26

Comparative Effectiveness Research

Soko Setoguchi[1] *and Ian Chi Kei Wong*[2,3]

[1] *Rutgers Robert Wood Johnson Medical School, Rutgers Center for Pharmacoepidemiology and Treatment Science, New Brunswick, NJ, USA*
[2] *Department of Pharmacology and Pharmacy, University of Hong Kong, Hong Kong*
[3] *UCL School of Pharmacy, London, UK*

Comparative Effectiveness Research in the US

Clinical and scientific communities have a long-standing desire to use scientific evidence to optimize clinical decisions for patients. While randomized controlled trials (RCTs) and meta-analyses of RCTs are generally considered to constitute the highest level of evidence, they have also been criticized for several aspects inherent to their design: comparison to placebo rather than alternative treatment; nonrepresentativeness of patient populations that tend to exclude older and multimorbid adults and children; controlled settings that differ from real-world care settings; relatively short follow-up; frequent use of surrogate endpoints rather than hard endpoints such as clinical events or death; and insufficient sample size to assess subgroup effects. Efforts to produce evidence that overcomes these limitations and is directly applicable to real-world patients as well as making clinical care more rational have been referred to at different times as *outcomes research*, *effectiveness research*, *evidence-based research*, *health technology assessment*, and, most recently, *comparative effectiveness research* (CER) [1]. To reduce the perception that the main agenda behind the push for CER is cost containment for healthcare, at least one government agency has begun relabeling CER as *patient-centered health research* [2].

CER is not a new concept, but has existed for the last several decades under various labels, and its popularity in the US had risen in response to several government initiatives. Earlier government initiatives for CER in the US were attempted first by the Congressional Office of Technology Assessment (established in 1972), then by the National Center for Health Care Technology (1978–1982), and then by the Agency for Health Care Policy and Research (established in 1989 and later renamed the Agency for Healthcare Research and Quality, AHRQ) [3]. The most recent impetus for CER came from the 2009 American Recovery and Reinvestment Act (ARRA Stimulus), with an appropriation of $1.1 billion "to study the comparative effectiveness of healthcare treatments" [4].

Furthermore, in 2010, the Patient Protection and Affordable Care Act (PPACA) authorized

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

the establishment of the Patient-Centered Outcomes Research Institute (PCORI) to carry out CER and improve its quality and relevance [5]. PPACA established new requirements for the Department of Health and Human Services (HHS) to disseminate findings from federally funded CER, including findings published by PCORI, and to coordinate with relevant federal health programs to build data capacity for this research. To fund CER activities, PPACA established the Patient-Centered Outcomes Research Trust Fund (PCORTF), from which PCORI and HHS are expected to receive an estimated $4 billion from fiscal years 2010 through 2019. As of November 2017, PCORI had disbursed more than $2 billion for approximately 600 CER-related projects [6].

CER in Europe and Other Countries

Europe

In recent years a number of European Union (EU) countries have introduced so-called health technology assessments (HTA). HTA includes not only assessment of clinical effectiveness, but cost-effectiveness as well [7]. Publicly funded healthcare systems are the main healthcare providers in a number of EU countries, and these systems are under substantial financial pressure to make the best use of available resources. Assessing cost-effectiveness as part of HTA is therefore critical in the evaluation of health technology.

The National Institute for Health and Care Excellence (NICE) in England and Wales, created in 1999, represents one model for using CER primarily to inform policy and practice, but also to develop research recommendations [8]. Since April 2013, NICE has gained new responsibilities for providing guidance to those working in social care. Accordingly, NICE guidance documents are used by the National Health Service (NHS), local government, employers, volunteer groups, and others involved in delivering care or promoting wellbeing [9]. NICE guidance takes several forms, including NICE guidelines, technology appraisals guidance, medical technologies, and diagnostics guidance, as described shortly.

NICE guidelines make evidence-based recommendations on a wide range of topics, from preventing and managing specific conditions, improving health, and managing medicines in different settings, to providing social care to adults and children and planning broader services and interventions to improve the health of communities. These guidelines aim to promote integrated care where appropriate. NICE has provided a substantial number of evidence-based guidelines for clinical practice [10], though not without controversy and challenge [11].

Technology appraisals guidance assesses the clinical and cost-effectiveness of health technologies, such as new pharmaceutical and biopharmaceutical products, but also procedures, devices, and diagnostic agents. For example, recent guidance recommends ixazomib with lenalidomide and dexamethasone for use within the Cancer Drugs Fund (a central funding source for cancer drugs in England) as an option for treating multiple myeloma in adults only if patients have already had two or three lines of therapy [12]. Technology appraisals guidance is intended to ensure that all NHS patients have equitable access to the most clinically and cost-effective treatments that are viable.

Medical technologies and diagnostics guidance helps to ensure that the NHS is able to adopt clinically and cost-effective technologies rapidly and consistently. For example, Neuropad is a technology that aims to detect preclinical diabetic peripheral neuropathy. However, its use is not supported by evidence [13]. Interventional procedures guidance provides recommendations on whether interventional procedures are effective and safe enough for use in the NHS. For example, NICE recently

recommended that intravesical microwave hyperthermia and chemotherapy for non-muscle-invasive bladder cancer be used only with special arrangements for clinical governance, consent, and audit or research [14].

For their evaluations, NICE's advisory committees use objective evidence provided by academic institutions in the UK, such as the Royal College of Physicians, under contract with NICE to perform evidence syntheses and to conduct small-scale studies entailing primary data collection [8]. The explicit use of cost-effectiveness data to evaluate and choose among medical interventions is viewed in the UK "as a tool to ensure fair shares for all in a resource-limited system," according to Chalkidou and Walley [8].

A six-country comparison by Sorenson [15] and a similar three-country comparison by Evans [16] illustrate the considerable efforts extended by European governments to incorporate CER into health policy decisions and the different approaches used for organizing these efforts. In France (the National Authority for Health – Haute Autorité de Santé or HAS [17]), Germany (the Institute for Quality and Efficiency in Healthcare – Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen or IQWiG [18]), and the Netherlands (Commissie Farmaceutische Hulp or CHF, Committee for Pharmaceutical Aid), the entities responsible for CER, act in an advisory role to the government, making recommendations on reimbursements and pricing. This is in contrast with the UK (NICE), Denmark (Reimbursement Committee of the Danish Medicines Agency or DKMA), and Sweden (Dental and Pharmaceutical Benefits Board or TLV), where the CER entities have regulatory authority and are directly responsible for prioritizing reimbursements for drug and devices [15,16]. Cost-effectiveness data are formally incorporated in evaluations and recommendations about coverage and pricing by most CER entities (UK, Germany, the Netherlands, and Sweden) [15].

Another five-country comparison by Levy *et al.* [19] that also includes Canada and Australia (the Pharmaceutical Benefits Advisory Committee, or PBAS [20]) noted that in each of the countries surveyed, the health technology evaluation committees (conceptually comparable to CER) retain their independence regarding decisions about which technologies are included in the formulary, despite receiving government funding. Members of these committees are primarily health professionals, with only Canada, Australia, and Scotland also including public representatives, and only Scotland permitting industry representation as well [19].

Other Countries

Healthcare systems and their financing mechanisms outside Western Europe and North America are very diverse, and it is impossible to comprehensively discuss the applications (or potential applications) of CER in all countries. According to Bloomberg Health Care Efficiency, Hong Kong was ranked the most efficient healthcare system in the world in 2017 and 2018 [21]. Hong Kong has a universal, publicly funded healthcare system which does not formally apply CER in decision-making. In parallel, Hong Kong also has a very well-developed private healthcare system, funded by insurance and patient out-of-pocket payments. The two sectors complement each other, in that the private sector is the major provider of primary healthcare, while the public sector is the predominant provider of secondary and tertiary healthcare services. About 70% of outpatient consultations are provided by the private sector, while over 90% of inpatient services (in terms of the number of bed-days) are provided by public hospitals. This system presents substantial difficulties for effectiveness research, as the two sectors cannot use the same CER for evaluation. At present, neither the public nor the private sector has adopted formal assessment of CER for evaluation of treatment. This raises important

questions regarding what factors beyond CER play significant roles in the efficiency of the healthcare system.

African countries in general are facing significant issues in healthcare financing and are struggling to provide sufficient publicly funded healthcare services. In South Africa, the National Department of Health includes a series of explicit references to HTA in a white paper setting out the government's 10-year vision for high-quality universal healthcare coverage. A dedicated taskforce has been set up to consider HTA and other tools in order to design high-quality, affordable packages of health services [22].

Patel *et al.* describe the healthcare and government environment and the use (and potential use) of CER to control healthcare expenditures in China, India, and South Korea [23]. This report demonstrates the diversity of the healthcare systems and potential uses of CER in these three countries. CER will clearly be of increasing importance to aid government agencies in healthcare resource allocation. While the use of CER by government agencies has been well established for a substantial period outside the US, much of the recent activity is occurring within the US, and that will be the primary focus of this chapter.

Efficacy vs. Effectiveness

A study of *treatment efficacy* investigates whether a drug *has the ability to* bring about a given intended effect in ideal (controlled) settings. For example, a drug efficacy study would be centered on the question: "In an ideal world, with perfect adherence, no interactions with other drugs or other diseases, etc., *could* the drug achieve its intended effects?" In contrast, a study of *treatment effectiveness* investigates whether, in real-world patients and settings, a treatment *in fact* achieves its desired effect. For example, a drug given in a controlled setting may be shown to reduce glucose levels in younger patients having no major co-morbidities, but it might not achieve good glucose control in older patients with heart failure if it causes even mild water retention that leads to nonadherence or premature discontinuation. To answer questions about effectiveness, studies need to include representative real-world patients and assess effectiveness in real-world care settings.

Definitions, Key Components, and Goals of CER

CER seeks to assist stakeholders, for example patients, clinicians, insurers, the medical products industry, and policymakers to make informed decisions to improve healthcare at both individual and population levels. Several definitions of CER have been proposed by US government and nongovernment organizations and are summarized in Table 26.1. In Europe, the term CER is not commonly used, but HTA describes similar though not identical research. Several definitions of HTA are also provided in Table 26.1. HTA as defined by the UK National Institute for Health Research (NIHR) is actually broader than CER, since it formally includes cost-effectiveness evaluation, whereas CER generally does not.

For CER to assist in clinical decision-making, it must include three key components: (i) evidence synthesis (identifying and summarizing already existing data addressing a question); (ii) evidence generation (creating new data addressing a question); and (iii) evidence dissemination (distributing the available data with the goal of informing healthcare decision-making). In other words, for some decisions, existing evidence from individual studies may be controversial or insufficient to support specific clinical decisions. In such cases, the evidence must be synthesized (evidence synthesis), which may then provide a sufficient basis to support the decision or identify knowledge gaps to guide

Table 26.1 Definitions of comparative effectiveness research and health technology assessment proposed by US and other government and nongovernment organizations.

Agency/report	Definition
Comparative effectiveness research (CER)	
US Congressional Budget Office report, December 2007 [24]	"A rigorous evaluation of the impact of different options that are available for treating a given medical condition for a particular set of patients. Such a study may compare similar treatments, such as competing drugs, or it may analyze very different approaches, such as surgery and drug therapy."
Institute of Medicine report [25]	"The 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 the delivery of care. The purpose of CER is to assist patients, clinicians, purchasers, policy makers, and the public to make informed decisions that will improve health care at both the individual and population levels."
US Federal Coordinating Council [26]	"CER is the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in 'real world' settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances."
Patient-Centered Outcomes Research Institute [27]	"CER is a field of research designed to compare the effectiveness of two or more interventions or approaches to health care, examining their risks and benefits. CER findings assist clinicians, patients and other stakeholders in making informed decisions that improve health care for both individuals and populations. The direct comparison of two or more interventions distinguishes CER from studies explor[ing] outcomes related to one intervention alone. CER can not only validate a particular intervention but also identify which of available treatments best meet the needs of a given population."
Health technology assessment (HTA)	
National Institute for Health Research, UK [7]	"HTA research is undertaken when evidence exists to show that a technology can be effective. The purpose of an HTA study is to establish the clinical and cost-effectiveness for the NHS in comparison with the current best alternative(s). A study may also investigate uncertainty around a technology's place in the existing care pathway. 'Technologies' in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease."
European Commission [28]	"HTA measures the added value of a new health technology compared to existing ones. Examples of health technologies include medicinal products, medical equipment, diagnostic and treatment methods, rehabilitation, and prevention methods."
International Network of Agencies for Health Technology Assessment (INAHTA) [29]	"HTA is the systematic evaluation of the properties and effects of a health technology, addressing the direct and intended effects of this technology, as well as its indirect and unintended consequences, and aimed mainly at informing decision making regarding health technologies. HTA is conducted by interdisciplinary groups that use explicit analytical frameworks drawing on a variety of methods."

further evidence generation. For some decisions faced by patients, clinicians, insurers, and policymakers, there may be insufficient evidence from individual studies to inform the decision. In these cases, new CER studies must be conducted to generate evidence (evidence generation). Generated or synthesized evidence must be disseminated for decision-makers of healthcare and CER to make informed decisions. It is also important to note that CER may assess both benefit and harms. Therefore, traditional pharmacoepidemiologic studies assessing the safety (harms) of medications in postmarket settings fall under the umbrella of CER.

The most important gaps in the current knowledge base about treatment interventions are lack of information about how a treatment works in actual clinical practice in contrast to the artificial settings of clinical trials, lack of information about the comparative effectiveness of treatment options, and lack of information about how variation in patient characteristics affects treatment effectiveness [30]. CER has the potential to fill important evidence gaps associated with the limitations of a predominantly RCT-driven drug and device approval pathway. The RCT pathway speaks to efficacy rather than effectiveness, because (i) placebo is often used rather than an active comparator agent; (ii) RCT study populations are not representative of the medication users post-approval (i.e., RCTs tend to exclude older and multimorbid adults and children); and (iii) the controlled settings used in clinical development programs often differ substantially from real-world care settings (e.g., they use short follow-up and surrogate endpoints).

In summary, the goals of CER are (i) to inform decisions on interventions or approaches to health care in real-world settings with regard to their intended and unintended outcomes that are relevant to patients; (ii) to put new technology into proper perspective in relation to older technology; and (iii) to identify patients who are more or less likely to respond to some interventions than others [31]. As a result, CER is expected to increase the use of more effective clinical options and decrease the use of less effective treatments [1,32–34]. Another consequence of achieving these goals could be a reduction in healthcare costs through avoidance of treatments that do not work or are less effective than alternatives.

CER and Pharmacoepidemiology

The concept of CER is in fact very familiar to pharmacoepidemiologists. Classic pharmacoepidemiologic studies that assess postmarket safety of medications constitute CER as defined earlier. Also, soon after the field of pharmacoepidemiology emerged in response to the need to study drug safety after marketing of medications, pharmacoepidemiologists recognized the need for postmarketing "efficacy" assessment (now defined as "effectiveness") and debated the challenges of assessing "intended" effects or benefits [35–37]. Despite the concern that non-experimental studies may not be useful in studying the intended effects of drugs, Strom *et al*. showed that of the 100 most recently approved drugs with 131 potential drug uses, only 28% would require experimental designs [38]. In the field of pharmacoepidemiology, we developed a research framework for experimental and non-experimental studies, knowledge of study designs, data sources, and analytic strategies, and faced various new methodologic challenges when studying unintended and intended effects in real-world patients. As described in the previous section, CER became a popular concept and a well-funded field as a result of the most recent government initiative in 2008 and the subsequent establishment of PCORI. The majority of CER has dealt with the effectiveness of medications, surgical procedures, and

medical devices, which is another reason why the field is of great relevance and interest to pharmacoepidemiologists.

In the context of pharmacoepidemiology and especially in the drug development process, CER covers the tail end of the pathway that begins with bench research (characterized by preclinical research to qualify for Phase I regulatory approval), moving to bedside research (characterized by proof of concept and efficacy research to qualify for Phase II regulatory approval), and ending with population research (characterized by clinical efficacy to qualify for Phase III regulatory approval and Phase IV clinical safety and effectiveness in postmarket settings), and finally research on the effect of policies (characterized by postmarketing surveillance and pharmacoeconomic research). A schematic illustration of this process is presented in Figure 26.1. In one sense, the full scope of CER is much broader than pharmacoepidemiology, as CER covers a range of clinical modalities for prevention, diagnosis, and treatment (drugs, medical devices, procedures, behavioral and other complex social interventions, as well as health delivery systems and policies) [39]. It also covers strategies for implementation. In another sense, however, CER is narrower than pharmacoepidemiology, not only because it covers the tail end of the pharmacoepidemiology spectrum (Figure 26.1), but also because it emphasizes "head-to-head" comparisons of the safety and benefits of treatments and diagnostic strategies to identify "best-in-class" treatments in the real world [40], whereas pharmacoepidemiology can compare users to nonusers or to alternative treatments.

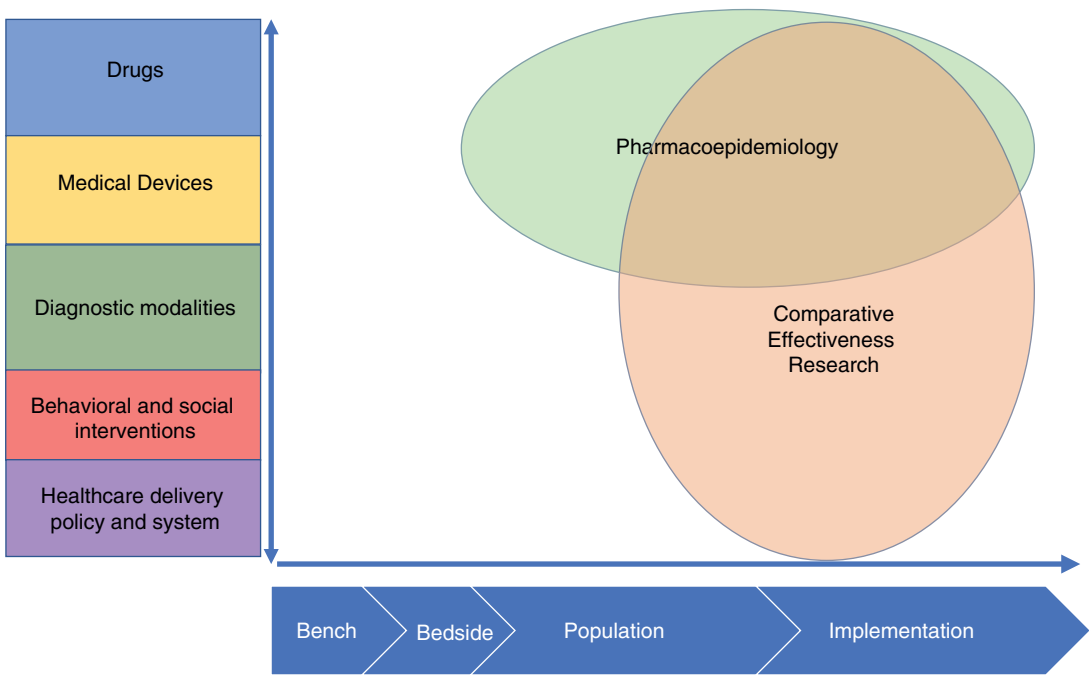

Figure 26.1 Schematic representation of overlap between pharmacoepidemiology and comparative effectiveness research.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Scope of CER

CER is broad in scope and addresses the continuum of medical and surgical interventions, including drugs, biologics, devices, medical procedures, technologies, behavioral interventions, prevention strategies, talk therapies, diagnostics, complex social interventions, and health delivery systems [25,41]. In addition, as characterized by Lauer and Collins [42], CER should utilize an array of technologies that enable quality and efficient healthcare delivery, and should account for the wide range of infrastructure of integrated healthcare systems. CER also encompasses beneficial and adverse effects as well as economic implications. It focuses attention not only on knowledge creation, but also on strategies for implementation. This broad scope can only be addressed by a diverse research portfolio that employs multiple study designs and analytic techniques (randomized trials, observational studies, and meta-analyses), as well as diverse data from primary data collection, preexisting data, and hybrid approaches linking different data sources [32,42].

Key Attributes of CER

The key attributes of CER that are embedded explicitly or implicitly in the aforementioned definitions and goals are as follows:

- It studies effectiveness in real-world patients and settings.
- It directly compares alternative methods to prevent, diagnose, treat, and monitor clinical conditions (rarely comparing alternatives to placebo, as "doing nothing" is often not a real-world clinical decision).
- It involves stakeholders, including patients and caregivers, in the research process.
- It uses clinically relevant and patient-centered outcomes.
- It assesses subgroups and different care settings in which differential effects may be observed (so that the evidence is more applicable to individual patients and is useful in various clinical settings).

Key attributes and related goals of CER studies are presented in Table 26.2.

The first attribute (inclusion of real-world patients and settings) is necessary to increase the direct applicability of the evidence generated from CER. Traditional efficacy trials are typically conducted by investigators affiliated with tertiary care hospitals. In contrast, CER should include data from patients and physicians from a wide range of care settings. The vision for CER is that it will provide opportunities for community hospitals and practices to become involved [43].

In real-world clinical practice, clinicians and patients need information to understand the comparative benefit or safety of two or more alternative treatments and to choose the best option. Therefore, "doing nothing" (placebo) is infrequently a viable alternative to treatment. However, traditional efficacy trials compare an intervention to nonintervention, for example treatment A to placebo, and are thus not informative on the comparative effect of different treatments. The goal of CER, as already mentioned, is to inform clinical or policy decisions *among alternative options*. Therefore, head-to-head comparison of alternative methods (including nonintervention if that is a legitimate option in clinical practice) is the second key attribute of CER. This attribute also addresses the goal of putting new technology into proper perspective in relation to older technology. The importance of comparing alternative healthcare options is highlighted by the 2007 Institute of Medicine (IOM) report [44], which points out that "the rate with which new interventions are introduced into the medical marketplace is

Table 26.2 Attributes and corresponding goals of comparative effectiveness research studies.

Desired/necessary attributes	Corresponding CER goals
Real-world patients and settings	To inform decisions on interventions or approaches to healthcare in real-world settings with regard to their intended and unintended outcomes that are relevant to patients.
Head-to-head comparison of various treatment/diagnostic/ implementation strategies	To inform decisions on interventions or approaches to healthcare in real-world settings with regard to their intended and unintended outcomes that are relevant to patients. To put new technology into proper perspective versus older technology.
Inclusion of all stakeholders of healthcare (including patients/ caregivers) in the research process	To inform decisions on interventions or approaches to healthcare in real-world settings with regard to their intended and unintended outcomes that are relevant to patients.
Use clinically relevant and patient-centered outcomes	To inform decisions on interventions or approaches to healthcare in real-world settings with regard to their intended and unintended outcomes that are relevant to patients.
Assess heterogeneity of effects by patient variability, including phenotype and genotypes	To identify patients who are more or less likely to respond to some interventions than others.

currently outpacing the rate at which information is generated on their effectiveness and circumstances of best use" [3], and that "less than half of all medical care is based on or supported by adequate evidence about its effectiveness" [44]. In addition, wide variation in practice [45–49] as well as geographic variations in the utilization of certain treatments and procedures [50–55] suggest a lack of "sufficient evidence to determine which approach is most appropriate" [44].

The third key attribute invoked by the US CER initiatives is involving stakeholders, including patients and caregivers, in the research process [25,30,56]. As conceived in the IOM's recommendations for "a robust national CER enterprise" [25], this should involve a continuous process that considers and prioritizes topics for CER research and funding to address current knowledge gaps about diseases and conditions, and that consistently includes participation of patients, caregivers, and consumers to provide input regarding issues of public concern [25]. According to Slutsky *et al.* [30], priorities for CER must be based on input from all healthcare stakeholders, research and synthesis must apply to a wide range of healthcare services, and the results must be made accessible to multiple audiences. Stakeholders in healthcare are generally categorized as consumers (patients, caregivers, and the public), providers (clinicians), payers (health insurance programs and patients/ caregivers), policymakers, product makers (pharmaceutical industry), and researchers. A systematic review [57] assessing stakeholder engagement in CER and patient-centered outcomes research (PCOR) in published articles from 2003–2012 found that reports on stakeholder engagement were highly variable in content and quality. In this review, the most frequent engagement was with patients, engagement with clinicians was modestly frequent, and engagement with other groups was infrequent. Stakeholder engagement was more

common in the prioritization of CER than in its implementation and dissemination.

The fourth attribute of CER is that effectiveness and safety should be addressed using outcomes of interest and importance to patients and clinicians. This attribute also addresses the gap of traditional evidence based primarily on efficacy trials, many of which focused on surrogate outcomes instead of hard clinical outcomes or patient-reported outcomes [58] of most interest to patients and clinicians, such as quality of life or functional status.

Traditional efficacy trials typically report average effects and are usually underpowered to detect variability in patient responses. However, clinicians must make decisions about choices for patients whose profiles are similar to the average of study participants in the trials. Therefore, the fifth attribute of CER is exploration of heterogeneity to identify subgroups of patients who benefit more (or less) from a given intervention. While CER explores patient variability, it assesses treatment effects in subgroups that are not typically narrow enough to reflect differences in how individual patients respond to therapies [59]. Better practices are needed to evaluate treatment heterogeneity, accounting for more precise individual-level factors and preferences as well as genetic information, such as the conditional average treatment effect [59]. Developments in molecular biology and genomics will increasingly make it possible to assess genetic variation in individual responses to different treatment interventions, with the goal of individualized and predictive medicine [25] (see Chapter 30).

Methodologic Problems to Be Solved by Pharmacoepidemiologic Research

All three components of CER (evidence synthesis, evidence generation, and evidence dissemination) are relevant in the field of pharmacoepidemiology. This section will cover all three components but will focus most on evidence generation, as it is the core field in pharmacoepidemiology and its methods are directly relevant to CER.

Issues for Evidence Synthesis

Systematic Reviews and Meta-Analyses in CER

The synthesis of evidence features prominently in definitions of CER, and systematic reviews and meta-analyses are the central approaches in evidence synthesis. To clarify common terminology following the Cochrane Collaboration and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [60,61], a systematic review refers a collection of all empirical evidence that fits prespecified eligibility criteria to answer a specific research question. Meta-analysis is the use of statistical methods to summarize and combine the results of independent studies (see Chapter 36). Therefore, many systematic reviews contain meta-analyses, but not all. Systematic review and meta-analyses can be used to discover patterns among study results and to provide reproducible summaries of study findings. In CER, systematic reviews may provide direct answers to CER questions, or may elucidate the need for more evidence generation when the results from individual studies are contradictory or when the magnitude of the underlying risk is small. Also, systematic reviews have been used in combination with clinical guidelines as a framework to identify knowledge gaps and to set research priorities [62].

The strengths of meta-analyses are mitigated by several methodologic challenges. The methodologic issues of systematic reviews and meta-analyses described in Chapter 36 are also relevant in evidence synthesis in CER. Briefly, the results of a meta-analysis are often highly subject to decisions made by the investigator: which studies to include or exclude from a meta-analysis, which outcome endpoints to

consider, and how to pool studies that differ in design and methods. Simmonds *et al*. [63] identified several sources of disagreement among experts that can affect the summary findings of meta-analyses. In addition, any limitations of the original studies will influence conclusions from the analysis of the pooled studies. This can be problematic in meta-analyses of observational studies. Consequently, some researchers have argued that only randomized trials should be meta-analyzed [64–67] (further discussion about meta-analyzing observational studies is found later in this chapter). It has also been argued that the outputs of meta-analyses may not provide greater insights than the results of individual studies [63,68].

Another limitation is that reviewers of the same studies may reach different conclusions, because of varying expertise in the topic of the review or in the technical skill of performing meta-analyses [69], or because of differences in values and orientations held by different investigators. The value of meta-analyses may also be seriously limited by publication bias, which can take several forms [70]. Studies with statistically nonsignificant or negative results are less likely to be published, and studies with statistically significant results and with stronger treatment effects tend to be published with less delay than studies with nonsignificant results. In addition, findings in some areas of research, such as complementary and alternative medicine, are less likely to be published. The summary conclusions from pooled published results will thus tend to be biased because of this preferential selectivity [70]. Problems stemming from publication bias may be amplified in meta-analyses of observational data. In addition, meta-analyses commonly combine the summary statistics from individual studies, whereas stronger results could be produced by obtaining and aggregating individual patient data from the separate studies analyzed [63,71,72]. However, issues of access, privacy, and ownership of original data make it difficult for investigators to obtain individual-level data.

Meta-Analyses of Observational CER

Generated evidence for CER can take the form of either observational studies or clinical trials. Therefore, longstanding debates about meta-analysis of observational studies are particularly relevant to CER. While some commentators have argued that meta-analyses of randomized trials are preferred to meta-analyses of observational studies [64–67], meta-analyses of observational studies are as common as those of randomized trials [68,73]. Reviews and practice guidelines on meta-analyzing observational data [68,73,74] show some disagreement with regard to this message. Some common opinions distilled include:

- Observational studies are more diverse in their designs and populations.
- Publication bias may be more problematic in observational studies [75,76].
- Biases are more problematic in observational studies.
- Therefore potential biases in the original studies make the calculation of a single summary estimate of effect of exposure potentially misleading, creating more precise but equally spurious effect estimates.
- More is gained by carefully examining possible sources of heterogeneity between the results of different observational studies.
- Concerns related to methodology and interpretation make the clear and thorough reporting of meta-analyses of observational studies absolutely essential (one guide provides a draft checklist summarizing recommendations for reporting meta-analyses of observational studies [73]).

From the point of view of pharmacoepidemiology, *a priori* exclusion of observational studies from meta-analyses would constitute a major loss. In fact, in some circumstances, meta-analysis of observational studies may be the only option to quantitatively synthesize current evidence. For example, Man *et al*. investigated the long-term effectiveness of methylphenidate in

the reduction of physical injuries. Harm reduction is a very important clinical outcome for patients and the healthcare system due to the high personal and economic cost of injuries. No clinical trials for methylphenidate were sufficiently long or even measured this outcome; hence, meta-analysis of observational studies was the only available option [77]. A study by Kirtane *et al.* [78] provides an example of a comprehensive meta-analysis that included both RCTs and observational studies, but analyzed them separately because of the differences in these types of study designs. Regardless of which types of studies are included in a meta-analysis, we agree with the need for a careful and systematic examination and reporting of observational studies, and for using epidemiologists' and clinicians' judgments to reach decisions about whether meta-analyses should be performed, and, if so, what studies should be included.

The expertise and effort required to perform a well-conceived and credible meta-analysis are not trivial. The AHRQ and IOM have published recommended standards for performing and reporting systematic reviews [79,80]. Nevertheless, the conclusions obtained by a rigorous meta-analysis cannot be deemed to provide a lasting answer to a clinical question, because new information may continuously become available. Therefore, the meta-analysis will require regular updates to keep it relevant for clinical guidelines [81].

Issues for Evidence Generation

Observational (Nonexperimental) Studies in CER

Observational studies have an important place in CER. First, observational studies provide data on real-world patients in usual clinical practice, which is one of the required attributes of CER evidence. Second, observational studies can provide larger samples and/or longer follow-up more easily than experimental studies. These are features that will be needed, as CER compares two options head to head, and this type of comparison will result in smaller effect sizes than comparing one treatment to placebo [42,82–84]. To date, the majority of CER studies have been conducted using observational study designs. Observational study designs and the methodologic issues they raise [85] are directly applicable to CER. In this section, we will summarize methodologic issues of particular importance in observational CER.

Confounding by Indication

As mentioned earlier, observational CER studies of intended effects are more susceptible to confounding by indication than observational studies of unintended effects (e.g., studies evaluating adverse drug events). While confounding by indication is covered in greater detail in Chapter 43, this bias is especially prominent when studying beneficial effects of treatments. In clinical practice, if one assumes prescribers are rational, one would expect treated patients to differ from untreated patients, as the former have an indication for the treatment. To the extent that the indication is related to the outcome variable as well, the indication can function as a confounding variable. On the other hand, confounding by indication for the treatment is less of a problem when a study is focused on unintended drug effects (side effects), regardless of whether those effects are harmful or beneficial. In this situation, the indication for treatment is less likely to be related to the outcome variable under study. However, this is sometimes not the case for studies of intended beneficial effects, which are the focus of many CER studies.

Confounding by indication (for the treatment) may also appear to be less of a problem when making comparisons between therapeutic alternatives for the same condition, since both study groups have the indication for treatment. However, nonrandomized studies comparing therapeutic alternatives are not necessarily free from confounding by indication, because the

true indication for a given treatment is often more subtle than the regulator-approved indication. For example, patients prescribed an angiotensin-converting enzyme (ACE) inhibitor as initial treatment for hypertension are likely to be different from those prescribed a thiazide diuretic for the same condition, as the former are more likely to have diabetes with nephropathy, myocardial infarction, and heart failure, whereas the latter being are more likely to have uncomplicated hypertension. As a second example, patients prescribed a combination of methotrexate and tumor necrosis factor alpha inhibitor as initial treatment for rheumatoid arthritis are likely to have more severe and active disease than those prescribed methotrexate monotherapy. Unless the choice between treatment alternatives is effectively random given measured variables, confounding by indication remains an issue in comparative studies. Disease severity is often associated with the outcomes of interests in studies assessing intended effects of treatment (e.g., improvements in rheumatoid arthritis symptoms or disease activity in the second example). Therefore, inability or limited ability to control for disease severity will results in bias (confounding by severity). The subtler examples of confounding by indication in the aforementioned scenarios are directly pertinent in CER, as these are exactly the types of questions CER addresses (comparing alternative options head to head). Furthermore, CER studies generally aim to detect differences that are likely to be smaller than in studies comparing exposed subjects to unexposed subjects. Accordingly, subtle instances of confounding by indication or confounding by severity can be especially problematic in CER.

Considerable effort has been undertaken to develop more effective methods for control of confounding in studies based primarily on administrative data (see also Chapters 12 and 43). Common approaches include propensity score-based methods, disease risk score-based methods, doubly robust methods, and instrumental variable analyses. These methods may, under certain conditions, provide better control of confounding than standard multivariable adjustment. However, it is important to keep in mind that most of these approaches (including propensity score but not instrumental variable analysis) are dependent on identifying and measuring those variables that are the true predictors of therapeutic choice in the databases.

For example, we conducted a study assessing the comparative effectiveness of carotid stenting versus carotid endarterectomy in older patients with carotid stenosis. Substantial confounding by indication would be suspected in this study, as carotid stenting is indicated and reimbursed only for patients with high surgical risks due to age, anatomic characteristics of carotid stenosis, or other cardiovascular or noncardiovascular co-morbidities (which are the predictors of worse prognosis). In this study, we demonstrated that propensity score-based methods (including high-dimensional propensity score methods [86]) using only claims data are insufficient to control for confounding by indication, and additional clinical information from vascular registries was necessary to achieve adequate control of confounding [87]. While these approaches provide sufficient confounding control in certain situations, they are generally not sufficient unless important variables related to the indication for treatment are available in the data [88]. Instrumental variables are promising alternatives if a valid instrument can be found for the clinical question. However, finding valid instruments in pharmacoepidemiology is extremely difficult – some would say impossible [89]. Design-based approaches such as restriction [90,91] and use of active comparators [92] may work in certain situations, and self-controlled methods can sometimes help to control time-invariant confounders [93], but it must be kept in mind that such approaches are directly linked to how the research question is defined. Accordingly, investigators must ensure

that the research question and the design are consistent and as intended. Much more work is needed in these areas to advance the field of CER.

Healthy User/Candidate Bias

Healthy user effect or bias has been observed among users of some medications, especially preventive medications such as hormone replacement therapies, statins, and certain antihypertensive medications [94–99]. For example, cardiovascular event reduction is consistently smaller in clinical trials compared with observational studies of antihypertensive medications [100,101], which suggests healthy user bias plays a role in observational studies of these preventive therapies. In CER comparing medical devices or interventions to a pharmacologic treatment, the healthy user effect (or more precisely for interventions, the healthy candidate effect) will be a major concern, as interventions typically pose short-term risks in exchange for long-term benefits, and patients at high risk of complications or deemed too sick to benefit are thus less likely to be selected for interventions. The healthy candidate effect is one of the biggest threats to validity in CER when comparing different treatment modalities. For example, we previously demonstrated the existence of healthy candidate bias in older heart failure patients who received implantable cardioverter-defibrillators (ICDs) versus those on medical therapies only by showing that patients with implanted ICDs had drastically better outcomes that were very unlikely to be attributed to the effects of ICDs (e.g., nursing home admissions, hip fracture, and short-term mortality) [102]. This highlights the utility of including falsification outcomes in observational studies.

Healthy user or candidate bias is thought of as a mix of confounding and selection bias. It arises when users of certain medications or candidates for invasive interventions have better outcomes due to factors other than effects of the treatment. While the factors associated with healthy user/candidate bias are not fully understood, many factors suggested to be responsible for healthy user/candidate effects are typically unmeasured in most databases. These include healthier lifestyles (healthy diets, regular exercise, and being less prone to using tobacco or alcohol) [99,103], higher socioeconomic status, better adherence to screenings and other preventive therapies [96], better physical [104] and cognitive function, less frailty [105], better social support, and stronger willingness to live. The effect size of healthy user/candidate bias can be quite substantial and is often as strong or stronger than the effect of the treatment itself [94–97,106]. Most importantly, healthy user bias may be refractory to analytic solutions unless prevented by thoughtful study design (e.g., self-controlled) and/or availability of extensive data on lifestyle and behavioral factors, thus resulting in inflation of the apparent benefits of preventive and other medications or invasive interventions.

Data Sources, Record Linkage, and Multidatabase Studies

As noted by the IOM report [25], CER studies should rely on multiple types of data sources, including primary data sources (medical and pharmacy records, electronic medical records, and *de novo* data generated through clinical trials or observational studies) and secondary data sources (administrative claims and clinical registries). Most CER studies to date have used the same data resources described in Chapters 12–14 of this book. As in usual pharmacoepidemiology practice, the data sources should be selected based on the study question and to maximize the internal and external validity of the results. To overcome the biases mentioned as well as others, including misclassification bias and selection bias, linking multiple data sources through record linkage can be a powerful tool, as it enriches the information for the given study subjects. Also, multidatabase studies within or across countries can potentially

enhance observational CER studies by enlarging the sample size for statistical power, assessment of effect heterogeneity, and improving generalizability. The methods, applications, and challenges of record linkage and multidatabase studies are described in other chapters.

A distinction must be made between data or record linkage (linking multiple data sources to enrich information) and multidatabase studies (using multiple databases for mostly nonoverlapping individuals). These two approaches are often confused, but each has a distinct goal. Data linkage is conducted in order to enrich information using a record linkage method, a computer-based technique to identify and link records from different databases that refer to the same entity or individual [107]. The data required for impactful and valid CER studies may be spread across two or more databases. Linking records across databases can transform ordinary individual datasets into powerful new platforms from which to perform timely and valid CER. For example, linkages between administrative claims databases and clinical or device registries can add longitudinal follow-up to registry data and add clinical details to administrative data. In addition to answering clinical questions, data linkage can also be used to address various methodologic issues in CER. For example, a linked database can be used to study data quality (e.g., by assessing agreement between two sources of the same data) and to validate claims-based endpoint ascertainment algorithms (e.g., by comparing a claims-based variable to a clinical gold standard) [108]. In addition to facilitating observational CER, record linkage can improve randomized trial evidence by linking patients in the trial to complementary data. For example, linking patients in a trial to Medicare claims can be a relatively inexpensive and effective way to extend the follow-up period of a clinical research study. Data linkage combining two or more data sources has enabled the conduct of more observational CER and/or more valid observational CER that would not be possible using a single data source [106,109–111]. Recent development of data linkage in Scandinavian countries has provided exciting opportunities to evaluate effectiveness beyond medical care. For example, Lichtenstein *et al*. linked the use of attention deficit hyperactivity disorder (ADHD) medication with criminal justice system records. They found that among patients with ADHD, rates of criminality were lower during periods when they were receiving ADHD medication. These findings raise the possibility that the use of medication reduces the risk of criminality among patients with ADHD [112].

The challenges of data linkage are both methodologic and ethical. Methodologic challenges include unavailability of linkage variables to researchers, especially unique identifiers such as names or social security numbers, incompleteness or inaccuracy of linkage variables due to poor data quality, nonoverlap or relatively small overlap of populations covered in each database, general misconceptions about linkage methods (especially probabilistic linkage methods), and understanding when to use what linkage method [113,114]. When unique identifiers of subjects are not available, at least for certain databases and populations (e.g., linking inpatient or outpatient claims data to clinical registry data for patients with heart failure, device implantations or surgeries, rheumatoid arthritis, and atrial fibrillation) [115,116], it is possible to conduct record linkage with high accuracy using multiple nonunique identifiers [115,116]. The primary ethical challenges of data linkage are ensuring patient privacy, which can be achieved by removing or limiting access to patient identifiers for research use including record linkage. However, this can make the linkage more difficult and sometimes impossible.

In our recent study of patients in a US-based online community, most reported that they were comfortable with researchers accessing their de-identified data for research purposes. Our study indicated that patient comfort levels

may be improved by better communication and transparency around specific research goals and how they may be beneficial to patient communities. In addition to mitigating re-identification risk, developing and improving methods to link databases through use of multiple nonunique identifiers may also improve patient comfort with secondary use of health data for research [117]. In a survey commissioned by the Wellcome Trust (a UK medical charity), 53% of respondents in the UK indicated that they would be happy for their data to be used by commercial organizations if it was for research purposes. Interestingly, over 60% of respondents indicated they would prefer that commercial research organizations have access to health data than that society miss out on the benefits these companies could potentially create. One of the most significant findings from the survey is that respondents considered academic researchers, charities, and organizations working in partnership with the public sector to be the most acceptable users of health data [118]. Patients' understanding and perceptions of the use of health data are still evolving, and it is important to continue to maintain communication and transparency regarding the use of health data for research.

In the last 10–15 years, networks of national and multinational database studies relevant to CER have been established in the US, Canada, Europe, and Asia [119–124]. Multidatabase study networks have been used to conduct observational CER or to provide a platform for CER trials [125]. The advantages of multidatabase studies in CER are that they capture diverse patient populations and/or increase the number of patients to detect relatively small effect sizes that can be expected in head-to-head comparisons and to assess the heterogeneity of effects. In multidatabase studies, data linkage methods are not necessary, as their intention is usually to bring databases together for nonoverlapping populations. However, structure, governance, and methods to manage and conduct a study using data from multiple sites and to synthesize results are needed. The structure, governance, and methods have to meet data management policies and data safety and privacy standards that may be unique to each database and can vary substantially, especially in international contexts [126,127].

Many networks employ a distributed network approach with a common data model (CDM), where the ownership and management of the database are left with individual data partners participating in the network. This approach is often preferred, as it mitigates most of the ethical and political issues with data privacy, governance, and ownership. In this model, each database is converted using a CDM so that its structure and coding are fully standardized. Multiple CDMs have been developed and modified to date [122,128–130]. To conduct analyses, researchers create a single statistical program that can be run against any database in the network with minimum or no modification. Another common approach is a distributed network with a common protocol rather than a CDM. The Canadian Network for Observational Drug Effect Studies (CNODES) operates using this approach, which eliminates the need to convert data from each site to a CDM [123]. In a distributed network with a CDM, a standardized coding language and format are needed to permit identical computerized queries to be submitted and executed across data resources, as well as standardized formats for returning responses from different databases [128,131,132].

In any approach for conducting multidatabase studies, understanding and dealing with variability in results across databases are especially challenging, particularly when the data come from different countries or diverse geographic regions/populations with differing healthcare systems, policies, and patient and clinician behaviors [126]. When large or unexpected variability in the results from each database is observed, researchers must first exclude the

possibility that the observed variability is due to technical issues arising from mapping codes or converting to a CDM, and/or from biases that are unique to each database (e.g., poor data quality, existence of and/or lack of understanding of unique features or idiosyncrasies in the data). When this possibility has been excluded, considerations must still be given as to when it is appropriate to combine results from different databases. This is especially important in CER, as understanding heterogeneity of effects is a major attribute of CER, and combining results that exhibit significant variability is not desired. The methodologic issues involved in combining results for meta-analyses discussed earlier are directly applicable to multidatabase studies as well, since it is generally not possible to analyze patient-level data to synthesize the results from each database, due to concerns about data security and privacy and/or restrictions of policies for data access and use.

Common challenges for data linkage and multidatabase studies include (i) logistical problems in accessing data sources, including issues of ownership of data, infrastructure, governance, data security, and data privacy (see also Chapters 12–14); and (ii) the required familiarity with the logical organization and content of disparate databases, including features or quirks in the data that are unique to each database. Needless to say, the aforementioned methodologic issues in observational CER (e.g., confounding by indication) can also affect the conduct and validity of results in linked database or multidatabase studies.

Experimental Studies

As already discussed, the goals of CER are to fill gaps in evidence that is traditionally and heavily based on premarketing RCTs. Observational studies leveraging existing data sources or primary data collection can be used as a valid and more cost-efficient approach for CER when available data include the necessary fields, and/or when researchers employ design-based or analytic methods to overcome potential biases. However, there are situations where bias is intractable and randomized trial designs are needed to obtain valid results. Large simple trials such as pragmatic trials or cluster randomized trials can determine the effects of an intervention under the usual conditions in which it will be applied, and therefore can assess real-world treatment effectiveness [133]. For clinical trials to be used in CER, researchers must focus on using trial designs that are flexible, adaptive, pragmatic, practical, and efficient, in contrast to traditional randomized, blinded, placebo-controlled clinical trials [134–137] (see also Chapter 32 for a discussion of large simple trials).

Briefly, pragmatic clinical trials are intended to overcome the limitations of traditional RCTs in order to answer CER questions. Pragmatic trials include real-world patients such as those with co-morbid conditions and those from diverse demographic backgrounds [138], and providers from community settings instead of only tertiary settings. In pragmatic CER trials, comparator treatments should be those in use in clinical practice (rather than placebo controls), outcomes should be those that matter to patients and clinicians rather than investigators or drug companies, and variations in patient responses to the treatment (treatment heterogeneity) should be explored [134,137,139].

One example of a pragmatic trial for CER is the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) [140], a $120 million NIH-funded trial comparing three antihypertensive medications and evaluating more than 42 000 patients in 600 clinics and centers in the US, Canada, Puerto Rico, and the US Virgin Islands [141]. A more recent example of a pragmatic trial for CER is the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC) study, which compared ziprasidone and olanzapine for their risk of nonsuicide mortality [142,143]. This large randomized pragmatic trial included

approximately 18 000 patients in 18 countries and cost $85 million.

As in observational CER studies, using clinical trial designs for CER presents its own challenges. First, demonstrating clinically meaningful effect sizes is often challenging for several reasons: (i) the liberal inclusion criteria needed to assure generalizability of the study groups, although this increased heterogeneity can decrease the probability of detecting a given treatment effect as statistically significant, requiring an even larger sample size [134]; (ii) head-to-head comparison of commonly used clinical strategies; and (iii) modern medical interventions showing less dramatic benefits. The effect sizes seen even in placebo-controlled trials have been decreasing over time, and this pattern is considered to stem from the increasing rarity of discoveries of transformational medical interventions [144]. To detect a relatively small effect size and evaluate long-term and clinically relevant outcomes including hard endpoints, larger samples make CER trials more expensive [145]. The emphasis on usual care settings and the less-controlled nature of the trials lead to problems that are well known in observational studies. For example, loss to follow-up and/or nonadherence over time can introduce bias [135]. The lack of blinding in pragmatic trials creates the potential for biased observations and a threat to internal validity [134]. Another similar limitation of this type of trial results from the flexible treatment protocols that are preferred, as they are closer to what happens in real-world settings. Specifically, pragmatic trials involve the participation of community providers in their usual practice. Accordingly, providers can vary the treatment process, dose, and regimen given to different patients, depending on differing responses to therapy over time. This flexibility permits assessment of the outcomes of the composite treatment, but not of particular components within the treatment process [135].

As shown in the earlier examples of CER, the high cost of pragmatic trials is a major obstacle. However, conducting a simple randomized trial in usual clinical care conditions using routinely collected data, what are called "electronic point-of-care trials," could minimize the cost burden [146]. Attempts to conduct such trials with information technology (IT) tools to facilitate timely and efficient point of care (POC) recruitment have been reported by UK researchers using the General Practice Research Database® (GPRD®) [146,147]. In the US, researchers at Veterans Affairs (VA) hospitals have a head start taking advantage of the VA's sophisticated electronic health records (EHR) system [148]. The UK experience showed that the recruitment of clinicians and patients was a major challenge: the investigators observed that the number of interested clinicians/practices dropped substantially with each stage of the governance process, including site contracts, local approval forms, web-based good clinical practice, and protocol training [146]. A successful implementation of electronic POC trials will require three conditions at minimum: (i) a well-connected community with a network of practices and patients; (ii) data and IT infrastructure that enables patient recruitments at POC and captures clinically important outcomes; and (iii) readiness of clinicians and patients to accept "randomization" in routine clinical practice when there was equipoise among therapeutic options. A recent review article described attempts to conduct POC trials and integrate comparative effectiveness trials into patient care, illustrating challenges and limitations specific to POC trials [149]. Obviously, use of EHR poses limitations on the questions that can be addressed, processes that can be implemented, and outcomes that can be assessed.

Choosing the Right Methods: Experimental vs. Observational CER

There are inherent limitations in both experimental and observational designs, as discussed in the previous sections. While an experimental design is generally accepted and considered to

yield higher internal validity than observational CER, one must carefully examine how either of these types of studies are conducted in order to assess validity (both internal and external), interpret results, and draw meaningful conclusions. Several considerations can guide the decision of which study design to use for CER. First, some interventions cannot be investigated with clinical trials because of ethical considerations, even though such trials may otherwise be preferred scientifically. Second, it is not practical to employ pragmatic trials for most CER research questions due to their prohibitively high cost. Thus, observational studies using the techniques of nonexperimental pharmacoepidemiology will continue to play a role in CER, because some questions cannot be answered in clinical trials or because observational studies provide a cost-effective approach (when they can provide results with high internal validity).

Finally, there are some research questions that cannot be answered in observational CER due to intractable biases that severely compromise the internal validity of the results. Questions related to bias are not binary (existence or nonexistence of bias) but rather quantitative (degree of bias), and researchers must consider all potential biases and their quantitative impacts on the results before conducting observational CER studies. As one of the goals of CER is to produce results that are applicable and generalizable to real-world patients and practices, attempts to achieve higher generalizability may compromise features that are favorable to internal validity [84]. Nonetheless, results from CER studies that have significant bias and therefore have poor internal validity cannot be generalized. Once observational CER studies are completed, when intractable and significant biases are suspected in the results, researchers must provide a fair and honest assessment of study validity and must attempt to publish the results including this assessment in order for the research community to learn from their experiences.

Published Guides for CER Studies for Evidence Generation

Standards for performing and reporting observational studies have been provided by several professional associations [150–152]. While these guides are not specific to CER, they are relevant and directly applicable to observational CER. In addition, several other guides specifically targeting observational CER have been published through initiatives of AHRQ, PCORI, other governments, and professional societies (e.g., International Society of Pharmacoepidemiology [ISPE], International Society of Pharmacoeconomics and Outcomes Research [ISPOR], American Heart Association [AHA]) [151–160].

A recent systematic review of these CER-related guidance documents assessed shared expectations for quality CER [161]. The review identified nine documents with over 300 recommendations for designing and conducting CER. The most frequently shared recommendations included transparency and adaptation for relevant stakeholders in the interpretation and dissemination of results. Other frequently shared CER methods recommendations included developing an *a priori* study design and operational definitions that allow for replication (n=8 documents); focusing on areas with gaps in current clinical knowledge that are relevant to decision-makers (n=7); assessment and discussion of validity of measures, instruments, and data (n=7); and clinically meaningful and objectively measured outcomes, including benefits and harms (n=7). Additional commonly shared recommendations included assessment for and strategies to minimize bias (n=6 documents), confounding (n=6), and heterogeneity (n=4). Pragmatism in the design of experimental CER trials has been widely discussed [139,162–177], and there are proposed tools to assess pragmatism in clinical trials that researchers and clinicians can use when designing or evaluating pragmatic trials, especially for CER [178–181].

Issues for Evidence Dissemination

The ultimate goal of CER to improve clinical care will not be achieved without successful dissemination and adaptation of CER evidence. Evidence dissemination has several distinct goals. One goal involves identifying priority topics, comprehensively identifying available information on these topics, and developing objective interpretations of the information [182] (for example, as provided by Cochrane Collaboration reviews [183]). The output from this research then becomes the source information for dissemination to clinicians, patients, and policymakers. This goal will be achieved through expanding efforts on systematic reviews and studies using novel research designs (see earlier discussion), focusing on the priority research areas that were identified by the IOM as having knowledge gaps.

Another goal involves knowledge translation; namely, using research findings as the basis for drafting clinical guidelines. Achieving this goal will require qualified review panels that have scientific and clinical expertise in the content areas of the topics for which guidelines are developed, and who can develop clinical practice guidelines. Ideally, clinical guidelines should also be both comprehensive for general patient care and specific for particular patient circumstances – a very demanding specification. Furthermore, to remain relevant, guidelines need to be updated periodically to incorporate new information about existing interventions and new treatments. The IOM recently proposed standards for developing trustworthy clinical practice guidelines [184].

A third goal involves knowledge exchange and utilization, achieved by the actual distribution of information and the education of clinicians, patients, and policymakers about current knowledge and best practices. This goal may be attained by more intensive use of technology and/or social interventions. Examples of such tools include computerized physician order entry (CPOE) systems, supplemented by computerized clinical decision support systems (CDSSs) that incorporate electronic reminders to comply with guidelines (e.g., reminders to perform screening tests or to order other tests or treatments, reminders to avoid co-prescribing interacting drugs, etc.). Other strategies for achieving knowledge exchange and knowledge utilization will require educating clinicians and patients about what treatments work best [2,26,185]. These efforts should include monitoring to ensure that information is integrated into the normal workflow and decision processes of clinicians and patients.

A final goal involves monitoring and assessment of whether these efforts translate into actual good practice and, if not, to identify which means of dissemination have a greater chance to create an impact. However, recent history suggests that scientific evidence is often slow to change clinical practice. For example, despite harms associated with overdiagnosis of prostate cancer with prostate-specific antigen screening [186], the test is widely utilized in general practice [187]. Also, after the aforementioned multimillion-dollar ALLHAT pragmatic trial showed that thiazide diuretics are more effective than ACE inhibitors or calcium channel blockers for patients with hypertension, no significant changes in practice were observed [188]. Timbie *et al.* [189] reviewed CER studies conducted in the 2000s, including the ALLHAT trial, and identified five root causes underlying the failure of many CE studies to alter patient care:

- Financial incentives, such as fee-for-service payment, that may go against the adoption of new CER evidence.
- Ambiguity or concerns about the validity of CER study results.
- Common cognitive biases, including confirmation bias [189], pro-intervention bias [190], and pro-technology bias [191] in the interpretation of new CER information.

- Failure of research to address the needs of end users of CER evidence (clinicians, patients, policymakers).
- Limited use of decision support tools by patients and clinicians.

The authors offered several suggestions that align with the four dissemination goals already described. In addition, they suggested that in developing guidelines based on CER evidence, adapting the standards proposed by IOM (one of which was that guideline development groups be "multidisciplinary and balanced" [192]) may overcome several of the root causes mentioned, such as ambiguity of CER results and cognitive bias in interpreting the evidence. Finally, the authors proposed that aligning the incentives of clinicians and patients by changing payment and insurance models may facilitate the adoption of CER evidence in clinical practice. However, a recent systematic review found that pay-for-performance programs in healthcare were associated with improved processes, but not patient outcomes [193].

Dissemination of CER evidence is a legal mandate of PCORI [194,195]. The federal AHRQ works to disseminate findings from patient-centered outcomes research funded by PCORI, as well as government agencies and other sources. PCORI's release of new evidence from the funded studies begins with translating all research findings into understandable summaries with the help of the Patient-Centered Outcomes Research Translation Center. PCORI funds support not only engagement activities and infrastructure development, but also research to bring findings from completed studies into practice in real-world settings, and to compare approaches to communicating and disseminating patient-centered outcomes research findings, as well as research on shared decision-making [194]. PCORI recently announced a dissemination initiative (Eugene Washington PCORI Engagement Awards) through which it was planning to award $20.5 million in fiscal year 2018. Between fiscal years 2011 and 2017, AHRQ committed about $260 million for the dissemination and implementation of CER findings.

CER and Cost-Effectiveness Analyses

The primary goal of CER is to inform decisions that lead to better care, not necessarily cheaper care [196]. This could result in abandoning expensive technologies that are no better than less expensive options. However, it could also result in paying for a more expensive technology because the evidence shows it is superior [33,196,197]. The relevance of including cost-effectiveness analyses in CER investigations (see Chapter 34) is unquestionable. However, CER should not be used for cost-containment decisions [33,40,198], and the experts conducting CER studies should not be placed in the position of using their findings about treatment effectiveness to make recommendations about reimbursement. Nevertheless, well-performed CER inevitably should and will affect reimbursement decisions. In some cases, CER studies will find that the more expensive treatment is preferable. Yet, over time, CER should ultimately save money by preventing wasteful spending on treatments that are less effective, especially if dissemination is successful and CER evidence is adapted into clinical practice [196]. Ultimately, CER alone will not solve the US's healthcare spending problem ($3.3 trillion in 2016 – 17.9% of GDP – and $5.7 trillion projected in 2026 – 19.7% of GDP [199]).

Currently Available Solutions

In a review of recently published studies of the comparative effectiveness of existing (rather than new) medications, Hochman and McCormick [200] compared active therapies to each other (rather than to placebo comparators); compared medications to nonpharmacologic

interventions such as surgery or lifestyle interventions; compared different pharmacologic strategies for medication use; and compared different medication doses, durations, or formulations. They found that only one-third of studies evaluating medications qualified as comparative effectiveness research, and only a minority compared pharmacologic and nonpharmacologic therapies, emphasizing the need to expand the scope of CER. Another study reviewed clinical trials conducted in the US between 2007 and 2010 addressing priority CER topics identified by the IOM [201]. Among 1035 studies found on clinicaltrials.gov, 231 (22%) were comparative effectiveness (CE) studies. The most common interventions examined in CE studies were drugs (37%), behavioral interventions (29%), and procedures (16%).

These studies show what is observed in major medical journals or on clinicaltrials.gov during the period 2007–2010, but more recent data are not available. As described in the next section, the predominant source of funding for CER in the US thus far has been the federal government. Since PCORI was established in 2010, it has funded 596 CER projects (approximately $1.7 billion) [5]. These include CER studies, projects to examine CER methods, and projects to build infrastructure for CER and PCOR. The most frequent disease conditions for funded studies include mental/behavioral health (115 studies), cancer (84), neurologic disorders (74), cardiovascular diseases (69), and multiple chronic conditions (58). Most-studied populations of interest include racial/ethnic minorities (290), individuals of low socioeconomic status (194), women (145), older adults (134), and patients with multiple chronic conditions (110). Most of PCORI's research projects awarded through fiscal year 2017 are still underway; only 53 projects had been completed at the end of fiscal year 2017, but many are projected to be completed between 2018 and 2020.

The initial wave of funding came from the President's budget proposal for fiscal year 2011, with $286 million for patient-centered health research (the rebranded term for CER) through AHRQ. Much more substantial funding continued to come from the nongovernmental, nonprofit PCORI and its Patient-Centered Outcomes Research Trust Fund, established in 2010. As of January 2018, PCORI had brought a total investment of over $2 billion in projects meeting its congressional mandates, including funding for nearly 400 CER studies ($1.7 billion), as well as projects to improve the methods ($129 million) and infrastructure for CER, including the National Patient-Centered Clinical Research Network (PCORnet; $374 million). PCORI is projected to commit an additional $721 million for awards in fiscal years 2018 through 2021 [5]. From fiscal years 2011 through 2017, HHS including AHRQ committed approximately $448 million from the Trust Fund. Of this amount, HHS committed approximately $260 million (or 58%) to the dissemination and implementation of CER findings. HHS is projected to commit an additional $120 million for these activities in fiscal years 2018 through 2020 [5].

The EU does not have a central budget for healthcare expenditure. However, the European Commission has provided funding to conduct studies in European countries via its research budget (e.g., the CEPHOS-LINK Project [202]). In individual countries such as England, the NIHR has also provided funding for CER, particularly via the HTA Programme [7]. Examples from Asia include the Hong Kong Government's funding of CER via the Health and Medical Research Fund [203].

The Future

Funding

Thus far, the predominant source of funding for CER in the US has been the federal government, whereas much of the funding for clinical efficacy

research (i.e., RCTs) comes from industry sources. Recently, real-world evidence (RWE) has been gaining in popularity in the US, since the FDA's leadership published its opinion [204] and guidance document [205] on RWE, defining terminology and discussing its use in regulatory decision-making. RWE is defined by the FDA as "the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of Real-World Data (RWD)" [206]. In the EU, according to the GetReal Glossary of Definitions of Common Terms [207], "RWE derives from the analysis and/or synthesis of RWD that can either be primary data collected in a manner which reflects how interventions would be used in routine clinical practice or secondary data derived from routinely collected data." There is substantial overlap between RWE and "evidence generation in CER," as RWE includes comparative evidence of the "potential benefits and risks of medical products."

With the emphasis on RWE among regulatory bodies in the US and Europe, there has been increasing interest in RWE/CER among producers of medical products. In this new CER/RWE environment where the results of RWE or CER studies can be used in regulatory decision-making, companies may invest resources in RWE/CER, thus allowing them to compete on the utility (e.g., comparative effectiveness) of their products rather than just on their marketing ability [196]. For example, two global pharmaceutical companies recently collaborated to invest in a set of RWD analyses, including CER studies, of direct oral anticoagulants [208,209]. As most head-to-head comparative randomized studies are already sponsored by industry [210], with the increasing demand for RWE in North America, Europe, and Asia, future comparative studies may be conducted using RWD to provide CER evidence for new medical products. In addition to the possibility of more funding from industry for CER, public funding agencies and charities in the US and Europe have supported and will continue to support nonpharmacologic trials and studies of off-patent medications, such as through the Better Medicines for Children Initiative [211].

Human Capital Development

As noted by the Federal Coordinating Council [26], training will be required for new researchers to apply the specialized methods of CER and to develop new CER methods [212]. Specialized skills are needed to perform both traditional RCTs and novel pragmatic trials. It should be noted that these are not necessarily new but specialized skills, as CER is not a novel concept. Specialized expertise is also needed to perform formal meta-analyses and nonexperimental studies, using either *ad hoc* data collection or existing databases, and to successfully access and link various databases and conduct multidatabase studies. Finally, the field needs individuals who are able to translate the findings into practice guidelines and for other dissemination channels. The emphasis of CER on community participation and inclusion will dictate that experts from many different fields and backgrounds will be required to communicate with each other, finding and developing a common language to permit productive interactions. Therefore, the research teams participating in CER will be composed of professionals from different disciplines and different settings, including pharmacoepidemiologists and practicing clinicians from specialties relevant to the given studies [40]. These teams will need to have the capacity to develop a shared understanding of basic scientific terminology and methods.

Accordingly, it is necessary to create and support training programs for researchers seeking careers in CER in order to develop capacity in the research community to conduct CER. In addition to preparing a cadre of researchers with expertise in CER methods, a critical mass of such researchers is required in order to undertake the large number of studies needed

to fill current knowledge gaps and to continuously update the knowledge inventory with systematic reviews. So far, PCORI has committed $30 million to workforce training awards for clinicians and researchers. For example, one of its career development programs, conducted in partnership with AHRQ, is designed to train clinician and research scientists to conduct patient-centered outcomes research and to actively engage stakeholders in efforts to improve the quality and safety of care [5]. AHRQ committed $94 million for efforts to train researchers on the conduct of CER [5] and plans to commit an additional $14 million by fiscal year 2020 for CER training. In Europe, the European Commission has provided funding for training in health and medical research including CER and pharmacoepidemiology in European countries (e.g., the Marie Skłodowska-Curie Fellowship [213]). In England, the NIHR has also provided funding for fellowships in health research including CER research, particularly through the NIHR Fellowship Programme [214]. In Asia, the Hong Kong Government funds health fellowships via the Health and Medical Research Fund [203].

CER and Clinical Practice

Though the research community continues to appear excited and energized, expectations must be tempered by several limits on what CER can realistically solve. It is unrealistic to expect that CER will address all therapeutic questions; healthcare is simply too complex. Sir William Osler, the Father of modern medicine, wrote, "The practice of medicine is an art, based on science" [215]. Even in the era of evidence-based medicine and CER, the practice of medicine is as much the application of art as the application of evidence. To reach an optimal decision in any given clinical situation, evidence must be applied to an individual patient who has her or his own values, preferences, life situations, and goals. Treating not only the disease but also the patient as a whole requires both understanding and application of the best evidence, as well as the skills and behaviors physicians bring to their own practice of the art of medicine [216]. Alternatively, if it were possible to base medical practice entirely on evidence, such evidence would consider not only complex pathophysiology, but also personal factors such as values, preferences, perceptions, and attitudes about risks, quality-of-life preferences, cost tradeoffs, as well as clinician–patient interactions. However, as the evidence underlying current medical practice consists of estimated averages from studied populations and is also not sufficiently complete, art comes into practice when subjective judgment is required. Therefore, even in situations where complete evidence-based information is available to guide clinical decisions, providers or patients may still opt for a decision based on personal choices that they value irrespective of the scientific evidence. As stated by Kerridge *et al*. [217]:

> Medical decision making draws upon a broad spectrum of knowledge – including scientific evidence, personal experience, personal biases and values, economic and political considerations, and philosophical principles (such as concern for justice). It is not always clear how practitioners integrate these factors into a final decision, but it seems unlikely that medicine can ever be entirely free of value judgments.

Overemphasis on scientific evidence can lead to therapeutic nihilism; that is, paralysis when such evidence is unavailable. Also, an overreliance on evidence-based guidelines can result in algorithmic care [218]. Ironically, this in turn may devalue individualized care, which is another goal and feature of CER. Subgroup analysis is the most commonly used approach for assessing heterogeneity in CER, but it has obvious limitations in providing sufficient evidence to "individualize" the care of each

patient. In the face of uncertainty, variation among reasonable but unproven options should be tolerated and even encouraged, as it will facilitate later evaluation. This runs contrary to the vision of a knowledge state that is sufficiently complete to guide all decisions about effective interventions at the individual patient level. We also need to be sure that the desire for scientific evidence does not paralyze medical practice when such evidence is absent. In such circumstances, the resulting variability in practice can provide the data that will underlie future CER studies.

As healthcare communities continue to embrace CER and demand better evidence to inform clinical decisions, CER will play an expanding role in healthcare research. It continues to establish its position as a central component of clinical research that is directly relevant to clinical practice and health policy; that is, CER is needed in order to practice the best evidence-based medicine and evidence-based policymaking. Nonetheless, creating sustainable funding sources, improving the quality of evidence generation and evidence dissemination through development of methods and better use of these methods, educating consumers about CER, and managing their expectation are ongoing challenges.

References

1 Luce BR, Drummond M, Jönsson B, *et al*. EBM, HTA, and CER: clearing the confusion. *Milbank Q* 2010; **88**: 256–76.

2 Conway PH, Clancy C. Charting a path from comparative effectiveness funding to improved patient-centered health care. *JAMA* 2010; **303**: 985–6.

3 Congressional Budget Office. Research on the comparative effectiveness of medical treatments. Nov 2007. http://www.cbo.gov/ftpdocs/88xx/doc8891/12-18-ComparativeEffectiveness.pdf (accessed January 2011).

4 H.R. 1 The Medicare Prescription Drug, Improvement, and Modernization Act of 2003. Public Law 108-173, enacted December 8, 2003.

5 United States Government Accountability Office. Comparative effectiveness research: activities funded by the Patient-Centered Outcomes Research Trust Fund. March 2018. https://www.gao.gov/assets/700/690835.pdf (accessed September 2018).

6 Patient-Centered Outcomes Research Institute. Research Funding. 2018. https://www.pcori.org/sites/default/files/PCORI-Research-Funding.pdf (accessed September 2018).

7 National Institute for Health Research. Health Technology Assessment. https://www.nihr.ac.uk/funding-and-support/funding-for-research-studies/funding-programmes/health-technology-assessment (accessed October 2018).

8 Chalkidou K, Walley T. Using comparative effectiveness research to inform policy and practice in the UK NHS: past, present and future. *Pharmacoeconomics* 2010; **28**: 799–811.

9 National Institute for Health and Care Excellence. What we do. https://www.nice.org.uk/about/what-we-do (accessed October 2018)

10 National Institute for Health and Care Excellence. NICE guidance. https://www.nice.org.uk/guidance (accessed October 2018).

11 Poole-Wilson PA. NICE: guessing is not guidance. National Institute for Clinical Excellence. *Heart* 2002; **87**(6): 510–12.

12 National Institute for Health and Care Excellence. Ixazomib with lenalidomide and dexamethasone for treating relapsed or refractory multiple myeloma. Technology appraisal guidance [TA505]. February 2018.

13 National Institute for Health and Care Excellence. Neuropad for detecting preclinical diabetic peripheral neuropathy. Medical

technologies guidance [MTG38]. September 2018.
14 National Institute for Health and Care Excellence. Intravesical microwave hyperthermia and chemotherapy for non-muscle-invasive bladder cancer. Interventional procedures guidance [IPG628]. September 2018.
15 Sorenson C. Use of comparative effectiveness research in drug coverage and pricing decisions: a six-country comparison. *Issue Brief (Commonw Fund)* 2010; **91**: 1–14.
16 Evans H. Comparative effectiveness in health care reform: lessons from abroad. Heritage Foundation. 2009. http://www.heritage.org/Research/Reports/2009/02/Comparative-Effectiveness-in-Health-Care-Reform-Lessons-from-Abroad (accessed October 2018).
17 Haute autorité de santé (HAS). https://www.has-sante.fr/portail/jcms/r_1455081/en/home-page?portal=r_1455081 (accessed May 2019).
18 Institute for Quality and Efficiency in Health Care. Responsibilities and objectives of IQWiG. https://www.iqwig.de/en/about-us/responsibilities-and-objectives-of-iqwig.2946.html (accessed October 2018).
19 Levy AR, Mitton C, Johnston KM, Harrigan B, Briggs AH. International comparison of comparative effectiveness research in five jurisdictions: insights for the US. *Pharmacoeconomics* 2010; **28**: 813–30.
20 Sweeny K. Key aspects of the Australian Pharmaceutical Benefits Scheme. Working Paper No. 35. Pharmaceutical Industry Project. November 2007. http://vuir.vu.edu.au/15944 (accessed May 2019).
21 Miller LJ, Lu W. These are the economies with the most (and least) efficient health care. https://www.bloomberg.com/news/articles/2018-09-19/u-s-near-bottom-of-health-index-hong-kong-and-singapore-at-top (accessed October 2018).
22 Chalkidou K. From methods to policy: comparative effectiveness research around the globe: a valuable tool for achieving and sustaining universal healthcare. *J Comp Eff Res* 2018; **7**(5): 503–15.
23 Patel I, Rarus R, Tan X, *et al*. Investigation of comparative effectiveness research in Asia, Europe, and North America. *Indian J Pharmacol* 2015; **47**(6): 585–93.
24 Congressional Budget Office. Research on the comparative effectiveness of medical treatments: issues and options for an expanded federal role. 2007. www.cbo.gov/ftpdocs/88xx/doc8891/12-18-ComparativeEffectiveness.pdf (accessed October 2018).
25 Institute of Medicine. Initial national priorities for comparative effectiveness research. Washington, DC: National Academies Press, 2009. http://www.nationalacademies.org/hmd/reports/2009/comparativeeffectivenessresearchpriorities.aspx (accessed December 2010).
26 Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and the Congress. Washington, DC: Department of Health and Human Services, June 2009. https://osp.od.nih.gov/wp-content/uploads/FCCCER-Report-to-the-President-and-Congress-2009.pdf (accessed December 2010).
27 Patient-Centered Outcomes Research Institute. Research we support. https://www.pcori.org/research-results/research-we-support (accessed September 2018).
28 European Commission. Health technology assessment. https://ec.europa.eu/health/technology_assessment/overview_en (accessed October 2018).
29 International Network of Agencies for Health Technology Assessment. Welcome to INAHTA. http://www.inahta.org (accessed October 2018).
30 Slutsky JR, Clancy CM. Patient-centered comparative effectiveness research. *Arch Intern Med* 2010; **170**: 403–4.
31 Benner JS, Morrison MR, Karnes EK, *et al*. An evaluation of recent federal spending on

comparative effectiveness research: priorities, gaps, and next steps. *Health Aff (Millwood)* 2010; **29**: 1768–76.

32 Sox HC, Helfand M, Grimshaw J, *et al*. Comparative effectiveness research: challenges for medical journals. *Trials* 2010; **28**(11): 45.

33 Selker HP. Comparative effectiveness research: medical practice, payments, and politics: the need to retain standards of medical research. *J Gen Intern Med* 2009; **24**: 776–8.

34 Martin DF, Maguire MG, Fine SL. Identifying and eliminating the roadblocks to comparative-effectiveness research. *N Engl J Med* 2010; **8**(363): 105–7.

35 Slone D, Shapiro S, Miettinen OS, *et al*. Drug evaluation after marketing. *Ann Intern Med* 1979; **90**(2): 257–61.

36 Siskind V. The need for randomization in the study of intended effects. *Stat Med* 1985; **4**(3): 387–9.

37 Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? *Am J Med* 1985; **78**(3): 475–80.

38 Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? *Am J Med* 1984; **77**(4): 703–8.

39 Iglehart JK. Prioritizing comparative-effectiveness research: IOM recommendations. *N Engl J Med* 2009; **361**: 325–8. DOI:10.1056/NEJMp0904133

40 Manchikanti L, Falco FJ, Boswell MV, *et al*. Facts, fallacies, and politics of comparative effectiveness research: part I. *Basic considerations. Pain Physician* 2010; **13**: E23–E54.

41 Volpp KG, Das A. Comparative effectiveness: thinking beyond medication A versus medication B. *N Engl J Med* 2009; **361**: 331–3.

42 Lauer MS, Collins FS. Using science to improve the nation's health system: NIH's commitment to comparative effectiveness research. *JAMA* 2010; **303**: 2182–3.

43 Aston G. Comparative effectiveness (why does it matter to you?). *Hosp Health Netw* 2009; **83**: 22–5.

44 Institute of Medicine. Learning what works best: the nation's need for evidence on comparative effectiveness in health care. September 2007. https://www.ncbi.nlm.nih.gov/books/NBK64784 (accessed May 2019).

45 Institute of Medicine. Knowing what works in health care: a roadmap for the nation. January 2008. https://www.nap.edu/catalog/12038/knowing-what-works-in-health-care-a-roadmap-for-the (accessed May 2019).

46 Tang A, Hooper T, Hasan R. A regional survey of chest drains: evidence-based practice? *Postgrad Med J* 1999; **75**: 471–4.

47 Likosky DS, FitzGerald DC, Groom RC, *et al*. Effect of the perioperative blood transfusion and blood conservation in cardiac surgery clinical practice guidelines of the Society of Thoracic Surgeons and the Society of Cardiovascular Anesthesiologists upon clinical practices. *Anesth Analg* 2010; **111**: 316–23.

48 Goldberger JJ, Bonow RO, Cuffe M, *et al*. Beta-blocker use following myocardial infarction: low prevalence of evidence-based dosing. *Am Heart J* 2010; **160**: 435–42 e1.

49 Phillips KA, Marshall DA, Haas JS, *et al*. Clinical practice patterns and cost effectiveness of human epidermal growth receptor 2 testing strategies in breast cancer patients. *Cancer* 2009; **115**: 5166–74.

50 Piccini JP, Hernandez AF, Dai D, *et al*.; Get With the Guidelines Steering Committee and Hospitals. Use of cardiac resynchronization therapy in patients hospitalized with heart failure. *Circulation* 2008; **118**: 926–33.

51 Matlock DD, Peterson PN, Heidenreich PA, *et al*. Regional variation in the use of implantable cardioverter-defibrillators for primary prevention: results from the National Cardiovascular Data Registry. *Circ Cardiovasc Qual Outcomes* 2011; **4**(1): 114–21.

52 Sung JC, Curtis LH, Schulman KA, *et al*. Geographic variations in the use of medical and surgical therapies for benign prostatic hyperplasia. *J Urol* 2006; **175**(3 Pt 1): 1023–7.

53 Warren OJ, Wallace S, de Wit KL, *et al*. Variations in the application of various perfusion technologies in Great Britain and Ireland: a national survey. *Artif Organs* 2010; **34**: 200–5.

54 Lougheed MD, Garvey N, Chapman KR, *et al*. Variations and gaps in management of acute asthma in Ontario emergency departments. *Chest* 2009; **135**: 724–36.

55 Luo X, Pietrobon R, Curtis LH, *et al*. Prescription of nonsteroidal anti-inflammatory drugs and muscle relaxants for back pain in the United States. *Spine* 2004; **29**: E531–E537.

56 Wu AW, Snyder C, Clancy CM, *et al*. Adding the patient perspective to comparative effectiveness research. *Health Aff (Millwood)* 2010; **29**(10): 1863–71.

57 Concannon TW, Fuster M, Saunders T, *et al*. A systematic review of stakeholder engagement in comparative effectiveness and patient-centered outcomes research. *J Gen Intern Med* 2014; **29**(12): 1692–701.

58 Howie L, Hirsch B, Locklear T, *et al*. Assessing the value of patient-generated data to comparative effectiveness research. *Health Aff (Millwood)* 2014; **33**(7): 1220–8.

59 Basu A, Jena AB, Goldman DP, *et al*. Heterogeneity in action: the role of passive personalization in comparative effectiveness research. *Health Econ* 2014; **23**(3): 359–73.

60 Green S, Higgins JPT, Alderson P, *et al*. What is a systematic review? In *Cochrane Handbook for Systematic Reviews of Interventions*, version 5.0.0. London: Cochrane Collaboration, 2008, Chapter 9.1. https://community.cochrane.org/book_pdf/764 (accessed May 2019).

61 Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700.

62 Li T, Vedula SS, Scherer R, *et al*. What comparative effectiveness research is needed? A framework for using guidelines and systematic reviews to identify evidence gaps and research priorities. *Ann Intern Med* 2012; **156**(5): 367–77.

63 Simmonds MC, Higgins JP, Stewart LA, *et al*. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials* 2005; **2**(3): 209–17.

64 Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994; **140**(3): 290–6.

65 Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998; **351**(9096): 123–7.

66 Shapiro S. Meta-analysis/shmeta-analysis. *Am J Epidemiol* 1994; **140**(9): 771–8.

67 Blettner M, Sauerbrei W, Schlehofer B, *et al*. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* 1999; **28**(1): 1–9.

68 Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; **316**(7125): 140–4.

69 Manchikanti L, Datta S, Smith HS, *et al*. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: part 6. Systematic reviews and meta-analyses of observational studies. *Pain Physician* 2009; **12**: 819–50.

70 Song F, Parekh S, Hooper L, *et al*. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010; **14**: iii, ix–xi, 1–193.

71 Broeze KA, Opmeer BC, Bachmann LM, *et al*. Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. *BMC Med Res Methodol* 2009; **9**: 22.

72 Koopman L, van der Heijden GJ, Hoes AW, *et al*. Empirical comparison of subgroup effects in conventional and individual patient data meta-analyses. *Int J Technol Assess Health Care* 2008; **24**: 358–61.

73 Stroup DF, Berlin JA, Morton SC, *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; **283**(15): 2008–12.

74 Shrier I, Boivin JF, Steele RJ, *et al*. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007; **166**(10): 1203–9.

75 Easterbrook PJ, Berlin JA, Gopalan R, *et al*. Publication bias in clinical research. *Lancet* 1991; **337**(8746): 867–72.

76 Schlesselman JJ. Risk of endometrial cancer in relation to use of combined oral contraceptives: a practitioner's guide to meta-analysis. *Hum Reprod* 1997; **12**(9): 1851–63.

77 Man KKC, Ip P, Chan EW, *et al*. Effectiveness of pharmacological treatment for attention-deficit/hyperactivity disorder on physical injuries: a systematic review and meta-analysis of observational studies. *CNS Drugs* 2017; **31**(12): 1043–55.

78 Kirtane AJ, Gupta A, Lyengar S, *et al*. Safety and efficacy of drug-eluting and bare metal stents: comprehensive meta-analysis of randomized trials and observational studies. *Circulation* 2009; **119**: 3198–206.

79 AHRQ's methods research paper on comparative effectiveness review methods: clinical heterogeneity. AHRQ publication no. 10-EHC-070-EF. September 2018. https://www.ncbi.nlm.nih.gov/books/NBK53310 (accessed May 2019).

80 Institute of Medicine. Standards for systematic reviews of clinical effectiveness research. finding what works in health care: standards for systematic reviews. September 2018. https://www.ncbi.nlm.nih.gov/books/NBK209526 (accessed May 2019).

81 Shojania KG, Sampson M, Ansari MT, *et al*. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007; **147**: 224–33.

82 Dreyer NA, Tunis SR, Berger M, *et al*. Why observational studies should be among the tools used in comparative effectiveness research. *Health Aff (Millwood)* 2010; **29**: 1818–25.

83 Teutsch SM, Berger ML, Weinstein MC. Comparative effectiveness: asking the right questions, choosing the right method. *Health Affairs* 2005; **24**: 128–32.

84 Sox HC, Goodman SN. The methods of comparative effectiveness research. *Annu Rev Public Health* 2012; **33**: 425–45.

85 Strom BL. Methodologic challenges to studying patient safety and comparative effectiveness. *Med Care* 2007; **45**(10 Suppl 2): S13–S15.

86 Schneeweiss S, Rassen JA, Glynn RJ, *et al*. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009; **20**(4): 512–22.

87 Kumamaru H, Jalbert JJ, Setoguchi S. Incremental utility of additional clinical information and propensity score methods in controlling for severe confounding by indication. *Pharmacoepidemiol Drug Saf* 2018; **27**: 489.

88 Green SB, Byar DP. Using observational data from registries to compare treatments: the fallacy of omnimetrics. *Stat Med* 1984; **3**: 361–70.

89 Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**: 360–72.

90 Schneeweiss S, Patrick AR, Sturmer T, *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care* 2007; **45**(10 Suppl 2): S131–S142.

91 McGrath LJ, Ellis AR, Brookhart MA. controlling time-dependent confounding by health status and frailty: restriction versus statistical adjustment. *Am J Epidemiol* 2015; **182**(1): 17–25.

92 Setoguchi S, Gerhard T. Comparator selection. In Velentgas P, Dreyer NA, Nourjah P, *et al*., eds., *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Rockville, MD: AHRQ, 2013, Chapter 5. https://www.ncbi.nlm.nih.gov/books/NBK126184 (accessed May 2019).

93 Lao KS, Chui CS, Man KK, *et al*. Medication safety research by observational study design. *Int J Clin Pharm* 2016; **38**(3): 676–84.

94 Majumdar SR, McAlister FA, Eurich DT, *et al*. Statins and outcomes in patients admitted to hospital with community acquired pneumonia: population based prospective cohort study. *BMJ* 2006; **333**(7576): 999.

95 Humphrey LL, Chan BK, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med* 2002; **137**(4): 273–84.

96 Brookhart MA, Patrick AR, Dormuth C, *et al*. Adherence to lipid-lowering therapy and the use of preventive health services: an investigation of the healthy user effect. *Am J Epidemiol* 2007; **166**(3): 348–54.

97 Ray WA, Daugherty JR, Griffin MR. Lipid-lowering agents and the risk of hip fracture in a Medicaid population. *Inj Prev* 2002; **8**(4): 276–9.

98 Setoguchi S, Avorn J, Schneeweiss S. Statins and the risk of colorectal cancer. *N Engl J Med* 2005; **353**(9): 952–4.

99 Kinjo M, Chia-Cheng Lai E, Korhonen MJ, *et al*. Potential contribution of lifestyle and socioeconomic factors to healthy user bias in antihypertensives and lipid-lowering drugs. *Open Heart* 2017; **4**(1): e000417.

100 Blood Pressure Lowering Treatment Trialists Collaboration, Turnbull F, Neal B, *et al*. Effects of different regimens to lower blood pressure on major cardiovascular events in older and younger adults: meta-analysis of randomised trials. *BMJ* 2008; **336**(7653): 1121–3.

101 Mazzaglia G, Ambrosioni E, Alacqua M, *et al*. Adherence to antihypertensive medications and cardiovascular morbidity among newly diagnosed hypertensive patients. *Circulation* 2009; **120**(16): 1598–605.

102 Setoguchi S, Warner Stevenson L, Stewart GC, *et al*. Influence of healthy candidate bias in assessing clinical effectiveness for implantable cardioverter-defibrillators: cohort study of older patients with heart failure. *BMJ* 2014; **348**: g2866.

103 Halava H, Korhonen MJ, Huupponen R, *et al*. Lifestyle factors as predictors of nonadherence to statin therapy among patients with and without cardiovascular comorbidities. *CMAJ* 2014; **186**(12): E449–E456.

104 Dublin S, Jackson ML, Nelson JC, *et al*. Statin use and risk of community acquired pneumonia in older people: population based case-control study. *BMJ* 2009; **338**: b2137.

105 Glynn RJ, Knight EL, Levin R, *et al*. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology* 2001; **12**(6): 682–9.

106 Chen CY, Stevenson LW, Stewart GC, *et al*. Real world effectiveness of primary implantable cardioverter defibrillators implanted during hospital admissions for exacerbation of heart failure or other acute co-morbidities: cohort study of older patients with heart failure. *BMJ* 2015; **351**: h3529.

107 Christen P. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. New York: Springer, 2012.

108 Setoguchi S, Solomon DH, Glynn RJ, *et al*. Agreement of diagnosis and its date for hematologic malignancies and solid tumors between Medicare claims and cancer registry data. *Cancer Causes Control* 2007; **18**(5): 561–9.

109 Douglas PS, Brennan JM, Anstrom KJ, *et al*. Clinical effectiveness of coronary stents in elderly persons: results from 262,700 Medicare patients in the American College of Cardiology-National Cardiovascular Data Registry. *J Am Coll Cardiol* 2009; **53**(18): 1629–41.

110 Weintraub WS, Grau-Sepulveda MV, Weiss JM, *et al*. Comparative effectiveness of revascularization strategies. *N Engl J Med* 2012; **366**(16): 1467–76.

111 Jalbert JJ, Gerhard-Herman MD, Nguyen LL, *et al*. Relationship between physician and hospital procedure volume and mortality after carotid artery stenting among Medicare beneficiaries. *Circ Cardiovasc Qual Outcomes* 2015; **8**(6 Suppl 3): S81–S89.

112 Lichtenstein P, Halldner L, Zetterqvist J, *et al*. Medication for attention deficit-hyperactivity disorder and criminality. *N Engl J Med* 2012; **367**(21): 2006–14.

113 Zhu Y, Chen CY, Matsuyama Y, *et al*. Comparative validity of methods to select appropriate cutoff weight for probabilistic linkage without unique personal identifiers. *Pharmacoepidemiol Drug Saf* 2016; **25**(4): 444–52.

114 Zhu Y, Matsuyama Y, Ohashi Y, *et al*. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *J Biomed Inform* 2015; **56**: 80–6.

115 Setoguchi S, Zhu Y, Jalbert JJ, *et al*. Validity of deterministic record linkage using multiple indirect personal identifiers: linking a large registry to claims data. *Circ Cardiovasc Qual Outcomes* 2014; **7**(3): 475–80.

116 Curtis JR, Chen L, Bharat A, *et al*. Linkage of a de-identified United States rheumatoid arthritis registry with administrative data to facilitate comparative effectiveness research. *Arthritis Care Res (Hoboken)* 2014; **66**(12): 1790–8.

117 O'Brien E, O'Brien S, Setoguchi S. Patient perspectives on the linkage of health data for clinical research: insights from a survey in the United States. *World Congress of Epidemiology*. International Epidemiological Association. Saitama, Japan, August 2017.

118 Ipsos MORI. The one-way mirror: public attitudes to commercial access to health data. Report prepared for the Wellcome Trust. March 2016. https://wellcome.ac.uk/what-we-do/our-work/our-policy-work-using-patient-data-research (accessed October 2018).

119 Coloma PM, Schuemie MJ, Trifiro G, *et al*. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf* 2011; **20**(1): 1–11.

120 Platt R, Wilson M, Chan KA, *et al*. The new Sentinel Network: improving the evidence of medical-product safety. *N Engl J Med* 2009; **361**(7): 645–7.

121 Andersen M, Bergman U, Choi NK, *et al*. The Asian Pharmacoepidemiology Network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf* 2013; **22**(7): 700–4.

122 Lai EC, Ryan P, Zhang Y, *et al*. Applying a common data model to Asian databases for multinational pharmacoepidemiologic studies: opportunities and challenges. *Clin Epidemiol* 2018; **10**: 875–85.

123 Suissa S, Henry D, Caetano P, *et al*. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med* 2012; **6**(4): e134–e140.

124 PCORnet Front Door. https://pcornet.org/frontdoor (accessed May 2019).

125 ADAPTABLE. The Aspirin Study. PCORnet demonstration trial. http://theaspirinstudy.org (accessed May 2019).

126 Lai E, Stang P, Kao Yang Y, *et al*. International multi-database pharmacoepidemiology: potentials and

pitfalls. *Curr Epidemiol Rep* 2015; **2**(4): 229–38.

127 Lai EC, Shin J, Kubota K, *et al*. Comparative safety of NSAIDs for gastrointestinal events in Asia-Pacific populations: a multi-database, international cohort study. *Pharmacoepidemiol Drug Saf* 2018; **27**(11): 1223–30.

128 Sentinel Initiative. Distributed database and common data model. https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model (accessed September 2018).

129 Observational Health Data Sciences and Informatics. OMOP common data model. https://www.ohdsi.org/data-standardization/the-common-data-model (accessed September 2018).

130 PCORnet. PCORnet Common Data Model (CDM). https://pcornet.org/pcornet-common-data-model (accessed September 2018).

131 Maro JC, Platt R, Holmes JH, *et al*. Design of a national distributed health data network. *Ann Intern Med* 2009; **151**(5): 341–4.

132 Chan KA, HMO Research Network CERT Patient Safety Study Investigators. Development of a multipurpose dataset to evaluate potential medication errors in ambulatory settings. In Henriksen K, Marks ES, Lewin DI, eds, *Advances in Patient Safety: From Research to Implementation. Vol. 2: Concepts and Methodology*. AHRQ Publication No. 05-0021-2. Rockville, MD: AHRQ, 2005, pp. 225–38.

133 Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967; **20**(8): 637–48.

134 Macpherson H. Pragmatic clinical trials. *Complement Ther Med* 2004; **12**: 136–40.

135 Relton C, Torgerson D, O'Cathain A, *et al*. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. *BMJ* 2010; **340**: c1066.

136 Brass EP. The gap between clinical trials and clinical practice: the use of pragmatic clinical trials to inform regulatory decision making. *Clin Pharmacol Ther* 2010; **87**: 351–5.

137 Alford L. On differences between explanatory and pragmatic clinical trials. *New Zeal J Physiotherapy* 2006; **35**(1): 12–16.

138 Tinetti ME, Studenski SA. Comparative effectiveness research and patients with multiple chronic conditions. *N Engl J Med* 2011; **364**(26): 2478–81.

139 Luce BR, Kramer JM, Goodman SN, *et al*. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med* 2009; **151**: 206–9.

140 Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA* 2002; **288**(23): 2981–97.

141 Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) Data Coordinating Center. Dose of reality: study questions high-cost drugs for hypertension. https://ccct.sph.uth.tmc.edu/ALLHAT/News/HighCostDrugs.aspx (accessed September 2018).

142 Strom BL, Eng SM, Faich G, *et al*. Comparative mortality associated with ziprasidone and olanzapine in real-world use among 18,154 patients with schizophrenia: the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC). *Am J Psychiatry* 2011; **168**: 193–201.

143 Strom BL, Faich GA, Reynolds RF, *et al*. The Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC): design and baseline subject characteristics. *J Clin Psychiatry* 2008; **69**: 114–21.

144 Olfson M, Marcus SC. Decline in placebo-controlled trial results suggests new directions for comparative effectiveness

research. *Health Aff (Millwood)* 2013; **32**(6): 1116–25.

145 Charlson ME, Horwitz RI. Applying results of randomised trials to clinical practice: impact of losses before randomisation. *Br Med J (Clin Res Ed)* 1984; **289**: 1281–4.

146 van Staa TP, Dyson L, McCann G, *et al*. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess* 2014; **18**(43): 1–146.

147 van Staa TP, Klungel O, Smeeth L. Use of electronic healthcare records in large-scale simple randomized trials at the point of care for the documentation of value-based medicine. *J Intern Med* 2014; **275**(6): 562–9.

148 Fiore LD, Brophy M, Ferguson RE, *et al*. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. *Clin Trials* 2011; **8**(2): 183–95.

149 Fiore LD, Lavori PW. Integrating randomized comparative effectiveness research with patient care. *N Engl J Med* 2016; **374**(22): 2152–8.

150 ISPE. Guidelines for Good Pharmacoepidemiology Practices (GPP). *Pharmacoepidemiol Drug Saf* 2008; **17**: 200–8.

151 Berger ML, Mamdani M, Atkins D, *et al*. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part I. *Value Health* 2009; **12**: 1044–52.

152 Cox E, Martin BC, Van Staa T, *et al*. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report – Part II. *Value Health* 2009; **12**: 1053–61.

153 Velentgas P, Dreyer NA, Nourjah P, *et al*., eds. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Rockville, MD: AHRQ, 2013. https://www.ncbi.nlm.nih.gov/books/NBK126190 (accessed May 2019).

154 PCORI (Patient-Centered Outcomes Research Institute) Methodology Committee. The PCORI methodology report. https://www.pcori.org/research-results/about-our-research/research-methodology/pcori-methodology-standards (accessed September 2018).

155 European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on methodological standards in pharmacoepidemiology (Revision 3). EMA/95098/2010. 2014. http://www.encepp.eu/standards_and_guidances (accessed October 2018).

156 Gibbons RJ, Gardner TJ, Anderson JL, *et al*. The American Heart Association's principles for comparative effectiveness research: a policy statement from the American Heart Association. *Circulation* 2009; **119**(22): 2955–62.

157 Johnson ML, Crown W, Martin BC, *et al*. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part III. *Value Health* 2009; **12**(8): 1062–73.

158 Willke RJ, Mullins CD. "Ten commandments" for conducting comparative effectiveness research using

"real-world data." *J Manag Care Pharm* 2011; **17**(9 Suppl A): S10–S15.

159 Dreyer NA, Schneeweiss S, McNeil BJ, *et al.* GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care* 2010; **16**(6): 467–71.

160 Luce BR, Drummond MF, Dubois RW, *et al.* Principles for planning and conducting comparative effectiveness research. *J Comp Eff Res* 2012; **1**(5): 431–40.

161 Morton JB, McConeghy R, Heinrich K, *et al.* Consensus of recommendations guiding comparative effectiveness research methods. *Pharmacoepidemiol Drug Saf* 2016; **25**(12): 1354–60.

162 Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998; **316**(7127): 285.

163 Godwin M, Ruhland L, Casson I, *et al.* Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003; **3**: 28.

164 Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet* 2005; **365**(9453): 82–93.

165 Devereaux PJ, Bhandari M, Clarke M, *et al.* Need for expertise based randomised controlled trials. *BMJ* 2005; **330**(7482): 88.

166 Ernst E, Canter PH. Limitations of "pragmatic" trials. *Postgrad Med J* 2005; **81**(954): 203.

167 Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 2009; **10**: 37.

168 Kent DM, Kitsios G. Against pragmatism: on efficacy, effectiveness and the real world. *Trials* 2009; **10**: 48.

169 Eldridge S. Pragmatic trials in primary health care: what, when and how? *Fam Pract* 2010; **27**(6): 591–2.

170 Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 2011; **13**(2): 217–24.

171 Ware JH, Hamel MB. Pragmatic trials: guides to better patient care? *N Engl J Med* 2011; **364**(18): 1685–7.

172 Mitka M. FDA advisory decision highlights some problems inherent in pragmatic trials. *JAMA* 2011; **306**(17): 1851–2.

173 Chalkidou K, Tunis S, Whicher D, *et al.* The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clin Trials* 2012; **9**(4): 436–46.

174 Ratner J, Mullins D, Buesching DP, *et al.* Pragmatic clinical trials: U.S. payers' views on their value. *Am J Manag Care* 2013; **19**(5): e158–e165.

175 Kim SY, Miller FG. Informed consent for pragmatic trials: the integrated consent model. *N Engl J Med* 2014; **370**(8): 769–72.

176 Sugarman J, Califf RM. Ethics and regulatory complexities for pragmatic clinical trials. *JAMA* 2014; **311**(23): 2381–2.

177 Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; **3**(4): 409–22.

178 Thorpe KE, Zwarenstein M, Oxman AD, *et al.* A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009; **62**(5): 464–75.

179 Anderson ML, Griffin J, Goldkind SF, *et al.* The Food and Drug Administration and pragmatic clinical trials of marketed medical products. *Clin Trials* 2015; **12**(5): 511–19.

180 Loudon K, Treweek S, Sullivan F, *et al.* The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015; **350**: h2147.

181 Zwarenstein M, Treweek S, Gagnier JJ, *et al.* Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008; **337**: a2390.

182 Carey TS, Melvin CL, Ranney LM. Extracting key messages from systematic reviews. *J Psychiatr Pract* 2008; **14**(Suppl 1): 28–34.

183 Waters E, Armstrong R. Knowledge synthesis, translation and exchange in public

health and health promotion: the role of a Cochrane Review Group. *Promot Educ* 2007; **14**: 34–5.

184 Institute of Medicine. IOM Standards for Developing Trustworthy Clinical Practice Guidelines. February 2011. http://www.nationalacademies.org/hmd/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx (accessed May 2019).

185 Kanouse DE, Kallich JD, Kahan JP. Dissemination of effectiveness and outcomes research. *Health Policy* 1995; **34**: 167–92.

186 Fenton JJ, Weyrich MS, Durbin S, *et al*. Prostate-specific antigen-based screening for prostate cancer: a systematic evidence review for the US Preventive Services Task Force: Evidence synthesis No. 154. AHRQ publication 17-05229-EF-1. Rockville, MD: AHRQ, 2018.

187 Etzioni R, Penson DF, Legler JM, *et al*. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *J Natl Cancer Inst* 2002; **94**(13): 981–90.

188 Pollack A. The minimal impact of a big hypertension study. *New York Times, Nov* **28**, 2008. https://www.nytimes.com/2008/11/28/business/28govtest.html (accessed May 2019).

189 Timbie JW, Fox DS, Van Busum K, *et al*. Five reasons that many comparative effectiveness studies fail to change patient care and clinical practice. *Health Aff (Millwood)* 2012; **31**(10): 2168–75.

190 Ayanian JZ, Berwick DM. Do physicians have a bias toward action? A classic study revisited. *Med Decis Making* 1991; **11**(3): 154–8.

191 Emanuel EJ. What cannot be said on television about health care. *JAMA* 2007; **297**(19): 2131–3.

192 Institute of Medicine. Clinical practice guidelines we can trust. Washington, DC: National Academies Press, 2011. http://www.nationalacademies.org/hmd/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx (accessed May 2019).

193 Mendelson A, Kondo K, Damberg C, *et al*. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. *Ann Intern Med* 2017; **166**(5): 341–53.

194 Patient-Centered Outcomes Research Institute. Putting evidence to work. https://www.pcori.org/research-results/putting-evidence-work (accessed September 2018).

195 Patient-Centered Outcomes Research Institute. Dissemination and implementation. https://www.pcori.org/sites/default/files/PCORI-Dissemination-Implementation.pdf (accessed September 2018).

196 Gaige M. Comparative effectiveness research: a valuable tool for health-care reform? *Penn Medicine* 2009/2010; **Winter**: 24–31, p. 29.

197 Mitka M. US government kicks off program for comparative effectiveness research. *JAMA* 2010; **304**: 2230–1.

198 Wilensky GR. Cost-effectiveness information: yes, it's important, but keep it separate, please! *Ann Intern Med* 2008; **148**: 967–8.

199 Centers for Medicare & Medicaid Services. Projected national health expenditure data. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsProjected.html (accessed September 2018).

200 Hochman M, McCormick D. Characteristics of published comparative effectiveness studies of medications. *JAMA* 2010; **303**: 951–8.

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

202 CORDIS. Comparative Effectiveness research on Psychiatric HOSpitalisation by record LINKage of large administrative data sets. https://cordis.europa.eu/project/rcn/185457/factsheet/en (accessed May 2019).

203 Health and Medical Research Fund. About HMRF. https://rfs1.fhb.gov.hk/english/funds/funds_hmrf/funds_hmrf_abt/funds_hmrf_abt.html (accessed October 2018).

204 Sherman RE, Anderson SA, Dal Pan GJ, *et al*. Real-world evidence: what is it and what can it tell us? *N Engl J Med* 2016; **375**(23): 2293–7.

205 US Food and Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices: guidance for industry and Food and Drug Administration Staff. 2017. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM513027.pdf (accessed November 2018).

206 US Food and Drug Administration. Real world evidence. https://www.fda.gov/scienceresearch/specialtopics/realworldevidence/default.htm (accessed November 2018).

207 Plueschke K, McGettigan P, Pacurariu A, *et al*. EU-funded initiatives for real world evidence: descriptive analysis of their characteristics and relevance for regulatory decision-making. *BMJ Open* 2018; **8**(6): e021864.

208 Pfizer Inc. BMS-Pfizer alliance to unveil real-world data analyses – cost, safety and comparative effectiveness findings associated with oral anticoagulants in non-valvular atrial fibrillation. 2017. https://investors.pfizer.com/investor-news/press-release-details/2017/BMS-Pfizer-Alliance-to-Unveil-Real-World-Data-Analyses---Cost-Safety-and-Comparative-Effectiveness-Findings-Associated-with-Oral-Anticoagulants-in-Non-Valvular-Atrial-Fibrillation/default.aspx (accessed November 2018).

209 Tepper PG, Mardekian J, Masseria C, *et al*. Real-world comparison of bleeding risks among non-valvular atrial fibrillation patients prescribed apixaban, dabigatran, or rivaroxaban. *PLoS One* 2018; **13**(11): e0205989.

210 Flacco ME, Manzoli L, Boccia S, *et al*. Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor. *J Clin Epidemiol* 2015; **68**(7): 811–20.

211 European Commission. Medicines for children: major developments. https://ec.europa.eu/health/human-use/paediatric-medicines/developments_en (accessed October 2018).

212 Murray MD. Curricular considerations for pharmaceutical comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2011; **20**(8): 797–804.

213 Marie Skłodowska-Curie Actions. Research fellowship programme. https://ec.europa.eu/research/mariecurieactions (accessed October 2018).

214 National Institute for Health Research. NIHR fellowships. https://www.nihr.ac.uk/funding-and-support/funding-for-training-and-career-development/fellowship-programme.htm (accessed October 2018).

215 Osler W. *Aequanimitas: With Other Addresses to Medical Students, Nurses and Practitioners of Medicine*. Birmingham, AL: Classics of Medicine Library, 1987. Originally published 1904.

216 Egnew TR. The art of medicine: seven skills that promote mastery. *Fam Pract Manag* 2014; **21**(4): 25–30.

217 Kerridge I, Lowe M, Henry H. Ethics and evidence based medicine. *BMJ* 1998; **316**: 1151–3.

218 Greenhalgh T, Howick J, Maskrey N, *et al*. Evidence based medicine: a movement in crisis? *BMJ* 2014; **348**: g3725.