USE OF CLINICALTRIALS.GOV REGISTRY IN SYSTEMATIC REVIEWS AND META-ANALYSES: A MASTER'S THESIS

A Master's Thesis Presented

By

RICHEEK PRADHAN

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

(November 30th 2017)

Biomedical Sciences

USE OF CLINICALTRIALS.GOV REGISTRY IN SYSTEMATIC REVIEW AND META-ANALYSES: A MASTER'S THESIS

A Masters Thesis Presented

By

RICHEEK PRADHAN

The signatures of the Master's Thesis Committee signify completion and approval as to style and content of the Thesis

Dr. Jerry Gurwitz, M.D., Chair of Committee

Dr. Bruce Barton, Ph.D., Member of Committee

Dr. Mara Epstein, Ph.D., Member of Committee

Dr. Mary Ellen Lane, Ph.D., Member of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all master's degree graduation requirements of the school.

Dr. Anthony Carruthers, Ph.D.,

Dean of the Graduate School of Biomedical Sciences

Interdisciplinary graduate program

(November 30th 2017)

Acknowledgments and dedication

I would like to acknowledge my Thesis Mentor, Dr. Hong Yu for her guidance throughout my time at UMass Medical School. Her open-minded and enthusiastic approach to research will always be an inspiration for me. I also would like to thank Dr. Robert Goldberg, my co-mentor, whose support and guidance has helped me navigate many scientific hurdles in this time.

Working on his dissertation let me work with three professors who I consider myself lucky to have been associated with: Dr. David Hoaglin, whose emphasis on methodological accuracy and rigor I hope to carry with me to future projects, Dr. Arlene Ash, whose elegant approach to statistics and writing I have had the chance to learn from, and Dr. Sonal Singh, whose kind and prescient mentorship I have benefitted from greatly.

Heartfelt thanks to the members of my lab, Victoria, Weisong, Matt, and Feifan who I have learnt from everyday. Also thanks to Kyle and Bikramjit who helped me collect data for what constitutes Chapter 2 in the dissertation.

I also would like to thank Dr. Bill Jesdale, conversations with whom have molded my thoughts on social medicine, Drs. Jennifer Tjia, Stavroula Chrysanthopoulou, and Catherine Dube whose courses on observational research have strengthened my concepts, and Dr. David Chiriboga, whose views on global health have been both eye opening and inspiring. Thanks to Drs. Mitali Chatterjee, Suparna Chatterjee, and Avijit Hazra, my teachers during my residency at IPGMER, India, who helped me build a solid foundation for clinical research.

On a personal front, I want to thank Sambuddha, Isaac, Ribhu, Aditya, Pritikanta, Harsh, and Elaine whose friendships have made living on unaccustomed shores easier, happier, and more worthwhile. Thanks to Rena and Aveek for being ever ready to provide lengthy counseling-sessions over Whatsapp.

Hugs for my parents, for keeping relatively healthy, and for being just as present in my life as they were back home. Hugs also for Mimmi, Puntuli, Chhoto mimmi, Anup mama, Shikha mimmi, Jhumjhumi didi, Sadhona mashi, and Monu mashi.

And finally, this work is dedicated to my grandma, who, through her wit, humor, resilience, and her ability to embrace darkness, taught me why life is worth enduring.

Abstract

Ensuring the objectivity of systematic reviews and meta-analyses (SRMA) begins with comprehensive searches into diverse resources mining primary studies. Guidelines for systematic reviews recommend authors to routinely search of trial registries to identify unpublished studies. In this dissertation, I investigated the utilization of ClinicalTrials.gov (CTG), the world's largest clinical trial registry that contains data from clinical trials of products that are subject to United States Food and Drug Administration (FDA) regulation, as an information resource in SRMAs. First, I examined the use of various information resources including CTG in SRMAs published from 2005-2016, and identified the factors associated with their use. Thereafter, to determine the accuracy of trial safety data reported at CTG, I compared the data at CTG with that in corresponding journal articles and FDA drug reviews. I found that trial safety data at both CTG and articles differed frequently from FDA drug reviews, but the differences were modest in magnitude. Finally, I repeated published meta-analysis (conducted using data from primary study articles) with data at CTG to find that most meta-analysis results were reproduced using CTG data. Taken together, this work suggests that CTG should not only be searched more often to find primary research for systematic reviews, but that data at CTG can also be used to conduct quantitative data synthesis.

Table of contents

Front Matter

Signature Page i

Acknowledgements. ii

Abstract iii

Table of Contents iv

List of Tables v

List of Figures vi

Preface vii

Body Matter

Chapter 1:Introduction 1

Chapter 2: Patterns in use of information resources for evidence synthesis in systematic reviews and meta-analyses: 2005-2016 6

Chapter 3: Concordance of data on serious adverse events and death between ClinicalTrials.gov, journal articles and FDA medical reviews 30

Chapter 4: Examining reproducibility of published meta-analyses using data at ClinicalTrials.gov downloaded by an automatic extraction tool 42

Chapter 5: Conclusion 55

Back Matter

References 56

Appendices Provided separately

List of Tables

Chapter 2.

Table 1. Characteristics of systematic reviews and meta-analyses (SRMA) included in the study.

Table 2. Information sources that, when searched alongside Medline, are negatively associated with publication bias.

Chapter 3.

Table 1. Differences in death and serious adverse event reporting between Food and Drug Administration (FDA) reviews, ClinicalTrials.gov, and published articles.

Table 2. Mismatches in reporting of safety outcomes mentioned in "Adverse Reaction" section of product labels for anticancer drugs between FDA reviews, ClinicalTrials.gov, and published articles.

Chapter 4.

Table 1. Selected systematic reviews.

Table 2. Results of meta-analyses from published systematic review and meta-analyses (SRMA) using data manually extracted from original articles and from ClinicalTrials.gov using EXACT.

List of Figures

Chapter 2.

Figure 1. Study flow diagram.

Figure 2. Percentage of SRMAs using each information resource in the years 2005 and 2016.

Figure 3. Factors associated with use of various information resources.

Figure 4. Network of data sources in 2005 and 2016.

Figure 5. Interrupted time series analysis showing changes in data-resources before and after 2014.

Chapter 3.

Figure 1. Selection trials of new molecular entities (NMEs) approved by the United States Food and Drug Administration (FDA).

Chapter 4.

Figure 1. Diagram showing development of EXACT.

Figure 2. EXACT's user interface.

Figure 3. Flowchart for selection of meta-analyses reproduced using data manually extracted from original articles and data extracted by EXACT from ClinicalTrials.gov.

Preface

Parts of this thesis are modified from drafts that either have been or will be sent to journals for peer review.

1. Pradhan R, Garnick K, Barkondaj B, Jordan H, Ash A, Yu H. Trends in use of information resources for evidence synthesis in systematic reviews and meta-analyses: 2005-2016. Planned submission to Journal of Medical Internet Research in December 2017.

(Modification of this constitutes Chapter 2)

Kyle Garnick and Bikramjot Barkondaj helped review the meta-analyses in this work.

2. Pradhan R, Singh S. Concordance of data on serious adverse events and death between ClinicalTrials.gov, journal articles and FDA medical reviews. Planned submission to Drug Safety in December 2017.

(Modification of this constitutes Chapter 3)

3. Pradhan R, Hoaglin D, Cornell M, Liu W, Wang V, Yu H. Examining reproducibility of published meta-analyses using data at ClinicalTrials.gov downloaded by an automatic extraction tool. Submitted to Journal of Clinical Epidemiology as of November 2017.

(Modification of this constitutes Chapter 4)

Matthew Cornell and Weisong Liu developed the software EXACT described as a part of this work

CHAPTER 1

Introduction

Publications of systematic reviews and meta-analyses (SRMAs) have tripled over the last decade, without comparable improvement in methodological rigor or reporting standards.[1] Meanwhile, studies linking favorable meta-analysis outcomes with financial conflicts of interest of authors have highlighted the susceptibility of SRMAs to manipulation.[2–4] Amidst these findings, while some have questioned the position of SRMAs as the highest level of evidence, [5]others have called for ensuring greater objectivity and reproducibility in producing them.[6] A way to ensure the objectivity of SRMA results is to perform comprehensive searches into resources mining primary literature.[7] However, what constitutes a comprehensive search remains open to interpretation,[8] and potential manipulation.

The purpose of a SRMA is to summarize *all* scientifically generated evidence on a topic of interest. A systematic search for data from a diverse body of evidence is fundamental to serve that purpose. In addition to extracting data from published studies, it is important to search unpublished studies (also known as grey literature) as research shows that the latter have smaller treatment effects than published studies,[9,10] and that inclusion of unpublished results can change conclusions of meta-analyses.[11,12] Conversely, failure to include unpublished data biases the results towards a positive treatment effect (also known as publication bias).[13,14] Research into the evolving use and relative

importance of information resources mining published and unpublished research can improve the scientific rigor of evidence synthesis.

Although including both published and unpublished data is important for validity, in practice, neither is searching for the unpublished data easy,[15–17] nor are guidelines suggesting resources to be searched for unpublished data consistent among each other.[18] Most popular biomedical search engines mine only published studies, and little consensus exists regarding which resources to look into for unpublished data.[8,19] Among the sources of unpublished studies, trial registries established to mine data from clinical studies have emerged as a rich source of information over the last decade. Registries store information from studies regardless of the success of their outcomes, making them an important source for unpublished research.[7] Clinical trials of drugs, biologics, and devices must be registered in study registries including ClinicalTrial.gov (CTG).[20] CTG, launched in 2000, is currently the world's largest clinical study registry and contains information from clinical trials of products that are subject to Food and Drug Administration (FDA) regulation.[21] While studies have documented the underuse of registries,[1,22–25] factors that may lead to more widespread inclusion of such resources in search strategies have not been studied.

Querying the registry is important for meta-analysis research as almost half of large trials reported in CTG remain unpublished in journals.[26] However, since trialists are also required to report data on trial results in CTG, apart from being just a source of information on unpublished trials, CTG also acts as a publicly available source of trial data. CTG is particularly important in safety data, as recent research has shown that CTG

covers a greater number of and a wider range of adverse events (AEs) than published articles on primary studies, the traditionally used source of data for quantitative data synthesis in secondary research.[27]

On the other hand, works showing inconsistencies in efficacy[28] and safety[29] results published in articles and those reported in registries have questioned the quality of the data in registries. Since peer reviewers reviewing articles typically do not have access to patient-level data, the reviewers are unable to assess the provenance of the study results. Raw data from trials submitted for drug licensing, however, are routinely reviewed by the US-FDA, and these medical reviews are made public at Drugs@FDA.com.[29,30] Since trialists are legally mandated to submit accurate data to the FDA,[31,32] the FDA medical reviews can serve as reference standards to validate the data at CTG and articles. A recent study focusing on primary outcome comparison between the medical reviews conducted by the FDA, [30] and CTG results found that most primary efficacy reports are concordant between the two sources.[33] However, they did not compare the FDA reviews and CTG reports with the corresponding published articles. Thus, the need to evaluate the concordance between FDA review (the gold standard) and the two test resources (CTG and articles) on adverse events remains unaddressed. This is an important lacuna, preventing the use of CTG data in meta-analysis research.

A final question that needs answering before data at CTG can be used to perform secondary analysis instead is whether differences in data between two sources cause meta-analyses of data at CTG to differ qualitatively from meta-analyses using original-article data. No consensus exists on the use of the trial data reported at CTG for

quantitative data synthesis in meta-analyses.[34] Use of data at CTG, however, might provide certain logistic benefits over use of data from published articles. For example, a SRMA is a labor-intensive and time-consuming endeavor; one of its most protracted and error-prone steps involves extraction of data from reports on primary trials.[35–37] Given that the median time between publication of protocols of Cochrane SRs and their final publication is 2.4 years,[38,39] and that errors in extracting numerical data pose are common,[40] an automated methods of data extraction for meta-analysis research can be of benefit. Although recent attempts have automated extraction of qualitative and design-related variables from published articles,[41] few have automated extraction of numerical results, possibly because of heterogeneous publishing formats and reporting standards. CTG forms an alternative source that is more amenable to automatic extraction because of its uniform reporting format. However, if data at CTG, given its differences from article data, qualitatively alters meta-analysis conclusions, its utility in meta-analysis research will be limited. Thus, the validation of use of CTG data in meta-analysis research is necessary before tools to automatically extract data from CTG can be routinely used.

In chapter 1 of this dissertation, I conduct a situational analysis into the use of CTG and other information resources searched to find primary studies for evidence synthesis in SRMAs. For this, I track the self-reported use of various information resources in SRMAs published between 2005-2016, and identify the resources that are associated with low publication bias.

In chapter 2, using FDA drug reviews, I examine the discrepancies between published article and CTG data, specifically in context of safety endpoints. I examine the frequency and magnitude of differences in safety data from clinical trials reported across three sources: CTG and journal published articles as index sources and corresponding FDA medical reviews as the reference.

In chapter 3, I repeat meta-analyses from published SRMAs using CTG data instead of data from primary articles to investigate whether the use of CTG data materially alters conclusions of the SRMAs.

CHAPTER 2

Patterns in use of information resources for evidence synthesis in systematic reviews and meta-analyses: 2005-2016

Introduction

Commercial publishers are likely to be interested in publishing studies that have positive results in alignment with the hypothesis, and thus are "news worthy". Similarly, industry-based or academic researchers are likely to be interested in publishing positive results to more efficiently use their time in talking about successful products or research, respectively. In fact, not only are negative studies published on an average two years later than their positive counterparts, they are also are less likely to be published at all.[42] Although ignored during publication, however, negative studies assume just as much importance as published ones when the aim is to summarize evidence on a topic of interest. Importantly, inclusion of results from unpublished studies has been shown to alter the results of meta-analysis.[11,12] Thus, searching unpublished studies is important for valid meta-analyses. Study registries like ClinicalTrials.gov (CTG) store data from both published and unpublished research. For example, Riveros et al. 2013 found that up to 50% of trials posted in CTG remained unpublished in peer reviewed journals.[26] Registries should thus constitute an important resource for unpublished literature. Accordingly, 2012 Cochrane guidelines mandate inclusion of study registries and grey sources in SRMA search strategies. CTG contains information from clinical trials of products that are subject to Food and Drug Administration (FDA) regulation.[21] Other

trial registries can be accessed through the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP), a portal to 16 trial registries managed by various national regulatory bodies.[43] Additional repositories of trial reports can include regulatory body reports (FDA databases), grant databases, and manufacturer websites.[7] Sources of grey literature that are not clinical trials include conference abstracts, dissertations, book chapters, policy documents, and specialized grey literature databases, among others.[17] In order to avoid skewing meta-analysis results towards positive outcomes, a systematic search for both published and unpublished studies are necessary. However, systematic evidence on the resources searched, and what constitutes a comprehensive search, is lacking.

In this chapter, I examined the use of various information resources by randomly sampling SRMAs published in each year from 2005-2016, and identified the factors associated with their use. Using network analysis, I explored which resources were simultaneously searched in each year. I used interrupted time series to look for changes in the use of various information resources following the promulgation of a guideline (in late 2012) urging the use of diverse resources in SRMAs. Using logistic regression, I sought to identify information sources associated with a reduced likelihood of publication bias.

Methods

Searching and eligibility

I searched PubMed for SRMAs published from 2005 through 2016. I chose the year 2005 to start search in order to specifically evaluate CTG usage in SRMAs after International Committee of Medical Journal Editors (ICMJE) publication rules (2005) which caused substantial increase in the trial registration in CTG, an converted CTG into a repository of clinical trials with worth looking into before evidence synthesis. Using search terms aimed at finding systematic reviews with quantitative meta-analyses I extracted all articles indexed in each calendar year. To focus on articles likely to search CTG for unpublished studies, I restricted the search to include articles with human subjects and those authored by US-based investigators. The complete search term was: "systematic[sb] AND USA[ad] AND (Meta-Analysis[ptyp] AND ("respective year/01/01"[PDAT]: "respective year/12/31"[PDAT]) AND "humans"[MeSH Terms])".

From those, I randomly selected 100 articles for each year to review manually. I excluded articles if they (1) did not find full-texts, (2) did not have a defined and documented search strategy, (3) did not perform quantitative data synthesis and meta-analysis, (4) did not involve humans, (5) specifically excluded the USA population, (6) used a single case experiment design, or (7) had duplicates in another year. Exclusion criteria 1, 2 and 3 were deemed necessary to allow for imperfect indexing in PubMed. The study flow diagram is represented in Fig 1.

Data extraction

Each eligible SRMA article was manually reviewed to extract the data regarding publication characteristics and search resources by one of three investigators (me/Kyle /Bikramjit: please see preface). Three investigators independently extracted data from a set of 10 SRMAs to help achieve uniform extraction. For all reviewed SRMAs, after one investigator extracted the data, a second investigator validated the extraction, and disagreements were resolved by mutual discussion.

I collected the information on all online databases and other information resources utilized by authors of each SRMA to search for primary studies, as mentioned in the search strategy of the SRMAs. Whether or not SRMA authors searched from the information resources was noted as a binary variable. I identified 30 information resources and retrospectively classified them into four categories: 1. Study registries [CTG, ICTRP, regulatory databases (eg, FDA, EMEA), manufacturer database, grant websites for funded studies (eg, NIH, Wellcome trust), others (ISTRN, HTA, HSRProj, C2-SPECTR, PROSPERO)];2. Resources mining general published literature without special focus [Medline, EMBASE, others (Scielo, JSTOR, etc)]; 3 Resources mining specialized published literature [Cochrane library, psycINFO, CINAHL, POPLINE, regional/language specific database, review collections (eg, EBMR, ACP journal club), SportDiscus, HealthStar, International pharmaceutical abstracts, PILOTS]; 4. Resources other than registries that include unpublished literature [Conference proceedings, dissertations, Scopus, Web of Science, Search engines (eg, Google Scholar), ProQuest, BIOSIS, ERIC, Clinical query applications, designated grey literature databases

(OpenGrey, Sigle, NTIS), and others (MedScape, manufacturers package inserts, etc)]. Salient features of each information resource are described in Appendix 1.

Information was also collected on each SRMA regarding the nature of primary studies (interventional or observational), whether the study involved a pharmacological treatment, and the outcomes studied (efficacy outcomes, safety outcomes, others). In order to determine the nature of the primary studies, the search criteria in each SRMA were looked into: when the specific nature of the primary studies searched was not mentioned, primary studies directly were directly examined. In cases where the specific nature of the primary studies was not mentioned in the search criteria, and yet only one type of primary study (say, interventional) was included in the SRMA, the SRMA was classifies to have interventional studies only. The logic for this was that the SRMA authors would have anticipated the nature of their primary studies, and searched resources accordingly. The nature of the journal in which the SRMA was published were noted [general med/surgery journals *(general medical journals)*, journals specializing in epidemiology/public health/research methods *(methods journals)*, psychiatry and psychology journals *(psychiatry journals)*, specialized journals on sub-disciplines within medicine/surgery *(specialized medical journals)*] along with their impact factors each year. A separate class for psychiatry journals was formed because of the fairly distinct information resources accessed by this medical discipline. A complete list of all journals classified into each of the four groups is provided in Appendix 2. Additionally, it was noted if publication bias was statistically assessed, and, if assessed, whether the authors deemed publication bias to be present or not.

Analysis

I summarized the characteristics of the SRMAs over the years using percentages of total for the year for categorical parameters or mean and standard deviations for numerical parameters. Differences in parameters between years were examined using chi-square statistics for categorical variables and ANOVA for numerical variables. I tracked the use of all 30 information resources over 12 years using bar diagrams. To identify the factors predicting study registries, and CTG in particular, I used a logistic regression model with their characteristics as independent variables. To identify concurrent use of similar types of information resources, I constructed the network geometry of articles and the information resources in each year. Each information resource formed one node in each network, and each node was connected to another by a line if at least one article searched both information resources. The thickness of the connecting lines denotes the number of articles searching the two resources. Distance between two nodes is given by their Euclidian distance, which represents how frequently they are co-searched. To detect the presence of homophily (tendency of resources to be associated and, in our scenario, searched concomitantly), I used the ANOVA density model.[44] The network densities for a year (number of actual ties between nodes/number of all possible ties between all nodes with theoretical extremes between zero and one) were calculated. I also used interrupted time series analysis to see if there was a change in the use of registries after December 2012, when Cochrane guidelines [Methodological Expectations of Cochrane Intervention Reviews (MECIR)][34] first mandated using registries as an information resource in SRMAs. Assuming a time lag of one year between promulgation of MECIR

and publication of SRMAs that might have followed the guideline, I selected 2014 as the year from which I could expect the promulgation guideline to show results. Last, I used logistic regression adjusted for SRMA characteristics and other resources to identify the information resources negatively associated with publication bias. Data were analyzed using STATA version 14[45] and UCINET version 6.[46]

Results

Characteristics of SRMAs and information sources used

I retrieved a total of 11,868 PubMed entries using the search terms, and randomly selected 100 entries each for each year (1200 entries between 2005-2016). A total of 817 SRMA articles with an average of 68 SRMAs per year (range: 59-79) were included in the final analyses (Fig 1). Table 1 describes the included SRMAs. Of the journals publishing the SRMAs, about two-thirds (65.2%) were specialized medical journals, while the rest were approximately similar proportions of general medical journals (9.2%), psychiatry journals (14.7%), and methods journals (10.9%). The mean impact factor of these journals was 4.7 (standard deviation: 4.5). The primary research analyzed by the SRMAs included interventional studies in 61.0% cases and observational studies in 62.9% cases. Almost one third of the SRMAs had a safety parameter as an endpoint (32.4%), and one third of the SRMAs examined a pharmacological agent (33.6%). Publication bias was statistically assessed in about 45.9% of the SRMAs, of which a third (33.1%) found statistical evidence of such bias. Fig 2 represents a comparative picture of information resource use in 2005 and 2016. The most common sources used overall were

Medline/PubMed, EMBASE, and the Cochrane databases in both years. Use of registries and Scopus increased substantially from 2005 to 2016. Trends in utilization of registries, general and specialized published literature databases, and grey literature sources are provided in Appendix 3.

Table 1. Characteristics of systematic reviews and meta-analyses (SRMA) included in the study

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Number of SRMAs included (n)	66	59	65	74	71	69	65	64	62	68	75	79
General medical journals (% of n)	15.2	6.8	15.4	10.8	7.0	10.1	12.3	4.7	8.1	5.9	5.3	8.9
Methods journals (% of n)	10.6	10.2	9.2	9.5	12.7	14.5	9.2	6.3	14.5	5.9	17.3	10.1
Psychiatry journals (% of n)	19.7	20.3	18.5	17.6	16.9	13.0	4.6	17.2	16.1	13.2	9.3	11.4
Specialized journals (% of n)	54.6	62.7	56.9	62.2	63.4	62.3	73.9	71.9	61.3	75.0	68.0	69.6
Mean (SD) impact factors of journals	4.8 (4.7)	5.0 (5.0)	5.6 (7.3)	5.3 (6.0)	4.6 (2.9)	5.1 (5.6)	4.8 (3.9)	5.2 (4.0)	4.8 (3.5)	3.3 (2.0)	4.7 (3.9)	3.9 (2.9)
Efficacy endpoints (% of n)*	42.4	28.8	35.4	37.8	40.9	53.6	53.9	42.2	30.7	48.5	54.7	27.9
Safety endpoints (% of n)	37.9	30.5	35.4	33.8	38.0	37.7	30.8	26.6	32.3	32.4	32.0	22.8
Others endpoints (% of n)	54.6	59.3	55.4	52.7	42.3	47.8	53.9	53.1	67.7	42.7	42.7	51.9
Pharmacological agent studied (% of n)	40.9	27.1	33.9	35.1	42.3	34.8	38.5	15.6	30.7	30.9	37.3	34.2
Primary studies: Interventional (% of n)	65.2	54.2	58.5	60.8	60.6	59.4	56.9	65.6	59.7	63.2	66.7	60.8
Primary studies: Observational (% of n)	56.1	61.0	53.9	60.8	67.6	63.8	63.1	68.8	64.5	67.7	60.0	67.1
Publication bias assessed (% of n)*	39.4	37.3	30.8	48.7	47.9	40.6	56.9	54.7	45.2	39.7	58.7	48.1
Publication bias identified (% of n)	12.1	11.8	6.1	20.3	15.5	17.4	21.5	12.5	11.3	11.7	22.6	20.2

Footnote: None of the variables were significantly different between the years except * frequencies of SRMAs with an efficacy endpoint (X^2, P=0.002) and **frequencies of SRMAs where publication bias was assessed statistically (X^2, P=0.030).

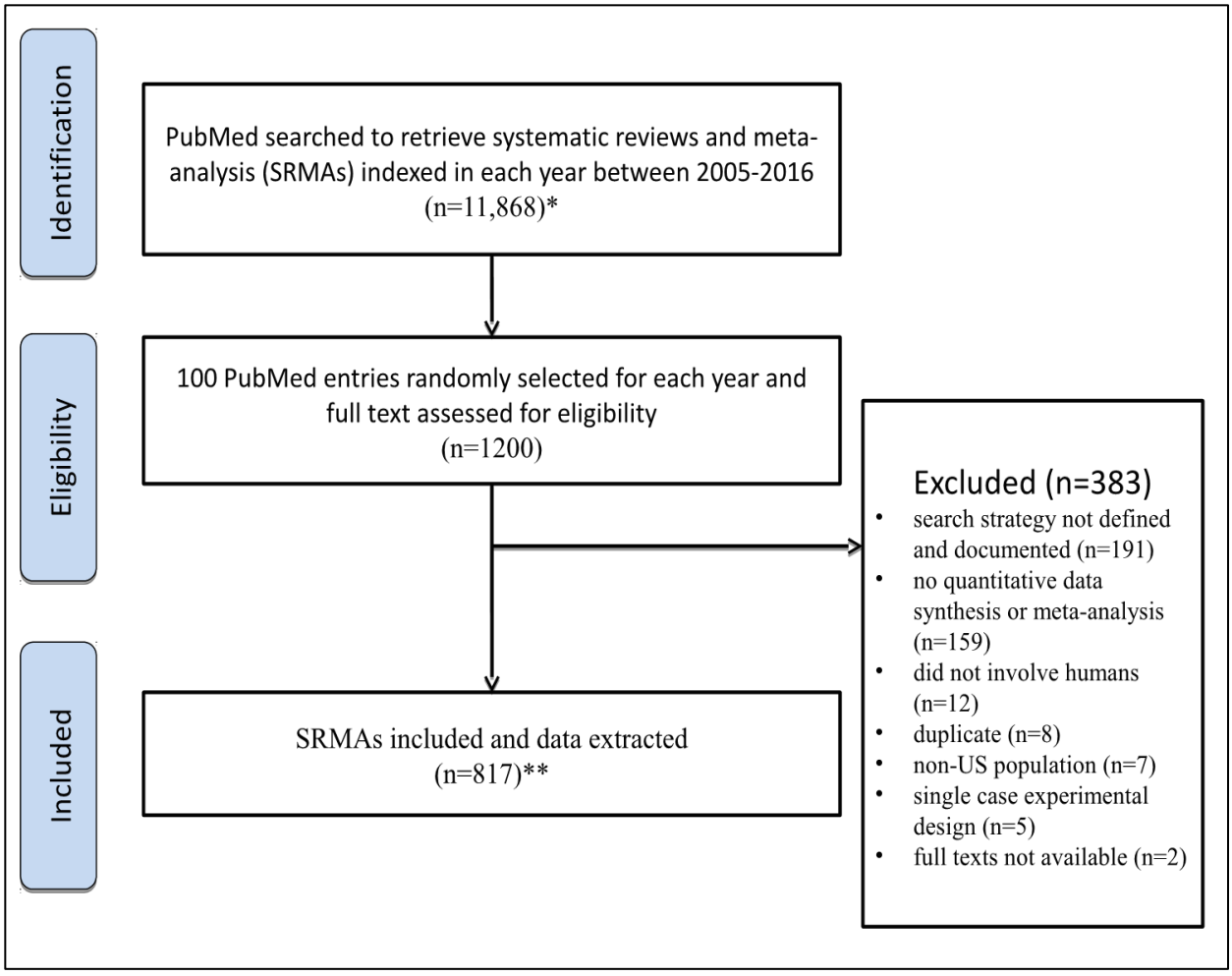


Figure 1. Study flow diagram.

[Footnote

* 2005 - 2016 n's: 550, 694, 720, 778, 810, 963, 1107, 1301, 1198, 1071, 1507, 1169

** 2005 - 2016 n's: 66, 59, 65, 74, 71, 69, 65, 64, 62, 68, 75, 79]

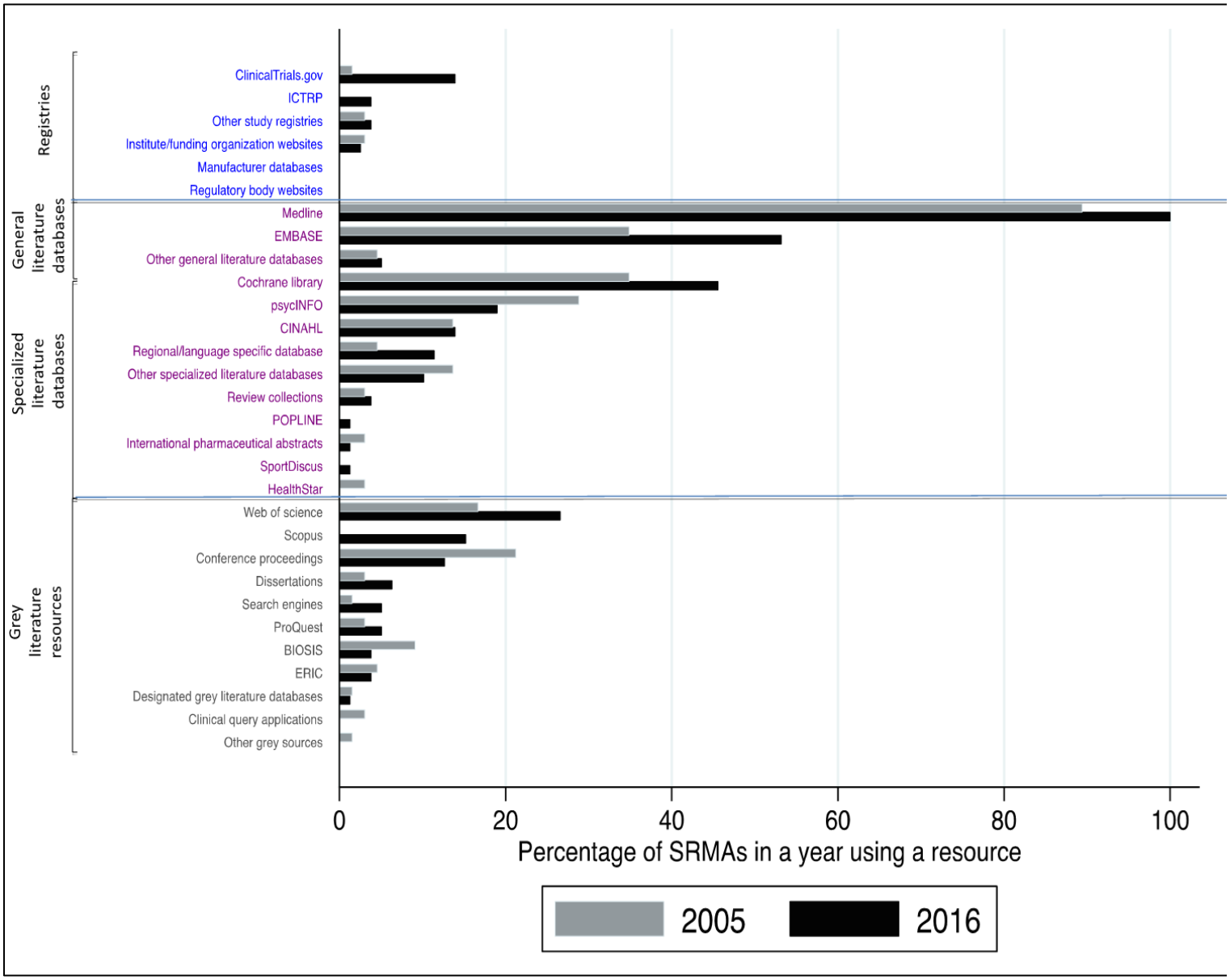


Figure 2. Percentage of SRMAs using each information resource type in the years 2005 and 2016: Grey bars represent percentage of SRMAs using a resource in 2005 (n=66) and black bars, 2016 (n=79). Within each class (registries/general literature database/specialized literature database/grey literature resource), resources are arranged in decreasing order of usage frequency in 2016. Percentages of SRMAs using each information resource are shown in Appendix 3.

Predictors of use of different information resources

I examined factors predictive of searching various types of information resources (Fig 3). Appendix 4 describes the factors predictive of searching individual databases. Methods journals were strongly associated with the use of registries, grey resources, and specialized literature resources. SRMAs used registries more often if a pharmacological agent was being studies. There overall use of registries or grey sources did not increase over the years. Use of specialized literature databases was more prominent in psychiatry journals. Perhaps encouragingly, there was no correlation of journal impact factor and greater use. Interestingly, general medical journals showed a correlation with lower use of general literature databases. Upon examining the 34 SRMAs that did not use any generalized literature database, I found that three were in specialized medical journals, 26 in psychiatry journals and five in general medical journals. For the five in general medical journals, the following resources were used: Cochrane library, Web of Science, regulatory and manufacturer websites, sponsoring website, and specialized resources. All the five SRMAs were published on or before 2012.

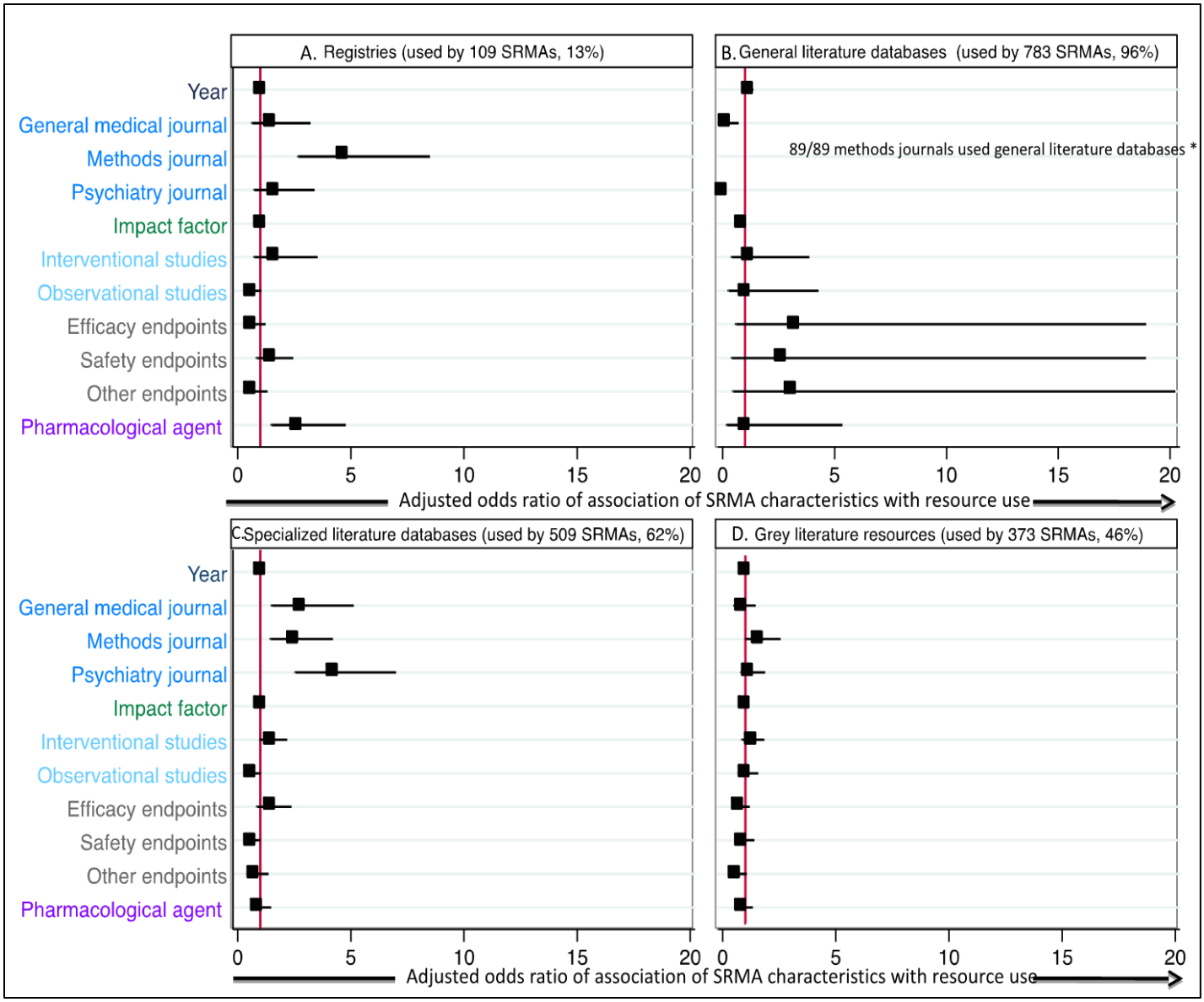


Figure 3. Factors associated with use of various information resources: Using a bivariate logistic regression model, we derived adjusted odds ratios (boxes) and their 95% confidence limits for association of various SRMA characteristics and use of A. registries, B. general literature databases, C. specialized literature resources, D. grey literature resources. SRMA characteristics are color coded to represent year (dark navy), journal type (blue) with specialized medical journals as reference, impact factor (green), primary study type (turquoise), endpoint examined in SRMA (grey), and whether pharmacological agents were examined in SRMA (purple). Factors associated with each individual data source are shown in Appendix 4.

* Methods journal always searched general literature databases, causing the odds ratio to be much greater than 20, the x-axis scale limit.

Networks of information resources

The motivation behind searching multiple resources for SRMAs is to cast a wider net for finding relevant studies. Search strategies are thus most successful when multiple, distinctive resources are searched, and searching many resources is not particularly helpful unless they have substantial non-overlap. I performed network analysis to ascertain the degree of clustering of the different resources searched by the SRMAs over the years. Figs 4A and 4B show representative resource networks formed by the SRMAs in the years 2005 and 2016, respectively. Each node represents an information resource that has been used by an SRMA. A line joins two nodes if any SRMA has used the two information resources concurrently, while the thickness of the line denotes the number of SRMAs that have searched both resources. The closeness of any two nodes in the network depends upon the number of SRMAs that have searched both resources.

In both networks, the three published literature databases Medline, EMBASE, and Cochrane databases form the central resources and are closest to each other, denoting the highest degree of co-occurrence in SRMA searches. The network connections increase substantially from 2005 through to 2016. However, the presence of published literature resources (combining generalized and specialized literature databases, both denoted by purple squares) clustered towards the center and registries (denoted by cyan circles) and other grey literature resources (denoted by grey triangles) towards the network periphery indicates significant amount of co-searching of published literature resources by the same SRMAs.

ANOVA density models constructed to detect homophily (tendency of resources to be associated and, in our scenario, searched concomitantly) confirms that while querying multiple databases authors mostly just search analogous resources likely to have high degree of overlap. The standardized regression coefficient of co-searching of published resources is highest among the possible permutations (0.22, P=0.04). The ranking of information resource type co-searches is as follows:1.two published literature databases (Coefficient: 0.22, P=0.04); 2. A published literature database and a grey literature database (Coefficient: 0.01, P=0.06); 3. A published literature database and a registry (Coefficient: - 0.01, P=0.59); 4. Two registries (Coefficient: - 0.007, P=0.43); 5. A grey literature database and a registry (Coefficient: - 0.04, P=0.43) (reference combination: two grey literature databases).

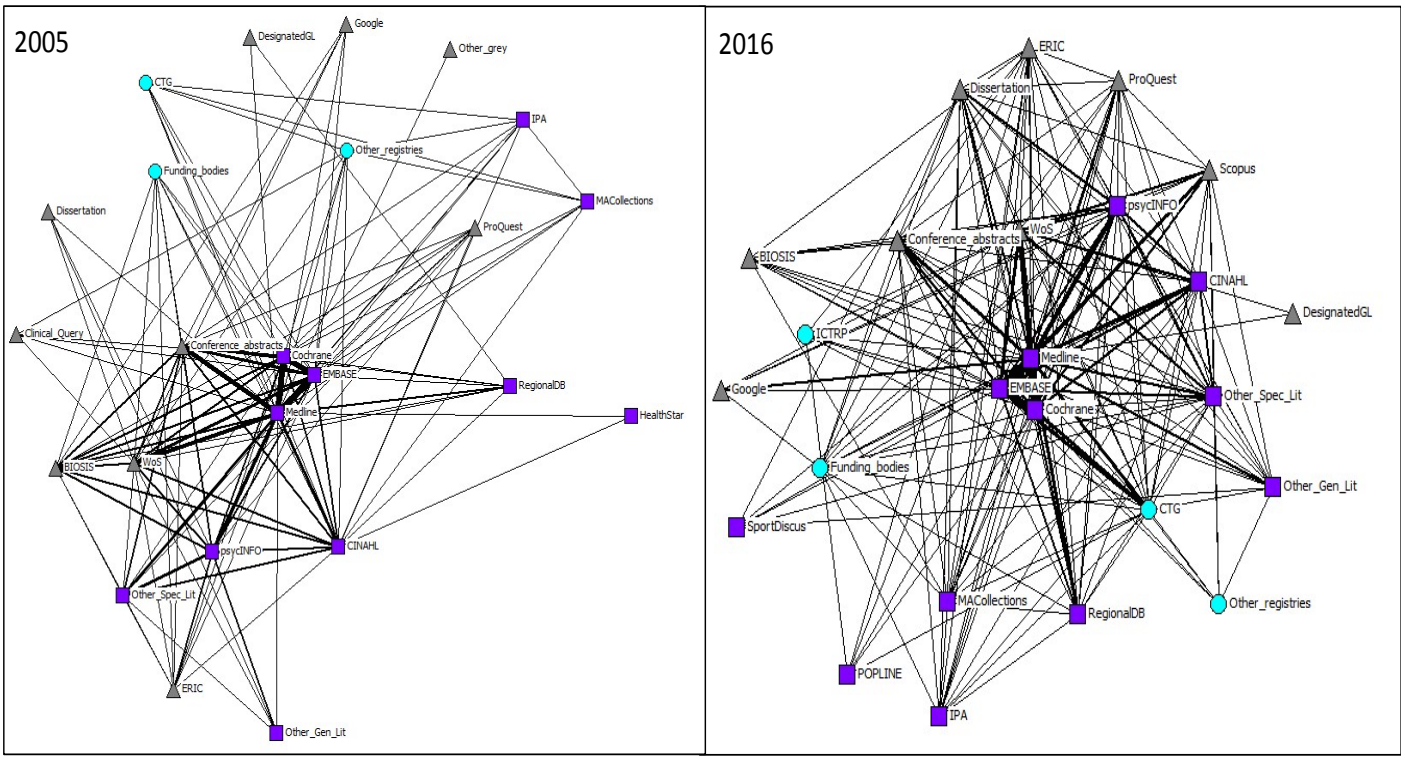


Figure 4. Network of data sources in 2005 and 2016.
Nodes denoted by cyan circles ( ) for registries, purple squares ( ) for general or specialized literature resources, grey triangles ( ) for grey literature resources. Each information resource forms one node in each network, and each node was connected to another by a line if at least one article searched both information resources. The thickness of the connecting lines denotes the number of articles searching the two resources. [Nonstandard abbreviations: CTG=ClinicalTrials.gov, other_spec_lit= other specialized literature database, other_gen_lit=other general literature database, DesignatedGL= designated grey literature database, RegionalDB= regional/language specific database, MAcollections= Collections of EBM reviews/ACP journal clubs]

Effect of registry search guidelines on CTG usage in SRMAs

In 2012, the Cochrane guideline for systematic review and meta-analyses methodology first mandated searching from study registries like CTG and ICTRP as well as grey resources over and above Cochrane databases. I conducted an interrupted time series analysis to see whether or not there was any increase in registry usage after the implementation of the guideline (2014-2016) as compared to before its implementation (2005-2013). I did not find a significant effect of the guideline on the usage of either use of general literature databases, registry, or grey sources (Fig 5A, B, D). The use of specialized literature databases (Fig 5C) and the yearly network densities (number of actual ties between nodes/number of all possible ties between all nodes with theoretical extremes between zero and one) increased from 2014 onwards (Fig 5E). Increasing network density denotes an increase in diversity in the usage of information resources in SRMAs. Results for individual resources are in Appendix 5.

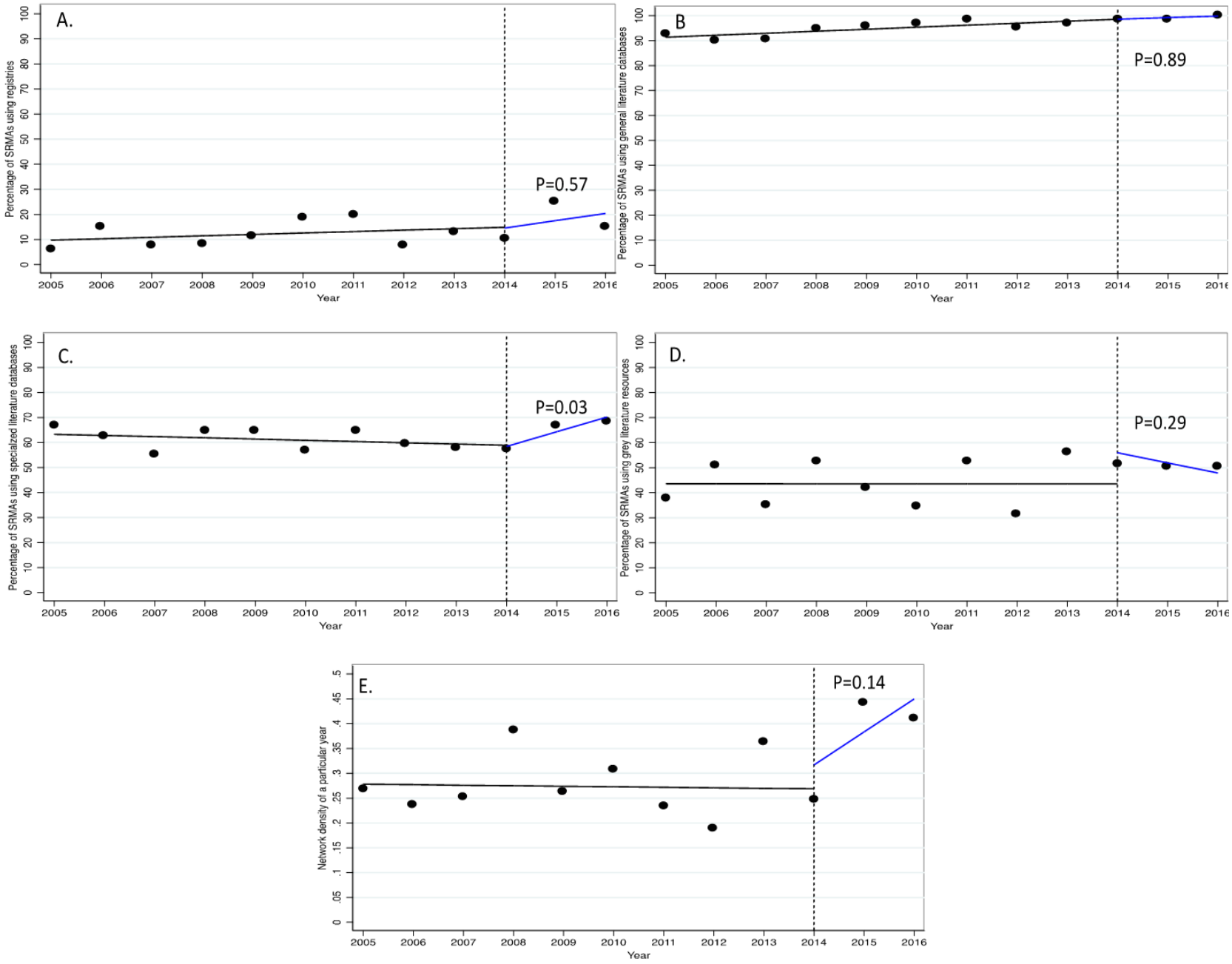

A.
Percentage of SRMAs using registries
P=0.57
Year
B
Percentage of SRMAs using general literature databases
P=0.89
Year
C.
Percentage of SRMAs using specialized literature databases
P=0.03
Year
D.
Percentage of SRMAs using grey literature resources
P=0.29
Year
E.
Network density of a particular year
P=0.14
Year

Figure 5. Interrupted time series analysis showing changes in data-resources before and after 2014: Analysis for A. registries, B. general literature databases, C. specialized literature resources, D. grey literature resources, and E. network densities. For analysis, we used a Praise Weinstein autocorrelation model with lag of one year. Black dots represent the actual percentage of SRMAs using the resource, while the connecting solid line is the model prediction. P values represent the significance of change in resource use after 2014. In each graph the two lines are fit to the first 9 data points (black) and again to the last 3 (blue). Results of interrupted time series analysis for each individual data source are shown in Appendix 5.

Relative importance of information resources in avoiding publication bias

363 SRMAs out of the 375/817 SRMAs that statistically examined publication bias searched Medline. To understand the additional benefit of resources when searched alongside Medline, I identified the resources that were negatively associated with publication bias using a logistic regression model (n=363). Table 2 lists the resources negatively associated with publication bias (adjusted odds ratio <1) in all SRMAs significant at P=0.20, stratified by endpoint type. Appendix 6 provides a list of all resources and their associated adjusted odds ratios.

Of the information resources, Scopus was significantly associated with low publication bias, while CTG exhibited a strong trend towards low publication bias. On repeating the procedure for SRMAs looking at only safety outcomes, low publication bias was associated with CTG.

Table 2. Information sources that, when searched alongside Medline, are negatively associated with publication bias.

Resource	Adjusted odds ratio (95% confidence limits)	P value
For all endpoints (n=363)		
Scopus	0.32 (0.12- 0.88)	0.03
ClinicalTrials.gov	0.33 (0.09-1.20)	0.09
Conference proceedings	0.56 (0.27-1.15)	0.12
Other specialized literature database	0.37 (0.10-1.40)	0.14
For safety endpoints (n=120)		
ClinicalTrials.gov	0.02 (0.00-0.56)	0.02
Scopus	0.17 (0.02-1.72)	0.13
For other endpoints (n=188)		
Conference proceedings	0.14 (0.01-1.28)	0.08
Scopus	0.25 (0.05-1.25)	0.09
SportDiscus	0.10 (0.01-1.52)	0.10

Using a bivariate logistic regression model, adjusted for SRMA characteristics and use of other resources, we derived odds ratios (OR) for association of information resources and the finding of publication bias in the SRMAs that already searched Medline. We list the information resources negatively associated with publication bias, significant at P=0.2, for SRMAs examining any endpoint (n=363), SRMAs examining safety endpoints (n=120), and SRMAs examining other endpoints (n=188). No resource was significantly negatively associated with publication bias at P=0.2 for only efficacy endpoints (n=148). Odds ratios for all resources are shown in Appendix 6.

Discussion

Searching from a large and diverse body of studies is key to valid evidence generation in systematic reviews and meta-analyses (SRMA). However, little evidence-based guidance exists on which resources to search, resulting in haphazard searches of cherry-picked resources. I conducted a systematic review of the information resources searched by SRMAs published from 2005 to 2016. The best predictor of use of alternate resources is the type of journal publishing the SRMA. Although search guidelines have not directly contributed to the increase in registries or grey resources, the diversity of searches after

propounding of 2012 guidelines has increased. However, overall, there is still considerable sequestration in use of published literature databases. Among the resources, searching CTG and Scopus were significantly associated with low publication bias.

Although many studies have shown that registries are underutilized in SRMAs, factors associated with greater use have not been studied. I found that, adjusted for other covariates, methods journals are most likely to use trial registries. The methods journals were also associated with increased use of other resources including conference proceedings, dissertations and all grey literature resources combined. This implied that the policies implemented by editors of such journals significantly affect SRMA search strategies. At the same time, SRMAs in both general and specialized medical/surgical journals rarely searched trial registries and other grey literature resources. This is concerning because such journals not only form the majority of journals publishing SRMAs and the likeliest to be read by clinicians who might practice healthcare based on the results of SRMAs. Encouragingly, I did find an increasing trend in the use specifically of CTG over time.

Although several guidelines exist directing the use of information resources in conducting SRMAs, advice regarding use of resources storing unpublished data is mostly ambiguous.[18] In their review examining the clarity of SRMA guidelines, Boden et al. find that only three provide procedural guidance on using databases such as registries to identify unpublished trials: the Cochrane MECIR guideline (2012),[34] the Centre for Reviews Dissemination (CRD) guideline (2009)[47], and the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA-P) guideline (2015).[48] Of the

three, although CRD was the earliest to suggest use of registries to find unpublished studies, the Cochrane MECIR guideline published in December, 2012 were the first to "mandate" searching CTG, ICTRP and other grey literature resources. Hence, to examine whether the guidelines had influenced the practice of CTG search in SRMAs I tried identifying differences in trends of database usage before and after 2014 (having a one year lag after promulgation of Cochrane MECIR guideline in December 2012). I found that in spite of a general increase from 2005 to 2016, there was no significant effect of the guidelines on the absolute rate of CTG search 2014 onwards. This suggests that mere guideline recommendations are insufficient to ensure a wider use of registries as information resources in SRMAs. While a survey by Tetzlaff et al., 2006 finds a majority of meta-analysts and editors agree that unpublished data should be searched, such practice is not implemented routinely.[49] All journal editors, mirroring the practices in methods journals, should insist upon searching registries. However, search guidelines did seem to increase the use of diverse resources as evidenced by a progressively higher network density after 2014. This can explain Kicinski and colleagues' finding of a decrease in publication bias evidenced in relatively recent SRMAs.[50]

The purpose of searching multiple resources for SRMAs is to avoid missing relevant research that may not have been indexed in one particular database. Little is gained by searching multiple databases if there is significant overlap in content covered by them. Efficient search strategies should aim to retrieve the maximum number of studies using only the most essential information resources: not only because several data resources require paid subscriptions, but also searching each resource requires substantial time,

skilled manpower, and eventual elimination of duplicate studies. Therefore, searching data resources diverse in their study coverage should be the aim of the review strategy. However, I found evidence of sequestration of similar resources in the SRMA searches. For example, the most common two resources to co-occur in searches were Medline and EMBASE. Medline, EMBASE, and the Cochrane libraries form the most used triad. Similar sequestration was noted among all databases focusing on published literature. Conversely, co-searching diverse resources such as databases mining published literature and registries or databases mining published literature and grey literature sources were relatively rare. Such practice is questionable because there is considerable overlap between published literature resources like Medline and EMBASE, and studies suggest searching EMBASE in addition to Medline provides little incremental benefit.[19] On the other hand, Cochrane CENTRAL databases mainly mine data from Medline and EMBASE (only 0.02% of studies deposited in CENTRAL were not from either Medline or EMBASE).[51] Such sequestration of search resources seems reflective of blindly following prevalent practice, rather than well thought-out strategies. In such scenarios, instead of investing time searching studies from similar types of resources, researchers should diversify the nature of their information resources based on the type of studies covered (published/unpublished).

Finally, of the resources used in SRMAs, searching Scopus and CTG were associated with statistical findings of no publication bias. The benefit of Scopus may be explained by the fact that it has 100% coverage of multiple resources including Medline, EMBASE, Compendex, World textile index, Fluidex, Giobase, and Biobase. Additionally, it has

approximately 20% more citation coverage than Web of Science, and covers patents and web literature as sources of grey literature.[52] CTG usage showed a trend towards low publication bias in general, while this link was stronger in studies with safety endpoints. The benefit of CTG can be explained by its coverage of unpublished trial reports as evidenced in previous research.[26] The fact that CTG provides a wider coverage of safety data than published articles may explain its benefit in SRMAs with safety studies. [27,29]

This study has several limitations. First, I look at use of information resources in search strategy and link this to publication bias. This is, however, an indirect linkage, as mere searching of a certain resource may not lead to finding additional studies in individual cases. Moreover, I use statistically measured publication bias results as end points in the above analysis. However, many studies do not report statistical tests to identify publication bias, simply because determination of publication bias by most statistical tests is imperfect. However, due to our relatively large sample size, I aimed to find an association between finding publication bias and information resources. Also, an ideal scenario for an interrupted time series analysis is to have more endpoints both before and after the intervention than I did. However, since the Cochrane guideline was a recent implementation, meaningful distal time points were unavailable.

In conclusion, this study provides evidence that registries and other forms of grey literature in SRMA are mostly used in methods journals only, and recommendations in best practice guidelines may not be sufficient for widespread use of such resources. I additionally found that even SRMAs that search multiple sources tend to search similar

resources, rather than databases with diverse coverage. This study provides supports using Scopus and CTG in addition to Medline search to reduce publication bias.

CHAPTER 3

Concordance of data on serious adverse events and death between ClinicalTrials.gov, journal articles and FDA medical reviews

Introduction

The issue of discrepancy of data between CTG and articles reach beyond concerns about lack of research reproducibility: it also prevents building confidence regarding the usability of such data in secondary research. Journal articles have limited publishing space, and do not report all safety data.[53,54] Since secondary researchers systematically review published articles, the unpublished safety data often remain ignored in meta-analysis research, resulting in a falsely skewed impression favoring drug efficacy that may ignore existing (yet under-published) safety issues. Increased use of wider safety data reported in registries may pose a solution, provided such data is correct.

Importantly, previous works reporting discrepancy between CTG and published articles implicitly assume the veracity of article data, just because they are peer reviewed.[29] However, the peer reviewers do not get access to the patient level data, and do not perform re-analysis of the data.[55] Thus, the peer review process cannot guarantee correctness of article data. On the other hand, the FDA reviews patient level data, and makes these reviews publicly available.[30] Moreover, while the consequences of data manipulation in articles can mostly be academic or ethical, manipulated data presented to the FDA will have legal implications.[31,32] Thus, FDA reviews are more appropriate gold standards to judge to data accuracy in registries.

In this chapter, I compared the concordance of safety data among the three resources: FDA medical reviews (the reference standard), CTG and journal articles, (the test resources), in a random sample of new molecular entities (NMEs) approved by the FDA between 2013-2015.

Methods

Searching and eligibility

Of the new molecular entities (NMEs) approved from 2013 to 2015, I randomly sampled 30% of the medications, stratified by their year of approval. Thereafter, I extracted safety information from trials considered "pivotal" for the FDA approval of the respective drugs. I excluded a drug from the final analysis if 1. Relevant trial reports were available within all three resources, 2. Cohort sizes of the trials described in all three resources were comparable, and 3. Data from two or more pivotal trials were pooled to evaluate safety in the FDA reports (since pooled data would increase uncertainty about the inconsistency within the sources).

Data sources and extraction

I manually extracted data from the three resources: the FDA, the CTG, and the published articles. Data from the FDA was extracted from the website Drugs@FDA, from the medical reviews enlisted within the drug approval notices. The pivotal trials used to assess safety by the FDA reviewer were identified and relevant data were extracted. All safety information was taken from the most recent NME submission cycle, or if enough detail wasn't provided in that, the last cycle report where safety details were provided.

Thereafter, CTG was searched for the corresponding trial reports using the drug names and trial acronyms. Published articles were identified from the Medline indexed articles that are automatically linked to the relevant CTG report. That the reports extracted from the three resources described the same trial was confirmed by comparing the cohort sizes in the three reports.

Outcomes of interest

For all drugs, I noted the numbers of deaths (from all causes), and number of serious adverse events (SAE) noted in each pivotal trial reports from the three resources. These were the primary outcomes of interest since these safety endpoints have uniform definitions across the resources. From each resource, I noted the outcome denominators, i.e., the total number of individuals exposed to each treatment group (receiving NME/control), and the numerators, i.e., the number (or proportion) of patients with at least one outcome. If the numerators were reported as percentages in any report, those percentages were converted to full numbers and rounded up. I calculated outcome rates by determining the percentage of patients exposed to treatment who had outcomes. Additionally, I noted the total period of observation for which each endpoint within each report was recorded. I calculated the observation period as the time difference between the trial initiations till the cut-off date for database closure for the analysis, as noted in each of the three resources.

To look at reporting concordance among the resources for more granulated safety endpoints, I selected a subgroup of cancer NMEs. This is because the secondary safety

endpoints are better described across resources for cancer medications. For this evaluation, I selected all medications included in our final analysis that were L (anticancer and immunomodulatory) drugs according to the Anatomic Therapeutic Chemical drug classification.[56] I selected the safety endpoints mentioned in the "Adverse reactions" section described the product label associated with the NME. My rationale was that: 1. Since the risk-benefit assessments of the NMEs of these medications of concern, the safety endpoints of these drugs are likelier to be subject of quantitative meta-analysis in future assimilations, and should thus be reported consistently across the trial reports, and 2. These safety endpoints are described using "Preferred terms" from the MedDRA vocabulary across the resources, thus making the case definitions uniform across the data sources. I compared the reported numbers for these safety events of all severities; for CTG, I added the numbers reported as "Serious AEs" and "Other AEs" to arrive at the total numbers for "all grade AEs".

Analysis

I assessed discordance between the point estimates of the extracted numerators, denominators, and calculated AE rates as reported within the three resources. I estimated any mismatch between the index and reference resources on numerators, denominators, outcome rates, and observation time as a percentage change from the value in the reference standard:

$$\frac{(\text{Value}_{\text{FDA}} - \text{Value}_{\text{test_resource}}) \times 100}{\text{Value}_{\text{FDA}}}$$

I considered mismatches over ±30%[1] of the FDA value as being significant.

I then measured if there was any consistent pattern in over/under reporting of endpoints in either CTG or articles as compared to the FDA reviews. To explain differences in reporting, I compared the cut-off time frames used for reporting of the safety endpoints. Numerical data were summarized as medians and interquartile ranges while categorical data were presented as frequencies. To compare between groups of numerical data I used Mann-Whitney test, whereas to compare between categorical data, chi-squared analysis was used. I calculated linear correlations between numerical data using Spearman's Rho. I considered a P value lower than 0.05 to be statistically significant. All statistical analyses were performed using Stata version 14.[45]

Results

Out of a total of 113 NMEs approved between 2013-2015, I selected a 30% random sample of 38 NMEs stratified over approval year. Of these, 13 NMEs met our eligibility criteria. The sampling process and reasons for exclusion are described in Figure 1, the most common cause of exclusion being the pooling of safety data across trials in the FDA medical reviews. Four medications were approved in 2013 (Riociguat, Conjugated estrogens/Bazedoxifene, Afatinib, Trametinib), four in 2014 (Ceftolozane/Tazobactam, Ramucirumab, Peginterferon beta-1a, Siltuximab), and five in 2015 (Edoxaban, Ivabradine, Sacubitril/Valsartan, Ixazomib, Aripiprazole lauroxil). Six out of the 13 medications were ATC class L (anticancer and immunomodulatory drugs), three were class C (cardiovascular drug), and one each of classes B (drug for blood and blood forming organs), G (sex hormone), J (anti-infective), and N (drug for nervous system). Eleven of the 13 drugs contributed one trial each, while two had two trials each. Of the

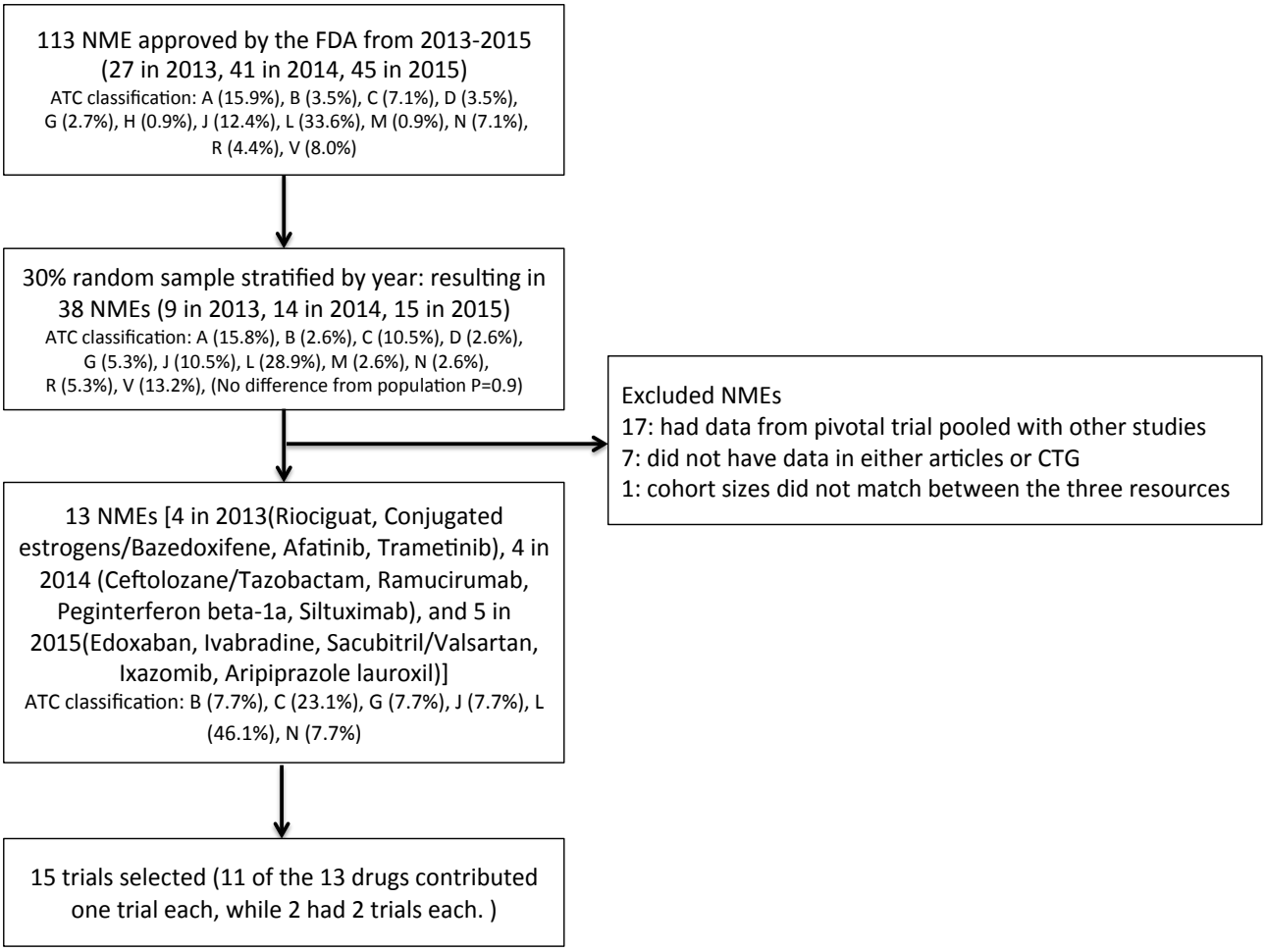


Figure 1. Selection trials of new molecular entities (NMEs) approved by the United States Food and Drug Administration (FDA). [Abbreviations: ATC= Anatomic Therapeutic Chemical; CTG= ClinicalTrials.gov; Drug classes: A: Alimentary tract and metabolism; B: Blood and blood forming organs; C: Cardiovascular system; D: Dermatologicals; G: Genito-urinary system and sex hormones; H: Systemic hormonal preparations, excluding sex hormones and insulins; J: Antiinfectives for systemic use; L: Antineoplastic and immunomodulating agents; M: Musculo-skeletal system; N: Nervous system; R: Respiratory system; V: Various]

15 trials, four trials had three arms each [two test arms (with a different dose of test drug each) and a control arm], while the rest 11 two arms each (one test and one control arm). For the ease of analysis, in trials that had three arms, I converted the two active drug arms into one by adding across the numerators and denominators. A total of 60 point-estimates (30 numerators and 30 denominators) were to be compared across three resources per endpoint.

There were frequent mismatches between the index and reference sources in data (either numerators or denominators) in NME arms for both death (72% for ClinicalTrials.gov, 53% for articles) and SAEs (30% for ClinicalTrials.gov, 30% for articles). A similar pattern of frequent but minor deviations in mismatch for death and SAE was also seen for control arms as shown in Table 1. However, mismatches of >30% from FDA values were infrequent for outcome rates in in both resources [for death rates, 18% cases in ClinicalTrials.gov and 6% cases in articles; for SAE rates, never in either resource]. When CTG and articles were compared with FDA reviews, I did not find any systematic tendency towards over- or underreporting in the endpoints death or SAE [median change in risk ratio of death rate from FDA vs CTG as opposed to that in FDA vs article= 0% vs 0%, P= 0.87; median change in risk ratio of SAE rate from FDA vs CTG as opposed to that in FDA vs article= 0% vs 0%, P= 0. 0.51].

Table 1. Differences in death and serious adverse event reporting between Food and Drug Administration (FDA) reviews, ClinicalTrials.gov, and published articles.

Index resource compared with FDA reviews	Number of trials out of 15[a] that had data in both resources compared	Any mismatch in "Numbers of outcomes" (numerators)				Any mismatch in "Numbers at risk" (denominators)				>30% mismatch in outcome rates [d]			
		NME[b] arm		Control arm		NME arm		Control arm		NME arm		Control arm	
		Number of trials with any mismatch	Median % mismatch (IQR[c])	Number of trials with any mismatch	Median % mismatch (IQR)	Number of trials with any mismatch	Median % mismatch (IQR)	Number of trials with any mismatch	Median % mismatch (IQR)	Number of trials with any mismatch	Median % mismatch (IQR)	Number of trials with any mismatch	Median % mismatch (IQR)
Death													
ClinicalTrials.gov	11	7/11	0 (0 to 4.2)	6/11	0 (0 to 5.6)	4/11	0 (0 to 0)	4/11	0 (0 to 0)	2/11	0 (-0.1 to 4.2)	3/11	0 (0 to 5.6)
Article	15	6/15	0 (0 to 0)	5/15	0 (0 to 0)	7/15	0 (-0.3 to 0)	7/15	0 (0 to 0.3)	1/15	0 (-0.1 to 0.3)	2/15	0 (-0.01 to 0.7)
Serious adverse events													
ClinicalTrials.gov	15	5/15	0 (0 to 0)	5/15	0 (0 to 0)	2/15	0 (0 to 0)	2/15	0 (0 to 0)	0/15	0 (-0.1 to 0)	1/15	0 (-0.7 to 0)
Article	10	3/10	0 (0 to 0)	4/10	0 (-5 to 0)	2/10	0 (0 to 0)	2/10	0 (0 to 0)	0/10	0 (0 to 0)	1/10	0 (-5.6 to 0)

None of the comparisons were significant at P=0.05.

Table 2 shows the mismatches in data on safety endpoints mentioned in product labels of the six anticancer/immunomodulatory drugs in our sample. Mirroring the pattern in death and SAE, frequent but minor deviations were seen for "Adverse reactions" for the selected anticancer drugs. Frequencies of any mismatches for adverse reactions were similar in ClinicalTrials.gov and articles for NME arms (80% in ClinicalTrials.gov, 62% in articles, P=.11) as well as controls (64% in ClinicalTrials.gov, 44% in articles, P=.13).

I explored whether discrepancies between the data sources can be explained by differences in time frames of cohort observation. In general, such data was poorly reported in the articles (unclear in 4/15 articles). The time frames were available from all three resources in 10/15 trials. In those ten trials, the percentage change in observation period from FDA reviews and CTG or articles were comparable (median percentage change from FDA in CTG 2.2% vs that in articles 1 %, P= 0.47). While the time frames were rarely exactly same between FDA and either of the test sources, there were no significant correlations between the changes in time frame with respect to FDA review and changes in event rates (correlation coefficient between percentage difference in time frame vs percentage difference in event rates active arms: FDA vs CTG for death Rho= -0.20, P= 0.58; FDA vs articles for death Rho= 0.006, P= 0.98; FDA vs CTG for SAE, Rho= 0.08, P= 0.77; FDA vs article for SAE Rho= -0.20, P= 0.65). This makes differences in time frames of cohort observation unlikely to be the chief source of the discordances in event rates.

Table 2. Mismatches in reporting of safety outcomes mentioned in "Adverse Reaction" section of product labels for anticancer drugs between FDA reviews, ClinicalTrials.gov, and published articles.

New molecular entity (Approval year)	Adverse reaction (in MedDRA preferred term)	Percentage mismatch in outcome rate at ClinicalTrials.gov compared to FDA		Percentage mismatch in outcome rate in articles compared to FDA	
		NME[a] arm	Control arm	NME arm	Control arm
Trametinib (2013)	Rash	0.0	0.0	0.0	0.0
	Diarrhea	0.0	0.0	0.0	0.0
	Lymphedema	-9.1	0.0	Not reported	Not reported
Afatinib (2013)	Diarrhea	-129.2	-8.7	-127.1	26.1
	Rash/Dermatitis acneiform	-60.0	0.0	-126.7	36.4
	Stomatitis	-22.5	26.7	-132.4	-13.3
	Paronychia	-124.1	0.0	-124.1	0.0
	Dry skin	-122.6	0.0	-116.1	0.0
	Decreased appetite	-127.6	0.0	-62.1	0.0
	Pruritus	-119.0	0.0	-104.8	0.0
Siltuximab (2014)	Pruritus	4.3	0.0	4.3	0.0
	Increased weight	0.0	0.0	0.0	0.0
	Rash	-17.4	-33.3	21.7	0.0
	Hyperuricemia	0.0	100.0	0.0	100.0
	Upper respiratory tract infection	26.9	20.0	26.9	20.0
Ramuciruab (2014)	Hypertension	10.3	-22.2	2.6	0.0
	Diarrhea	0.0	10.0	Not reported	Not reported
PEGylated-Interferon Beta-1a (2014)	Injection site erythema	0.1	-0.2	0.0	0.0
	Influenza-like illness	0.3	-0.2	0.0	0.0
	Pyrexia	0.3	1.1	0.0	0.0
	Myalgia	-0.9	60.4	0.0	60.5
	Chills	0.1	-0.2	0.0	0.0
	Injection site pain	-0.6	-0.2	0.0	0.0
	Injection site pruritus	-1.5	-0.2	0.0	0.0
Ixazomib (2015)	Diarrhea	95.4	97.7	-8.3	-7.2
	Constipation	99.2	98.9	-3.0	-4.7
	Thrombocytopenia	94.5	86.1	-53.0	-58.8
	Peripheral neuropathy	0.0	0.0	-161.4	-169.7
	Nausea	97.8	100.0	-12.7	-7.1
	Peripheral edema	98.9	98.5	-10.7	-10.9
	Vomiting	97.5	100.0	-6.0	-8.0
Median (interquartile ranges) of percentage mismatches		0.0 (-17.4 to 10.2)	0.0 (-0.2 to 60.4)	0.0 (-62.1 to 0.0)	0.0 (-4.7 to 0.0)

[a]NME= new molecular entity

Discussion

Built with an original aim to enhance research transparency, trial registries that mandate reporting of trial results also lend themselves as sources of publicly available trial data.[57] However, accuracy of such data has been called into suspicion upon findings that they do not always match corresponding reports in published articles, thus questioning the usability of registry data for secondary research.[28,29] I compared safety data in CTG and articles with FDA reviews to find that the discrepancies in both test resources, while equally prevalent, usually are of modest magnitude.

This work has implications for the use of use of publicly available safety data in registries for secondary research. First, since the differences are small, and not due to biased over or under-reporting, safety data from registries can be considered as an alternative to published article data. There are several reasons to encourage such practice: since accessing articles often require paid subscriptions to journals, performing secondary research becomes a monopoly of resource-rich settings, a present-day reality that ignores scientific perspective of the underprivileged.[58] Since data in registries are publicly available, such inequitable situations would be challenged. Moreover, since a significant proportion of large trials remain unpublished in journals while being reported in registries, increased use of registries in secondary research may help avoid publication bias, particularly in safety data.[26,27]

This work has several limitations. First, to use FDA reviews as gold standards, I limit our sample to only NMEs. Since any report on NMEs are under considerable scrutiny from

regulators, researchers, clinicians and the public, the slightness of differences between FDA reports and CTG/articles in case of NMEs may not extrapolate to trials performed in later stages of a drugs life cycle. Second, the observation periods I attempted to extract from all three resources are only rough estimates, and may not be accurate. Given that the event rates change with time frames, all resources should precisely identify the observation periods for reporting safety data to ensure greater transparency. Third, the sample is small. However, the randomized sampling procedure may ensure external validity of this research.

In conclusion, thisresearch shows that differences in data between FDA reviews, CTG, and articles are prevalent but small. If FDA reviews are considered gold standards, neither CTG nor articles systematically over- or underreport rates of safety events. Future work needs to identify whether the small differences in safety data between the resources qualitatively or quantitatively affect results of meta-analyses.

CHAPTER 4

Examining reproducibility of published meta-analyses using data at ClinicalTrials.gov downloaded by an automatic extraction tool

Introduction

The finding of frequent, minor, yet equally prevalent differences in data between CTG and FDA reviews as opposed to journal articles and FDA reviews raises the question that whether it would make a difference if data at CTG is used instead of article data to conduct of met-analysis research. If such interchangeability is established, it can potentially make meta-analysis research faster and more accurate, since the uniformly reported data at CTG can be automatically downloaded in analysis-ready formats, eliminating the necessity of the error-prone and time consuming process of manual data extraction from articles.

In this chapter, I describe whether the data at CTG can reproduce the results in the published meta-analysis articles. As a part of this work, I led the development of web-based interactive tool (EXACT, **Ex**tracting **A**ccurate efficacy and safety information from **C**linical**T**rials.gov: http://bio-nlp.org/EXACT/) that allows automated extraction of trial data by end-users.

Methods

Selection of published meta-analysis articles for reproduction

Because mandatory reporting of trial results at CTG is relatively recent,[59] I searched for meta-analyses of drugs recently approved by FDA.[60] I listed the new molecular entities (NMEs) approved by FDA in the year 2013, and randomly selected three (10% sample) of those NMEs. I then searched PubMed for meta-analyses involving those drugs. As a proof of concept, I selected one meta-analysis article per drug. I excluded meta-analyses that 1. Used Bayesian or other model-based methods, 2. Did not include both efficacy and safety endpoints, 3. Did not have trial reports available in both published articles and CTG results, or 4. Had ambiguous endpoints. When more than one paper for a drug satisfied all these criteria, I selected the one that had the largest number of endpoints.

Ensuring validity of our meta-analysis methods

To ensure I was using the same methods as the published meta-analyses, I repeated the meta-analyses of endpoints in the published articles using data manually extracted from the primary study articles. I considered a published meta-analysis to have been reproduced by us using data from primary study articles if 1. The relative risk (RR) was within +/-20% of the RR in the published SR, and 2. The P-value remained on the same side of 0.05 as in the published SR. This validated I were using the same methods as the published meta-analyses. Having established that I was using identical methods as the

published meta-analyses for the reproduced endpoints, I repeated the meta-analyses for those endpoints using data extracted from CTG.

Extracting data from CTG automatically using EXACT, a web-based tool

To automatically extract data from CTG, I led the development of a web-based tool called EXACT, (**Ex**tracting **A**ccurate efficacy and safety information from **C**linical**T**rials.gov: http://bio-nlp.org/EXACT/) with Matthew Cornell, our software developer (Please see preface for credits). The EXACT implementation is a Python program[61] comprising a library to parse records expressed in the CTG XML format, a Flask web application[62] that allows the user to customize which data are desired, and library routines to extract those data in a structured format. The XML parsing library contains 30 functions in categories corresponding to general information in the CTG website (trial title, study type, conditions, interventions, and design) and routines to extract data from the sections on baseline, outcome, participant flow, and reported events. The library consists of about 1800 lines of code, about one-third for internal tests to ensure proper functioning.

On the server side, the application requires two databases: A MySQL[63] database that contains the extracted numerical data, and an Apache Solr[64] instance that indexes the fields to be searched for by one of the application's two search features (indication and/or intervention search, and trial body search). The trial body search covers the fields 'arm_group', 'brief_summary', 'brief_title', 'condition', 'intervention', and

'official_title'. Information on the intervention or the indication is automatically downloaded for the user. Figure 1 describes the steps involved in developing the tool.

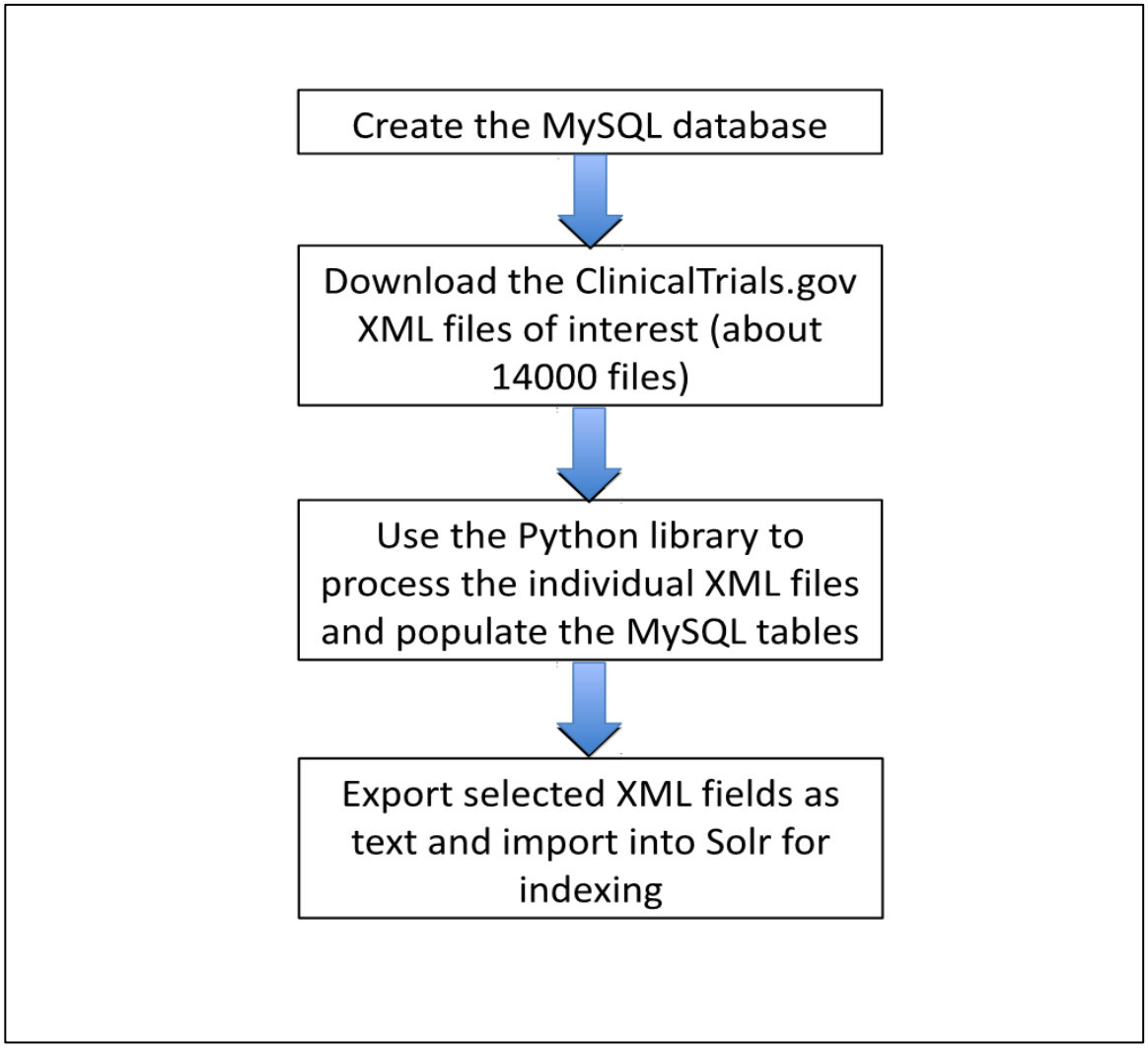


Figure 1. Diagram showing development of EXACT

Sources of data for EXACT

CTG allows users to download data either via the "Clinical Trials Transformation Initiative Aggregate Analysis of ClincalTrials.gov" (CTTI AACT) database[65] or as XML files from a search result.[66] we found it difficult to process the CTTI database (which is updated infrequently), so we chose to program from the search list. We used our Python library[61] to process the XML files and populate a MySQL database with the extracted trial information and data.

Internal validation of the tool

The program was developed using Extreme Programming's Test Driven Development methodology,[67] which resulted in extensive coverage of all library functions via 25 tests to ensure proper functioning (about 600 lines of code). Test inputs were an arbitrary selection of eight actual CTG XML files chosen to represent a range of trials, along with hand-collected expected outputs for them, obtained from each trial's CTG results page.

With the resulting application, a user of EXACT can initiate a search with the unique CTG identifier for a particular trial. Thereafter, assuming the trial results have been reported at CTG, the user can download all trial data, or can select any of 1. reporting groups, 2. period (main trial period/follow-up period) and participant flow, 3. outcome measures, 4. serious adverse events, and 5. other adverse events. Figure 2 shows the user interface and Appendix 7 is a manual to use EXACT.

A. Trial result in ClinicalTrials.gov
Serious Adverse Events
participants affected / at risk
Imatinib 400 mg QD
Nilotinib 300 mg BID
0/280 (0.00%)
1/279 (0.36%)
B. Using EXACT
Select reporting groups (Step 1)
Reporting groups
Imatinib 400 mg QD
Nilotinib 300 mg BID
Nilotinib 400 mg BID
Next
Select serious adverse events (Step 4)
Serious Adverse Events
Blood and lymphatic system disorders
Anaemia
Bone marrow failure
Febrile neutropenia
Iron deficiency anaemia
Leukocytosis
Leukopenia
Neutropenia
Pancytopenia
Splenomegaly
Thrombocytopenia
Cardiac disorders
Acute myocardial infarction
Angina pectoris
Select period and participant flow (Step 2)
Periods
Period 1: Overall Study
STARTED
SAFETY ANALYSIS SET
COMPLETED
NOT COMPLETED
Adverse Event
Select other adverse events (Step 5)
Other Adverse Events
Blood and lymphatic system disorders
Anaemia
Leukopenia
Neutropenia
Thrombocytopenia
Cardiac disorders
Palpitations
Eye disorders
Conjunctival haemorrhage
Conjunctivitis
Dry eye
Eyelid oedema
Periorbital oedema
Gastrointestinal disorders
Select outcome measures (Step 3)
Outcome Measures
1. Primary: Molecular Response Rate (MMR)
Number of Participants [units: participants]
Molecular Response Rate (MMR) at 12 Mo
2. Secondary: Rate of Durable MMR at 24 Mo
3. Secondary: Rate Reduction in BCR-ABL
Confirm entire query (Step 6)
Query Summary
NCT ID: NCT00471497
Reporting Groups:
Imatinib 400 mg QD
Nilotinib 300 mg BID
Period: Overall Study
Participant Flow > Milestones:
SAFETY ANALYSIS SET
COMPLETED
NOT COMPLETED
Participant Flow > Drop Withdraw Reason
Adverse Event
Outcome Measures:
1. Primary: Molecular Response Ra
2. Secondary: Rate of Durable MMR
Adverse Events:
(Serious) Blood and lymphatic syste
C. Trial data in analysis-ready spreadsheet format
AF
AG
AH
AI
Adverse Events
Outcome Measure 1
Group 1: Affected
Group 1: At Risk
Group 2: Affected
Group 2: At Risk
0
280
1
279

Figure 2. EXACT's user interface: A. Screenshot of trial result reported at ClinicalTrials.gov; B. Screenshot of six steps through which the user specifies the items to be downloaded (Appendix describes a user manual); C. Screenshot of data extracted in excel format.

Use of the CTG data for meta-analysis

To validate the use of CTG data thus extracted, I re-conducted meta-analyses published in three peer-reviewed articles with data extracted from CTG. I considered a meta-analysis to have been reproduced if 1. The relative risk (RR) was within +/-20% of the RR in the published SR, and 2. The p-value remained on the same side of 0.05 as in the published SR. All analyses used Stata version 14 and used the same random-effects or fixed-effect model as in the published SR.

Results

From the 27 NMEs approved by the FDA in 2013, I randomly selected three: Simeprevir, Trametinib, and Vortioxetine. Figure 3 describes the article selection, Table 1 describes the articles selected, and Appendix 8 describes the reasons for excluding other SR articles.

The three SR articles contained meta-analyses of a total of 28 endpoints. From manually extracted data I were able to reproduce results for 25 endpoints (details of the other three endpoints are in Appendix 9). Then I sought to reproduce the meta-analysis results for these 25 endpoints using the CTG data extracted by EXACT. The 25 outcomes required extraction of 498 pairs of data elements from original articles and from CTG via EXACT. [Besides 480 data points for 23 endpoints (4 from Qu et al., 2015, 7 from

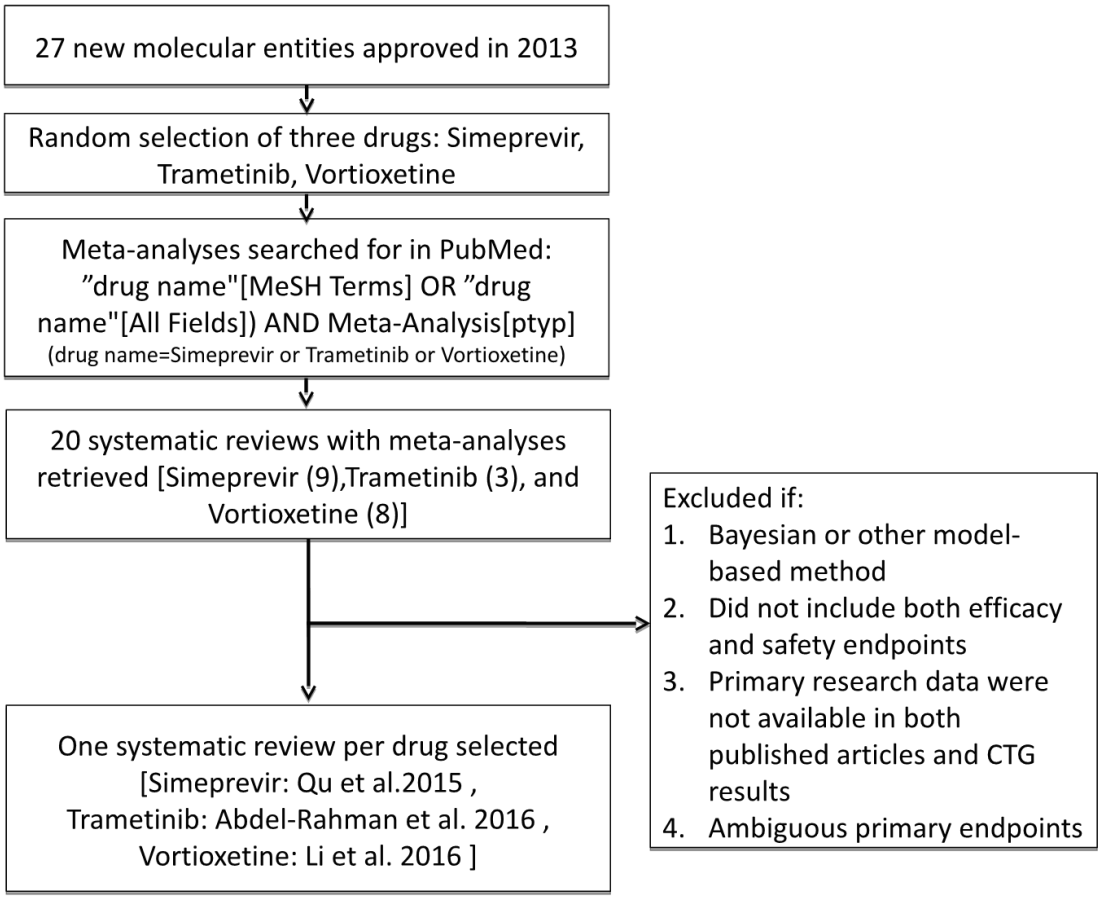


Figure 3. Flowchart for selection of meta-analyses reproduced using data manually extracted from Original articles and data extracted by EXACT from ClinicalTrials.gov

Table 1. Selected systematic reviews.

Systematic review	Comparison	Number of RCTs pooled	Endpoints selected for meta-analysis
Qu et al., 2015	Comparison of Simpeprevir + Peginterferon + Ribavirin versus Peginterferon + Ribavirin in HCV infection	6	Sustained virological response at 12 weeks, Rapid virological response, Serious adverse events, Discontinuation
Abdel-Rahman et al., 2016	Comparison of BRAF inhibitors +Trametinib versus BRAF inhibitors in BRAF-mutated melanoma	4	Hazard ratio for progression-free survival, Hazard ratio for overall survival, Overall response rate, Diarrhea, Hypertension, Decreased ejection fraction, Acneiform dermatitis, Pyrexia, Squamous cell carcinoma
Li et al., 2016	Vortioxetine versus Duloxetine in major depressive disorder	5	Response rate, Remission rate*, Nausea, Constipation, Hyperhydrosis, Diarrhea , Dizziness , Dry mouth, Fatigue, Insomnia, Somnolence, Nasopharyngitis*, Decreased appetite*, Headache, Vomiting

RCT= randomized controlled trial, HCV= Hepatitis C virus, BRAF= a gene encoding B-Raf protein.
* Endpoint that could not be reproduced

Abdel-Rahman et al., 2016, and 12 from Li et al., 2016), 18 data elements (hazard ratio and its two 95% confidence limits from each of six trials (four for progression-free survival and two for overall survival) used in Abdel-Rahman et al., 2016)]. 87% of the extracted numbers matched between the two sources (details in Appendix 10). An equal amount of trial data was available from original articles and CTG for 20 of the 25 outcomes. More trial data were available from original articles for one efficacy outcome (hazard ratio for overall survival in Trametinib), and CTG provided more data for four safety outcomes (hypertension and acneiform dermatitis for Trametinib; hyperhidrosis and somnolence for Vortioxetine).

Using the data extracted from CTG, I was able to reproduce results for 22 endpoints (88%). The median difference in the risk ratio between published meta-analysis and reproduction using CTG data was 0.005 (Interquartile range: minus 0.0015 to 0.0175) for the efficacy endpoints, and 0.01 (Interquartile range: 0 to 0.02) for the safety endpoints. The results are shown in Table 2. [I was unable to reproduce meta-analyses for weighted mean differences in Li et al. 2016 because the primary articles for this SR did not consistently report measures of dispersion, Appendix 11. Replication using data at CTG, where standard deviations were reported, is presented in Appendix 12.]

Approximately 10 hours were spent collecting data from published articles, and approximately 4 hours were spent using EXACT. Thus, EXACT reduced the time of data extraction by 60%.

Table 2. Results of meta-analyses from published systematic review and meta-analyses (SRMA) using data manually extracted from original articles and from ClinicalTrials.gov using EXACT.

Outcomes	Type of outcome	Relative risk (95% confidence limits)			P values			I^2 statistic		
		Published SRMA	Data from original article	Data from CTG via EXACT	Published SRMA	Data from original article	Data from CTG via EXACT	Published SRMA	Data from original article	Data from CTG via EXACT
Qu et al.2015, Simeprevir										
Sustained Virological Response at 12 weeks	Efficacy	1.69 (1.37-2.08)	1.692 (1.37-2.08)	1.692 (1.37-2.08)	<0.001	<0.001	<0.001	0.824	0.824	0.824
Rapid Virological Response	Efficacy	9.57 (5.82-15.73)	9.57 (5.82-15.73)	9.68 (5.88-15.95)	< 0.001	<0.001	<0.001	0.636	0.636	0.639
Serious Adverse Events	Safety	0.67 (0.47-0.94)	0.67 (0.47-0.94)	0.65 (0.46-0.92)	0.023	0.024	0.017	0	0	0
Discontinuation	Safety	1.26 (0.58-2.74)	1.033 (0.65- 1.73)	0.98 (0.44-2.19)	0.566	0.899	0.970	0	0	0
Abdel-Rahman et al., 2016, Trametinib										
Hazard ratio for Progression-Free Survival	Efficacy	0.56 (0.49-0.64)	0.54 (0.47-0.62)	0.54 (0.47-0.62)	< 0.00001	<0.001	<0.001	NA	0.548	0.548
Hazard ratio for Overall Survival	Efficacy	0.7 (0.58-0.84)	0.57 (0.46-0.68)	0.67 (0.52-0.83)	0.00001	<0.001	<0.001	NA	0.486	0
Overall Response Rate	Efficacy	1.35 (1.16-1.58)	1.34 (1.23-1.45)	1.34 (1.23-1.45)	0.0002	<0.001	<0.001	NA	0.13	0.13
Diarrhea	Safety	1.3 (1.3-1.49)	1.3 (1.13-1.48)	1.29 (1.13-1.47)	0.0002	<0.001	<0.001	NA	0.909	0.895
Hypertension*	Safety	1.22 (0.99-1.52)	1.22 (0.98-1.51)	1.32 (1.08-1.62)	0.07	0.068	0.005	NA	0.247	0.458
Decreased ejection fraction*	Safety	4.63 (2.56-8.37)	4.63 (2.56-8.36)	3.35 (2.02-5.55)	<0.00001	<0.001	<0.001	NA	0.629	0.642
Acneiform dermatitis	Safety	1.61 (1.03-2.53)	1.61 (1.02-2.53)	1.58 (1.12-2.21)	0.04	0.038	0.008	NA	0.436	0.15
Pyrexia	Safety	1.98 (1.72-2.27)	1.97 (1.71-2.27)	2.18 (1.91-2.49)	<0.00001	<0.001	<0.001	NA	0.822	0.87
Squamous cell carcinoma*	Safety	0.16 (0.1-0.25)	0.16 (0.1-0.25)	0.2 (0.11-0.36)	<0.00001	<0.001	<0.001	NA	0.382	0.224
Li et al., 2016, Vortioxetine										
Response rate	Efficacy	0.83 (0.77-0.89)	0.83 (0.78-0.89)	0.83 (0.77-0.89)	<0.001	<0.001	<0.001	0	0	0
Nausea	Safety	0.7 (0.56-0.87)	0.75 (0.61-0.94)	0.7 (0.61-0.81)	0.001	0.01	<0.001	NA	0.854	0.754
Constipation	Safety	0.47 (0.34-0.64)	0.45 (0.32-0.65)	0.45 (0.32-0.65)	<0.001	<0.001	<0.001	NA	0	0
Hyperhydrosis	Safety	0.35 (0.23-0.55)	0.31 (0.19-0.5)	0.29 (0.18-0.45)	<0.001	<0.001	<0.001	NA	0	0
Diarrhea	Safety	0.74 (0.57-0.97)	0.72 (0.53-0.97)	0.72 (0.53-0.97)	0.03	0.03	0.03	NA	0	0
Dizziness	Safety	0.51 (0.37-0.69)	0.51 (0.33-0.8)	0.51 (0.33-0.8)	<0.001	0.004	0.004	NA	0.624	0.624
Dry mouth	Safety	0.5 (0.39-0.63)	0.48 (0.35-0.65)	0.52 (0.31-0.85)	<0.001	<0.001	0.01	NA	0.295	0.79
Fatigue	Safety	0.45 (0.32-0.64)	0.44 (0.29-0.67)	0.44 (0.29-0.67)	<0.001	<0.001	<0.001	NA	0.163	0.163
Insomnia	Safety	0.65 (0.46-0.92)	0.64 (0.42-0.96)	0.64 (0.42-0.96)	0.016	0.033	0.033	NA	0	0
Somnolence	Safety	0.33 (0.21-0.52)	0.31 (0.19-0.5)	0.33 (0.21-0.5)	<0.001	<0.001	<0.001	NA	0	0
Headache	Safety	0.93 (0.77-1.13)	0.93 (0.74-1.16)	0.93 (0.74-1.16)	0.468	0.523	0.523	NA	0	0
Vomiting	Safety	0.7 (0.45-1.09)	0.72 (0.45-1.16)	0.72 (0.45-1.16)	0.110	0.176	0.176	NA	0	0

Outcomes marked with * could not be reproduced.

NA= Not available

Discussion

This research suggests that, although mismatches occur between data published in original articles and data at CTG, they seldom qualitatively alter the results of meta-analyses. This aligns with studies showing that most mismatches in data at CTG are minor.[68] The strength of our results lies in the random selection of the SR articles and the wide range of endpoints, encompassing both efficacy and safety outcomes. Additionally, we developed a web-based tool to automatically extract data from CTG.

Although previous reports have emphasized the necessity of searching CTG alongside original articles as a source of additional data,[12,69] our study suggests that data from CTG, when used as the primary source, reproduce similar meta-analysis results. Thus, this work recommends use of CTG data particularly for safety results, which are often reported in greater detail at CTG.[27]

Our web-based tool also provides a way to accelerate the process of data extraction for meta-analyses research. Currently users have no publicly available method to extract the vast amounts of trial data available at CTG without advanced programming. Cepeda et al., constructed such an implementation, but it is proprietary.[70] EXACT fills this gap and reduces the time required for primary data extraction. Our tool can enhance the use of CTG by making data extraction from it substantially less laborious. Because EXACT makes no extraction errors, compared with many in manual extraction of data from published articles, it should also reduce errors in primary data extraction.[37] As a verification tool, EXACT could also be useful to journal reviewers and regulatory

authorities who might want to check the CTG site for the validity of data entered,[71] and tally CTG reports with the articles published and regulatory submissions corresponding to these studies. This capability will add to research transparency and reproducibility, areas of specific concern at present.[72]

Data from CTG reproduced the results most meta-analyses of data from original articles. EXACT fills a current gap in medical informatics tools, helping meta-analysis research by providing an application that automatically extracts results data from CTG.

CHAPTER 5

Conclusions

In conclusion, the work described in this dissertation looks into the current scenario of use of information resources in SRMAs to find that registries, although is an underutilized resource, is associated with the finding of low publication bias in SRMAs with safety endpoints. I verify the accuracy of the safety data at CTG by examining its differences with FDA drug reviews. Finally, I repeat meta-analysis using CTG data to show that most meta-analysis results are reproduces using CTG data. This work supports the use of CTG data in meta-analyses of safety endpoints.

References:

1. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. PLoS Med. 2016;13: e1002028. doi:10.1371/journal.pmed.1002028

2. 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: 513–518. doi:10.7326/M14-0933

3. 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: e1001578; dicsussion e1001578. doi:10.1371/journal.pmed.1001578

4. Ebrahim S, Bance S, Athale A, Malachowski C, Ioannidis JPA. Meta-analyses with industry involvement are massively published and report no caveats for antidepressants. J Clin Epidemiol. 2016;70: 155–163. doi:10.1016/j.jclinepi.2015.08.021

5. Honório HM. Should all systematic reviews be at the top of the pyramid? Arch Oral Biol. 2017;73: 321–322. doi:10.1016/j.archoralbio.2016.10.001

6. Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. Milbank Q. 2016;94: 485–514. doi:10.1111/1468-0009.12210

7. Citrome L. Beyond PubMed: Searching the "Grey Literature" for Clinical Trial Results. Innov Clin Neurosci. 2014;11: 42–46.

8. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. The contribution of databases to the results of systematic reviews: a cross-sectional study. BMC Med Res Methodol. 2016;16: 127. doi:10.1186/s12874-016-0232-1

9. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. Cochrane Database Syst Rev. 2007; MR000010. doi:10.1002/14651858.MR000010.pub3

10. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet Lond Engl. 1991;337: 867–872.

11. Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. BMJ. 2012;344: d7202.

12. 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.

13. 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: 252–260. doi:10.1056/NEJMsa065779

14. Conn VS, Valentine JC, Cooper HM, Rantz MJ. Grey literature in meta-analyses. Nurs Res. 2003;52: 256–261.

15. Mahood Q, Van Eerd D, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. Res Synth Methods. 2014;5: 221–234. doi:10.1002/jrsm.1106

16. Saleh AA, Ratajeski MA, Bertolet M. Grey Literature Searching for Health Sciences Systematic Reviews: A Prospective Study of Time Spent and Resources Utilized. Evid Based Libr Inf Pract. 2014;9: 28–50.

17. Paez A. Gray literature: An important resource in systematic reviews. J Evid-Based Med. 2017;10: 233–240. doi:10.1111/jebm.12266

18. Boden C, Bidonde J, Busch A. Gaps exist in the current guidance on the use of randomized controlled trial study protocols in systematic reviews. J Clin Epidemiol. 2017;85: 59–69. doi:10.1016/j.jclinepi.2017.04.021

19. Halladay CW, Trikalinos TA, Schmid IT, Schmid CH, Dahabreh IJ. Using data sources beyond PubMed has a modest impact on the results of systematic reviews of therapeutic interventions. J Clin Epidemiol. 2015;68: 1076–1084. doi:10.1016/j.jclinepi.2014.12.017

20. Zarin DA, Tse T, Williams RJ, Carr S. Trial Reporting in ClinicalTrials.gov - The Final Rule. N Engl J Med. 2016;375: 1998–2004. doi:10.1056/NEJMsr1611785

21. Home - ClinicalTrials.gov [Internet]. [cited 10 Jul 2017]. Available: https://clinicaltrials.gov/

22. 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: 126. doi:10.1186/2046-4053-3-126

23. Sinnett PM, Carr B, Cook G, Mucklerath H, Varney L, Weiher M, et al. Systematic Reviewers in Clinical Neurology Do Not Routinely Search Clinical Trials Registries. PloS One. 2015;10: e0134596. doi:10.1371/journal.pone.0134596

24. Bibens ME, Chong AB, Vassar M. Utilization of Clinical Trials Registries in Obstetrics and Gynecology Systematic Reviews. Obstet Gynecol. 2016;127: 248–253. doi:10.1097/AOG.0000000000001204

25. Yerokhin VV, Carr BK, Sneed G, Vassar M. Clinical trials registries are underused in the pregnancy and childbirth literature: a systematic review of the top 20 journals. BMC Res Notes. 2016;9: 475. doi:10.1186/s13104-016-2280-3

26. 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: e1001566; discussion e1001566. doi:10.1371/journal.pmed.1001566

27. 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: e1002127. doi:10.1371/journal.pmed.1002127

28. 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. Ann Intern Med. 2014;160: 477–483. doi:10.7326/M13-0480

29. 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: 189. doi:10.1186/s12916-015-0430-4

30. Drugs@FDA: FDA Approved Drug Products [Internet]. [cited 10 Jul 2017]. Available: https://www.accessdata.fda.gov/scripts/cder/daf/

31. Affairs O of R. Criminal Investigations - July 2, 2012: GlaxoSmithKline to Plead Guilty and Pay $3 Billion to Resolve Fraud Allegations and Failure to Report Safety Data [Internet]. [cited 19 Oct 2017]. Available: https://www.fda.gov/ICECI/CriminalInvestigations/ucm311065.htm

32. George SL, Buyse M. Data fraud in clinical trials. Clin Investig. 2015;5: 161–173. doi:10.4155/cli.14.116

33. Schwartz LM, Woloshin S, Zheng E, Tse T, Zarin DA. ClinicalTrials.gov and Drugs@FDA: A Comparison of Results Reporting for New Drug Approval Trials. Ann Intern Med. 2016;165: 421–430. doi:10.7326/M15-2658

34. Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological standards for the conduct of new Cochrane Intervention Reviews Version 2.2,. 2012.

35. Horton J, Vandermeer B, Hartling L, Tjosvold L, Klassen TP, Buscemi N. Systematic review data extraction: cross-sectional study showed that experience did not increase accuracy. J Clin Epidemiol. 2010;63: 289–298. doi:10.1016/j.jclinepi.2009.04.007

36. Jones AP, Remmington T, Williamson PR, Ashby D, Smyth RL. High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews. J Clin Epidemiol. 2005;58: 741–742. doi:10.1016/j.jclinepi.2004.11.024

37. Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. J Clin Epidemiol. 2006;59: 697–703. doi:10.1016/j.jclinepi.2005.11.010

38. 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: 531–536. doi:10.1016/j.jclinepi.2008.02.004

39. Tricco AC, Brehaut J, Chen MH, Moher D. Following 411 Cochrane protocols to completion: a retrospective cohort study. PloS One. 2008;3: e3684. doi:10.1371/journal.pone.0003684

40. Gøtzsche PC, Hróbjartsson A, Maric K, Tendal B. Data extraction errors in meta-analyses that use standardized mean differences. JAMA. 2007;298: 430–437. doi:10.1001/jama.298.4.430

41. Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. Syst Rev. 2015;4: 78. doi:10.1186/s13643-015-0066-7

42. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev. 2009; MR000006. doi:10.1002/14651858.MR000006.pub3

43. WHO | Welcome to the WHO ICTRP. In: WHO [Internet]. [cited 10 Jul 2017]. Available: http://www.who.int/ictrp/en/

44. Hanneman R, Ridde M. Introduction to social network methods [Internet]. Available: http://faculty.ucr.edu/~hanneman/nettext/C18_Statistics.html

45. Stata 14 | Stata [Internet]. [cited 10 Oct 2017]. Available: https://www.stata.com/stata14/

46. Borgattu SP, Everett MG, Freeman LC. UCINET for Windows: Software Package for the Social Network Analysis [Internet]. Harvard, MA: Analytical Technologies; 2002. Available: https://sites.google.com/site/ucinetsoftware/home

47. Systematic Reviews: CRD's guidance for undertaking systematic reviews in health care [Internet]. University of York; 2009. Available: https://www.york.ac.uk/inst//crd/index_guidance.htm

48. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;349: g7647.

49. Tetzlaff, Moher, Pham, Altman. Survey of views on including grey literature in systematic reviews.

50. Kicinski M, Springate DA, Kontopantelis E. Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews. Stat Med. 2015;34: 2781–2793. doi:10.1002/sim.6525

51. Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. Eval Health Prof. 2002;25: 38–64. doi:10.1177/016327870202500104

52. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J Off Publ Fed Am Soc Exp Biol. 2008;22: 338–342. doi:10.1096/fj.07-9492LSF

53. Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. Arch Intern Med. 2009;169: 1756–1761. doi:10.1001/archinternmed.2009.306

54. Vera-Badillo FE, 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 Off J Eur Soc Med Oncol. 2013;24: 1238–1244. doi:10.1093/annonc/mds636

55. Smith R. Peer review: a flawed process at the heart of science and journals. J R Soc Med. 2006;99: 178–182.

56. WHOCC - ATC/DDD Index [Internet]. [cited 23 Feb 2017]. Available: https://www.whocc.no/atc_ddd_index/

57. Zarin DA, Tse T, Williams RJ, Rajakannan T. Update on Trial Registration 11 Years after the ICMJE Policy Was Established. N Engl J Med. 2017;376: 383–391. doi:10.1056/NEJMsr1601330

58. Academic Journals are too Expensive For Harvard, Elsevier is Mega Greedy, and Why this Stinks for Future Librarians. In: Information Space [Internet]. 29 May 2012 [cited 19 Oct 2017]. Available: https://ischool.syr.edu/infospace/2012/05/29/academic-journals-are-too-expensive-for-harvard-elsevier-is-mega-greedy-and-why-this-stinks-for-future-librarians/

59. Anderson ML, Chiswell K, Peterson ED, Tasneem A, Topping J, Califf RM. Compliance with results reporting at ClinicalTrials.gov. N Engl J Med. 2015;372: 1031–1039. doi:10.1056/NEJMsa1409364

60. FDA. Available: http://wayback.archive-it.org/7993/20161022052135/http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm381263.htm

61. Welcome to Python.org. In: Python.org [Internet]. [cited 17 Jun 2016]. Available: https://www.python.org/

62. Welcome | Flask (A Python Microframework) [Internet]. [cited 17 Jun 2016]. Available: http://flask.pocoo.org/

63. MySQL [Internet]. [cited 17 Jun 2016]. Available: https://www.mysql.com/

64. Solr. http://lucene.apache.org/solr/. 2009.

65. State of Clinical Trials | Clinical Trials Transformation Initiative [Internet]. [cited 17 Jun 2016]. Available: http://www.ctti-clinicaltrials.org/what-we-do/analysis-dissemination/state-clinical-trials/aact-database

66. Downloading Content for Analysis - ClinicalTrials.gov [Internet]. [cited 17 Jun 2016]. Available: https://clinicaltrials.gov/ct2/resources/download

67. Meyer B, Oriol M, Leitner A. XP and TDD [Internet]. 2006. Available: http://se.inf.ethz.ch/old/teaching/ws2006/0239/slides/TC2006_XP_TDD.pdf

68. 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: 189. doi:10.1186/s12916-015-0430-4

69. Wilson LM, Sharma R, Dy SM, Waldfogel JM, Robinson KA. Searching ClinicalTrials.gov did not change the conclusions of a systematic review. J Clin Epidemiol. 2017;90: 127–135. doi:10.1016/j.jclinepi.2017.07.009

70. Cepeda MS, Lobanov V, Berlin JA. From ClinicalTrials.gov trial registry to an analysis-ready database of clinical trial results. Clin Trials Lond Engl. 2013;10: 347–348. doi:10.1177/1740774513475849

71. Mayo-Wilson E, Li T, Fusco N, Dickersin K, MUDS investigators. Practical guidance for using multiple data sources in systematic reviews and meta-analyses (with examples from the MUDS study). Res Synth Methods. 2017; doi:10.1002/jrsm.1277

72. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? Sci Transl Med. 2016;8: 341ps12. doi:10.1126/scitranslmed.aaf5027