Publication Rates of Randomized Comparative Effectiveness Research Relevant to Primary Care: A ClinicalTrials.Gov Analysis

Sara Riaz, HBSc

Department of Family Medicine
Biomedical Ethics Unit

McGill University, Montreal

June 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Family Medicine

© Sara Riaz 2020

Abstract

Introduction: Pragmatic studies such as the comparative effectiveness research (CER) trial, are increasingly being used to support evidence-based medicine (EBM). The CER trial aims to generate evidence by directly comparing two or more alternative interventions that are used in routine practice; benefit to primary care is substantial. As a result, there is need to monitor publication of completed CER trials that are valuable to primary healthcare. The publication patterns of these trials and characteristics associated with the likelihood of publication have yet to be evaluated. There is also little known about time to publication after study completion for published studies. ***Objective:*** This thesis sought to evaluate what proportion of primary care CER trials registered on clinicaltrials.gov (CT.gov) were published and the time to publication following primary study completion as well as associated factors. ***Methods:*** The CT.gov database was searched for completed randomized CER trials relevant to primary care with a study start date no earlier than September 27, 2007 and a primary completion date no later than August 01, 2015, to ensure a minimum of four-years of follow-up time for publication. PubMed, Embase and for NIH-funded studies, the RePORTER online portal, were searched for matching publications. ***Statistical Analysis:*** Logistic regression was used to identify factors associated with published trials compared to unpublished ones. For published studies, descriptive statistics were used to identify characteristics of all studies, including reporting of significant primary endpoints and lastly, linear regression was used to explore factors associated with time lag from primary study completion to publication. ***Results****:* The study cohort included 122 trials. **56 (46%) trials had no matching full-text publication within four years and 39 (32%) trials had not been published in any form**. Factors associated with published trials compared to unpublished trials on univariate logistic regression were government-funding and trials studying

cardiovascular conditions seen in primary care. On multivariable analysis, the only significant factor associated with publication was a cardiovascular study condition, as compared to an infectious disease study condition (odds ratio [4.4], 95% confidence interval 0.04-2.9, *p*=0.04). Out of the 54% of studies that were published, median time lag between primary study completion and publication was 25 months (IQR, 16-34). Trials comparing interventions for infectious disease and psychiatry conditions were associated with a longer time lag in publishing compared to cardiovascular study conditions (14.9 months, p=0.001 and 11.1 months, p=0.019, respectively). More than half (54%) of published studies reported a significant primary endpoint. ***Conclusion:*** Timely publishing of randomized CER trial results is unsatisfactory. Primary care physicians are often involved in the care of patients with cardiovascular disease, and timely publication of evidence that can inform real-world medical decision-making for these patients is promising. However, without timely reporting of **all** CER trials, potential evidence is left undiscovered and as a result, little to no valuable contribution is made to the overall field of primary healthcare. Consequences of non-publication include lack of regard for the primary care patients who volunteered in these studies, a breach in society's trust in the biomedical research enterprise and a growing skepticism around the substantial efforts toward CER and EBM, given that the purpose of conducting comparative effectiveness studies is undermined when results fail to be reported.

Résumé

Introduction: Les essais pragmatiques comme les essais contrôlés randomisés de recherche d'efficacité comparative (CER) soutiennent la médecine factuelle. L'essai CER crée des preuves en comparant directement deux ou plusieurs interventions alternatives qui sont utilisées dans la pratique courante; les médecins de soins primaires en bénéficieront. En conséquence, il est nécessaire de surveiller la publication des essais CER achevés qui sont précieux pour les soins primaires. Le temps de publication après l'achèvement de l'étude est inconnu. ***Objectif:*** Cette thèse évalue quelle proportion des études comparatives d'efficacité enregistrés liés aux soins primaires ont été publiés et le temps à publication après l'achèvement de l'étude primaire ainsi que les facteurs associés. ***Méthodes:*** La base de données (clinicaltrials.gov) a été recherchée pour les essais CER randomisés terminés pertinents pour les soins primaires avec une date de début au plus tôt le 27 septembre 2007 et une date d'achèvement primaire au plus tard le 01 août 2015, afin d'assurer un minimum de quatre ans de suivi pour publication. PubMed, Embase et pour les études financées par les NIH, le portail en ligne RePORTER, a été recherché pour les publications correspondantes. ***Analyse statistique:*** une régression logistique a été utilisée pour identifier les facteurs associés aux essais publiés par rapport aux essais non publiés. Pour les études publiées, les statistiques descriptives ont été utilisées pour identifier les caractéristiques de tous les essais, la régression linéaire a été utilisée pour étudier les facteurs associés à décalage dans le temps de la fin à la publication. ***Résultats****:* Notre cohorte comprenait 122 essais. **56 (46%) essais n'avaient pas de publication correspondante dans les quatre ans et 39 (32%) essais n'avaient été publiés sous aucune forme.** Les facteurs associés aux essais publiés par rapport aux essais non publiés sur la régression logistique univariée étaient du financement du gouvernement et des essais qui étudient les maladies cardiovasculaires observés dans les soins

primaires. Sur l'analyse multivariable, le seul facteur significatif associé à la publication était une condition d'étude cardiovasculaire, par rapport à une condition d'étude de maladie infectieuse (odds ratio [4.4], intervalle de confiance à 95% 0,04-2,9, p = 0,04). Sur les 54% des essais publiés, le délai médian entre la fin de l'étude primaire et la publication était de 25 mois (IQR, 16-34). Les essais comparant les interventions pour les maladies infectieuses et la psychiatrie étaient associés à un délai de publication plus long par rapport aux conditions cardiovasculaires (14,9 mois, p = 0,001 et 11,1 mois, p = 0,019, respectivement). Plus de la moitié (54%) des essais publiés ont signalé un critère d'évaluation principal significatif. ***Conclusions:*** La publication des résultats des essais randomisés CER à temps n'est pas observée. Les médecins de soins primaires s'occupent des patients atteints de maladies cardiovasculaires, et la publication d'informations pouvant éclairer la prise de décisions médicales réelles pour ces patients est prometteuse. Cependant, sans rapport en temps opportun de tous les essais, les preuves potentielles ne sont pas découvertes et, par conséquent, peu ou pas de contribution précieuse n'est apportée aux soins primaires. Les conséquences de la non-publication comprennent le manque de respect pour les patients de soins primaires qui se sont portés volontaires pour ces études, une rupture de la confiance de la société dans l'entreprise de recherche biomédicale et un scepticisme croissant à l'égard des essais comparatifs (CER). L'objectif de mener des études comparatives d'efficacité est compromis lorsque les résultats ne sont pas communiqués.

1 TABLE OF CONTENTS

Abbreviations

ARRA	American Recovery and Reinvestment Act of 2009
CER	Comparative Effectiveness Research
CT.gov	ClinicalTrials.Gov
EBM	Evidence-based Medicine
FDAAA	Food and Drug Administration Amendments Act
NCT	National Clinical Trial identifier
NIH	National Institutes of Health
RCT	Randomized Controlled Trial
RePORTER	Research, Portfolio Online Reporting Tools Expenditures and Results

List of Tables and Figures

Acknowledgements

From the time I started my MSc to now, the world has quite literally changed. I entered as Class of 2019, but the hardships that have led me to graduating as Class of 2020 pale in comparison to what humanity is facing now. I am proud to be studying bioethics at a time when the world is making difficult decisions and hope that I can give back to society through research.

I would like to begin by acknowledging my undergraduate philosophy mentor, Professor Andrew Sepielli, and my statistics professors for encouraging me to pursue many courses in moral philosophy, math and statistics. The rigorous statistical preparation helped me greatly when writing R code and interpreting data analysis for my thesis manuscript. The advice and guidance I received from the Département de médecine de famille has been essential and invaluable to my MSc career. Graduate program coordinators Jamie and Sherrie have kept me on track for my coursework and thesis submission through the years. I was also privileged to have a role model in Dr. Gillian Bartlett, whose passion for research inspired me. During times of uncertainty, Dr. Mark Yaffe was there to support and help me make important choices.

A huge 'thank you' to the students and faculty at the Biomedical Ethics Unit, as well. Dr. Jennifer Fishman warmly welcomed me into the program. My supervisor, Dr. Jonathan Kimmelman, is not only a renown bioethicist who shares my passion for the ethics of clinical research but is also a phenomenal professor. His teachings in research ethics provided a strong foundation for this thesis. Thank you, Jonathan, for introducing me to Comparative Effectiveness Research as well as for all the consultations and writing critiques – it wasn't easy arriving at a research question to begin with! I am grateful to the BMEU for helping me connect with individuals such as Dr. Deborah A. Zarin as well as Dr. Steven Joffe. Both brilliant bioethics professionals whose tokens of wisdom regarding clinicaltrials.gov and comparative effectiveness research, respectively, helped shape my thesis.

My thesis committee has provided great editorial assistance and guidance throughout. Dr. Eugene Bereza provided clinical expertise and Dr. Tibor Schuster was patient enough to lend me his statistical expertise regarding the most appropriate research design for my thesis and reviewed my data analysis. Last but surely not least, I had the privilege and honour of having Dr. Carolyn Ells as not only the professor of my very first graduate bioethics course, but also as a member of my thesis committee. Whether we discussed moral psychology, nineteenth century philosophy, or even the weather, every conversation was just as heartfelt. The time you dedicated to mentoring me was inspiring and enlightening. Your advice has helped me become a better writer and a better ethicist, and it goes without saying, thank you Carolyn for your advice over the years. I hope to return to McGill some day and reconnect with you all.

Finally, there is no amount of gratitude that will do justice to my mother. The lifelong sacrifices my mother made so her children could pursue an education are immeasurable. I will also always be grateful to my co-workers who provided moral support throughout the writing phase of my thesis. Without this role, I would not have been able to support myself. I hope that my thesis makes a valuable contribution to knowledge because that is why we do research – for others. Although I was not awarded any funding for the completion of this MSc or thesis, I am grateful to have been a recipient of the renewable McGill Graduate Excellence Award.

PREFACE

This is a manuscript-based thesis with an additional "methods" chapter that justifies search methodology, describes esoteric terminology and improves the ambiguity that exists in the field of Comparative Effectiveness Research. The chapter outlining methodology was necessary as the data extraction and analysis, upon which this entire thesis is founded, took place independently.

CONTRIBUTION OF AUTHORS

Chapter I – Sara Riaz (SR) wrote the background of this thesis based on feedback from Dr. Carolyn Ells (CE) when CE reviewed the thesis protocol written by SR.

Chapter II – SR constructed the literature review based on literature searches performed independently and a thesis committee discussion held on July 09, 2018.

Chapter III – SR designed the study as lead author, from protocol to final report, with review from Dr. Jonathan Kimmelman (JK). SR performed data collection, data analysis and interpretation of statistical tests independently with assistance from JK for the sensitivity analysis. SR was the lead writer for the manuscript with review from JK and CE for intellectual content. Dr. Eugene Bereza (EB) provided secondary review of the final dataset from a clinical perspective. Dr. Tibor Schuster (TS) reviewed R code and final data analysis. Both authors will provide approval before submission for publication.

Chapter IV – SR was solely responsible for the methods section including identification of implications of data collection along with the study records listed in **Table 2** of this chapter.

Chapter V – SR crafted the arguments and overall summary discussion independently. JK identified some limitations of this thesis. Conclusion was written by SR.

CHAPTER I - The CER trial and research integrity: an introduction

Pragmatic research like the Comparative Effectiveness Research (CER) trial generates valuable evidence to inform medical and policy decisions alike by evaluating which standard of care interventions work better and for whom [1]. It makes sense to have a pool of unbiased and genuine research outcomes to make well-supported clinical decisions from. Primary care providers (PCPs) who are responsible for treating and providing follow-up care to both chronically ill and aging populations can benefit from CER trial outcomes. Results from comparative research have the potential to improve rates of informed and effective prescribing of common therapies [2,3]. However, research in the past has demonstrated that not all trial results are reported, and when they are reported, not all randomized controlled trial (RCT) results are reported honestly [4]. Investigators undertake the responsibility of modestly reporting their research via timely publications. This implies that CER RCTs do have the ability to add “good”, unbiased and timely published results into the healthcare evidence base, correcting inadequacies of former practice and treatment guidelines [5,6]. As the popularity of pragmatic research grows, the need to ensure unbiased outcomes are published when completed grows in parallel.

Biases that detract from meaningful research have commonly been cited in the biomedical literature to result in phenomena such as selective outcome-reporting and inadequate publication of results [7]. Overall, researchers label the withholding of certain study outcomes as “publication bias”, a hallmark finding for the need for research transparency [8]. Publication bias involves consequences greater than studies simply going unpublished. Many have interpreted this bias as the publication of only positive results without reporting negative results, because negative studies are perceived as "failures" [9]. Publishing research outcomes of clinical trials are important for several specific reasons beyond simply being “the right thing to do”. A lack of honest reporting delays timely improvements in medicine as well as promotes the unnecessary

repetition of similar trials. Apart from the potential for redundant studies, resources such as the time of the clinical support team and research subjects as well as the funding invested into research by sponsors such as non-profits, academic medical centers and the industry are at stake [8]. Further, patients are recruited into studies as research subjects based on the signing of an informed consent form (ICF), which explains the risks and benefits of trial participation. The ICF commonly mentions that the involvement of research subjects is integral to enhancing scientific discovery in hopes of helping future patients [10]. Research subjects who participate in trials where outcomes are not published make no considerable scientific contribution and as a result, were likely misled to participate. Lastly, withholding evidence from RCTs poses the risk that medical opinion could potentially be altered for reasons not grounded in science (i.e. personal gain). Thus, nonpublication is unjust to both trial participants and patients outside of these trials who will be exposed to policies that did not account for negative study outcomes due to a lack of transparency in outcomes reporting.

Given that a well-conducted study is important, the integrity of randomized comparative effectiveness research (CER) should be no exception. Being a pragmatic design, CER trials aim to evaluate head-to-head which therapies work compared to other commonly used standard of care therapies [11]. Moreover, there exist treatments and preventive approaches in medical care that have not yet been rigorously tested, which is problematic because trials of new interventions can be tested against control groups of "best-available" therapies to determine efficacy [12]. By testing the standard of care interventions already in use, physicians get a better idea of which standard therapy works better, rather than the knowledge that it has the potential to work in an ideal setting. The reasons to publish CER specifically are even more important than research in general. As described, comparative effectiveness trials study interventions that one could

typically access outside of the trial – such as a routinely used drug or medical device. Many suggest that this makes CER "minimal risk" to patients participating in the trial [13]. As a result, patients are not commonly subjected to ICF requirements to participate and ethics committees often provide these studies with expedited review rather than the thorough reviews that trials studying novel therapeutics receive [13]. Without needing to sign an ICF, patients in a doctor's office could unknowingly be involved in a trial, believing that they are receiving care tailored to their health when instead they are participating in research where the goals of future patients are prioritized. Not publishing CER RCTs may cause distrust in the patient-physician relationship and undermine patient autonomy, as outside of the trial these patients could potentially choose to receive custom care for their needs instead of "care" determined by a trial protocol. For these reasons, among others, the general reasons for withholding research results are amplified for CER trials. Effectiveness trials are aimed at generating evidence to inform policy, and the moral basis for conducting CER is not met when outcomes are not disseminated to the public.

Many CER RCTs test routine interventions by recruiting from primary care patient populations given the pragmatic nature of this research design [14]. This makes it important to ensure PCPs are effective and efficient when prescribing drug and treatment plans to their patients – one of the goals that randomized CER aims to achieve. What is missing from this picture is the need to monitor late phase trials, such as these CER trials, for unethical publication practices that can hold negative implications for primary care. A biased study, as discussed, can pose additional risks to study participants and down-stream information users regardless of the risks involved in participating in the study itself.

General Research Objectives

This thesis serves two main purposes. First, to empirically evaluate the publication patterns of randomized CER trials with potential to improve primary care. Second, to assess the ethical implications associated with non-publication of randomized CER and the important role publication holds within the big picture of research.

CHAPTER II – Defining the CER trial and the need for transparency: a literature review

What is Comparative Effectiveness Research?

How CER should be defined has been debated in the literature [15]. The Institute of Medicine (IOM), the Federal Coordinating Council (FCC), the Agency for Healthcare Research and Quality (AHRQ) as well as other researchers have outlined their own terms for defining comparative effectiveness [16]. The different definitions are compared in **Table 1**. What is agreed upon is that the CER trial, also referred to as the comparative effectiveness trial, is distinct from other "typical" RCTs that come to mind. This is because CER RCTs generate outcomes that are useful in resolving "common decisional dilemmas", rather than outcomes that are needed to assess a new drug's safety or are needed for a treatment to receive FDA approval [17]. Although *clinical* effectiveness studies also serve to evaluate the effectiveness of therapies and preventive approaches, they are not the same as *comparative* effectiveness studies. CER trials directly compare multiple standard of care interventions, or "active" comparators, against each other, which is something not *all* clinical effectiveness studies are designed to do [11]. Further, participants of comparative effectiveness studies are recruited from real world populations, such as primary care, without regard for restrictive inclusion criteria to help reflect routine care [1,11]. However, other studies that do not assess effectiveness not only differ because they enforce stricter criteria for participation but also because they evaluate the study intervention in combination with the standard of care available, typically against a placebo [18]. Thus, regardless of how CER is defined specifically, the overall goal is to conduct a study where outcomes can be generalized to the real world, resulting in informed decisions and improved healthcare**.**

Table 1: Comparing Existing Definitions of Comparative Effectiveness Research

Institutions	**Definition**
Federal Coordinating Council	"the conduct and synthesis of research comparing the benefits and harms of various interventions and strategies for preventing, diagnosing, treating, and monitoring health conditions in **real-world settings**" [19]
Agency for Health Care Research and Quality	"compares the results of one approach for **managing** a disease to the results of other approaches." [16] "head-to-head comparisons of treatment alternatives." [20] "generating new findings through scientific studies of different interventions" [20] "Results summarized in a **systematic review**." [20]
Institute of Medicine	"generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition, or to improve delivery of care." [11] "purpose to assist consumers, clinicians, and policy makers to make informed decisions" [11] "improve health care at **both** individual and population levels" [11]

Although all three widely accepted definitions of CER are similar, they are not the same. The FCC defines comparative effectiveness like the IOM does, focusing on generating and synthesizing research on harms and benefits of interventions whereas the AHRQ stresses the importance of disease management. Overall, CER helps establish standards for reliable practice guidelines that the medical community can trust. However, the FCC limits CER to studies conducted exclusively in real world settings. Although most CER is conducted in the community by non-expert health professionals to increase generalizability of evidence to routine clinical practice, a trial outcome that informs policy and clinical decisions need not have been conducted in a *representative* real-world setting [11]. For example, trials that inform decisions in a primary care setting could have been conducted by a specialist, although ideally these trials are conducted in the settings in which they are used. The IOM presents a broader definition of CER that includes research that will, for instance, "prevent, diagnose, treat and monitor" a disease through comparison of interventions regardless of setting [11]. IOM's definition is also the only one

targeting a research audience – personalizing individual care with specific evidence generation and/or improving overall care approaches for a certain population [11].

CER is a pragmatic research design based on a continuum of settings that may not always be purely indicative of the real world but may still be generating evidence to inform clinical decision making [21]. Moreover, the IOM supports the US congressional initiative for high priority CER topics, describing comparative effectiveness as the pursuit of understanding "what works in healthcare" [11]. Both the FCC and the IOM suggest that CER outcomes will eventually contribute to higher standards of evidence being synthesized. As mentioned, the IOM's definition of CER places emphasis on identifying areas in medicine that are a national priority. However, this does not necessarily hold true. Authors Bourgeois et al found only a minority of research addressing the IOM's high priority topics were studied through a CER design [22]. Lastly, AHRQ's representation of CER places less emphasis on benefits and harms and more emphasis on managing disease and/or treatment approaches. Thus, compared to the FCC and IOM, AHRQ likely values preventive care less than the other patient-oriented initiatives and stresses active disease management. The AHRQ also views CER differently because where FCC and IOM mention synthesis, AHRQ specifies systematic reviews summarizing outcomes to be an important component of comparative effectiveness [20].

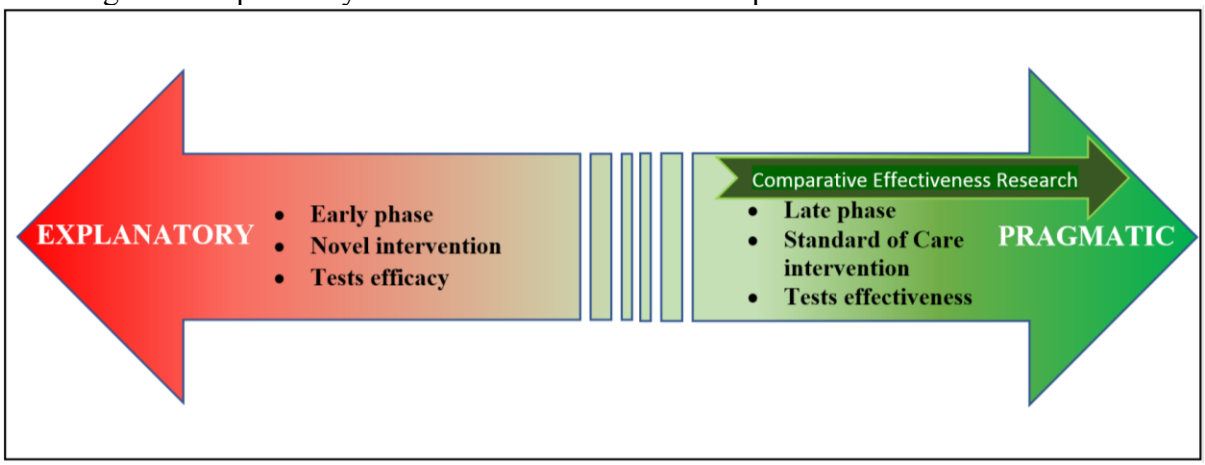
Other independent authors have also provided their own definitions of what comparative effectiveness studies are. For instance, the AHRQ explicitly states CER involves trials studying **different** interventions, but some argue that comparing the same interventions between different groups of patients as well as different clinical environments also constitutes CER [16,20]. Moreover, the same authors also argue that a single intervention could commonly be prescribed in different doses, and an RCT comparing varying doses of the same drug could also constitute

CER because of the greater aim of generating real-world evidence [16]. This would then imply that regardless of the specific definitions, any comparative study randomizing heterogenous, less fastidious study populations could possibly be CER since the study population represents the real world. However, this leads to the conversation of how different types of studies are designed and the complexities involved.

Not so clear-cut: The Pragmatic-Explanatory Continuum

The pragmatic-explanatory continuum that Thorpe et al. describe as the "PRECIS tool" evaluates the variation that exists between the explanatory trial and the pragmatic trial, illustrating that not all trials fall strictly under either type [21]. **Figure 1** below demonstrates where on the Continuum, modified from Thorpe et al, studies that represent CER could take place. Explanatory trials evaluate a hypothesis to understand whether an intervention can work in an "ideal" environment, which is a highly stylized setting with narrow inclusion criteria for trial enrolment [23]. Early phase studies that are designed to test experimental therapies for safety and efficacy would count as explanatory. Pragmatic trials typically study primary endpoints concerning patient-centered outcomes and rigorously evaluate interventions used in the "usual" setting of routine care [24,25]. Late phase effectiveness studies are examples of pragmatic research since logically, only the interventions approved for consumer use based on early phase study results can be used in routine clinical practice [21]. Pragmatic research, like CER, values the non-compliance involved in routine care, understanding that patients will not always follow their doctor's treatment and/or prevention plans. However, unlike explanatory trials, research with pragmatic qualities would not adjust study results for non-compliance, as this is what reflects care in the real world [23].

Figure 1:
The Pragmatic-Explanatory Continuum Modified for Comparative Effectiveness Research



CER seeks to find an appropriate balance between pragmatic and explanatory [26]. Not all pragmatic trials are CER trials but all randomized CER trials are pragmatic [24]. Finding the right balance between broad generalizability of results and robust scientific quality when a fastidious study is not being conducted is complex. Many confounders are present in CER trials, as they are the result of measures taken to increase external validity of research outcomes to generate real world evidence [1,23]. Patients enrolled in CER RCTS often have co-morbidities and varying health profiles compared to the backgrounds of participants of early phase studies. Moreover, many factors that may influence the results of a trial are not controlled for by CER investigators to help prevent the "ideal" settings of preapproval research [23]. No matter how pragmatic the CER trial design may be, explanatory components are bound to be present. Given the difficulties designing a methodologically robust CER study that is also representative of usual care, it's important to ensure these studies are not biased. As AHRQ's Effective Health Care Program mentions, "unbiased information" is what drives the head-to-head assessment of comparative effect for alternative treatments [20].

Primary Care and CER

The World Health Organization (WHO) defines primary care as "first-contact, accessible, continued comprehensive and coordinated care" provided by family physicians, general practitioners, pediatricians and/or other generalist health professionals [27]. Since patients present themselves to primary care practices with common conditions and similar underlying symptoms that physicians see and treat routinely, a pragmatic research design is important to answer field-specific questions [14].

As discussed, CER trials are a type of pragmatic research design but the distinction becomes increasingly complicated when studying interventions beyond the primary level of healthcare. One could even question whether pragmatic trials are worth conducting at tertiary level of care such as for surgical interventions, given between 2008 and 2009, one in five surgical RCTs were discontinued early [28]. For instance, let's look at a surgical trial in which two commonly used, complex surgery techniques are compared. The technique in question is most commonly used on patients in high-volume medical centers. This observation creates a conflict in the design of the pragmatic trial since performing such a technique under study protocol in an intense, high-volume environment may not be feasible. Moreover, investigators conducting highly technical and specialty pragmatic trials must now consider whether the trial should be conducted in the health care environment that is most feasible or in the health care environment that is *likely* to exist in the relevant specialist area [29,30]. Thus, primary care is the field of medicine where improving health outcomes for patients through pragmatic comparative effectiveness trials proves to be more valuable than other fields.

Smelt et al state, "many trials in primary care require a pragmatic design" [14]. These authors also argue that although research outcomes are useful, investigators conducting

pragmatic trials in primary care should be aware that the standard of care control arm that should reflect routine care may not always be representative of usual care in the real world. These authors describe the risk of "study-induced behavioural change", meaning differences may exist in the way patients within the trial are treated compared to the way routine patients are treated outside of the trial [14]. Although this risk may be present, primary care medicine remains a good fit for the CER trial design. Ford and Norrie confirm this observation stating that "by virtue" certain trials are designed to be "more" pragmatic [31]. An intervention that is relatively less expensive and subjected patients to lower levels of risk would be more representative of usual care than a trial studying a complex intervention where both costs and risks are high [31]. For example, primary care trials can be far more easily conducted by randomizing entire clinics to two different standard of care therapies in a cluster RCT, resulting in comparative outcomes that can help improve usual care [32]. Thus, primary care medicine can be thought of as an ideal reservoir for conducting pragmatic CER trials.

Importance of CER

Active-comparator trials are more clinically valuable in terms of improving healthcare. Pre-license studies performed to obtain approval of the FDA and/or other competent authorities are not robust enough to generate evidence that is applicable to the many subgroups of patients that are present in routine care [25]. Pre-approval studies are conducted using small study sample sizes, making it easy to miss rare adverse events and the endpoints chosen are unlikely to address long-term outcomes [11]. Moreover, strategic selection of comparators and/or use of a placebo in pre-license research also results in outcomes unrepresentative of clinical practice [11]. Further, Naci et al argue that founding the regulatory approval of drugs and devices on placebo-controlled trials discourages researchers from participating in clinically relevant late phase comparative

effectiveness trials later on in the intervention's lifespan [33]. These authors find this concerning because when a drug is thought to be efficacious, capable of providing benefit for a condition where approval was granted based on pre-license studies, post-market comparative studies of these interventions are unlikely to be conducted [33].

Although comparative effectiveness trials are critical, conducting early phase studies to assess safety and efficacy are crucial, as well. Testing a new and experimental therapy in a controlled setting helps evaluate whether the intervention is effective against a placebo, standard of care or no therapy [11]. For instance, the Ischemic Optic Neuropathy Decompression Trial found that the surgical technique they were studying was not beneficial and instead may actually have caused harm when compared to no therapy [34]. This highlights how important pre-approval research is to establish if an intervention is effective against no treatment and/or placebo. However, the use of suboptimal comparators during pre-licence research has been documented, leaving important pre-approval studies unjustified. For instance, authors found suboptimal standard of care comparators have been used when seeking initial FDA approval of anticancer drugs [35]. Around 17% of approved anticancer drugs used suboptimal control arms to compare the experimental drug with, leaving clinical data for these drugs up to cautious interpretation [35]. If a drug or device is approved without honest and appropriate demonstration of superiority to the best-available standard of care, physicians could be recommending better interventions in routine care already [12]. As Naci et al explain earlier, following the approval of an experimental agent, there is little push to monitor its effectiveness and eventually certain newly approved drugs become embedded into routine use, leaving clinicians to use their judgement when treating patients [33]. If an agent was FDA approved based on a trial that subjectively chose a comparator to speed up the drug's delivery to the consumer or to increase

options for drugs available, then there remains the need for robust comparative effectiveness trials to assess which standard of care therapies work and for whom.

Last, we note a criticism in the effectiveness of comparative effectiveness trials when it comes to observable effect sizes. Robert Temple, Deputy Center Director for Clinical Science at the FDA, describes how difficult observing a statistically significant difference is between a therapy and no therapy (or a placebo) even if such a difference may exist [36]. Further, observing this difference between two or more therapies that are already considered effective, as in the case of CER active-comparator RCTs, is far more difficult [36]. A remedy to preserving the value of CER trials would be to conduct studies of large sample sizes [37].

What are transparent reporting practices and why are they important?

William Shakespeare famously wrote in one of his tragedies, "*If I lose mine honor, I lose myself"* [38]. Honesty is important in all walks of life, as an academic or as a scientist, we all hold responsibilities to convey our findings modestly by neither over nor underreporting. Author John Ioannidis argues at length in his paper, "Why most clinical research is not useful", that biased, and inadequately reported results leave important research of little value [39]. As discussed in Chapter I, biases are not new to biomedical research and the discoveries of dishonest research practices have tainted society's trust in research at large. Attempts to enhance transparent reporting include development of clinical trial registries and stricter regulations for publishing in peer-reviewed journals. The database known as CT.gov was created so that investigators could prospectively register their trials and the International Committee for Medical Journal Editors (ICMJE) issued guidelines requiring trial registration prior to patient enrolment as a requisite to publication, a recommendation adopted by numerous journals [40]. In 2007, Section 801 of the Food and Drug Administration Amendments Act (FDAAA 801) regulation

was established to issue "civil monetary penalties" to investigators who failed to register their trials and those who did not comply with the mandated results reporting requirements [41]. Unfortunately, research is not always transparent despite the multiple safeguards in place.

Bias in clinical research can be described as faults, intentional or not, that "encourage one outcome over others" [42]. If an RCT is conducted and biased outcomes are reported, medical experts, policymakers and the public in general will be misinformed about the benefits, risks, safety profile, effectiveness and/or efficacy of the intervention in question [43]. This a disservice not only to those who devoted their time and physical involvement as research subjects but to the down-stream subpopulations who will eventually be subjected to policies and reform based upon biased trial outcomes [43]. Despite the importance of honest and ethically-sound research, fraud and misconduct exists. A systematic review found that medical researchers frequently testified to inappropriate reporting and publication practices, including falsifying data and altering genuine outcomes. Moreover, over 33% of scientists revealed "questionable research practices" while approximately 2% admitted to distorting their data for various reasons – an unacceptable breach of scientific conduct under any terms [44]. Prior to the enactment of FDAAA 801, Boutron et al analysed a group of 72 RCTs published in 2006 that reported nonsignificant primary outcomes and found results to be commonly distorted, also known as "spin" [45]. The spin of results may influence readers to misinterpret facts and misinform the actions taken based on the distorted evidence [46].

Furthermore, there are different types of misconduct that decrease the transparency of research beyond simply inappropriate reporting of results including, but not limited to, conflicts of interest and nonpublication. Given that one in every four clinical investigators have industry ties, Bekelman and colleagues observed in a comprehensive systematic review that industry-

sponsorship was significantly associated with study outcomes favourable to the industry [47]. However, this finding appears to vary given different fields of medicine. For instance, critical and intensive care RCTs are not commonly industry-sponsored but when they were funded by the industry, no significant association with pro-industry outcomes could be seen [48]. Other authors have found that research outcomes are unchanged when holding financial ties with the company sponsoring the trial, but the interpretation of those outcomes are noticeably pro-industry [49]. Despite ICJME registration requirements for RCTs to publish in member journals, not all investigators register their studies, contributing to a lack of transparency of ongoing trials and possible redundancy of trials already conducted when they are not registered prior to patient enrolment. Authors Milette et al found that in 63 RCTs published in top psychology and behavioural health journals between 2008 and 2009, 60% of reports improperly defined primary outcomes and only 13% were prospectively registered in a publicly available database [50]. To further propagate withholding of results beyond nonpublication, restrictive data-sharing policies limit widespread access to data from clinical trials that could be used for secondary analyses to improve health outcomes [51].

Underreporting and withholding of RCT results is a frequent contributor of bias in meta-analyses as well as a disregard for the ethically-sound research practices outlined in policies such as the Declaration of Helsinki [52]. In specific, when negative results are not reported as often as positive results, this leads to skewed information resulting from publication bias [8]. Non-publication is a relevant concern across multiple disciplines of medicine. A cross-sectional analysis of RCTs studying rare diseases that were registered on CT.gov between 2010 and 2012 found 66.5% remain unpublished within two years of completion and slightly above 30% remain unpublished within four years [53]. An analysis of large RCTs registered and completed before

early 2009 similarly observed a nonpublication rate of 29% and found 78% of unpublished studies did not report results on CT.gov either [54]. Digital health RCTs completed between 2010 and 2013 presented almost the same finding, resulting in 27% of trials remaining unpublished even 5 years following study completion [55]. A rigorous evaluation of nonpublication of pediatric RCTs also uncovered the same outcome; 30% of completed trials were not published in any journal [56]. Even in an analysis of RCTs for ophthalmology interventions completed as far back as 1972, a surprising 81% of trials remained unpublished [57]. Thus, across disciplines and different time periods, nonpublication is still a threat to research integrity and transparency.

Lastly, research design may also be a contributor of bias. The non-inferiority RCT is by virtue more vulnerable to misinterpretation and biased findings. This is because non-inferiority designs depend greatly on margins pre-determined by investigators and without legitimate reasoning, an erroneous margin of noninferiority could leave outcomes biased toward an outcome favoured by investigators [58]. Gyawali et al conducted an unrestricted PubMed search of non-inferiority RCTs of cancer drugs where 39% of RCTs used the non-inferiority design without proper justification and industry-sponsorship was significantly associated with unjustified use [59]. Inappropriate and fraudulent research practices hold consequences for both the overall research enterprise and the participants who volunteer in these trials.

Why are transparent reporting practices important for CER?

It is acknowledged that if the scientific gains from a trial are deceptively reported, distorted or unpublished, then the research-related risks posed to subjects will go unredeemed because of the biased trial's lack of genuine knowledge transfer to society [60]. CER trials are often classified as minimal risk because they offer standard therapies that PCPs can also make available to patients outside of a trial [13]. This "minimal risk" reputation is what establishes the

foundation for the more lenient regulations for CER trials, including expedited and/or delegated review, waivers of informed consent as well as less vigilant data monitoring [13]. What is not as obvious with CER trials is that if the evidence somehow becomes biased, then that evidence is unfit to inform future policy/clinical decisions. Consequently, if the social gains are not present in the form of publication, then the trial is not justified regardless of how little physical risk trial participants are subjected to. Moreover, if CER trials do not ask for informed consent, then patients participating in biased trials will have their autonomy compromised for having participated in a trial that resulted in neither a significant social nor scientific gain [61]. Many patients enroll in trials, despite the varying levels of risk involved, so that they can play their part in advancing knowledge [10]. If this genuine medical knowledge is not the result of a trial and/or not honestly disseminated, the moral basis for randomizing subjects is not met, and the integrity of the research enterprise is put at stake.

Additionally, since CER trials take place in pragmatic settings such as doctor's offices. If they do not require the explicit informed consent from research subjects, then there is a high likelihood that subjects will mistake their participation in research for care [61]. Research subjects can be rendered ignorant of the part they are playing in research without an ICF available to shake them out of their complacency. Although the interventions offered through the trial will be standard, subjects will receive protocolized-care. In a clinical research setting, this "protocolized-care" will be the treatment delivered to the subject as part of a rigorous and methodological routine, dependent on the trial arm to which the subject is assigned [62]. Since these subjects are part of a trial that needs to preserve internal validity, PCPs cannot deliver optimal care adjusted to the subject's needs [26]. This is so that investigators can meet the goal of a CER trial – generating scientifically robust evidence to support the effectiveness of a

commonly used intervention [63]. Attempts to enhance evidence-based approaches in treating and prescribing at both the physician-level and the policy-level is highly valued in primary care, but this evidence needs to be ethically generated. If CER trials are impacted by outcome-reporting and publication biases, then the participation of patients in CER trials may be unjustified for several reasons including: withholding patient autonomy, an unclear research-or-care distinction and inadequate scientific/social gains to redeem risks of randomization. Overall, EBM is only valuable if it aggregates the most reliable and trustworthy sources of research results and outcomes to inform medical decision-making [64].

Authors Naci et al highlight that post-marketing studies and post-approval studies like CER are often designed like pre-approval studies, undermining the purpose of completing them [65]. If a CER study uses inadequate clinical endpoints and/or lacks proper active comparators, then they are ill-suited to inform evidence-based decisions. The same authors also argue that industry-research partnerships in CER hold special obligations to advance EBM, recognizing the conflicts of interest that exist and efforts should be made to prevent undue influences from biasing research [66]. Dunn and Coiera even went as far as debating whether industry funded CER should be included as a source of genuine evidence in medical decision-making and policy reform without improved transparency in CER [67]. At last, it is worth exploring how frequently pragmatic trials like CER RCTs are sponsored by industry members, and according to Buesching and colleagues, the answer is rarely. Even when industry did fund CER RCTs, no significant differences were reported between generic drugs and the sponsor's drugs [68], illustrating that conflict of interest may not play a huge role in biasing CER trials.

Possible reasons reporting practices are not scrutinized for CER

CER tests therapies commonly used by physicians, which are typically already thought of as more or less equally effective by the expert medical community [69]. This makes the non-

inferiority design a popular choice for CER RCTs because they are in a unique position to compare the other attractive factors associated with standard therapies [69]. These factors include convenience, cost and tolerability, as long as one standard therapy is "not worse than" the other [70]. However, the non-inferiority trial poses a greater risk of producing biased results, as discussed earlier, if careful attention is not paid to conducting the trial and analyzing data [58]. Lack of participant adherence to a trial intervention in a superiority trial handicaps the study by diminishing its power – decreasing the study's ability to detect a difference if it exists [71]. This is not the case for a non-inferiority trial. Poor adherence will instead bias results in the direction of the non-inferiority outcome because both arms of the trial will appear more alike [71].

However, a closer look at what causes bias in CER trials that are *specific to primary care* suggests that this may not be the case [32]. Although non-adherence raises issues in the analysis of a non-inferiority trial design, it may not be a significant issue for a pragmatic trial. Non-adherence, if it is not limited and adjusted for, adds significant bias to a trial's results because it may be representative of the patients' response to the intervention being tested [71]. An experimental drug can only be offered to patients in a clinically controlled setting, such as an early phase study, because its safety, efficacy and adverse events data are in the process of being evaluated. A common therapy, such as Aspirin, is a product already on the market with extensive safety and clinical benefit data available for primary care patients. When these patients enrol in a CER trial, investigators will have a better picture of how the subjects will respond to the common therapy as opposed to the experimental therapy. This means bias from non-adherence to the therapy tested is not common for primary care patients in CER trials because research subjects are more likely to comply with these standard therapies than experimental ones [32, 72]. Major reasons for non-adherence would logically include adverse events and difficulties keeping

up with a new drug regimen, eventually leading to subjects in the novel therapy arm to be lost to follow up. The subjects in the control arm will most likely will not experience this, leading to bias favouring the novel drug. Even if the same number of subjects are lost to follow up in each trial arm, bias still exists because prognoses will likely differ between groups – the new drug may provide a better or worse outcome than the drug in the control arm [73]. The primary care patients enrolled in CER trials will most likely not be subjected to such biases because their prognoses and outcomes should be similar in both trial arms because both trial arms are a comparison of active controls [32] – not experimental therapies. Lastly, CER trial participants can remain in the trial even if they need a "rescue treatment" [74], since the trial is aimed at generating real world evidence and such additional treatments may be needed during routine care. All these instances describe why bias from research design may not be a concern for CER and why these RCTs are not under the scrutiny explanatory trials are.

Although trial registration and results submission to Clinicaltrials.gov is mandated for certain controlled clinical studies, investigators are not always compliant, resulting in delay as well as non-publication of results as described above [53-57]. In fact, a study done in 2011 found that non-industry sponsored trials were *less* compliant than industry-sponsored trials when submitting results to CT.gov [75]. A more recent analysis of CT.gov found that regardless of trial sponsor, late phase studies exhibited higher data submission rates [76]. These studies suggest that the structure and function of data monitoring and analysis is relatively more manageable for late phase studies. CER trial results are more likely to be available on public domain sooner and less likely to report distorted or biased outcomes. Also, if industry-sponsored studies are more compliant when it comes to results submission and higher data submission rates are limited to

late phase studies, then we are unlikely to observe pro-industry outcomes or other resulting unethical publication patterns for primary care CER trials.

To conclude, explanatory trials may pose more "risks" to subjects because of how risk and harm to trial participants are defined. Pragmatic trials in primary care pose comparably less risk [14, 32]. However, there are other risks present in pragmatic CER. For instance, if a CER trial is not published then patient autonomy is threatened, the risk of complacency exists and the line separating treatment from research is blurred without cause. Also, the trust instilled in users of medical research and policymakers counting on these outcomes to inform evidence for prescription guidelines and care of many downstream users including physicians, patients and drug policies are put at stake. Given the current landscape, it becomes vital to assess publication of CER RCTs to ensure safety of the primary care patients participating in these trials. How well registered trials abide by regulations is important because many trials following FDAAA 801 should be representative of investigators with good research practices. Overall, the CER trial being conducted transparently and in line with the ethical obligations of investigators is far more crucial than other types of research as this research is essentially conducted to inform policy and medical decisions. If the evidence generated is not reliable, or worse, not published at all – this renders the randomizing patients in the trial morally impermissible.

Specific Research Objectives

The current state of knowledge regarding the publication status of randomized CER trials is relatively unknown. Since CER compares the harms and benefits of two or more unique standard of care interventions tested head-to-head, patient care at the level of primary healthcare can benefit [31,32,77]. Thus, pragmatic studies play an important role in facilitating clinical decisions between seemingly analogous treatment choices by broadening the existing evidence base with research outcomes applicable to real-world practice [78].

The following chapters will thus address the general research objectives formed at the conclusion of Chapter I. The first objective is an empirical analysis of the CT.gov database to assess the proportion of completed randomized CER trials (with relevance to primary care) published within four years as well as identify the time lag following primary study completion until publication for published studies. The first objective will also evaluate factors associated with publication and time lag of CER trials included in the analysis. Following the CT.gov analysis, the second objective of the thesis will be met during the discussion, conceptually exploring the question: what are the ethical implications associated with non-publication of randomized CER trials within the context of primary care and medicine overall?

CHAPTER III – Publication rates of randomized comparative effectiveness trials relevant to primary care: A ClinicalTrials.Gov analysis

Preface to manuscript

The following study is an analysis of the CT.gov database, an approach commonly used in the biomedical literature to assess the present status of publication rates for completed studies [79]. In an attempt to increase transparency of trial reporting for both research subjects and the broader research community, FDAAA 801 was established [41]. By doing so, certain clinical trials became subject to prospective registration as well as mandated results reporting onto the public CT.gov database, increasing accessibility of this information. However, not all types of CER trials fall under the registration requirement and thus, some CER trials on CT.gov have been registered voluntarily [80]. Nonetheless, research integrity, including timely reporting of results, is an ethical obligation of all researchers regardless of whether they are mandated to by government regulations.

Our study was conducted to examine the rate of publication of completed CER trials relevant to primary care, as the publicly available registry records on CT.gov provide us with a time stamp of when a trial was completed as well as other information to help locate if a matching publication exists in the biomedical literature. These results add to the currently limited data available on the publication status of randomized comparative effectiveness trials. This manuscript will be submitted for publication in the peer-reviewed journal, *PLOS ONE* due to its outreach to active members of the medicine and health science community, who may benefit from the findings of our study.

Contribution of authors

Authors in order of publication

Sara Riaz (SR)
Jonathan Kimmelman (JK)

SR designed the study, from concept to protocol. SR performed data collection, data analysis and interpretation of statistical tests with assistance from JK for the sensitivity analysis. SR was the lead writer for the manuscript with review from JK for important intellectual content. Both authors will provide final approval before submission for publication.

Funding: This research was not funded.

Conflicts of Interest: The authors have no relevant disclosures.

Publication rates of randomized comparative effectiveness research trials relevant to primary care: A ClinicalTrials.Gov analysis

Brief Title

Publication of Primary Care CER Trials

Authors in Order of Publication

Sara Riaz (SR), ***Corresponding Author***

Department of Family Medicine

McGill University

5858, chemin de la Côte-des-Neiges, Suite 300

Montréal, QC H3S 1Z1

Sara.riaz@mail.mcgill.ca

Jonathan Kimmelman (JK), PhD

Biomedical Ethics Unit

McGill University

3647 Peel St.

Montreal QC Canada H3A 1X1

Jonathan.kimmelman@mcgill.ca

Abstract

Introduction Little is known about the publication patterns of randomized comparative effectiveness research (CER) trials, which play an important role in improving evidence-based medicine approaches and use of standard of care interventions. **Methods** Completed primary care CER trials conducted between September 27, 2007 and August 01, 2015 were extracted from Clinicaltrials.gov and searched for matching publications. Logistic and linear regression were used to explore factors associated with publication and time lag from primary study completion to publication. **Results** Out of 122 trials, 56 (46%) had no full-text publication within 4 years and 39 (32%) were not published in any form. Cardiovascular study conditions were significantly associated with publication (odds ratio 4.4, 95% confidence interval 0.04-2.9, p=0.04). Median time to publication was 25 months (IQR, 16-34). **Conclusion** Timely publishing is unsatisfactory, suggesting nonpublication of CER trials impacting primary care exists and should be stringently regulated as they generate evidence that can improve clinical decision-making and health outcomes.

Keywords: ClinicalTrials.Gov, Publication, Comparative Effectiveness Research, Primary Care, Randomized Controlled Trial, Real World Evidence

Introduction

As defined by the Institute of Medicine (IOM), the primary goal of Comparative Effectiveness Research (CER) is to broaden the existing evidence base by asking the question: "does this work?" instead of "can this work?" [1]. Many interventions are brought into routine practice based on findings from research conducted to obtain FDA and/or other regulatory approval. Once an intervention is approved for consumer use, there is limited evidence available regarding its use in routine care where varying populations of patients often present with co-morbidities, unlike the pre-approval trial participants who are subject to strict inclusion criteria [2]. The settings in which the approved interventions are typically used are also different from the pre-determined settings in which the pre-license evidence was generated. This makes it likely that the health outcomes observed in pre-market research differ from the ones observed in real-world settings.

The IOM defines CER as a strategy to conduct research with generalizable outcomes, comparing the risks and benefits of alternative interventions to either "prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care," [1]. The CER trial is a head-to-head comparison of interventions, including but not limited to prevention strategies, treatments, therapies and diagnostic tools that are commonly used to generate evidence that reflects the real world. Doing so enables the medical community, as well as downstream users of the evidence generated, to make informed decisions for individual patients along with broader patient populations. Adequately reported randomized CER trial results can be used to establish evidence-based guidelines for medical practice, since the information generated will be applicable to the real-world settings in which the interventions tested will be used [2]. Complete and timely reporting of CER trials remains significant to primary care given that improvements

in clinical decision-making paves the way for improvements in patient outcomes as well as a reduction in overall healthcare expenditure [3].

Timely publication is an ethical obligation of all researchers, yet not all research outcomes are adequately reported. Without publishing results from studies, future meta-analyses may incur biases [4], unnecessary time and resources are likely spent conducting studies already performed and human subjects who, often altruistically, participate in research despite risks have not contributed to genuinely advancing science as many are led to believe. Examples of publication bias – the lack of publishing negative research outcomes – has been discussed in the literature for decades, maintained by the concern of dismal publication rates for clinical research [5,6]. Although registries such as Clinicaltrials.gov (CT.gov) as well as policies such as that of the International Committee of Medical Journal Editors (ICJME) may have increased registration rates of trials, the rates of publication are still found to be unsatisfactory [7]. A recent CT.gov analysis of registered lung cancer studies found that 1 in 4 clinical trials remain unpublished, continuing to raise the concern of publication bias [8]. Another CT.gov analysis found that 29.5% of registered studies did not publicly disclose results within four years of completion [9]. A similar study showed that 48% of registered studies did not publish findings within two years of study completion [10]. Even alongside the addition of CT.gov, half of studies that do report results never publish [11]. Publication of clinical research outcomes remains inadequate.

Data regarding the publication rate of CER trials are lacking. The focus throughout the biomedical literature has often been on the underreporting of registered Randomized Controlled Trials (RCTs) overall [12-17]. The proportion of all registered randomized CER trials that are not published remains unknown. The primary objective of our study was to estimate the proportion of unpublished randomized CER trials addressing primary care questions, and to

identify factors associated with publication. Secondary objectives were to assess factors associated with time lag from study primary endpoint completion to full-text publication.

Methods

All interventional randomized CER trials completed between September 27, 2007 and August 1, 2015 were reviewed on CT.gov. Given that all analyses were based on trials registered in the publicly available CT.gov database, this study was exempted review from an ethics committee and/or competent authority. If the trial started before September 27, 2007 but was still ongoing as of December 2007, then it was included, as per FDAAA 801 [18], a regulation that subjected certain clinical trials to prospective registration as well as mandated results reporting onto the public CT.gov database. Manual abstraction took place in three steps: identifying the CT.gov registration records that matched the inclusion criteria in **Table 1**, assessing relevance of chosen CT.gov records to primary care and lastly, identifying matching publications. Two additional post-hoc reviews also took place. The first was a sensitivity analysis extending years to follow-up to identify if additional publications could be found and second, an additional review of the final dataset by a physician trained in family medicine who confirmed accuracy of data as well as identified relevance to primary care.

Search strategy for data extraction was as follows:

Clinicaltrials.gov Review

Search: "comparative effectiveness" | ***Recruitment****:* Completed | ***Eligibility Criteria:*** *Sex:* All | ***Study Type:*** Interventional (Clinical Trial) | ***Study Results:*** All | ***Study Phase:*** Phase 3, Phase 4, Not Applicable | ***Funder Type:*** [NIH, Other U.S. Federal Agency] OR [Industry] OR [All other (individuals, universities, organizations)]

The randomized CER trials matching **Table 1** characteristics, including the study start and primary completion dates discussed, were extracted from the CT.gov search above and

transferred to an Excel spreadsheet where they were assessed for their relevance to primary care before being searched for matching publication.

Table 1: Comparative Effectiveness Trial inclusion criteria for the first phase of ClinicalTrials.gov data extraction

Trial Inclusion Criteria	Description, if applicable
Start Date	After September 27, 2007* inclusive If started before this date* but was still ongoing as of December 2007, trial was included
End Date	Before August 1, 2015 inclusive
Status	Completed
Type	Prospective, Interventional Randomized Controlled Trial
Phase	Phase III, Phase IV, N/A
Study Design	Comparative Effectiveness Research: -two or more **different** standard of care interventions tested -head-to-head comparison -purpose to generate evidence to inform clinical decision-making

Study characteristics extracted from the CT.gov registry and onto the Excel spreadsheet included: Title of study, NCT (National Clinical Trial) identifier, location, start date, primary completion date, purpose, primary care category, study condition, type of intervention, funding, phase, sample size, blinding, trial type (if known), trial design, if significant primary outcome was reported and date of matching publication. For purposes of this study, source of funding was identified based on sponsorship listing in CT.gov. If the financial disclosure on a matching publication revealed that an entity different from the one identified through the CT.gov filter financially supported the randomized CER trial, then that funder type was recorded for the final analysis instead. If an industry funder was involved in collaboration with other entities (e.g. non-profit, academic), the registration record was extracted as an industry-funded study.

After all registration records meeting inclusion criteria were identified, we searched for the matching publication of each trial in the final study cohort. This process took place in three phases, following an approach used by Chen et al. [19]. First, the NCT identifier for each

registration record from the final study cohort was searched for on CT.gov to determine if any publication(s) were provided under the record's "More Information, Publication of Results" section. Each potential publication was assessed by lead author for information matching the original registration record including the NCT identifier, study start and primary completion dates, name(s) of investigators, location and/or study site(s), study sample size, study title, primary outcome measures specified, and descriptions of the interventions tested. Second, if a matching publication could not be identified through the CT.gov "publication of results" section, then PubMed and Embase were searched using the NCT identifier as well as the terms "clinical trial" AND "[study title]" AND "[intervention name]" in the "article title, abstract, keywords" field [19]. For NIH-funded studies, the corresponding Research Portfolio Online Reporting Tools (RePORT) project information "Query" section was searched using fields "[ClinicalTrials.gov ID]" AND "[Principal Investigator (PI)]" AND "[Text Search: Project Title]" [20]. Lastly, an additional Google and Google Scholar search was performed [21] using the same identifying information from the CT.gov study registry record used in the earlier steps if a matching publication still could not be retrieved. Principal investigator(s) were not contacted to confirm matching publication of their study's registration record.

Studies Reviewed

Studies completed after August 1, 2015 were excluded to allow for a minimum of 4-year follow-up-to-publication time period to avoid providing an unfair advantage to the earlier CT.gov registration records in our study cohort, as data extraction took place August 1, 2019. All studies included in the analysis belonged to the category of the randomized CER trial as defined by the IOM earlier [22]: "evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition or to improve the delivery of care"

[1]. To be included in analysis, the studies must have directly compared two or more *different*, existing healthcare interventions as well as have been either conducted by primary care physicians *or* have asked a research question relevant to improving evidence-based primary care. A study was considered to be conducted by a primary care provider if the principal investigator's institutional affiliation was representative of a community health or medicine, primary care and/or family medicine department. In many cases, investigators hold broader affiliations to departments other than primary care but conduct research that, according to region-specific biomedical literature, address important routine interventions used in primary care. For the purposes of this study, the IOMs definition of CER is used instead of the US Federal Coordinating Council's, allowing us to label research as comparative effectiveness without it necessarily being conducted in "real world" primary care settings [1]. If the results from CER trials were deemed to be of importance to primary care, then the CT.gov registration record was included in final analysis. Thus, interventions tested on and/or for hospitalized patients, surgical interventions, specialty care interventions such as chemotherapy, ophthalmic surgery, anesthesia as well as radiological interventions were excluded from the final data set for not being of significant relevance to primary care and/or primary care providers. Post-approval comparative studies submitted to the Food and Drug Administration (FDA) as part of post-marketing commitments were also excluded for not aligning with IOM's definition of CER. A clinician-ethicist trained in Family Medicine reviewed the final set of included studies to ensure relevance to primary care and CER.

Originally, only prospective interventional trials testing a biological product, drug and/or medical device for a primary care condition were included in the search because this is the criteria for an "Applicable Clinical Trial" to which FDAAA 801 applies [18]. However, due to

the scarcity of primary care CER trials registered on CT.gov that are testing biologics and/or drug/device products, the inclusion criteria were expanded to include primary care randomized prospective CER trials testing behavioural and other interventions.

Main outcome measures

This study defined abstract-only reporting and lack of full-text publication within 4 years of primary completion date on CT.gov as non-publication of results. Abstract-only publication was considered "non-publication" given that a minority of abstracts result in appropriate publication as full-text articles [23]. Full-text publication after 4 years of primary completion was also considered to be non-publication of results. Full-text publication in a peer-reviewed journal within 4 years (inclusive) of the primary completion date stated on the study's registry record was defined as adequate publication of results. The earliest electronic publication date was used as date of publication when a matching publication was located. If the electronic date of publication could not be determined, then date of the next available print publication was noted. The time elapsed between the primary endpoint completion date stated on the study's registration record and the earliest date of full-text publication identified was referred to in this paper as the "time lag" for published studies.

Statistical Analysis

A binary logistic regression model was applied to examine the trial characteristics associated with publication versus non-publication of results for completed studies. This categorical binary variable was used to determine 95% Confidence Intervals (CI) and Odds Ratio (OR) estimates for other trial characteristics of interest. First, a univariable logistic regression model was used to assess which independent predictor variables to include in the multivariable logistic regression model. Variables associated with a p-value <= 0.10 in the univariate model were included in the multivariable logistic model, for which statistical significance was defined

as two-sided $p<0.05$. For continuous variables, median and interquartile range (IQR) values were reported. For categorical variables, number count in data (N) and proportion (%) were reported. For published trials, a linear regression model was used to compare the time lag to publication with relevant trial factors, reporting standard error (SE) and p-value, where statistical significance was also defined as two-sided $p<0.05$. The number of published trials that reported a significant primary endpoint were presented as a count and proportion. All statistical analysis was completed on R Version 3.3.2.

Results

Trial Characteristics

As of August 01, 2019, the search term "comparative effectiveness" yielded 1883 study registry records on CT.gov. A restricted search identified 1014 of the 1883 studies were completed, prospective and interventional studies. Following the remaining inclusion and exclusion criteria, 122 studies were included in the study cohort as randomized, prospective CER trials relevant to primary care. The specific reasons for exclusion are described in **Fig. 1.** Study characteristics and their association with publication are outlined in **Table 2**. According to the pre-set filters on CT.gov, thirty-four studies (28%) were industry-funded, 31 studies (25%) were government-funded by NIH or US Federal agencies and 57 studies (47%) received funding from other sources, including academic medical centers and non-profit organizations.

Fig 1: Clinicaltrials.gov registry records screening flow chart

The reason most primary-care-relevant CER trials were conducted was treatment as opposed to prevention, health services research and screening (81 studies, 66%). Of the total registered trials in the study cohort, around one third had a focus on cardiovascular and infectious disease primary care (41 studies, 34%), as depicted in **Fig. 2**. A number of studies

were open-label (55 studies, 45%). The study design most commonly used was parallel assignment (107 studies, 88%). Fifty-two (43%) studies compared drug interventions, 42 (34%) compared behavioural interventions, 12 (10%) compared devices, and 16 (13%) compared other interventions. Most of the studies had at least one study site in the United States (64 studies, 52%). Median sample size was 192 patients enrolled (IQR, 86-379). Out of the 66 (54%) studies that were published, median time between primary study completion and publication was 25 months (IQR, 16-34) and 54% reported statistical significance on their primary endpoint.

Fig 2: Distribution of study conditions of primary care CER trials included in final analysis

Factors Associated with Published Trials

Thirty-nine (32%) trials were not published in any form, including abstract. Most unpublished trials (n=32, 57%) were conducted outside of the United States and half of unpublished studies (n=28, 50%) were open-label studies. On univariate analysis, factors associated with publication were funding from the federal government and a cardiovascular primary care study condition (**Table 2**). In the multivariable model, the only significant factor associated with publication was a cardiovascular study condition, as compared to an infectious disease study condition (OR 4.4, 95% confidence interval 0.04-2.9, p=0.04) (**Table 2**). The pre-defined statistical analysis did not demonstrate statistically significant differences between any of the other trial factors and publication status.

Table 2: Characteristics of included studies and their association with publication

Characteristic	All Trials	Published	Unpublished	Univariable Analysis		Multivariable Analysis*	
				P	OR (95% CI)	P	OR (95% CI)
Number of Trials	122	66 (54%)	56 (46%)				
Study Type, N (%)							
Phase 3	18 (15%)	9 (14%)	9 (16%)		REFERENCE		
Phase 4	31 (25%)	16 (24%)	15 (27%)	0.913	--	--	--
N/A	73 (60%)	41 (62%)	32 (57%)	0.638	--	--	--
Funding, N (%)							
Government	31 (25%)	23 (35%)	8 (14%)		REFERENCE	NS	
Industry	34 (28%)	17 (26%)	17 (30%)	0.048	0.35 [-2.1, -0.01]		
Other†	57 (47%)	26 (39%)	31 (55%)	0.012	0.29 [-2.2, -0.27]		
Intervention Type, N (%)							
Behavioural	42 (34%)	23 (35%)	19 (34%)		REFERENCE		
Device	12 (10%)	6 (9%)	6 (11%)	0.771	--	--	--
Drug	52 (43%)	28 (42%)	24 (43%)	0.929	--	--	--
Other	16 (13%)	9 (14%)	7 (12%)	0.919	--	--	--
Purpose, N (%)							
Treatment	81 (66%)	42 (63%)	39 (70%)	0.484	--	--	--
Other	41 (34%)	24 (37%)	17 (30%)				
Blinding, N (%)							
Double+	34 (28%)	21 (32%)	13 (23%)		REFERENCE		
Open-label	55 (45%)	27 (41%)	28 (50%)	0.245	--	--	--
Single	33 (27%)	18 (27%)	15 (27%)	0.550	--	--	--
PC Category, N (%)							
Cardiovascular	24 (20%)	10 (16%)	14 (25%)		REFERENCE		REFERENCE
Infectious Disease	16 (13%)	12 (18%)	4 (7%)	0.043	4.2 [0.04, 2.8]	0.044	4.4 [0.04, 2.9]
Psychiatry	13 (11%)	10 (16%)	3 (5%)	0.048	4.7 [0.02, 3.1]	NS	
Pain	12 (10%)	8 (12%)	4 (7%)	0.164	--	--	--
Respiratory	11 (9%)	4 (6%)	7 (12%)	0.766	--	--	--
Other	46 (38%)	22(30%)	24 (43%)	0.624	--	--	--
Location, N (%)							

United States	64 (52%)	40 (60%)	24 (43%)	0.052*	--	NS	
Other	58 (48%)	26 (40%)	32 (57%)				
Number of Patients Enrolled, Median [IQR]	192 [86-379]	206.5 [103-392]	169.5 [72-347]	0.977	--	--	--

*All variables associated with a P<=0.10 in the univariate analysis were also included in the multivariable model

† "Other" funding represents financial support received from academic medical centers and non-profitable organizations for purposes of conducting the study.

NS- Not Significant, IQR- Interquartile Range, N – Number count, OR – Odds Ratio, P – p-value, CI – Confidence Interval

Additional post-hoc reviews were completed, including a sensitivity analysis of studies that was done to determine if rate of publication improved upon allowing a greater follow-up time, as past authors have found that studies remain unpublished for years after study completion [15-17]. Initially, the study cohort included 122 trials, which allowed for a minimum of only 4-years to follow-up.

To address the concern of limited follow-up time, only trials that permitted at least a 6-year follow-up time after primary study completion were included in this analysis. This reduced the study cohort to 66 registered primary care-relevant comparative effectiveness trials, of which 28 (42%) had no matching full-text publication within 6 years. 22 (33%) trials were not published in any form. Overall, 38 (58%) trials (**Fig 3**) were published in the cohort of studies that allowed for a minimum of 6 years to follow-up. A univariate logistic regression model was fitted to confirm the statistically significant association observed between study conditions and publication in the original analysis. In line with the original analysis, a cardiovascular study condition compared to an infectious disease study condition showed a trend towards association with publication, (OR 6.6, p=0.054).

Fig 3: Sensitivity analysis comparing percentage of studies published within four years with studies published within six years

Second, an additional post-hoc review of the final dataset was performed by a physician trained in family medicine, who identified 10 out of 122 trials as "too specialized" for inclusion. Specialties included dermatology, infectious disease, psychiatry, pain as well as the neurology study described earlier. It was ultimately decided that these trials should be subject to inclusion in the final dataset as they represent an important subset of the primary care population and the data disseminated will help answer specific questions that remain unanswered in pre-license research. Last, when the physician-ethicist judged accuracy of the studies extracted, only one

study was found to be labeled as cardiovascular instead of respiratory, leading to an error rate of less than 1%.

Factors Associated with Time Lag from Primary Study Completion to Publication

The only factor associated with longer delay between study completion and full-text publication was study condition. On univariate analysis, time lag between study completion and publication was longer by 14.9 months (standard error (SE) 4.4, p=0.001) and 11.1 months (SE 4.6, p=0.019) for infectious disease and psychiatry trials, respectively, when compared to cardiovascular trials (**Table 3**). Similarly, on multivariable analysis, time lag was longer by 13.8 months (SE 4.4, p=0.003) and 11.5 months (SE 4.9, p=0.021) for infectious disease and psychiatry trials, respectively, when compared to cardiovascular trials (**Table 3**).

Median lag between primary study completion and publication for all studies included was 25 months (IQR, 16-34). Shortest median time lag was for trials with cardiovascular primary care study conditions at only 17 months delay (IQR, 15-20) and between funding types, industry funding had the longest median time lag to publication at 29 months (IQR, 17-38). A distribution of time lag from primary study completion to publication can be seen in **Fig 4**.

Fig 4: Histogram of time to publication for completed, primary care comparative effectiveness ClinicalTrials.gov registration records

Table 3: Factors associated with time lag (months) from primary study completion to publication of ClinicalTrials.gov registered comparative effectiveness trials testing primary care conditions

Characteristic	Published	Median Time to Publication: months (IQR)	P-Value: Time Lag to Publication (months)			
			Univariable		Multivariable	
			P	Lag (months), SE	P	Lag (months), SE
Number of Trials	66 (54%)	25 (16-34)				
Study Type, N (%)						
Phase 3	9 (14%)	25 (18-38)		REFERENCE		
Phase 4	16 (24%)	22 (16-35)	0.549	--	--	--
N/A	41 (62%)	25 (16-32)	0.454	--	--	--
Funding, N (%)						

Government	23 (35%)	26 (17-33)		REFERENCE		
Industry	17 (26%)	29 (17-38)	0.627	--	--	--
Other	26 (39%)	18 (15-31)	0.092 **	--	NS	
Intervention Type, N (%)						
Behavioural	23 (35%)	25 (16-32)		REFERENCE		
Device	6 (9%)	30 (18-34)	0.922	--	--	--
Drug	28 (42%)	18 (15-37)	0.557	--	--	--
Other	9 (14%)	25 (17-33)	0.842	--	--	--
Purpose, N (%)						
Treatment	42 (63%)	24 (16-35)	0.881	--	--	--
Other	24 (37%)	25 (16-32)				
Blinding, N (%)						
Double+	21 (32%)	18 (16-32)		REFERENCE		
Open-label	27 (41%)	25 (15-36)	0.384	--	--	--
Single	18 (27%)	26 (19-32)	0.204	--	--	--
Primary Care Study Condition, N (%)						
Cardiovascular	10 (16%)	17 (15-20)		REFERENCE		REFERENCE
Infectious Disease	12 (18%)	38 (26-40)	0.001	14.9 [4.4]	0.003	13.8 [4.4]
Psychiatry	10 (16%)	29 (25-33)	0.019	11.1 [4.6]	0.021	11.5 [4.9]
Pain	8 (12%)	28 (20-33)	0.145	--	--	--
Respiratory	4 (6%)	29 (25-31)	0.189	--	--	--
Other	22(30%)	18 (15-26)	0.518	--	--	--
Location, N (%)						
United States	40 (60%)	24 (15-35)	0.993	--	--	--
Other	26 (40%)	25 (17-32)				
Significant Primary Endpoint Reported	36 (54%)	24 (17-32)	0.939	--	--	--

**All variables associated with a P<=0.10 in the univariate analysis were also included in the multivariable model

SE – Standard Error

Discussion

Investigators have ethical obligations to present and submit their research outcomes, without regard for whether results are neutral, negative or positive [24]. Discussion regarding publication bias, the phenomenon of suppressing unfavourable research outcomes, is widespread in the biomedical literature [25-30]. Our study followed along similar lines, finding that 46% of the CER trials that were relevant to primary care and included in final analysis had no matching

full-text publication. Further, 32% of the trials in our cohort were not published in any form, including abstract. The only significant factor associated with published trials was a cardiovascular study condition. This could be in part due to larger research teams in heart and vascular research or investigators believing that the results from these studies were of greater importance to publish promptly.

However, previous meta-analyses of biomarkers potentially useful in establishing evidence-based practice guidelines for assessment, treatment and prediction of heart disease have been associated with publication bias [31]. Certain standard of care therapies established as reliable predictors of cardiovascular risk, such as measurement of carotid intima-medial thickness, turn out to be supported by 12 times the number of significantly favourable findings than were expected for inclusion in establishing this type of guideline therapy [32]. A similar finding was observed in the evidence used to establish a guideline for identifying which biomarkers should be measured to predict coronary heart disease. Current routine practice is measurement of apolipoprotein B compared to low-density lipoprotein cholesterol, which are both biomarkers that predict heart disease. However, a meta-analysis presented significant results 1.5 times more than what would have been expected when comparing studies of these two biomarkers [33]. Thus, despite published evidence existing on the benefits of both heart disease predictors, it is often the one with more significant published findings that gets worked into patient care recommendations. Moreover, further analyses show that cardiovascular studies with unfavourable results are either suppressed through non-publication or are presented in biased ways to appear significant when reported [32]. This evidence of selective-reporting in published cardiovascular research makes us question not only the integrity of timely reported results in our study but also the translation of findings from CER trials to evidence-based practice. As

mentioned in the literature on publication bias, the quality of care guidelines can only be as robust as the published evidence they are founded on [32].

Overall, our results suggest that roughly half of CER trials (n=56, 46%) conducted did not publish results within four years. The sensitivity analysis of extended follow-up time (**Figure 3**) presented a slight improvement in the publication rate (from 54% to 58%) when allowing for a minimum of six years to publish post-primary study completion instead of four. However, rate of publication remained unsatisfactory. The publication rate of 54% in our original analysis, while being far from acceptable, is an improvement from the lower publication rates of early phase study analyses performed in the past, where only 17% of phase I studies had been published [34]. Moreover, a CT.gov analysis of all recently completed academic medical center studies found a publication rate of 66% [19]. However, it is plausible that late phase trials, which often have larger enrollment numbers than early phase trials, are more likely to be published regardless of study outcome due to the absence of experimental comparators and decreased likelihood of adverse events.

Moreover, our study also found median time to publication following study completion to be 25 months, in line with the literature, where approximately 2 years to publication has been observed in previous CT.gov analyses [35,36]. On both univariable and multivariable analysis, a cardiovascular study condition was associated with a shorter lag to publication following primary study completion compared to infectious disease and psychiatry conditions. A study evaluating time to publication for cardiovascular interventions funded by the National Heart, Lung and Blood Institute (NHLBI) found median time to publication to be 25 months as well as a shorter time to publication for studies with "positive" findings [37]. In our analysis, CER trials studying cardiovascular conditions seen in primary care had a 17-month median time to publication,

amongst the shortest duration of all study CER characteristics, which may be linked to the concept of publication bias in cardiovascular studies mentioned earlier [31,37].

Our study had several limitations. First, an analysis of publication rate alone cannot determine "publication bias" in our study. Although most of the published primary care CER trials in our study reported a statistically significant primary endpoint (n=36, 54%) in their results, we cannot associate non-publication with the withholding of unfavourable, null or negative findings. Lack of publication given neutral and/or unfavourable findings has previously been associated with studies of small sample sizes, industry-funding as well as, less commonly, rejection of manuscript by journals [27, 29, 30].

Moreover, we assessed the publication patterns of prospectively registered CER studies on CT.gov only, which means we did not account for randomized CER studies that were not registered, but instead only evaluated non-publication of registered studies. Our assumption was that if investigators are willing to follow FDAAA regulations, and in some instances voluntarily prospectively register their studies, then the same group of investigators are also likely to publish their study results/outcomes in a timely manner. Consequently, limiting inclusion criteria to exclusively CT.gov registration records may have deflated publication rates compared to examining study protocols for publication instead.

Second, classification of studies as CER required judgment. CER may be the comparison of standard of care interventions that are routinely used, but to determine which interventions primary care physicians routinely use or most benefit from in practice is not straightforward. For instance, for countries outside of the United States, therapies considered "experimental" in the U.S. could be routine if local regulations differ from the FDA. Even if the drug had not been approved for local consumer use for the indication in question – certain independent clinical

investigators may be inclined to conduct a study of an intervention that is routinely used off-label to generate evidence of the common practice against an approved, standard of care intervention. The data from such a registration record and/or subsequent publication were included in analysis. Lastly, the matching publication of the registration records included in the final analysis were determined, albeit comprehensively, by lead author only. Since there was also no contact to investigator(s) to confirm the publication status of registered studies if they could not be found in the databases mentioned in ***Methods***, some publications, for instance non-English articles, may have been missed.

What counts as CER relevant to primary care?

The conclusions of this paper are heavily based on the criteria used to select which studies qualify as both CER and significant to primary care. First, we will discuss the specific considerations, documented beforehand, that were applied consistently when deciding which studies to extract. As mentioned, the authors followed IOM's definition of CER with the understanding that comparative effectiveness studies are most commonly conducted in "community" settings by non-expert providers because these less specialized settings contribute to increased external validity – a reason why CER provides more generalizable evidence than pre-license studies [2]. Although the sub-group of disciplines within primary care are usually limited to family medicine, pediatrics and internal medicine [38], the types of patients seen within these settings are diverse, creating their own patient populations at the primary level of healthcare. CER is an opportunity to tailor research questions and primary endpoints to these subsets of populations that are understudied in fastidious pre-license research.

When considering which CER studies to include as relevant to primary care, specialty studies of patients that could be among subsets of primary care populations were included. For

instance, a CER trial that studied a neurology condition was included despite neurology being a non-primary care specialty. The American Recovery and Reinvestment Act of 2009 (ARRA) devoted generous research dollars to CER to help improve routine care for patients with various conditions and a basis for doing so included unaffordability of specialty care [22]. The study in question was conducted by neurologists, testing *active* comparator drugs for urinary incontinence in patients with spinal cord injury [39]. However, since not everyone can access a specialty care visit to a neurologist, primary care physicians who are trained to manage overactive bladder and urinary incontinence would benefit from the outcomes of this trial when treating patients with spinal cord injury [40]. Other studies, regardless of specialty, were also included if they were determined to be a disservice to primary care if not published. Further examples include oncology as a specialty that impacts primary care patients at the level of cancer screening decision making as well as cancer survivorship care [41]. Thus, this process of vetting CER trials relevant to primary care did not exclude all specialty studies as patients with certain conditions can present themselves in primary care settings and it would be considered advantageous for their care providers to be enabled with the outcomes of these studies.

Conclusions

In conclusion, unbiased reporting of research outcomes in a timely manner fulfills ethical obligations toward the research subjects who voluntarily participate in studies under the prospect of making a valuable contribution to the scientific community. ARRA devoted $1.1 billion in funding for CER [22] due to its potential to accumulate evidence that will directly improve health outcomes for patients – the majority of whom exist at the primary level of healthcare. Where typical explanatory studies set out to answer research questions requested by the FDA for approval or regulatory purposes, CER gives the expert medical community the opportunity to ask

questions that are important to real-world patient health as well as tailored to specific sub-sets of the population. Any percentage of randomized CER that goes unpublished implies that the purpose of generating real-world evidence is not met, an idea that challenges the moral permissibility of conducting CER in the first place. Our study encourages specific mandatory reporting of *all* late phase studies in addition to mandatory prospective registration to a public clinical trials database.

Acknowledgements: The authors would like to thank Dr. Carolyn Ells, PhD for editorial assistance and advice at all stages of writing, Dr. Tibor Schuster, PhD for statistical analysis review and Dr. Eugene Bereza, MD CM CCFP for confirming relevance to CER and primary care for the data used in this study.

See Below for Figures in the Manuscript:

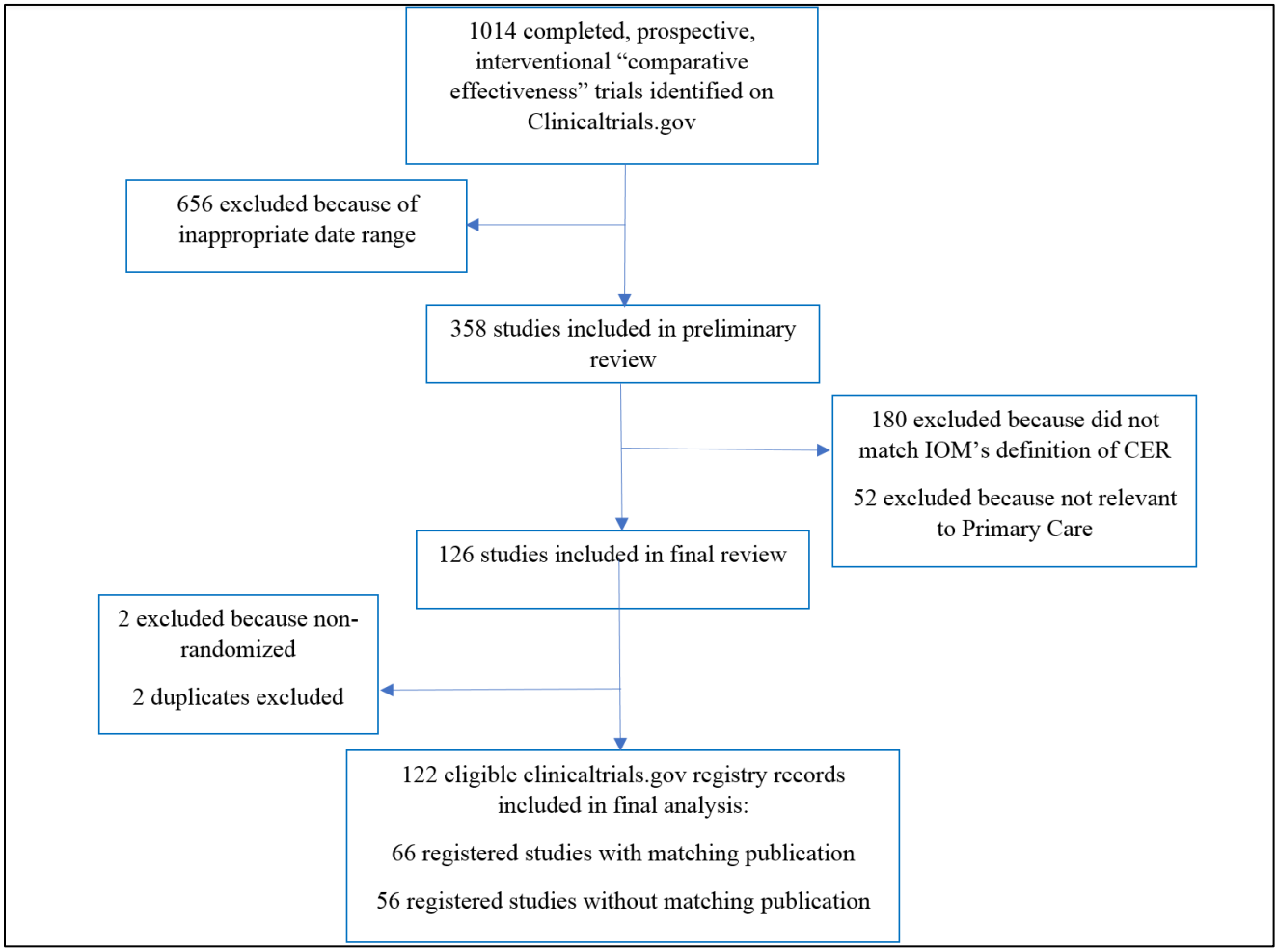


Figure 1: Clinicaltrials.gov registry records screening flow chart

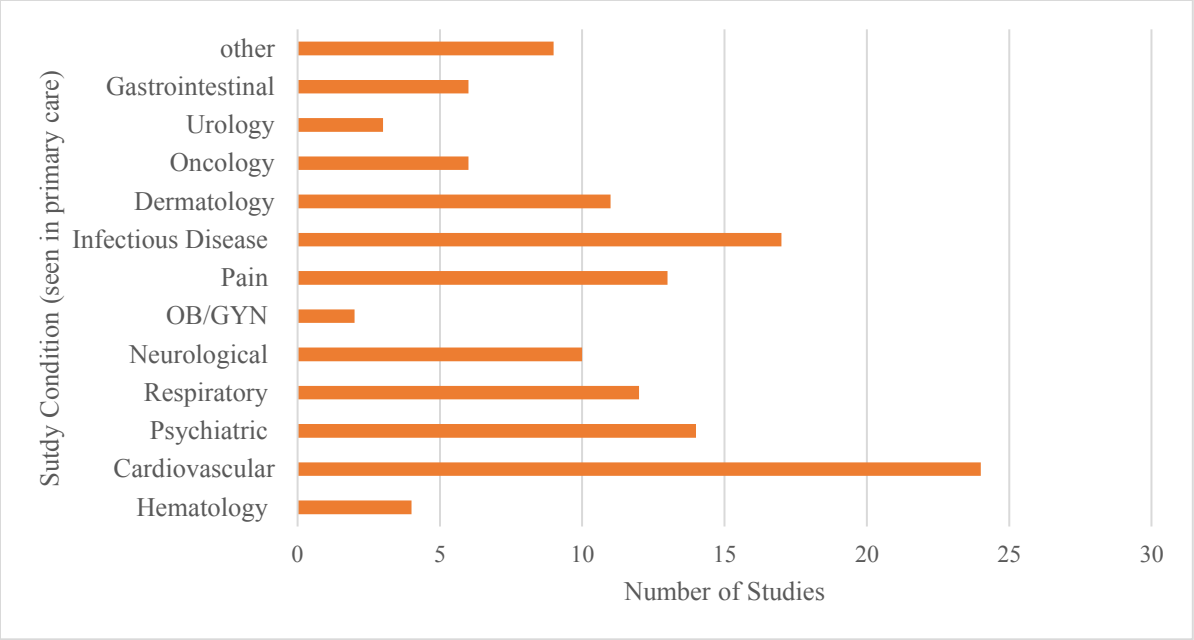


Figure 2: Study conditions of randomized CER trials with potential to improve patient outcomes in primary care that were included in final analysis

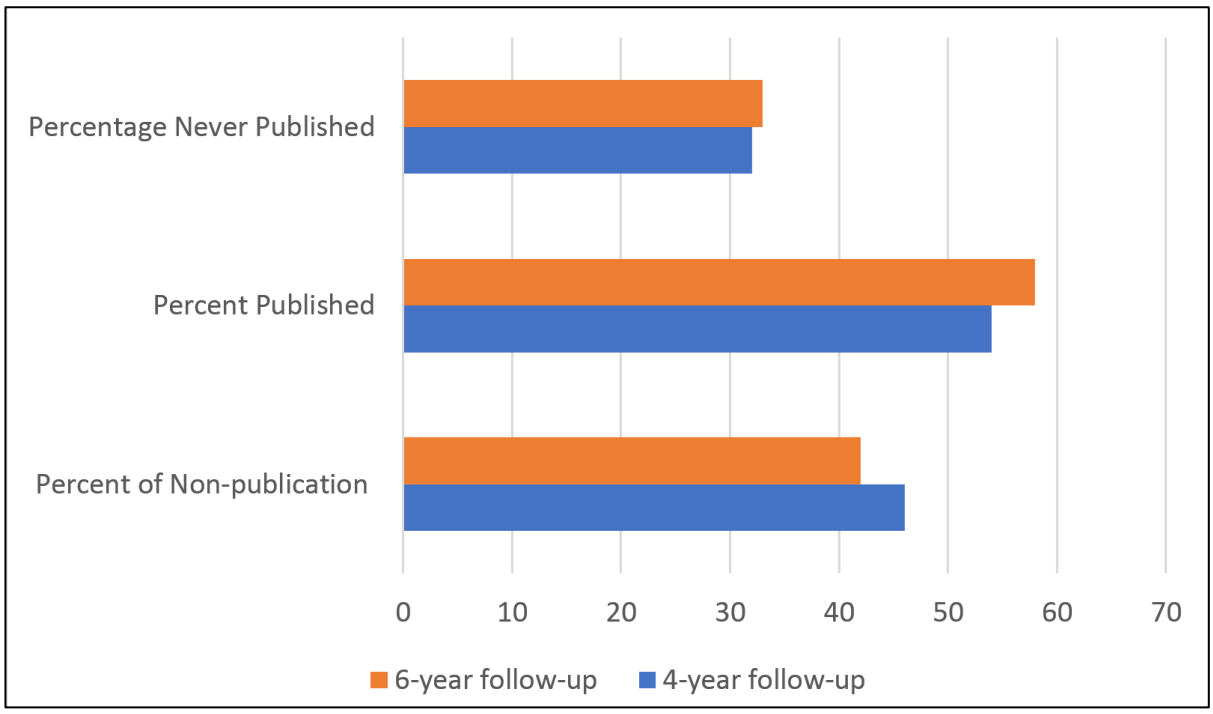


Figure 3: Comparing percentage of studies published within four years in original analysis with studies published within six years in sensitivity analysis

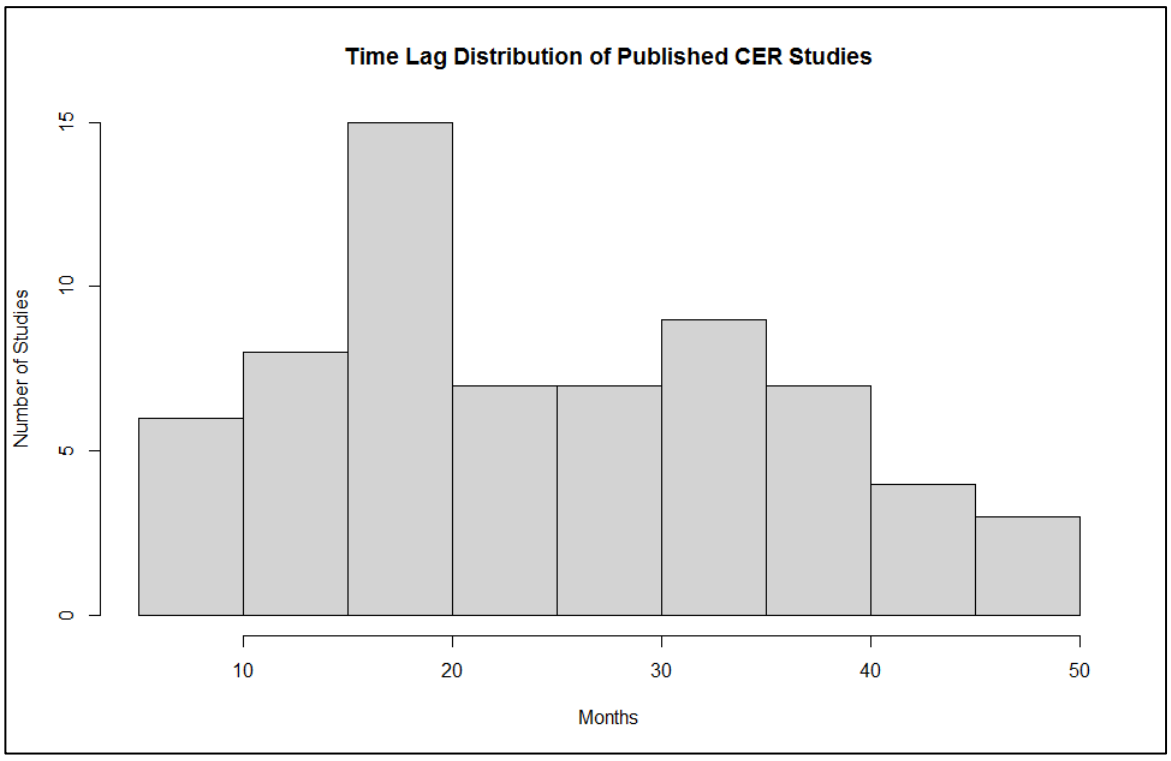


Figure 4: Histogram of time to publication for completed, primary care CER CT.gov registry records

References:

1. Institute of Medicine (U.S.). Committee on Comparative Effective Research Prioritization. Initial national priorities for comparative effectiveness research [Internet]. Washington, D.C.: National Academies Press; 2009. Available from: http://site.ebrary.com/id/10347043
2. Patsopoulos NA. A pragmatic view on pragmatic trials. Dialogues Clin Neurosci. 2011 Jun;13(2):217–24.
3. Lancet T. Prioritising primary care in the USA. The Lancet. 2019 Jul 27;394(10195):273.
4. Mlinarić A, Horvat M, Šupak Smolčić V. Dealing with the positive publication bias: Why you should really publish your negative results. Biochem Med (Zagreb). 2017 Oct 15;27(3):030201–030201.
5. Easterbrook PJ, Gopalan R, Berlin JA, Matthews DR. Publication bias in clinical research. The Lancet. 1991 Apr 13;337(8746):867–72.
6. Murad MH, Chu H, Lin L, Wang Z. The effect of publication bias magnitude and direction on the certainty in evidence. BMJ EBM. 2018 Jun 1;23(3):84.
7. Zarin DA, Tse T, Williams RJ, Rajakannan T. The Status of Trial Registration Eleven Years after the ICMJE Policy. N Engl J Med. 2017 Jan 26;376(4):383–91.
8. Al-Shbool G, Latif H, Farid S, Wang S, Ahn J, Giaccone G, et al. Publication Rate and Characteristics of Lung Cancer Clinical Trials. JAMA Network Open. 2019 Nov 6;2(11):e1914531–e1914531.
9. Saito H, Gill CJ. How frequently do the results from completed US clinical trials enter the public domain?--A statistical analysis of the ClinicalTrials.gov database. PLoS One. 2014 Jul 15;9(7):e101826–e101826.
10. van de Wetering FT, Scholten RJPM, Haring T, Clarke M, Hooft L. Trial Registration Numbers Are Underreported in Biomedical Publications. PLOS ONE. 2012 Nov 14;7(11):e49599.
11. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov Results Database — Update and Key Issues. N Engl J Med. 2011 Mar 3;364(9):852–60.
12. 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 Medicine. 2013 Dec 3;10(12):e1001566.
13. Camacho LH, Bacik J, Cheung A, Spriggs DR. Presentation and subsequent publication rates of phase I oncology clinical trials. Cancer. 2005 Oct 1;104(7):1497–504.
14. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial Publication after Registration in ClinicalTrials.Gov: A Cross-Sectional Analysis. Sim I, editor. PLoS Med. 2009 Sep 8;6(9):e1000144.
15. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials: Comparison of Protocols to Published Articles. JAMA. 2004 May 26;291(20):2457–65.
16. Turner EH, Tell RA. Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. n engl j med. 2008;9.
17. Ross JS, Mocanu M, Lampropulos JF, Tse T, Zarin DA, Krumholz HM. TIME TO PUBLICATION AMONG COMPLETED CLINICAL TRIALS. JAMA Intern Med. 2013 May 13;173(9):825–8.
18. FDAAA 801 and the Final Rule - ClinicalTrials.gov [Internet]. [cited 2020 Feb 23]. Available from: https://clinicaltrials.gov/ct2/manage-recs/fdaaa

19. Chen R, Desai NR, Ross JS, Zhang W, Chau KH, Wayda B, et al. Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. BMJ. 2016 Feb 17;352:i637–i637.
20. Query Form - NIH RePORTER - NIH Research Portfolio Online Reporting Tools Expenditures and Results [Internet]. [cited 2020 Feb 21]. Available from: https://projectreporter.nih.gov/reporter.cfm?def=1&frs=
21. Anders ME, Evans DP. Comparison of PubMed and Google Scholar Literature Searches. Respiratory Care. 2010 May 1;55(5):578.
22. Sox HC. Comparative Effectiveness Research: A Report From the Institute of Medicine. Ann Intern Med. 2009 Aug 4;151(3):203.
23. Grover S, Dalton N. Abstract to publication rate: Do all the papers presented in conferences see the light of being a full publication? Indian J Psychiatry. 2020/01/03. 2020;62(1):73–9.
24. Macrae DJ. The Council for International Organizations and Medical Sciences (CIOMS) Guidelines on Ethics of Clinical Trials. Proc Am Thorac Soc. 2007 May 1;4(2):176–9.
25. 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 of Systematic Reviews [Internet]. 2009 [cited 2020 Feb 24];(1). Available from: https://www-cochranelibrary-com.proxy3.library.mcgill.ca/cdsr/doi/10.1002/14651858.MR000006.pub3/full
26. Perlis RH, Perlis CS, Wu Y, Hwang C, Joseph M, Nierenberg AA. Industry Sponsorship and Financial Conflict of Interest in the Reporting of Clinical Trials in Psychiatry. AJP. 2005 Oct 1;162(10):1957–60.
27. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. Control Clin Trials. 1987 Dec;8(4):343–53.
28. Dickersin K, Min YI. Publication bias: the problem that won't go away. Ann N Y Acad Sci. 1993 Dec 31;703:135–46; discussion 146-148.
29. Dickersin K. The Existence of Publication Bias and Risk Factors for Its Occurrence. JAMA. 1990 Mar 9;263(10):1385–9.
30. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet. 1991 Apr 13;337(8746):867–72.
31. Nissen SE. Biomarkers in Cardiovascular Medicine: The Shame of Publication Bias Comment on "Bias in Associations of Emerging Biomarkers With Cardiovascular Disease." JAMA Internal Medicine. 2013 Apr 22;173(8):671–2.
32. Tzoulaki I, Siontis KC, Evangelou E, Ioannidis JPA. Bias in Associations of Emerging Biomarkers With Cardiovascular Disease. JAMA Internal Medicine. 2013 Apr 22;173(8):664–71.
33. Thompson A, Danesh J. Associations between apolipoprotein B, apolipoprotein AI, the apolipoprotein B/AI ratio and coronary heart disease: a literature-based meta-analysis of prospective studies. J Intern Med. 2006 May;259(5):481–92.
34. Decullier E, Chan A-W, Chapuis F. Inadequate Dissemination of Phase I Trials: A Retrospective Cohort Study. PLOS Medicine. 2009 Feb 17;6(2):e1000034.
35. Ross JS, Mocanu M, Lampropulos JF, Tse T, Krumholz HM. Time to Publication Among Completed Clinical Trials. JAMA Internal Medicine. 2013 May 13;173(9):825–8.
36. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. BMJ. 2012 Jan 3;344:d7292–d7292.

37. Gordon D, Taddei-Peters W, Mascette A, Antman M, Kaufmann PG, Lauer MS. Publication of Trials Funded by the National Heart, Lung, and Blood Institute. N Engl J Med. 2013 Nov 13;369(20):1926–34.
38. Institute of Medicine (US) Committee on the Future of Primary Care; Donaldson MS, Yordy KD, Lohr KN, et al., editors. Primary Care: America's Health in a New Era. Washington (DC): National Academies Press (US); 1996. 2, Defining Primary Care. Available from: https://www.ncbi.nlm.nih.gov/books/NBK232631/
39. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000 Feb 29 - Identifier NCT00800462, Comparative Study of Muscarinic M3 Receptors Antagonists in the Treatment of Neurogenic Detrusor Overactivity; 2008 Dec 2 [cited 2020 Jan 18]; [about 3 screens]. Available from: https://clinicaltrials.gov/ct2/show/NCT00800462
40. Imam KA. The Role of the Primary Care Physician in the Management of Bladder Dysfunction. Rev Urol. 2004;6(Suppl 1):S38–44.
41. Nekhlyudov L, O'Malley DM, Hudson SV. Integrating Primary Care Providers in the Care of Cancer Survivors: Gaps in Evidence and Future Opportunities. Lancet Oncol. 2017 Jan;18(1):e30–8.

S1 Data Table. Excel spreadsheet of ClinicalTrials.Gov registration record and publication characteristics extracted for final data analysis

Chapter IV – A detailed review of methods used during ClinicalTrials.Gov analysis of publication rates

This chapter will outline the approach to methodology in Chapter III, including justification of search strategy, interpretation of inclusion criteria as well as its purpose and lastly, the relevance and significance of the data extracted considering the real-world implications of this thesis.

1. Search Strategy

As the methods section in Chapter III indicates, search methodology was intended to result in a list of trials that were broadly defined but still well delineated. This was achieved by searching for the key term "comparative effectiveness", limiting the search criteria and restricting the number of relevant study records for data collection. The advanced search filters on CT.gov helped account for inclusion criteria with ease. For instance, during the piloting stage of this thesis, the key word "comparative effectiveness research" was considered more feasible than "comparative effectiveness", narrowing the results from 1856 registered studies to only 310. After applying the initial inclusion criteria search filters on CT.gov, the results were reduced to 105 completed, interventional, phase 3, 4 or N/A RCTs, out of which only 40 trials met the pre-specified date range requirements and were relevant to primary care. However, after realizing that there may be other trials registered on CT.gov that fit the definition of a comparative effectiveness research trial, but may have been mislabeled by investigators, the previous "narrow" selective search was expanded. For this reason, a thorough "wide" search of CT.gov was conducted by manually assessing each of the 1856 trials that resulted from a search of the terms "comparative effectiveness" in the CT.gov database using the Chapter III strategy.

2. Interpreting Inclusion Criteria

This section discusses the rationale behind the inclusion criteria in **Table 1** of Chapter III.

2.1 Date Range

Date range inclusion criteria is based on FDAAA 801, a regulation mandating registration of drug, biological and device trials within 21 days of first patient-subject enrollment and submission of results to the CT.gov database within one year of primary completion [41,81]. Since FDAAA 801 has been implemented beginning September 2007, this date was chosen as the study start date for data extraction purposes as well. The end date was chosen to provide up to four years of time elapsed between primary study completion and full-text publication, consistent with past CT.gov publication rate analyses[76]. Data extraction took place August 01, 2019, which is why the latest study completion date acceptable for inclusion in the data analysis was August 01, 2014, leaving four years to follow up. The reasoning for using FDAAA 801 as a guideline was the assumption that investigators who started and completed a trial while this regulation was in effect should have been inclined to not only register their trial but also report results in a timely manner compared to a time period when registration and results reporting was voluntary.

2.2 Types of Interventions Tested

Continuing our discussion about basing inclusion criteria on FDAAA 801, this specific regulation only impacts certain "Applicable Clinical Trials" (ACTs) which CER trials may or may not fall under [41]. ACTs include drug, biologic and device intervention trials but many CER trials test behavioural interventions as well, including cognitive therapies and educational approaches [81]. This means that these studies were voluntarily registered, and results were reported on CT.gov by choice, without behavioural interventions being subjected to Final Rule FDAAA 801. However, NIH-funded studies, Declaration of Helsinki, World Health Organization and ICJME beginning 2004, all included behavioural interventions within the scope of their respective registration policies [82]. Thus, considering the existence of other overarching

research ethics policies that instruct reporting of research beyond FDAAA 801's definition of the "ACT", behavioural intervention trials were added to the inclusion criteria for extracting studies, which also increased our sample size of studies for analysis. Given that honest and timely reporting is an ethical obligation of researchers, trial outcomes should be published, regardless of whether they are mandated to by regulations.

2.3 What is a primary care CER trial?

The studies that presented a challenge during data extraction were those studying interventions used in specialty care but possibly also important for a patient's primary care physician (PCP) to be aware of to provide that patient with optimal continuity of care. For example, studies comparing two or more routine Botox procedures would not be considered as "primary care" because Botox is regularly administered to patients by dermatologists and specialty care practitioners [83]. However, before undergoing a cosmetic procedure, patients should consult with their PCPs and before prescribing certain treatments, PCPs should be made aware of their patients' previous use of Botox, fillers and other cosmetic procedures.

Similarly, although patients with gastrointestinal (GI) disorders are commonly referred to GI specialists for treatment, primary care plays a key role in the referral process as a patient's first point of contact [84]. For instance, when extracting the studies NCT01976494 [85] and NCT01964417 [86], both CER trials met date range for inclusion, but one was testing a patient-controlled analgesic administered to patients with GI cancer and the other was testing two different methods of bowel preparation prior to colonoscopy. Although the colonoscopy would be performed by a GI specialist, the referring primary care physician (PCP) should be knowledgeable about the procedure in question to help their patient prepare for the procedure. However, even if the primary level of healthcare commonly refers a patient to certain

procedures, such as a chest x-ray or an ultrasound, but the comparative effectiveness study was conducted to generate field-specific evidence, then it was excluded from the final dataset. For instance, the study NCT01654887 compared effectiveness of a chest X-ray versus an ultrasound to diagnose pneumonia [87], which is a common primary care concern, but was conducted in the emergency department – a setting that differs greatly from outpatient primary care. Since CER is meant to generate pragmatic, real world evidence, the outcomes of this study would not be applicable to patients with pneumonia in primary care settings, ultimately leading to an emergency room referral. Thus, specialty-focused studies were excluded from this thesis, unless directly applicable for inclusion to the primary healthcare evidence base and/or were specialty care follow-up studies.

3. Heterogeneity of registration records: active comparators, efficacy and effectiveness

Through the laborious data extraction process, I noticed that not all investigators who registered a CER trial labeled their study as such. Conversely, not all studies with "comparative effectiveness" in the title and/or key words were CER trials according to IOM's definition. A few measures were taken to identify which studies were, in fact, CER studies according to the IOM definition most consistently utilized throughout the biomedical literature [11]. Although, general rule of thumb to identify randomized CER is to look for a study comparing two or more active standard of care interventions. However, when extracting data from what is publicly available on CT.gov, many investigators registered their studies in ways subject to interpretation, as seen in **Table 2**.

3.1 The Active Versus Experimental Comparator

As discussed briefly in Chapter III, a genuine CER trial would randomize patients to different active comparator interventions to assess comparative effect between the two. An active

comparator is an intervention that is already regarded as effective by health care providers [88]. However, certain interventions compared head-to-head in registered "CER" trials labeled an "active" comparator therapy as being tested against an "experimental" therapy within the CT.gov "Arms and Interventions" field. This, if labeled correctly, would count as a clinical effectiveness study, not a comparative effectiveness study as many registration records suggested. Rarely, investigators did correctly identify their CER trial interventions as a comparison of "active" comparators. Others described their CER trials as an experimental comparator versus an active comparator being tested. In such a case, one can identify via the biomedical literature if the experimental comparator is in fact a part of routine care. If so, the "experimental" study arm does not indicate an experimental therapy but instead, indicates that the therapy in question is also an active comparator that the investigators have good reason to believe is either equivalent, superior or non-inferior to the other standard of care therapy tested. This could be a logical assumption as to why some investigators may choose the term "experimental" instead of "active" for their CER trial to indicate it is the focus of their clinical study. However, this assumption, although applicable to certain studies, was not universally applicable. Certain investigators have labeled their studies as CER *incorrectly* and their identification of their comparator interventions as experimental helped determine this. In a few cases, the experimental comparator was a novel therapy compared against a standard care therapy, which does not count as CER.

Table 2: Differences in "arms and interventions" labels for "comparative effectiveness" trials from ClinicalTrials.Gov search results

Study Record	NCT Identifier	Comparators?	Comments
Comparative Effectiveness of Boceprevir vs Telaprevir	NCT02113631	Active vs active	Title and study arms correctly labeled. Government and non-profit sponsors better understood CER
Comparative Effectiveness Study	NCT01587274	Active vs active	

of Medications for Low Back Pain			
Treating Pain to Reduce Disability	NCT01614340	Usual care vs usual care + intervention	Did not use conventional arm types but is a CER trial
INSPIRE Diabetes Study	NCT01087567	Experimental vs active	Approved diabetes drug Lantus labeled experimental because it was the trial focus to compare against routine care
Comparative Efficacy of Low-Dose Estradiol and the SNRI Venlafaxine XR	NCT01418209	Active vs active	One intervention less established in routine use than other; good reason to determine comparative effect of both drugs -borderline efficacy/effectiveness
Comparative Study To Evaluate Efficacy of Peginterferon Alfa Plus Ribavirin	NCT02339337	Experimental vs active	Pegylated interferon plus Ribavirin is standard of care for HCV in China, despite experimental and efficacy label, this is a CER trial
Effectiveness of 3% Boric Acid in 70% Alcohol Versus 1% Clotrimazole Solution	NCT01547221	Experimental vs active	Standard of care treatments with determined clinical effectiveness for otomycosis but no consensus for most effective antifungal; both active comparators
Nebivolol Vs. Metoprolol: Comparative Effects on Fatigue	NCT00999102	Experimental vs experimental	Comparative effectiveness trial of two routine beta blockers, incorrectly labeled active as experimental
Comparative-effectiveness Study Comparing Epidural to Gabapentin	NCT01495923	Experimental vs active	Both standard interventions compared in a CER trial but active mislabeled as experimental
Comparative-effectiveness Study Comparing Epidural to Conservative Management	NCT01144923	Experimental vs active	Same as above
Different Types of Manual Therapy Techniques	NCT01792895	Active vs active	Not labeled as CER but compares all routine therapies, can be identified as such
Comparative Effectiveness Randomized Trial in Primary Care for Type 2 Diabetes	NCT01440530	Experimental vs active	A CER trial comparing routine interventions alongside an add on behavioural intervention; possible use of experimental to indicate intervention that is focus of study

Comparative Effectiveness Into Practice	NCT01293578	No intervention vs experimental	In this case, lack of specific intervention **is** standard of care against an active therapy; CER
CER for Two Medical Home Models for ADHD	NCT01275378	Active vs active	Academic, non-profit trial with CER labeled correctly and active comparators used
Comparative Evaluation and Cost-effectiveness of Two Interventions for Smoking Cessation	NCT00799279	Experimental vs active	Example of correctly labeled experimental arm indicating this is a study establishing efficacy – **not** a CER trial
Comparative Study of the Efficacy and Safety of Muscarinic M3 Receptors Antagonists	NCT00800462	Active vs active	Despite study titled "efficacy", active comparator arms of routine therapies indicate CER trial
Comparing the Efficacy and Safety of Ovarian Stimulation With Pergoveris® and Menopur®	NCT01623570	Experimental vs Experimental	Two standard treatments compared but labeled as experimental. Titled as an "efficacy" study incorrectly
Comparative Study of Efficacy and Safety of Licefreee Spray Against Nix 1% Permethrin	NCT01514513	Experimental vs active	Incorrectly labeled as "efficacy" study. Compares approved over-the-counter drugs that are standard of care

3.2 The efficacy versus effectiveness distinction

As discussed above and in Chapter III, effectiveness studies are often mistaken for efficacy and vice versa. The terms effectiveness and efficacy are not interchangeable. Studying efficacy is to answer the question, "can this work?" in ideal settings. However, studying effectiveness is to answer the question, "does this work?" in real world settings [11]. For this thesis, comparative effectiveness studies should set out to answer "what works better?" and "for which subset of the population?". Singal et al explain that CER trial design is far more complicated than establishing efficacy in stylized settings, but CER, if designed well, has a different aim – to enhance evidence-based medicine [18].

3.3 Off label studies

Since the experimental and active comparator terms were dependent on investigator interpretation, the studies that involved experimental interventions had to be analyzed carefully. For studies conducted outside the United States, determining whether an intervention was routinely used was not as straightforward. An intervention considered "experimental" in the US could be an "active" comparator in a different country, with local regulations differing from the FDA. Particularly, this was an issue when identifying industry-sponsored drug and device comparative effectiveness trials located in foreign countries.

In order to identify if the interventions tested in these studies were standard therapies within their respective countries but instead labeled "experimental", I had to conduct further research into the government regulatory bodies to assess whether the drug or device had been approved for entry into the consumer market or was commonly used off-label in the setting the trial had been conducted in. For example, if the study took place in India, I would have to confirm with the Government of India's Central Drugs Standard Control Organization to determine if the drug or device in question was approved for consumer use on the market or not. In the final dataset, a phase IV comparative study testing the drug Fenoverine was included despite the investigators labeling it as the "experimental" arm. This study was included because Fenoverine, according to authors Schmulson et al's four country comparison study of functional GI disorders, is a drug prescribed for treatment of IBD in Korea, where the study was conducted [89]. Lastly, even if the drug had not been approved for consumer use for the indication in question – certain independent clinical investigators may have conducted a study of an intervention that is routinely used off label to generate evidence of the common practice against

an approved, standard care intervention. According to this thesis, certain off-label trials are still CER and the data from those study records and/or subsequent publications were included.

Chapter V – Discussion and Conclusion

1. Summary of Results

With reasonable confidence, the original research investigation in Chapter III is the first CT.gov analysis to explore the publication rate of registered CER trials that had potential to improve primary care. Although many authors have used CT.gov as a tool for evaluating publication patterns, none have specifically put the spotlight on comparative effectiveness trials as a subset of clinical research that holds important pragmatic implications for health, medicine and society overall. One study has looked at the prevalence of CER that studied one of the IOM's 15 high priority research areas and were registered on CT.gov from 2007 to 2010 [22]. However, this study did not follow-up on the publication of the CER study records but instead only aimed at evaluating if the comparative effectiveness study design was used to test conditions where generalizable research outcomes were considered most needed [22]. Moreover, the assessment of time to publication following primary completion of registered CER trials can also be considered a continued contribution to knowledge. Many studies have explored time to publication for completed trials, but again to my knowledge, no analyses were specific to CER trials relevant to primary care.

Comparing characteristics of the primary care CER trials included in analysis, similarities were seen with descriptions already in the literature. The empirical analysis in Chapter III of this thesis found 45% of trials to be open label, meaning no blinding was used, and another 27% to be single-blind trials. The finding that around 72% of primary care CER trials were not adequately blinded mirrors the statement by the European Medicines Agency, where they

revealed that out of 42% of their post-approval studies (such as CER trials), 73% were open label [33,65]. As seen in Chapter IV **Table 1**, the confusion surrounding use of active comparators in CER trials screened for inclusion matches the findings of Naci et al, who likened post-approval studies to pre-approval studies, one basis being the lack of proper comparators used [33].

From our set of 122 CER trials judged to be important to primary care, 46% of CT.gov study records did not report a full-text publication within 4 years – indicating that non-publication is not limited to early phase research or pre-license research but is also present in primary care CER. Moreover, 32% of study records were not published in any form, despite several years elapsed following primary study completion. Cardiovascular study conditions were significantly associated with publication. Moreover, time from primary completion to publication was about an entire year longer for infectious disease and psychiatry CER trials compared to cardiovascular CER trials. However, in line with the findings of researchers in the past, median time to publication was 25 months [33-35]. Again, in this thesis, CER trials of cardiovascular study conditions held the shortest median time to publication at 17 months following primary study completion compared to 38 months for infectious disease CER trials.

A majority of patients at risk for cardiovascular disease (CVD) receive screening, treatment and follow-up in primary care settings [91,92]. Our finding that a CVD study condition increases likelihood for publication is encouraging for this large subset of patients whose initial point of contact and continued follow-up care is in the hands of their PCP. In an evidence-based guideline for the PCP community, Wallace et al emphasize that a "lack of current evidence must not be equated with evidence against effectiveness" [93]. The latter if believed by the preventive medicine and primary care community is worrying. As the authors discuss, many methods for CVD screening exist [93]; having robust evidence of comparative effect published and available

to evaluate both high and general risk populations is not only beneficial but crucial. When linking the strong motivations for an EBM approach to primary care in CVD prevention and treatment with our Chapter III findings, the significant association between cardiovascular study conditions and publication becomes unsurprising.

2. Significance to Bioethics and Family Medicine

At the core of a randomized primary care CER trial are the primary care patients that comprise the subgroups that the research outcomes are meant to benefit. However, when these results reach neither the public nor the medical community, how are CER trials particularly at fault? To begin, CER is a late phase research design as discussed in Chapter II, often gaining expedited status during ethics committee reviews along with waivers of informed consent from participants [13]. To highlight the concern of waiving informed consent for comparative effectiveness trials, I will discuss how non-publication threatens the moral basis for conducting CER, centering the argument around the differences between the relationship of the primary care research subject and their investigator as compared to the patient and their physician.

2.1 Informed Consent

While some argue that CER trials are minimal risk and only pose the added risk of protocolized care [94.95], others believe that the judgement of risk should include the risks of the standard of care therapies tested in the trial [13]. However, for the purposes of this thesis, most of the standard of care interventions tested in the cohort of primary care CER trials that reflected a 46% nonpublication rate would not be considered potentially high risk to begin with. This section will discuss the debate around waiving informed consent for CER trials focusing on the patient-physician relationship, which plays a significant role in primary care, instead of the level of risk involved [96].

To begin, the investigator-participant relationship is not parallel to the patient-physician relationship. The defense that CER involves interventions that a physician may offer in a routine clinic setting is an unjust comparison. The investigator, even if offering the same intervention that they would offer a patient in primary care, does not owe the subject, as they would owe their patient, an undivided duty of loyalty [97]. The hallmark of a PCP, especially true in family medicine, is the continuity and quality of care they provide to their routine patients. As Joffe and Truog state, the fundamental difference between these two distinct relationships suggests that the obligations of investigators to obtain informed consent are in fact *more* demanding than the lenient discussion in the doctor's office when patients receive care [97].

Proponents of waiving informed consent for CER and similar studies may also argue that the average patient cannot autonomously authorize their own medical decisions. Faden and Beauchamp acknowledge that "many decisions about routine and low-risk aspects of the patient's medical care [should] remain the exclusive province of the physician" [98]. Supporting this claim, Braddock et al describe in a finding from their large study of patient-physician encounters that PCPs "infrequently had complete discussions of clinical decisions with their patients" [99]. If the ethical model of decision making, as Braddock et al call it [99], is not a prominent occurrence in a primary care office setting, then it makes sense to question why the practice of "minimized" informed consent experienced in care, known as "passive acquiescence", cannot be translated to research of similar nature, such as CER [97].

We argue that passive acquiescence should not be the acceptable precursor to enrolment in a study, even if the same practice of passive acquiescence is routine in a primary care setting when it comes to the same interventions. As Joffe et al also posit, when patients rely on physicians to make decisions for them, they typically view those decisions as not implicating

their values and beliefs but instead decisions that are best informed by medical expertise [97]. This physician-patient relationship is built on the foundation of the physician's fiduciary duty to the patient to provide optimal medical care and this relationship facilitates trust and reliance. However, in research, the physician is an investigator, not a fiduciary agent for the patient-subject, since the ends of research and medical care differ greatly [97]. As previously noted, an investigator holds different obligations to society and the patient, who is now the research subject, by acting to generate evidence that researchers have good reason to believe will benefit future patients and inform the expert medical community. All these factors help stress how crucial informed consent is for a pragmatic trial such as CER, where patients need to be shaken from their complacency to realize they are committing to be research subjects.

All investigators are obligated to conduct ethical research, which includes unbiased studies, genuine and honest research outcomes and reporting of these unbiased results in a timely manner. To ensure transparency, FDAAA 801 mandated registration of studies with the expectation that the results of their registered studies be published in a timely manner since specific details of the research will already be available in a publicly accessible database [41].

Although we would like to encourage honest and complete reporting of CER outcomes, at present, a crucial addition to the CER landscape would be mandatory informed consent. Not only does a lack of informed consent lead to the ethical shortcomings explained above but these faults are amplified when results fail to be published. A dismal publication rate of 54% among primary care CER trials indicates that only around half of the patients who participated in these randomized trials had their rights redeemed. The other half of unpublished studies withheld important research outcomes, and if informed consent was waived before enrolling these subjects then they were not only deceived into participating in research, but their participation did not

make any meaningful contribution to society [60]. The common argument defending waiving informed consent for CER - claiming that interventions are similar to the interventions offered in routine medical practice under minimally informed acquiescence – does not hold for CER as general outcomes of research differ from medical care as well as the specific outcomes of CER differ from other types of research.

Thus, waiving informed consent in an era where nonpublication is common is not advisable. In particular, the objective of CER is unique from that of other research, which is to generate outcomes that inform policy and medical decisions within the doctor's office and other EBM initiatives. When research is registered, as FDAAA 801 requires for certain studies, and is not published within a respectable time frame, the CER study was not justified and patient-subject rights were not honored.

When a CER trial goes unpublished, or when any study goes unpublished, the research did not honour its commitment to the biomedical research enterprise but ethical safeguards such as IRB review and informed consent support the patient-subject's rights to some regard. Yet, in CER if a study goes unpublished, the entire trial becomes ethically unacceptable. The moral sustenance of randomizing patients in a CER study is drawn from the investigators believing that clinical equipoise has been met and the study is reasonably needed to inform policy and/or clinical decisions by uncovering evidence that is directly applicable to real world use of the interventions tested [11,15,37]. None of the above goals can be met from an unpublished study where research outcomes were never widely disseminated. Without informed consent – the patient-subject was unknowingly placed in research that never fulfilled its reasons for being conducted in the first place. To conclude, it is agreed that the interventions tested in CER do not pose the same absolute risks as those tested in novel studies as well as requesting informed

consent of all subjects may induce phenomena such as “mission creep” and other unnecessary obstacles to pursue research [100]. Although some of these “obstacles” like informed consent may be impossible to obtain in certain situations, investigators must always be reminded that the ends of medicine and research differ, and nonpublication exists, to the extent that the protection of informed consent is necessary whenever possible.

3. Limitations

A number of limitations were discussed in Chapter III within the manuscript and methodological flaws were discussed at length in Chapter IV, but here we will review overall limitations. Lack of education and awareness in the expert medical community regarding comparative effectiveness trials and what constitutes CER was the major growing pain of this thesis. Many researchers incorrectly used the term “efficacy” when referring to what was a comparative effectiveness design and certain investigators employed the term “CER” when no such research design was used for the study titled so. Several mis-labeled studies (see **Table 1**, **Chapter IV**) had errors in the title of the registration record and/or the matching publication – which may add to the propagation and continuation of the erroneous understanding of CER, as readers often notice the title of the research first [101]. It is important to acknowledge that studies identified in **Chapter IV** as mislabeled CER trials could in fact be poorly designed CER trials. Previous authors report post-approval studies lacking active comparators, poor choice of clinical endpoints as well as lack of adequate blinding – resulting in CER trials that resemble the pre-approval research that they are designed to supplement [65]. The difficulty identifying whether to include poorly designed CER trials or categorize them as mislabeled CER, added additional complexity to data collection. Other mentionable limitations include having only myself as a data extractor with a secondary review from a clinician. Also, principal

investigator(s) were not contacted to confirm if a publication matching their study registration existed. Since search strategy was limited to English language results only, those that published non-English articles may have been missed.

Moreover, this thesis only assessed non-publication of registered trials, which leaves primary care CER trials that were not registered unaccounted for in the publication rate of 54%. It is likely that investigators who did not voluntarily register their studies still pursued publication, resulting in possible deflation compared to the publication rate of all CER trials relevant to primary care. A better approach to assessing publication rate would have been to track protocols submitted to a local ethics committee to determine what percentage resulted in a full-text publication within a reasonable timeframe, identified in this thesis as "within four years". Lastly, although this thesis discusses publication bias as an unethical research practice, the empirical assessment of publication rate alone is descriptive and cannot draw any association to whether publication bias occurred. There are multiple reasons for non-publication, which were discussed in ***Chapter III***, within the manuscript limitations, including editorial decisions at journals, professional pressures suppressing null findings, and so forth.

4. Future Directions for CER? The Complicated World of "Standard of Care" Research

As reflected in Chapter II, measuring comparative effect of interventions that healthcare providers view as effective is a concept that holds great promise for informing clinical decisions with "real world" evidence instead of research conducted in highly controlled settings [3,11,15]. This head-to-head comparison of standard of care therapies helps identify which one is better and for whom, an approach that should not only result in better outcomes for patients but will also limit the use of ineffective therapies brought into routine practice prior to consulting robust evidence [11]. However, what happens when circumstances arise that restrict the use of a

standard of care? What is the value of evidence in the face of depleted resources? The only way to benefit from an intervention that is supported by high quality evidence is for there to be enough of it for the subset of the population that it is comparatively better for. Yet, in certain situations, healthcare systems must deviate from standard of care interventions, no matter how well-established they are, due to limited supply. This unveils another dimension to consider for the CER trial.

CER evidence surrounds research into endpoints that are not requested by the FDA, clinical endpoints that subgroups of patients not involved in pre-approval studies may benefit from as well as endpoints that physicians find worthwhile to inform optimal clinical decisions [11]. It could be argued that investigators planning a CER trial should ask the question "*is my healthcare system in a position to implement the evidence this trial serves to generate?*". This would be in addition to ensuring that clinical equipoise exists amongst the expert medical community regarding which approved intervention works better and for whom. From a theoretical perspective, it makes sense to have a robust toolbox of research outcomes to make data-driven decisions from for comparable treatment options, but the SARS-CoV-2 pandemic has taught us that this is not always practical.

If a treatment is in low supply, should we be conducting a CER trial? A CER trial, like any other RCT, randomizes human subjects, disenfranchises them from their right to analogous civilian freedoms and leaves them at the will of study design – as mentioned earlier, sometimes even without informed consent [102]. A CER trial also takes significant resources to conduct, time and effort of healthcare staff as well as patients and disrupts genuine routine practice in hopes of boosting external validity of results [26]. CER exists to generate evidence to improve

patient and provider outcomes, thereby limiting the "gut-driven" decisions that take place in healthcare when seemingly comparable standard of care interventions exist.

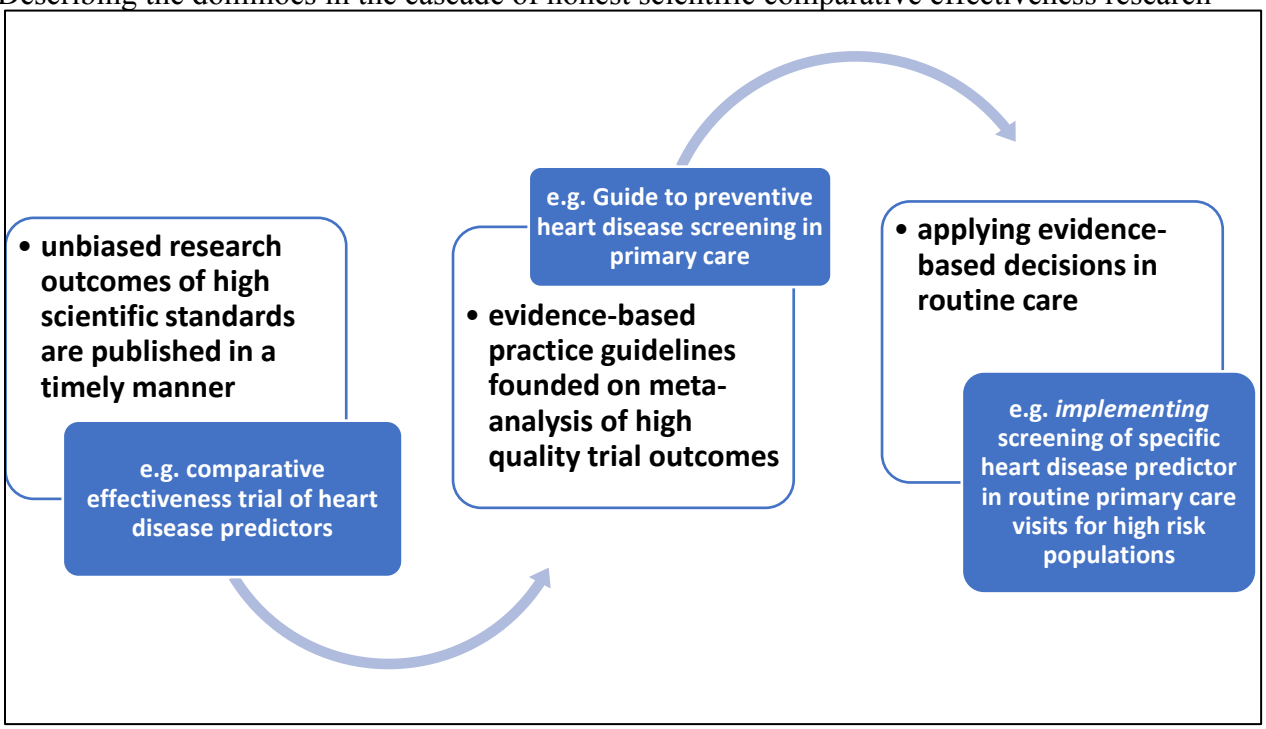
When crises occur, the availability of data to inform clinical decisions that yield optimal patient outcomes is not enough without reasonable resources to allocate to the individuals for whom those CER trials were conducted. For instance, before the pandemic and the global shortage in Personal Protective Equipment (PPE) [103], a joint initiative of multiple academic medical centers in the United States conducted a pragmatic, cluster randomized comparative effectiveness trial. Their aim was to assess comparative effect of medical (surgical) masks versus N95 respirators in preventing healthcare providers from contracting viral respiratory infections, or "flu-like" illnesses. The authors acknowledge that "pragmatic effectiveness trials are increasingly recognized as an essential component of medical evidence" [104], highlighting the need for their study. The CER trial concluded that there was no significant difference in laboratory-confirmed influenza between the two PPE methods, despite the N95 respirator's superior ability to filter out small airborne particles [105]. The study's primary completion date was March 2018, results were made available on CT.gov by April 25, 2019 and the full-text publication took an additional 4-5 months [106] – far less than the median 25 months elapsed from completion to publication observed for CER trials in Chapter III. This illustrates the concern that CER generating evidence for data-driven decisions should also be supported by proactive approaches to secure resources as determined by study outcomes.

In 2007, the IOM found only a minority of the care provided was backed up by evidence [107]. This finding helped emphasize the essential role CER could play in helping address the need for better and more informed clinical decision-making. Further, Fisher et al. found that although patients in certain "high-spend" regions of the United States received an average of

60% more healthcare services compared to patients in other "low-spend" regions, the extra care did not translate to better outcomes [108]. A plausible reason could be that multiple ineffective interventions were recommended to patients, increasing costs not justified by an improvement of health outcomes in exchange [109]. More recently, the American Heart Association issued evidence-based practice guidelines that were based on routine standard of care, expert medical opinion and anecdotal evidence rather than outcomes of RCTs [5]. There exist other instances where the quality of evidence comprising practice guidelines was lacking [5,6]. However, even if the push toward CER does have the ability to add "good" unbiased and timely published research outcomes into the healthcare evidence base, correcting what we've seen in the past, without the translation of well-supported EBM guidelines to actual clinical care, the CER initiative will not proceed to fruition.

The crisis of the PPE shortage has emphasized that conducting research of high scientific standards, transparent reporting and timely publication signal a mere slice of good research – the whole pie is complete only through implementation. Thus, not acting on evidence-based guidelines in a timely manner has demonstrated economic, political and medical implications for everyone. Future avenues for assessing ethical implications of CER trials beyond publication rate analyses include following-up on not only if practice guidelines are supported by good evidence but assessing whether crucial evidence-based guidelines from CER trial outcomes are actually *translated into clinical practice* as illustrated in **Figure 2.** This figure depicts the big picture behind recognizing the implications of the non-publication of results: Research should be **1)** registered and **published** since they help **2)** inform evidence-based practice guidelines, which are only useful if they are **3)** implemented in real-world and/or routine care settings.

Figure 2:
Describing the dominoes in the cascade of honest scientific comparative effectiveness research



Conclusion

In summary, this thesis evaluated the publication patterns of comparative effectiveness trials that had the potential to be advantageous to primary care when published. An alarming publication rate of only 54% was observed, indicating that valuable research outcomes are being withheld from society, an action that has repercussions not only for the integrity of the research enterprise but for clinicians and patients who benefit from those findings. Within the overall biomedical research landscape, non-publication is not new, and many authors have described the problem of "publication bias" that results from it [4,8]. However, the CER trial is a stark contrast to the typical clinical trial that comes to mind, considering that the purpose of evaluating comparative effects is not to obtain treatment approval or to determine the safety of a therapy but simply to generate evidence of a routine intervention to better inform decision-making [15,21].

We have argued that the non-publication seen in CER trials disadvantaged primary care and its patient-subjects to a greater degree than non-publication of other types of clinical research. This is because the non-publication of a CER trial questions the moral permissibility of conducting it in the first place – as the research outcomes, in other words, the evidence, the trial was conducted for were never disseminated to the public. The ultimate goal is for patients to receive the best care possible and publishing evidence that directly informs clinical decisions is a step towards achieving that goal. Our hope is that better and more stringent regulations will be put into effect to ensure that CER trials are not only registered but also published within a reasonable period of time.

APPENDIX

Table 3: List of Variables for Data Collection

Study Details	**Intervention and Study Population**	**Study Outcome**
Study Name	Purpose	Matching Publication (Y/N)
NCT Identifier	Type of Intervention	Publication Date
Location	PC Category (Study Condition)	Publication Year
Start Date	Condition	Link to Publication
Primary Completion Date	Number of Patients Enrolled	Time to Publication (months)
Funding		Primary Outcome Reported
Sponsor		
Phase		
Study Design		
Masking/Blinding		

Selected Details from Final Thesis Dataset Provided Below:

Table 4: Selected Characteristics of CT.gov Records Published Within 4 Years

NCT	**Location**	**Purpose**	**Primary Care Study Condition**	**Type**	**sponsor**	**Blinding**	**Time until Published (months)**	**Significant Outcome Reported? (y/n)**
NCT01419184	United States	Treatment	Infectious Disease	Drug	industry	none	38	no
NCT01182493	United States	Treatment	Cardiovascular	Device	industry	none	5	yes
NCT00545753	United States	Treatment	dermatological	Drug	industry	triple	18	yes
NCT01294592	France	Treatment	Urological	Drug	industry	none	15	yes
NCT01184872	Germany	Treatment	Infectious Disease	Drug	industry	none	29	yes
NCT00711802	United States	Treatment	Infectious Disease	Drug	industry	single	41	no
NCT00538512	United States	prevention	Infectious Disease	other	industry	double	16	yes
NCT01014013	Korea	Treatment	Infectious Disease	Drug	industry	double	38	yes
NCT01911780	Japan	Treatment	Cardiovascular	Drug	industry	double	17	no

NCT01888536	Korea	Treatment	Pain	Drug	industry	quadruple	29	yes
NCT01696162	United States	Treatment	Pain	Behavioural	industry	single	38	no
NCT01524705	United States	Treatment	Cardiovascular	Drug	industry	none	13	no
NCT01378286	Brazil	Treatment	Infectious Disease	Drug	industry	none	40	yes
NCT01295580	China	Treatment	Orthopedics	Device	industry	quadruple	43	yes
NCT01131624	Australia	Treatment	hematology	Drug	industry	single	25	yes
NCT00964223	United States	Treatment	dermatological	Drug	industry	single	36	yes
NCT00640276	Korea	Treatment	Cardiovascular	Drug	industry	none	42	yes
NCT01347619	United States	Health Services	psychiatric	Behavioural	AHRQ	single	25	no
NCT02113631	United States	Treatment	Infectious Disease	Drug	US Fed	none	42	no
NCT02027636	United States	Treatment	psychiatric	other	NIH	none	48	no
NCT02108977	United States	Health Services	other	other	US Fed	none	33	yes
NCT01997723	United States	Diagnostic	respiratory	Device	NIH	single	28	no
NCT01670825	United States	Treatment	Neurological	Drug	US Fed	quadruple	10	no
NCT01614340	United States	Treatment	pain	Behavioural	US Fed	double	32	no
NCT01593111	United States	prevention	respiratory	Behavioural	US Fed	single	30	no
NCT01502891	United States	Health Services	psychiatric	Behavioural	US Fed	single	25	yes
NCT01495923	United States	Treatment	Pain	Drug	US Fed	triple	10	no
NCT01459783	United States	Health Services	psychiatric	Behavioural	NIH	single	16	no
NCT01418209	United States	Treatment	oncological	Drug	NIH	quadruple	18	yes
NCT01377857	United States	Screening	Infectious Disease	Behavioural	NIH	none	37	yes
NCT01344278	United States	Health Services	Cardiovascular	Behavioural	US Fed	double	16	yes
NCT01331304	United States	Treatment	psychiatric	Drug	AHRQ	none	33	no
NCT01306695	United States	Supportive Care	Neurological	Behavioural	NIH	none	17	yes
NCT01296906	United States	Health Services	other	Behavioural	NIH	single	18	yes

NCT01288612	United States	Screening	gastrointestinal	Device	NIH	none	14	no
NCT01277939	United States	Treatment	psychiatric	Behavioural	NIH	none	26	no
NCT01241656	United States	prevention	oncological	Behavioural	NIH	single	26	no
NCT01144104	United States	Screening	psychiatric	Behavioural	NIH	single	48	no
NCT01030419	United States	Treatment	other	Behavioural	NIH	none	15	no
NCT01142882	United States	prevention	other	Behavioural	NIH	none	47	yes
NCT02816866	United States	Health Services	oncological	Behavioural	Non-profit	double	34	no
NCT02561780	Canada	Health Services	psychiatric	Behavioural	academic	single	32	yes
NCT02339337	Taiwan	Treatment	Infectious Disease	other	Non-profit	none	17	no
NCT02278289	Taiwan	Treatment	other	Device	NGO	none	35	yes
NCT02167464	United States	Treatment	Cardiovascular	Drug	academic	single	15	no
NCT02063048	United States	Health Services	Cardiovascular	Behavioural	Non-profit	none	15	yes
NCT01792895	Spain	Treatment	pain	other	academic	double	26	yes
NCT01685853	Italy	Treatment	gastrointestinal	Drug	Non-profit	single	18	yes
NCT01637181	Netherlands	Treatment	dermatological	other	Non-profit	single	16	yes
NCT01587274	United States	Treatment	pain	Drug	Non-profit	quadruple	12	no
NCT01547221	Thailand	Treatment	other	Drug	Non-profit	double	17	yes
NCT01537510	United States	Treatment	Cardiovascular	Behavioural	Non-profit	double	22	yes
NCT01488955	Germany	Treatment	Urological	Drug	Non-profit	double	10	yes
NCT01476306	United States	Health Services	Neurological	Behavioural	academic	none	13	no
NCT01458457	Germany	Treatment	oncological	other	academic	none	25	yes

NCT01275378	United States	Health Services	psychiatric	Behavioural	non-profit	triple	10	yes
NCT01216761	United States	prevention	Infectious Disease	Drug	non-profit	double	38	no
NCT01051388	Japan	prevention	Cardiovascular	Drug	non-profit	single	21	yes
NCT01726803	United States	Treatment	Pain	other	Non-profit	double	23	yes
NCT02208492	Korea	Treatment	Neurological	Drug	academic	none	17	no
NCT02133716	Spain	Treatment	Pain	other	non-profit	quadruple	35	no
NCT01918449	Spain	Supportive Care	respiratory	Behavioural	non-profit	none	16	yes
NCT01653730	Canada	Supportive Care	respiratory	device	non-profit	none	32	yes
NCT01623570	Italy	Treatment	endocrinology	Drug	non-profit	none	5	no
NCT01370668	Spain	Treatment	psychiatric	Behavioural	non-profit	single	32	yes
NCT00857870	Germany	Treatment	Cardiovascular	Drug	non-profit	none	18	no

Table 5: Selected Characteristics of CT.gov Records with Abstract-only or Non-full-text Publications within 4 Years

NCT	Location	Purpose	Primary Care Study Condition	Type	sponsor	Blinding
NCT01187771	United States	Treatment	Cardiovascular	procedure	NIH	single
NCT01006967	United States	prevention	Neurological	Device	US Fed	none
NCT00983476	United States	Treatment	psychiatric	Behavioural	US Fed	single
NCT02701010	Turkey	Treatment	Urological	Behavioural	academic	none
NCT01524874	United States	Treatment	Cardiovascular	Supplement	Non-profit	single
NCT01461473	United States	Treatment	respiratory	Device	academic	none
NCT01440530	United States	Health Services	Cardiovascular	Behavioural	non-profit	none
NCT01409889	United States	prevention	Cardiovascular	Behavioural	academic	none
NCT01053273	United States	Treatment	Pain	procedure	non-profit	double

Table 6: Selected Characteristics of CT.gov Records with Full-text Publication *after* 4 Years

NCT	Location	Purpose	Primary Care Study Condition	Type	sponsor	Blinding
NCT01229735	Korea	Treatment	Neurological	Drug	industry	none
NCT01466673	Thailand	Treatment	dermatological	Drug	industry	single
NCT00852540	United States	Treatment	Infectious Disease	Drug	industry	double
NCT02010918	Brazil	Treatment	Orthopedic	Drug	industry	none
NCT01006967	United States	prevention	Neurological	Device	US Fed	none
NCT01391156	Thailand	Treatment	dermatological	Drug	non-profit	quadruple
NCT01048801	Uganda	Treatment	Infectious Disease	device	non-profit	none
NCT01123174	France	Supportive Care	psychiatric	Behavioural	academic	single

Table 7: Selected Characteristics of CT.gov Records Without any Full-text Publication ("Nonpublication")

NCT	Location	Purpose	Primary Care Study Condition	Type	sponsor	Blinding
NCT00999102	United States	Treatment	Cardiovascular	Drug	industry	quadruple
NCT01036438	Czech Republic	Treatment	dermatological	Device	industry	quadruple
NCT00742183	United States	Health Services	dermatological	Device	industry	none
NCT00926211	United States	Treatment	dermatological	Device	industry	single
NCT02130063	United States	Treatment	hematology	Drug	industry	none
NCT00718887	China	Treatment	Infectious Disease	Drug	industry	none
NCT01160198	India	Other	hematology	Supplement	industry	single
NCT00496834	Korea	Treatment	Cardiovascular	Drug	industry	none
NCT03335566	China	Diagnostic	nephrology	Drug	industry	none
NCT01915914	China	Treatment	dermatological	Drug	industry	none
NCT01682564	Korea	Treatment	Cardiovascular	Drug	industry	none
NCT01507922	Korea	Treatment	gastrointestinal	Drug	industry	quadruple
NCT01429701	Brazil	Treatment	dermatological	Drug	industry	quadruple
NCT01653951	United States	Health Services	Cardiovascular	Behavioural	NIH	double
NCT01424046	Rwanda	Treatment	Cardiovascular	Drug	NIH	none
NCT01244568	United States	Treatment	oncological	Behavioural	NIH	none

NCT01144923	United States	Treatment	pain	procedure	US Fed	none
NCT01817842	United States	prevention	respiratory	Behavioural	NIH	none
NCT02551536	India	Treatment	other	Drug	academic	double
NCT02260401	United States	Health Services	Pain	Behavioural	Non-profit	none
NCT02108535	Brazil	Treatment	dermatological	Drug	academic	triple
NCT01964417	Korea	unknown	gastrointestinal	Drug	academic	single
NCT01885481	Korea	Treatment	Pain	Drug	academic	triple
NCT01877018	Spain	Screening	gastrointestinal	Behavioural	Non-profit	single
NCT01697826	India	Treatment	OB/GYN	Drug	Non-profit	none
NCT01670864	China	Treatment	respiratory	Behavioural	academic	none
NCT01656876	Taiwan	Treatment	Cardiovascular	Behavioural	academic	single
NCT01550718	United States	Treatment	Neurological	Behavioural	academic	single
NCT01514513	United States	Treatment	Infectious Disease	Drug	Non-profit	none
NCT01506310	Italy	Treatment	Cardiovascular	Behavioural	academic	none
NCT01456494	United States	other	respiratory	other	academic	single
NCT01416766	United States	Health Services	Cardiovascular	other	academic	single
NCT01293578	United States	Supportive Care	Cardiovascular	Behavioural	non-profit	double
NCT01109797	United States	Treatment	Cardiovascular	Behavioural	academic	none
NCT00800462	Canada	Treatment	Neurological	Drug	non-profit	quadruple
NCT00799279	Canada	prevention	respiratory	Behavioural	academic	none
NCT01482338	Thailand	Treatment	OB/GYN	Drug	academic	single
NCT01170793	France	prevention	Neurological	Behavioural	non-profit	none
NCT01129999	Germany	Treatment	respiratory	Behavioural	academic	single

Thesis References

1. Patsopoulos NA. A pragmatic view on pragmatic trials. Dialogues Clin Neurosci. 2011;13(2):217–24.
2. Soumerai SB, McLaughlin TJ, Avorn J. Improving Drug Prescribing in Primary Care: A Critical Analysis of the Experimental Literature. Milbank Q. 2005 Dec;83(4):10.1111/j.1468-0009.2005.00435.x.
3. Simon GE, Platt R, Hernandez AF. Evidence from Pragmatic Trials during Routine Care — Slouching toward a Learning Health System. N Engl J Med. 2020 Apr 15;382(16):1488–91.
4. DeVito NJ, Goldacre B. Catalogue of bias: publication bias. BMJ EBM. 2019 Apr 1;24(2):53.
5. Tricoci, P., J. M. Allen, J. M. Kramer, R. M. Califf, and S. C. Smith, Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA. 2009. 301(8):831-841.
6. Harpole, L. H., M. J. Kelley, G. Schreiber, E. M. Toloza, J. Kolimaga, and D. C. McCrory. Assessment of the scope and quality of clinical practice guidelines in lung cancer. Chest. 2003. 123(1 SUPPL.):7S-20S.
7. Norris SL, Holmer HK, Ogden LA, et al. Selective Outcome Reporting as a Source of Bias in Reviews of Comparative Effectiveness [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Aug. Background. Available from: https://www.ncbi.nlm.nih.gov/books/NBK100617/
8. Mlinarić A, Horvat M, Šupak Smolčić V. Dealing with the positive publication bias: Why you should really publish your negative results. Biochem Med (Zagreb). 2017 Oct 15;27(3):030201–030201.
9. Elsevier. Why it's time to publish research "failures" [Internet]. Elsevier Connect. [cited 2020 Jun 4]. Available from: https://www.elsevier.com/connect/scientists-we-want-your-negative-results-too
10. Nijhawan LP, Janodia MD, Muddukrishna BS, Bhat KM, Bairy KL, Udupa N, et al. Informed consent: Issues and challenges. J Adv Pharm Technol Res. 2013 Jul;4(3):134–40.
11. Institute of Medicine (U.S.). Committee on Comparative Effective Research Prioritization. Initial national priorities for comparative effectiveness research [Internet]. Washington, D.C.: National Academies Press; 2009. Available from: http://site.ebrary.com/id/10347043
12. Castro M. Placebo versus best-available-therapy control group in clinical trials for pharmacologic therapies: which is better? Proc Am Thorac Soc. 2007 Oct 1;4(7):570–3.
13. Joffe S, Volume: 36 AWM-J 2014, Issue: 3. Determining Minimal Risk for Comparative Effectiveness Research [Internet]. The Hastings Center. [cited 2020 May 23]. Available from: https://www.thehastingscenter.org/irb_article/determining-minimal-risk-for-comparative-effectiveness-research/
14. Smelt AF, Weele GM van der, Blom JW, Gussekloo J, Assendelft WJ. How usual is usual care in pragmatic intervention studies in primary care? An overview of recent trials. Br J Gen Pract. 2010 Jul 1;60(576):e305–18.
15. Sox HC. Defining comparative effectiveness research: the importance of getting it right. Med Care. 2010 Jun;48(6 Suppl):S7-8.
16. Abraham I, Sally Reel EA. A Definition of Comparative Effectiveness Research. 2020 May 29 [cited 2020 Jun 4]; Available from: https://www.bmj.com/rapid-response/2011/11/03/definition-comparative-effectiveness-research
17. Cipriani A, Ioannidis JPA, Rothwell PM, Glasziou P, Li T, Hernandez AF, et al. Generating comparative evidence on new Drugs and devices after approval. The Lancet. 2020 Mar 21;395(10228):998–1010.

18. Singal AG, Higgins PDR, Waljee AK. A primer on effectiveness and efficacy trials. Clin Transl Gastroenterol. 2014 Jan 2;5(1):e45–e45.
19. Conway PH, Clancy C. Comparative-effectiveness research -implications of the Federal Coordinating Council's Report. N Engl J Med. 2009;361:328-330.
20. Slutsky JR, Clancy CM. AHRQ's Effective Health Care Program: Why Comparative Effectiveness Matters. Am J Med Qual. 2008 Dec 5;24(1):67–70.
21. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol. 2009 May;62(5):464–75.
22. Bourgeois FT, Murthy S, Mandl KD. Comparative Effectiveness Research: An Empirical Study of Trials Registered in ClinicalTrials.gov. PLOS ONE. 2012 Jan 9;7(1):e28820.
23. Wasan AD. Efficacy vs Effectiveness and Explanatory vs Pragmatic: Where Is the Balance Point in Pain Medicine Research? Pain Medicine. 2014;15(4):539–40.
24. Chalkidou K, Tunis S, Whicher D, Fowler R, Zwarenstein M. The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. Clin Trials. 2012 Aug;9(4):436–46.
25. Mullins CD, Sanchez RJ. It is important to note that RWD will never replace the more traditional and more robust RCT data; however, the emerging trend is to incorporate data that are more generalizable. Introduction. J Manag Care Pharm. 2011 Dec;17(9 Suppl A):S03-04.
26. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol. 2003 Dec 22;3:28–28.
27. Primary Health Care, Main Terminology [Website]. World Health Organization; 2004 [Available from: http://www.euro.who.int/en/health-topics/Health-systems/primary-healthcare/main-terminology#.]
28. Chapman SJ, Shelton B, Mahmood H, Fitzgerald JE, Harrison EM, Bhangu A. Discontinuation and non-publication of surgical randomised controlled trials: observational study. BMJ. 2014 Dec 9;349:g6870.
29. Demange MK, Fregni F. Limits to clinical trials in surgical areas. Clinics (Sao Paulo). 2011 Jan;66(1):159–61.
30. Ashton CM, Wray NP, Jarman AF, Kolman JM, Wenner DM, Brody BA. Ethics and methods in surgical trials. J Med Ethics. 2009 Sep;35(9):579–83.
31. Ford I, Norrie J. Pragmatic Trials. N Engl J Med. 2016 Aug 3;375(5):454–63.
32. Fransen GAJ, van Marrewijk CJ, Mujakovic S, Muris JWM, Laheij RJF, Numans ME, et al. Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. BMC Med Res Methodol. 2007 Apr 23;7:16–16.
33. Naci H, Salcher-Konrad M, Kesselheim AS, Wieseler B, Rochaix L, Redberg RF, et al. Generating comparative evidence on new Drugs and devices before approval. The Lancet. 2020 Mar 21;395(10228):986–97.
34. The ischemic optic neuropathy decompression trial (IONDT): design and methods. Control Clin Trials. 1998 Jun;19(3):276–96.
35. Hilal T, Sonbol MB, Prasad V. Analysis of Control Arm Quality in Randomized Clinical Trials Leading to Anticancer Drug Approval by the US Food and Drug Administration. JAMA Oncol. 2019 Jun 1;5(6):887–92.

36. ORI News Plus FDA Suggestions on Comparative Effectiveness Research [Internet]. Policy & Medicine. [cited 2020 Jun 11]. Available from: https://www.policymed.com/2013/02/ori-news-plus-skepticism-on-comparative-effectiveness-research-at-the-fda.html
37. Sox HC, Goodman SN. The methods of comparative effectiveness research. Annu Rev Public Health. 2012 Apr;33:425–45.
38. THOMAS CE. If I Lose Mine Honour, I Lose Myself: Honour among the Early Modern English Elite [Internet]. University of Toronto Press; 2017 [cited 2020 Jun 8]. Available from: https://www.jstor.org/stable/10.3138/j.ctt1v2xsd0
39. Ioannidis JPA. Why Most Clinical Research Is Not Useful. PLoS Med. 2016 Jun 21;13(6):e1002049–e1002049.
40. ICMJE | Recommendations | Clinical Trials [Internet]. [cited 2020 Jun 11]. Available from: http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html
41. FDAAA 801 and the Final Rule - ClinicalTrials.gov [Internet]. [cited 2020 Jun 11]. Available from: https://clinicaltrials.gov/ct2/manage-recs/fdaaa
42. Gluud LL. Bias in Clinical Intervention Research. Am J Epidemiol. 2006 Mar 15;163(6):493–501.
43. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. The Lancet. 2014 Jan;383(9913):267–76.
44. Fanelli D. How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. PLOS ONE. 2009 May 29;4(5):e5738.
45. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes. JAMA. 2010 May 26;303(20):2058–64.
46. Marco CA, Larkin GL. Research Ethics: Ethical Issues of Data Reporting and the Quest for Authenticity. Academic Emergency Medicine. 2000;7(6):691–4.
47. Bekelman JE, Li Y, Gross CP. Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review. JAMA. 2003 Jan 22;289(4):454–65.
48. Janiaud P, Cristea I-A, Ioannidis JPA. Industry-funded versus non-profit-funded critical care research: a meta-epidemiological overview. Intensive Care Med. 2018 Oct 1;44(10):1613–27.
49. Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. BMJ. 2007 Dec 6;335(7631):1202.
50. Milette K, Roseman M, Thombs BD. Transparency of outcome reporting and trial registration of randomized controlled trials in top psychosomatic and behavioral health journals: A systematic review. Journal of Psychosomatic Research. 2011 Mar 1;70(3):205–17.
51. Doshi P, Goodman SN, Ioannidis JPA. Raw data from clinical trials: within reach? Trends Pharmacol Sci. 2013 Dec;34(12):645–7.
52. WMA - The World Medical Association-WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects [Internet]. [cited 2020 Jun 13]. Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/
53. Rees CA, Pica N, Monuteaux MC, Bourgeois FT. Noncompletion and nonpublication of trials studying rare diseases: A cross-sectional analysis. PLoS Med. 2019 Nov 21;16(11):e1002966–e1002966.

54. Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF. Non-publication of large randomized clinical trials: cross sectional analysis. BMJ. 2013 Oct 29;347:f6104.
55. Al-Durra M, Nolan RP, Seto E, Cafazzo JA, Eysenbach G. Nonpublication Rates and Characteristics of Registered Randomized Clinical Trials in Digital Health: Cross-Sectional Analysis. Journal of Medical Internet Research. 2018;20(12):e11924.
56. Pica N, Bourgeois F. Discontinuation and Nonpublication of Randomized Clinical Trials Conducted in Children. PEDIATRICS. 2016 Sep 1;138(3):e20160223–e20160223.
57. Yilmaz T, Gallagher MJ, Cordero-Coma M, Gealy D, Loporchio S, Padula WV, et al. Discontinuation and nonpublication of interventional clinical trials conducted in ophthalmology. Can J Ophthalmol. 2020 Feb;55(1):71–5.
58. Bikdeli Behnood, Welsh John W., Akram Yasir, Punnanithinont Natdanai, Lee Ike, Desai Nihar R., et al. Noninferiority Designed Cardiovascular Trials in Highest-Impact Journals. Circulation. 2019 Jul 30;140(5):379–89.
59. Gyawali B, Tessema FA, Jung EH, Kesselheim AS. Assessing the Justification, Funding, Success, and Survival Outcomes of Randomized Noninferiority Trials of Cancer Drugs: A Systematic Review and Pooled Analysis. JAMA Netw Open. 2019 Aug 2;2(8):e199570–e199570.
60. Is There a Blind Spot in the Oversight of Human Subject Research? [Internet]. leapsmag. 2018 [cited 2020 Jun 8]. Available from: https://leapsmag.com/is-there-a-blind-spot-in-the-oversight-of-human-subject-research/
61. Nicholls SG, Carroll K, Zwarenstein M, Brehaut JC, Weijer C, Hey SP, et al. The ethical challenges raised in the design and conduct of pragmatic trials: an interview study with key stakeholders. Trials. 2019 Dec 23;20(1):765.
62. Clinical Research Versus Medical Treatment | FDA [Internet]. [cited 2020 Jun 13]. Available from: https://www.fda.gov/patients/clinical-trials-what-patients-need-know/clinical-research-versus-medical-treatment
63. Sox HC. Comparative Effectiveness Research: A Report From the Institute of Medicine. Ann Intern Med. 2009 Aug 4;151(3):203.
64. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. (Editorial). *BMJ* 1996;312:71–2.
65. Characteristics of preapproval and postapproval studies for Drugs granted accelerated approval by the US Food and Drug Administration. JAMA, 318 (2017), pp. 626-636
66. Singh I, Naci H, Miller J, Caplan A, Cipriani A. Ethical implications of poor comparative effectiveness evidence: obligations in industry-research partnerships. The Lancet. 2020 Mar 21;395(10228):926–8.
67. Dunn AG, Coiera E. Should comparative effectiveness research ignore industry-funded data? Journal of Comparative Effectiveness Research. 2014 Jul 1;3(4):317–20.
68. Buesching DP, Luce BR, Berger ML. The role of private industry in pragmatic comparative effectiveness trials. J Comp Eff Res. 2012 Mar;1(2):147–56.
69. Leon AC. Comparative Effectiveness Clinical Trials in Psychiatry: Superiority, Non-inferiority and the Role of Active Comparators. J Clin Psychiatry. 2011 Mar;72(3):331–40.
70. Angeli F, Verdecchia P, Vaudo G, Masnaghetti S, Reboldi G. Optimal Use of the Non-Inferiority Trial Design. Pharm Med. 2020 Jun 1;34(3):159–65.
71. Pocock SJ. The pros and cons of noninferiority trials. Fundamental & Clinical Pharmacology. 2003 Aug 1;17(4):483–90.

72. Apter AJ. Understanding adherence requires pragmatic trials: lessons from pediatric asthma. JAMA Pediatr. 2015 Apr;169(4):310–1.
73. Uschner D, Hilgers R-D, Heussen N. The impact of selection bias in randomized multi-arm parallel group clinical trials. PLoS One. 2018 Jan 31;13(1):e0192065–e0192065.
74. Coeytaux RR, Park JJ. Acupuncture research in the era of comparative effectiveness research. Ann Intern Med. 2013 Feb 19;158(4):287–8.
75. Prayle AP, Hurley MN, Smyth AR. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. BMJ. 2012 Jan 3;344:d7373.
76. Saito H, Gill CJ. How Frequently Do the Results from Completed US Clinical Trials Enter the Public Domain? - A Statistical Analysis of the ClinicalTrials.gov Database. PLOS ONE. 2014 Jul 15;9(7):e101826.
77. The Lancet. Prioritising primary care in the USA. The Lancet. 2019 Jul 27;394(10195):273.
78. Tunis SR, Stryer DB, Clancy CM. Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy. JAMA. 2003 Sep 24;290(12):1624–32.
79. Shepshelovich D, Goldvaser H, Wang L, Abdul Razak AR. Comparison of published and unpublished phase I clinical cancer trials: an analysis of the CliniclTrials.gov database. Invest New Drugs. 2018 Oct;36(5):933–8.
80. Frequently Asked Questions - ClinicalTrials.gov [Internet]. [cited 2020 Jun 14]. Available from: https://clinicaltrials.gov/ct2/manage-recs/faq#fr_30
81. Federal Register :: Clinical Trials Registration and Results Information Submission [Internet]. [cited 2020 Jun 14]. Available from: https://www.federalregister.gov/documents/2016/09/21/2016-22129/clinical-trials-registration-and-results-information-submission
82. Why Should I Register and Submit Results? - ClinicalTrials.gov [Internet]. [cited 2020 Jun 14]. Available from: https://clinicaltrials.gov/ct2/manage-recs/background#WhyRegister
83. Bennett JDM FRC, Miller TAM FRC, Richards RSM FRCSC. The Use of Botox® in Interventional Radiology. Techniques in Vascular and Interventional Radiology. 2006;9(1):36–9.
84. Selvig D, Sewell JL, Tuot DS, Day LW. Gastroenterologist and primary care perspectives on a post-endoscopy discharge policy: impact on clinic wait times, provider satisfaction and provider workload. BMC Health Serv Res. 2018 Jan 10;18(1):16–16.
85. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000 Feb 29 - Identifier NCT01976494, The Efficacy and Safety of Accufuser Omnibus®; Comparative Clinical Trial; 2013 Nov 5 [cited 2020 Jan 18]; [about 3 screens]. Available from: https://clinicaltrials.gov/ct2/show/NCT01976494
86. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000 Feb 29 - Identifier NCT01964417, The Comparative Study Between Bowel Preparation Method; 2013 Oct 17 [cited 2020 Jan 18]; [about 3 screens]. Available from: https://clinicaltrials.gov/ct2/show/NCT01964417
87. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000 Feb 29 - Identifier NCT01654887, Comparative Effectiveness of Lung Ultrasound vs. Chest X-ray for the Diagnosis of Pneumonia in the Emergency Department; 2012 Aug 1 [cited 2020 Jan 18]; [about 3 screens]. Available from: https://clinicaltrials.gov/ct2/show/NCT01964417
88. Glossary of Common Site Terms - ClinicalTrials.gov [Internet]. [cited 2020 Jun 15]. Available from: https://clinicaltrials.gov/ct2/about-studies/glossary

89. Schmulson M, Corazziari E, Ghoshal UC, Myung S-J, Gerson CD, Quigley EMM, et al. A four-country comparison of healthcare systems, implementation of diagnostic criteria, and treatment availability for functional gastrointestinal disorders. Neurogastroenterology & Motility. 2014 Oct 1;26(10):1368–85.
90. 25 median months to pub left off on pg 75
91. Ojji DB, Lamont K, Ojji OI, Egenti BN, Sliwa K. Primary care in the prevention, treatment and control of cardiovascular disease in sub-Saharan Africa. Cardiovasc J Afr. 2017;28(4):251–6.
92. McGrath ER, Glynn LG, Murphy AW, Conghaile AO, Canavan M, Reid C, et al. Preventing cardiovascular disease in primary care: Role of a national risk factor management program. American Heart Journal. 2012 Apr 1;163(4):714–9.
93. Wallace ML, Ricco JA, Barrett B. Screening strategies for cardiovascular disease in asymptomatic adults. Prim Care. 2014 Jun;41(2):371–97.
94. Faden R, Beauchamp T, Kass N. Informed Consent, Comparative Effectiveness, and Learning Health Care. New England Journal of Medicine. 2014;370(8):766-768.
95. Sox H, Rennie D. Seeding Trials: Just Say "No". Annals of Internal Medicine. 2008;149(4):279.
96. Parchman ML, Burge SK. The Patient-Physician Relationship, Primary Care Attributes, and Preventive Services. Family Medicine. :6.
97. Joffe, Steven, and Robert D. Truog. 2010. Consent to medical care: the importance of fiduciary context. In *The Ethics of Consent*, F. G. Miller and A. Wertheimer (eds.), New York: Oxford University Press.
98. Faden RR, Beauchamp TL, King NMP. A History and Theory of Informed Consent [Internet]. New York: Oxford University Press; 1986 [cited 2020 Apr 12]. Available from: http://search.ebscohost.com.proxy3.library.mcgill.ca/login.aspx?direct=true&db=nlebk&AN=150188&scope=site
99. Iii CHB, Edwards KA, Hasenberg NM, Laidley TL, Levinson W. Informed Decision Making in Outpatient Practice: Time to Get Back to Basics. JAMA. 1999 Dec 22;282(24):2313–20.
100. Reducing Research Waste | Bioethics Journal [Internet]. [cited 2020 Apr 22]. Available from: https://bioethicsjournal.hms.harvard.edu/summer-2018/reducing-research-waste
101. Tullu MS. Writing the title and abstract for a research paper: Being concise, precise, and meticulous is the key. Saudi J Anaesth. 2019 Apr;13(Suppl 1):S12–7.
102. Kim SYH, Miller FG. Informed Consent for Pragmatic Trials — The Integrated Consent Model. Hamel MB, editor. N Engl J Med. 2014 Feb 20;370(8):769–72.
103. Shortage of personal protective equipment endangering health workers worldwide [Internet]. [cited 2020 Apr 20]. Available from: https://www.who.int/news-room/detail/03-03-2020-shortage-of-personal-protective-equipment-endangering-health-workers-worldwide
104. Radonovich LJ Jr, Simberkoff MS, Bessesen MT, Brown AC, Cummings DAT, Gaydos CA, et al. N95 Respirators vs Medical Masks for Preventing Influenza Among Health Care Personnel: A Randomized Clinical Trial. JAMA. 2019 Sep 3;322(9):824–33.
105. Janssen L, Ettinger H, Graham S, Shaffer R, Zhuang Z. The use of respirators to reduce inhalation of airborne biological agents. J Occup Environ Hyg. 2013;10(8):D97–103.
106. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000 Feb 29 - Identifier NCT01249625, The Respiratory Protection Effectiveness Clinical Trial (ResPECT); 2010 Nov 30 [cited 2020 April 24]; [about 3 screens]. Available from: https://clinicaltrials.gov/ct2/show/NCT01249625

107. Institute of Medicine (IOM). 2007. Learning what works best: The nation's need for evidence on comparative effectiveness in health care [Internet]. [cited 2020 Apr 15]. Available from: http://www.iom.edu/ebm-effectiveness
108. Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder. 2003. The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care. Annals of Internal Medicine 138(4):288-298.
109. Fisher, E. S., J. P. Bynum, and J. S. Skinner. 2009. Slowing the growth of health care costs—Lessons from regional variation. *New England Journal of Medicine* 360(9):849-852.