTrials.gov.	2015	Spine studies	Automatic and Inferred=Unknown
Gopal [224]	6251	818	ClinicalTrials.gov	To evaluate the rate of compliance with the FDA mandatory results reporting in ClinicalTrials.gov	2012	Interventional studies	Automatic=818
Lampert [225]	76	40	ClinicalTrials.gov	To determine selective outcome reporting and delay of publication	2015	Epilepsy	Automatic=32, Inferred=7, Inquired=1
Gandhi [226]	46	2	ISRCTN, ClinicalTrials.gov, ANZCTR	To determine the extent to which ongoing and future RCTs in diabetes, will ascertain patient-important outcomes	2008	Diabetes	Unknown

Appendix Table 2.5: Characteristics of 39 studies identifying trial registry entries from cohorts of published article

Study	Published article cohort	Registry entries found	Trial registries included	Study purpose	Study publication year	Application domain	Proportion of links by process
Mathieu [60]	234	323	ClinicalTrials.gov, ISRCTN, ICTRP, national register based on country of first author	Outcome reporting bias	2009	Cardiology, Rheumatology, Gastroenterology	Automatic=205, Inferred=6, Inquired=23
Chowers [227]	49	60	Unknown	Outcome reporting bias	2009	Anti-retroviral therapy	Unknown
Rasmussen [121]	54	137	ClinicalTrials.gov, ISRCTN, ICTRP, NCI-PDQ	To determine association of trial registration with the results and conclusions of published trials	2009	Oncology drugs	Inferred=54
Kunath [120]	63	106	ICTRP	To observe trial registration in urology journals	2011	Urology	Automatic=48, Inferred=15
Ewart [57]	135	124	ISRCTN, ClinicalTrials.gov, ANZCTR, EU-CTR, National Research Register	Outcome reporting bias	2009	RCTs in five high impact factor journals	Unknown
You [228]	215	366	ClinicalTrials.gov, ISRCTN	Outcome reporting bias	2011	Oncology drugs	Unknown
Reveiz [229]	89	526	ICTRP	Outcome reporting bias	2012	RCT from Latin America and Caribbean	Unknown
Nankervis [230]	37	109	ICTRP	Outcome reporting bias	2012	Eczema treatment	Automatic=20, Inferred=17
Pinto [231]	67	200	ClinicalTrials.gov, ISRCTN, ANZCTR, national register based on country of first author	Completeness of clinical trial registration and the extent of selective reporting of outcomes in published trials	2013	Physical therapy	Automatic=48, Inferred=2, Inquired=17
Wetering [92]	185	302	ClinicalTrials.gov, ISRCTN, ICTRP,	To determine reporting of trial registration numbers in biomedical publications	2012	RCT from core clinical journals	Automatic=166, Inquired=19 (contact=136, responded=51, published=19)

			national register based on country of first author				
Hannink [232]	218	327	ClinicalTrials.gov, ISRCTN, ANZCTR and others	Outcome reporting bias	2013	Surgical interventions	Automatic=218
Huser [87]	661	698	ClinicalTrials.gov, ISRCTN	Evaluating adherence to ICMJE policy of mandatory and timely clinical trial registration	2013	Trials published in 5 ICMJE journals	Automatic=661
Rosenthal [233]	51	55	ClinicalTrials.gov, ISRCTN, ANZCTR, ChiCTR, UMIN	Outcome reporting bias	2013	Surgery	Automatic, Inferred and Inquired= Unknown
Hopewell [234]	30	69	Unknown	To observe reporting characteristics of non-primary publications of results of RCTs	2013	RCTs from National Library of Medicine's set of 121 core clinical journals	Automatic=30
Babu [235]	121	417	Unknown	To observe clinical trial registration in physical therapy journals	2014	Physical therapy journals	Automatic=121
Lee [236]	8	83	Unknown	Assessment of compliance of randomized controlled trials in trauma surgery with the CONSORT statement	2013	Trauma surgery	Automatic= 8
Li [237]	252	305	ClinicalTrials.gov, Current Controlled Trials, NTR, ANZCTR, UMIN CTR	Outcome reporting bias	2013	Gastroenterology and Herpetology	Automatic=212, Inferred=40
Norris [238]	50	107	ICTRP	To determine selective outcome reporting	2013	Pharmacotherapy	Automatic=30, Inferred=20
Hardt [239]	85	103	ICTRP (ClinicalTrials.gov, ISRCTN, EU-CTR, NTR, ANZCTR, DRKS, JPRNUMIN, ChiCTR, CTRI), Belgian register	To determine whether the results of registered surgical RCTs are published in journals requiring registration	2013	Ten highest rank surgery journals	Automatic=68, Inferred=17
Anand [240]	133	197	ClinicalTrials.gov, ISRCTN, ANZCTR	To determine the registration and design alterations of clinical trials in clinical care	2014	RCT in clinical care medicine	Automatic=105, Inferred=28

Mann [241]	140	220	ICTRP	To assess the registration status of RCTs and analyse the correspondence of registered outcomes with published outcomes.	2014	Clinical geriatrics	Unknown
Walker [64]	75	76	ISRCTN, ClinicalTrials.gov, national register based on country of first author	Outcome reporting bias	2014	RCTs published in British Medical Journal and the Journal of American Medical Association	Automatic and Inferred= Unknown
Dekkers [242]	29	54	ICTRP	To compare non-inferiority margins defined in study protocols and trial registry records with margins reported in subsequent publications	2015	Non-inferiority trials submitted 2001 to 2005 to ethics committees in Switzerland and Netherlands	Automatic and Inferred= Unknown
Østervig [243]	85	200	ISRCTN, IRCT, EU-CTR, ChiCTR, CRiS, UMIN CTR, ClinicalTrials.gov	To check registration of randomized clinical trials	2015	Trials in Acta Anaesthesiologica Scandinavica	Automatic=85
Scott [123]	160	181	ISRCTN, NTR, ANZCTR, ClinicalTrials.gov, national register based on country of first author	Selective outcome reporting	2015	Psychiatry journals	Automatic=150, Inferred=6, Inquired=4
De Oliveira [244]	107	201	ISRCTN, ClinicalTrials.gov, ICTRP	Outcome reporting bias	2015	Anaesthesiology	Automatic, Inferred and Inquired= Unknown
Rayhill [122]	58	225	ClinicalTrials.gov and others	To assess the registration status of RCTs and analyse the correspondence of registered outcomes with published outcomes	2015	Core headache medicine journals	Automatic=58
Dal-Ré [196]	175	178	ClinicalTrials.gov, ISRCTN, ANZCTR, NTR, EU-CTR, CTRI, DRKS	To evaluate adherence to ICMJE policy on prospective trial registration	2016	Trials in in high-impact journals	Unknown

Reveiz [245]	52	144	Registered in any International clinical trial registry	To evaluate the influence of trial registration on reporting quality of RCTs	2010	Highest rank journals	Unknown
Rongen [246]	90	362	ClinicalTrials.gov, ISRCTN, ANZCTR, NTR and others	Outcome reporting bias	2016	Orthopaedic surgical interventions	Automatic=90
Harriman [119]	105	108	ClinicalTrials.gov, ISRCTN, ANZCTR, UMIN CTR, NTR, ChiCTR, IRCT	To assess trial registration, analysis of prospective versus retrospective registration	2016	Clinical trials published in the BMC series	Automatic=105, Inquired=0
Vera-Badillo [247]	30	164	ClinicalTrials.gov	Outcome reporting bias	2013	Breast cancer	Automatic and Inferred= Unknown
McGee [248]	74	307	ICTRP	To determine whether trial is registered and declared registration in the publication	2016	Kidney transplantation	Automatic=44, Inferred=30
Huić [249]	149	152	ClinicalTrials.gov	To determine completeness and outcome reporting bias	2011	RCTs published in ICMJE journals	Automatic=149
Chan [250]	519	553	Unknown	Outcome reporting bias	2005	RCTs indexed in PubMed	Inferred and Inquired= Unknown
Korevaar [251]	52	351	ClinicalTrials.gov, ISRCTN, national register based on country of first author	To identify the proportion of articles for which the corresponding study had been registered	2014	Test accuracy studies	Automatic=27, Inferred=11, Inquired=14 (contact=324, responded=187, published=14)
Jones [252]	57	123	ClinicalTrials.gov, ISRCTN, ICTRP, national register based on country of first author	Outcome reporting bias	2012	Emergency	Automatic= 23, Inferred=34
Smaïl-Faugeron [253]	73	317	ICTRP	To assess the registration rate of RCTs	2015	Oral Health	Automatic=50, Inferred=23
Riehm [254]	40	76	ISRCTN, ClinicalTrials.gov, ICTRP	Outcome reporting bias	2015	Psychosomatic and Behavioral health	Automatic=33, Inferred=7

Appendix Table 4.2: Rules for extracting participants information from each of the included trial in systematic review and its update

	Rules

	`# Participant_column contains participants information`
	`# pre-process the text in participant_column to remove`
	`# extra numeric values (replacing with 'xx') given along`
	`# with participants information`
	`#----------------------------------------------`
	
	`preprocessed_text = re.sub(`
	`r'((\w+\s[72](=|:)\s{0,1}\d+\s+)*(exclu\w+|withd\w+|screen\w+|(control|treatment|compar\w+)(\s+group)*)(\s{0,1}(:|=)\s{0,1}\d+)*)|([0-9]+\.[0-9]+)|((age)(\s+|:|=)(\d+|\s+\d+))|\d+\s{0,1}-\s{0,1}\d+|\d+\s+(week|day|month|year)\w{0,1}|(\d+\s+(to)\s+\d+)|(\d+\s+(and)\s+(\d+|\w+))','xx'``, participant_column, flags=re.IGNORECASE)`
	
	`#----------------------------------------------`
	
	`# After pre-processing, apply the rules`
	`# 13 rules are constructed, and combination of these rules are`
	`# also used`

1	`Rule1= re.search(` **`r'(sampl\w+\s+(size)(:\s{0,1}|\s{0,1})\d+)|(random\w+:\s{0,1}\d+)($|\s+|\,|\;|\.|[\)])|(total\s+){0,1}(N.|N|No.|numb\w+|parti\w+)\s+(random\w+)\s{0,1}((assign\w+|\w+)\s{0,1}){0,1}(=|:)(\s{0,1}total\s{0,1}:){0,1}\s{0,1}\d+($|\s+|\,|\;|\.|[\)])|(numb\w+)(\s+of parti\w+){0,1}((\s+was){0,1}\s+\d+|\s{0,1}(=|:)\s{0,1}\d+)($|\s+|\,|\;|\.|[\)])|((total\s+){0,1}n\s{0,1}(=|:)\s{0,1}\d+($|\s+|\,|\;|\.|[\)]))'`**`,preprocessed_text,flags=re.IGNORECASE)`
2	`Rule2= re.search(` **`r'(total)*\s+(n)\s+(random\w+)\s{0,1}(:|=)\s{0,1}\d+($|\s+|\,|\;|\.|[\)])'`**`,preprocessed_text,flags=re.IGNORECASE)`
3	`Rule3= re.search(`**`r'[0-9]+\s*(part\w+|patie\w+|infan\w+|su\w+|chi\w+|\w+\s*chi|coupl\w+)'`**`,preprocessed_text, flags=re.IGNORECASE)`
4	`Rule4= re.search(`**`r'([0-9]+\s*(\w+\s*(peop\w+|pers\w+|patie\w+)|(peop\w+|pers\w+)))'`**`,preprocessed_text,flags=re.IGNORECASE)`
5	`Rule5= re.search(`**`r'(^[0-9]+\s+\w+)'`**`, preprocessed_text,`

```
flags=re.IGNORECASE)


Rule6= re.search(r'\w+\s*\:\s*[\(]\d+[\)]',
preprocessed_text,flags=re.IGNORECASE)


Rule7= re.search(r'((part\w+\s+|patie\w+\s+)[0-9]+)',
preprocessed_text,flags=re.IGNORECASE)


Rule8= re.search(r'[0-
9]+\s+(met\s+\w+)',preprocessed_text,flags=re.IGNORECASE)


Rule9= re.search(r'[0-9]+\s+(wom\w+)',
preprocessed_text,flags=re.IGNORECASE)


Rule10= re.search(r'[\(]\w+/\w+[\)]:\s{0,1}\d+/\d+',
preprocessed_text,flags=re.IGNORECASE)


Rule11= re.search(
r'(\d+\s+(men)((,|\s+)(\s+|and)(\s+)*(\d+)*(\s+)*(wom\w+)))',pr
eprocessed_text,flags=re.IGNORECASE)


Rule12= re.search(r'(partic\w+(:|=)\s{0,1}\d+)',
preprocessed_text,flags=re.IGNORECASE)


# Rule13
# Extracting participant information provided in number words
and converting into numeric digits

ones = {'one': 1, 'eleven': 11,
        'two': 2, 'twelve': 12,
        'three': 3, 'thirteen': 13,
        'four': 4, 'fourteen': 14,
        'five': 5, 'fifteen': 15,
        'six': 6, 'sixteen': 16,
        'seven': 7, 'seventeen': 17,
        'eight': 8, 'eighteen': 18,
        'nine': 9, 'nineteen': 19}

tens = {'ten': 10, 'twenty': 20, 'thirty': 30, 'forty': 40,
'fifty': 50,'sixty': 60, 'seventy': 70, 'eighty': 80, 'ninety':
90}

groups = {'thousand': 1000, 'million': 1000000, 'billion':
1000000000, 'trillion': 1000000000000}

groups_match =
re.search(r'(^\s?([\w\s]+?)(?:\s((?:%s?patients))))' %
```

```
('|'.join(groups)), preprocessed_text,flags=re.IGNORECASE)


hundreds_match =
re.search(r'(^([\w\s]+)\shundred(?:\s(.*?patients)))',
preprocessed_text,flags=re.IGNORECASE)

tens_and_ones_match = re.search(
r'(^((?:%s))(?:\s(.*?patients)))' % ('|'.join(tens.keys())),
preprocessed_text, flags=re.IGNORECASE)

if (groups_match):
replace_symbol = re.sub("-", '-', preprocessed_text,
flags=re.IGNORECASE)
remove_text = re.split('pati\w+', replace_symbol,
flags=re.IGNORECASE)
number_part = w2n.word_to_num(remove_text[0])

elif (hundreds_match):
replace_symbol = re.sub("-", '-', preprocessed_text,
flags=re.IGNORECASE)
remove_text = re.split('pati\w+', replace_symbol,
flags=re.IGNORECASE)
number_part = w2n.word_to_num(remove_text[0])

elif (tens_and_ones_match):
replace_symbol = re.sub("-", '-', preprocessed_text,
flags=re.IGNORECASE)
remove_text = re.split('pati\w+', replace_symbol,
flags=re.IGNORECASE)
number_part = w2n.word_to_num(remove_text[0])
```