How Good Is "Evidence" from Clinical Studies of Drug Effects and Why Might Such Evidence Fail in the Prediction of the Clinical Utility of Drugs?

Huseyin Naci[1] and John P.A. Ioannidis[2,3,4]

[1]LSE Health, London School of Economics and Political Science, London WC2A 2AE, United Kingdom; email: h.naci@lse.ac.uk

[2]Stanford Prevention Research Center, Department of Medicine, [3]Department of Health Research and Policy, and [4]Meta-Research Innovation Center at Stanford (METRICS), Stanford University School of Medicine, Stanford, California 94305; email: jioannid@stanford.edu

Annu. Rev. Pharmacol. Toxicol. 2015. 55:169–89

First published online as a Review in Advance on August 21, 2014

The *Annual Review of Pharmacology and Toxicology* is online at pharmtox.annualreviews.org

This article's doi: 10.1146/annurev-pharmtox-010814-124614

Keywords

treatment effectiveness, drug therapy, clinical trials, bias, precision medicine, prediction in pharmacology

Abstract

Promising evidence from clinical studies of drug effects does not always translate to improvements in patient outcomes. In this review, we discuss why early evidence is often ill suited to the task of predicting the clinical utility of drugs. The current gap between initially described drug effects and their subsequent clinical utility results from deficits in the design, conduct, analysis, reporting, and synthesis of clinical studies—often creating conditions that generate favorable, but ultimately incorrect, conclusions regarding drug effects. There are potential solutions that could improve the relevance of clinical evidence in predicting the real-world effectiveness of drugs. What is needed is a new emphasis on clinical utility, with nonconflicted entities playing a greater role in the generation, synthesis, and interpretation of clinical evidence. Clinical studies should adopt strong design features, reflect clinical practice, and evaluate outcomes and comparisons that are meaningful to patients. Transformative changes to the research agenda may generate more meaningful and accurate evidence on drug effects to guide clinical decision making.

BACKGROUND

There is widespread concern that initially promising evidence from clinical studies of drug effects is not invariably correlated with tangible gains in health (1). Benefit-harm assessments at the time of market entry are often not predictive of the observed effectiveness and safety of drugs once they have been adopted in clinical practice (2). As documented in the clinical literature, early studies tend to yield exaggerated drug effects (3); the effects become much smaller in later studies. Not infrequently, initial research findings are also found to be invalid (4), potentially resulting in reversals of standards of care (5). Although controversies most commonly arise from highly cited studies with weaker designs—the most important predictor of reversal of standards of care is the adoption of a practice based on evidence from studies with weak designs (6)—findings of studies with stronger designs are also sometimes challenged and refuted (7).

The dangers of failing to predict the utility of drugs on the basis of evidence from clinical studies include the adoption of not sufficiently effective (or even totally ineffective) and potentially harmful products and the resulting grave consequences for health, health-care costs, and the credibility of research (8). Recent widely publicized examples of drugs with discrepancies between their utility in early clinical studies and their utility in real-world settings include rofecoxib (Vioxx®) (9, 10), varenicline (Champix®/Chantix®) (11), and gabapentin (Neurontin®) (12), for which later evidence revealed a questionable benefit-harm profile in some patient groups. The rise and fall of these commonly used products underscore the problems in the clinical studies of drug effects.

Such problems stem from important flaws in the design, conduct, analysis, reporting, and synthesis of clinical studies of drug effects. In this article, we provide an overview of the limitations of evidence from clinical studies and discuss why such evidence often fails to predict the clinical utility of drugs. We offer potential remedies to improve the relevance of clinical evidence in predicting the real-world effectiveness and safety of drugs. To improve the utility of drugs, studies should be designed, conducted, analyzed, and interpreted in ways that minimize bias; include patient populations that resemble those in clinical practice; evaluate outcomes and head-to-head comparisons that matter to patients; and be fully disseminated and made available for further independent scrutiny in a timely manner—conditions that are unmet in the current research environment (13, 14).

PROBLEMS IN THE DESIGN OF CLINICAL STUDIES OF DRUG EFFECTS

Clinical studies are often not designed to evaluate the real-world effectiveness and safety of drugs; rather, they are designed to meet regulatory demands and commercial interests. Regulators and the pharmaceutical industry dictate the design of clinical studies, thereby shaping the nature of available evidence on drug effects. Such practices can create conditions for generating favorable estimates for drug effects (15). Under publish-or-perish pressures, researchers may contribute to the problem by failing to learn from existing and ongoing research activities and by prioritizing the practicality offered by studies with weak designs. Taken together, design features of studies of drug effects make resulting clinical evidence ill-suited to the task of predicting the therapeutic utility of drugs (**Figure 1**).

Insufficient Consideration of Prior Existing and Ongoing Evidence

Continuously updated systematic assessments of what is already known can serve as a guide for future research (16). Yet many clinical studies are designed, conducted, and interpreted without adequate consideration of prior existing and ongoing research (16–18). Published reports may

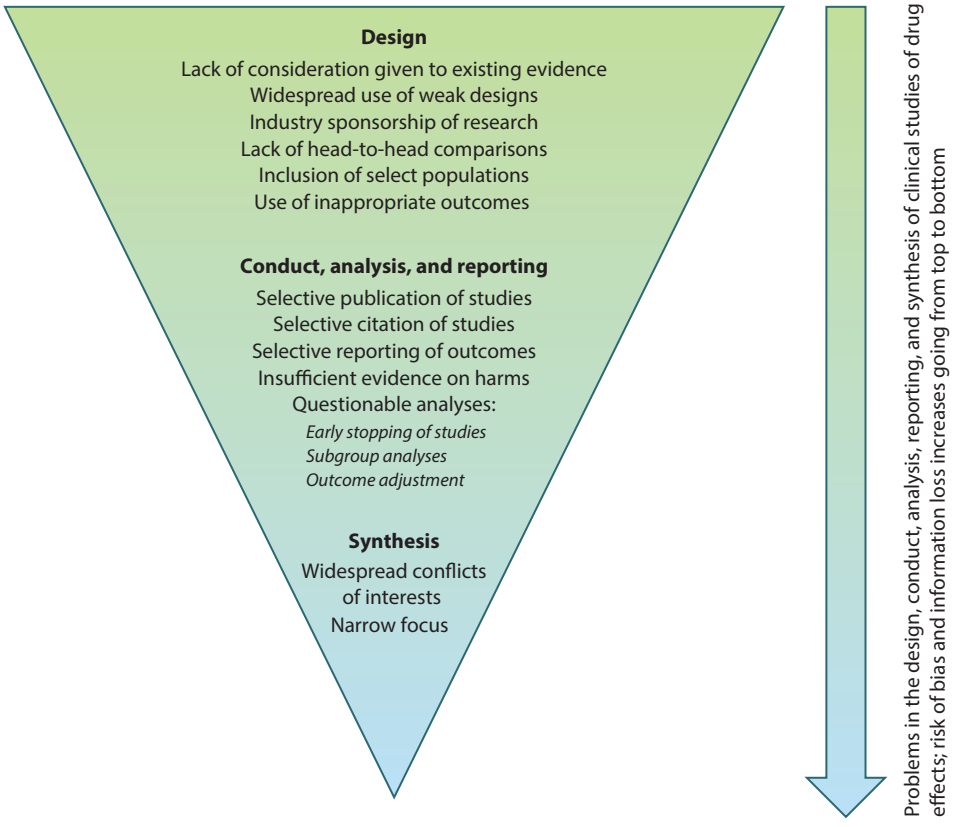

Figure 1

Current clinical research practices are geared toward showing favorable estimates of drug effects.

not systematically cite previous studies: According to a recent review of citation patterns, fewer than a quarter of previous findings were cited in more recent studies addressing similar clinical questions (17). In addition, fewer than half of researchers were even aware of relevant reviews of available, existing evidence when they designed their studies (19). Incomplete and selective citation of previous work may explain why dubious claims from highly cited research persist and continue to be supported in clinical practice despite contradictory evidence from later studies (20, 21).

A careful evaluation of prior knowledge is critical to ensuring that clinical studies have strong design features and sufficient statistical power to correctly estimate drug effects. Yet fewer than a quarter of studies consult the prior existing evidence for sample size calculations (22). Unsurprisingly, investigators embarking on new clinical studies without first consulting reviews and syntheses of existing evidence risk failing to address questions that are truly novel and important (16). Such insufficient consideration of existing evidence exacerbates the lopsided nature of the evidence base, with certain drugs and therapeutic areas receiving disproportionate research attention over others. For example, there are substantially more ongoing studies of anti–tumor necrosis factor-α (anti-TNF-α) agents than those that are completed or published, and the research agenda already comprises more than 200 trials (23, 24). In such cases, investing in yet another clinical study would more likely serve marketing purposes rather than address critical, unanswered clinical questions.

Use of Weak Study Designs

Randomized designs are widely considered the gold standard for determining whether a drug works or is more effective than an alternative (25). Randomization removes the potential for bias in the assignment of patients to one drug or another by introducing unpredictability. Randomized designs have high internal validity, i.e., can ensure that the observed treatment effect can be attributed to the drug of interest. In nonrandomized designs, however, the decision of which treatment to receive or administer is chosen by the patient or the clinician rather than randomly assigned. This can result in differences in the baseline characteristics of patient groups receiving different drugs. Such "confounding by indication" is a threat to the internal validity of nonrandomized studies, potentially rendering their findings biased.

Indeed, systematic assessments have documented that the findings of randomized and nonrandomized studies often differ, and the magnitude of this difference cannot be predicted (26). An empirical evaluation found that five of the six most cited nonrandomized studies were refuted or had exaggerated effects when tested in randomized studies (27). Strikingly, fewer than half of clinical studies used in early drug development are randomized despite such concerns (28). In addition, nonrandomized studies are commonly used to make inferences about drug effects after market entry (29). The reason for this strategy is that nonrandomized designs have several practical advantages over randomized studies, including speed, the potential availability of a large number of patients, and low cost. The authors of nonrandomized studies often extrapolate their results to make treatment recommendations without referring to (or calling for) randomized studies (30).

Although placed at the top of evidence hierarchies, randomized studies also have important limitations. Causal inferences from randomized studies can be jeopardized by limitations in their design, leading to bias (31–35). For example, failure to conceal allocation and to "blind" participants and personnel is associated with exaggerated treatment effects, on average, especially when subjective outcomes are measured (36, 37).

Industry Sponsorship of Research

Most clinical drug research is sponsored by the pharmaceutical industry (38, 39). Financial conflicts of interest exist with such sponsorship: Understandably, favorable study results present a strong financial incentive for pharmaceutical companies.

Do conflicts of interest that intertwine industry sponsors and researchers influence the outcomes of clinical studies? Industry-sponsored studies are more likely to favor the product than is research funded by other sources (40–42). Although it was hypothesized that pharmaceutical companies conduct poor-quality research, there is now conclusive evidence that the methodological quality of industry-sponsored studies is at least as good as (or even better than) that of non-industry-funded studies (40–43). Therefore, the apparent discrepancy seems to lie in the interpretation of the findings rather than in the conduct of the study or in the analysis of the results (24, 44, 45).

More problematic is that pharmaceutical companies may exclusively sponsor trials that they believe will have favorable conclusions for their products. Previous studies indicate that each pharmaceutical company generates a clinical research agenda that is strongly focused on its own products, whereas head-to-head comparisons with competitor drugs are uncommon (46). When different drugs are compared, companies preferentially sponsor trials that compare their products to "straw-man comparators" (41, 47, 48), thus distorting the evidence base in favor of the company's products.

Lack of Head-to-Head Comparisons

More than a thousand new drugs have reached the market in the United States during the past six decades (49). Despite the large numbers of available drug options (50), direct head-to-head comparisons are rarely available, leaving patients, clinicians, and policymakers without adequate information on the comparative clinical effectiveness and safety of seemingly similar drugs (51). Currently, the US Food and Drug Administration (FDA) does not require information regarding a drug's comparative effectiveness and safety against existing alternatives, instead requesting that pharmaceutical companies perform placebo-controlled studies (52). For pharmaceutical companies—likely reluctant to conduct head-to-head comparisons for fear of finding that a competitor's products might outperform their own—the lack of regulatory interest in comparative evidence at the time of drug approval serves as a disincentive. This interplay of regulatory expectations and pharmaceutical company interests may create asymmetries in the evidence base, favoring certain products.

As a result, there is a paucity of clinical studies that include head-to-head comparisons of new and existing drugs. Placebo-controlled trials dominate published clinical research (53, 54), in some cases constituting more than 90% of the available evidence (55). Although the situation seems to have improved over the past decade (56), partly owing to increasing pressure from third-party payers and pharmacy-benefit managers, comparative data are available for approximately half of new drugs at the time of market entry, with considerable variability across therapeutic areas (56–58).

Choice of Patient Populations

Early clinical studies—and particularly those conducted to achieve market entry—may have little relevance in real-world settings (59). When licensing a new drug, regulators rarely consider its effectiveness in the context of real-world use. Rather, drugs receive approvals on the basis of results from clinical studies in which patient enrollment is restricted by numerous inclusion and exclusion criteria. Such stipulations include, among many others, restrictions on comorbidities and concomitant medications, which eliminate the participation of certain populations (60). As a result, drugs [for example, cyclooxygenase inhibitors (61)] are approved on the basis of studies of very narrow clinical populations but are subsequently used much more broadly in clinical practice among populations for whom the benefit-harm balance may be much different. In the transition from such narrow samples to broad populations, beneficial effects of drugs often diminish while their adverse effects increase (2).

It is unknown whether (and sometimes highly unlikely that) drugs that are successful among select samples of patients will do well in real-world populations (62). Our ability to predict the clinical utility of drugs is particularly hindered for certain patient populations: Women, children, the elderly, and those with common comorbidities are frequently underrepresented or excluded from clinical studies (63). For example, elderly populations may be underrepresented in clinical studies of drugs that they are most likely to receive (64), including those for osteoarthritis (60), cancer (65), cardiovascular disease (66), and heart failure (67). As drug effects can vary depending on sex, age, and comorbidity status, study entry restrictions can undermine efforts to predict the clinical utility of drugs (68). Such exclusions impair the applicability of early study results and highlight a need for careful reconsideration of exclusion criteria in clinical studies (63).

Choice of Outcomes

The choice of outcomes in clinical studies of drugs adds to the complexity of predicting their real-world utility. Researchers rarely evaluate the impact of drugs on outcomes that matter to

patients, i.e., patient-centered outcomes. These include functional, social, and emotional well-being, in addition to survival (16, 69). Clinical studies commonly include composites of multiple outcomes that increase statistical power but are not clinically meaningful and that may complicate the interpretability of study findings (24, 70–73).

In addition, evaluations of drug effects are often based on short-term surrogate outcome measures (74). Such measures are sometimes warranted in predicting the long-term effects of drugs on clinical outcomes (75), but reliance on them does not necessarily result in accurate predictions (76). For example, accumulating evidence on ezetimibe has failed to demonstrate that effective control of low-density lipoprotein cholesterol (a widely used surrogate) translates to a reduction in the risk of cardiovascular events and mortality (77–80).

Systematic evaluations of emerging surrogate outcomes suggest that many findings are prone to bias (81). Indeed, studies using surrogate outcomes often exaggerate drug effects compared with those using patient-relevant clinical outcomes (82, 83). In addition, clinical studies evaluating surrogate outcomes are twice as likely to report positive treatment effects as those reporting clinical outcomes (82).

The use of surrogate measures is particularly problematic for predicting the long-term safety of drugs, especially because there are no good surrogates for harmful outcomes. Moreover, because the durations and sample sizes of studies using surrogate outcomes are generally small, many safety risks go undetected (84). Aprotinin (Trasylol®), doxazosin (Cardura®), erythropoietin (Epogen®), and rosiglitazone (Avandia®) are among many examples that, despite their early promising results for surrogate outcomes, were subsequently shown to have unfavorable harmful effects (73, 74, 85).

PROBLEMS IN THE CONDUCT, ANALYSIS, AND REPORTING OF CLINICAL STUDIES OF DRUG EFFECTS

Beyond design-related aspects, clinical studies of drug effects are subject to a range of variations in their implementation, potentially rendering their findings irrelevant for evaluating real-world effectiveness and safety. Together, problems in the conduct, analysis, and reporting of studies have important implications for the prediction of the clinical utility of drugs.

Publication and Selective Outcome Reporting Bias

The published literature includes only a subset of all clinical studies (14). Studies with positive or significant results are more likely to be reported than are those with negative or nonsignificant results—a phenomenon commonly referred to as publication bias (86, 87). Despite recent efforts to encourage the publication of completed studies, only about half of studies are reported within 4–5 years of their completion (14, 87). The results of many unpublished trials are now posted in ClinicalTrials.gov, so this proportion is probably gradually improving. However, the results of many trials remain unavailable. Moreover, even if studies with negative findings are published, their reports appear later than those with positive results, yielding a time-lag bias (88, 89).

Compounding the problem of publication bias is selective citation of clinical studies (90). Studies with statistically significant results are more than twice as likely to be cited as those with nonsignificant findings (91).

Publication biases also pose a threat to the validity of systematic reviews. An important example is the report by Kirsch and colleagues (92), who obtained unpublished data on new-generation antidepressants. Upon combining the published and unpublished evidence, they showed that drugs were no better than placebo for the treatment of mild and moderate depression. Similarly, another

review reported that antidepressants have an unfavorable benefit-harm profile when unreported studies were taken into account (93).

Further complicating matters is selective outcome reporting bias, whereby published reports describe an incomplete subset of outcomes evaluated in clinical studies (94). Such selective outcome reporting is widespread: Only a third to half of the outcomes are discussed in published reports, with significant outcomes being more than twice as likely to be fully reported as nonsignificant ones (95, 96). In general, outcomes in published reports differ from those defined in original study protocols (90, 97, 98), those submitted to the FDA (99), and those recorded in clinical study registries (100). Selective outcome reporting practices amplify the biases stemming from incomplete publication and selective citation of clinical studies and create a substantial barrier to performing objective assessments of clinical utility (14, 101, 102).

Paucity of Evidence on Harms

A balanced evaluation of both benefits and harms (i.e., harmful effects of drugs) is critical to successfully predicting the real-world utility of drugs. Yet clinical studies may fail to collect and report adequate data on adverse events, hindering a timely evaluation of long-term safety (103). Some studies do not report any evidence on harms, and those that do often lack adequate information on their definition (104), timing of onset (105), and severity (106). Most clinical studies of drug effects report only the most frequent events and/or limit their reporting to statistically significant results (107). Only a minority of studies report the reasons that patients discontinue treatment owing to adverse events (106, 108). According to some estimates from the past decade, adequate reporting of adverse events occurred in only 39% of clinical studies of drug effects (109). Despite the publication of guidelines to improve the quality of the reporting of safety in clinical studies (103), recent reporting of harm data remains inadequate: A third of studies do not properly report drug-related adverse events (107). Systematic reviews also may lack such information, compounding the inadequate reporting of harms in primary clinical studies (110). Most reviews include data on harms only to the extent that they are available from clinical studies (111), with approximately half of published reviews reporting adequate data on harms between 2007 and 2011 (112).

Questionable Analytical Practices

Measured against the recent efforts to standardize study reporting, current analytical practices in clinical studies of drug effects are heterogeneous, creating an opportunity for steering the results toward specific conclusions. Consequently, some of the reported findings in the published literature are spurious, owing to the use of poor methods for statistical analysis.

Assessment of subgroup effects is especially prone to such flawed practices. Because patients often do not respond uniformly to drugs, subgroup analyses are used to help tailor decisions for individual patients: Approximately 60% of clinical studies present subgroup findings (113). Looking for subgroup effects is important when there are potentially large differences among patient groups by racial, ethnic, and genetic characteristics. Despite this potential, initially compelling subgroup findings often prove spurious (114). This is perhaps inevitable, as researchers often undertake a large number of posthoc analyses and adopt inappropriate statistical methods for inferences (115, 116).

Even rigorous randomized designs are not immune to subjective judgments during statistical analysis, leaving room for so-called vibration effects—fluctuations in the study findings depending on the particular approach to analysis (18). Vibration effects are particularly problematic if investigators are influenced by conflicts of interest or optimism bias, the unwarranted belief in

the effectiveness and safety of interventions (24, 117). Empirical evaluations showed that flexibility in methods yields substantially different findings when researchers adopt different statistical analyses, for example, to adjust for imbalances between groups in randomized studies, as occurs in approximately 40% of clinical studies (118).

Other questionable research practices pose challenges to correctly interpreting the drug effects in clinical studies and their potential relevance in predicting clinical utility. Some studies are too small or poorly reported to be informative (albeit with apparent large effects) (24, 90, 119, 120); others have major statistical errors (121). Influential clinical studies may be stopped early because of apparent benefit in interim analyses, giving inflated estimates of drug effects (122–124). Despite the widespread adoption of guidelines aimed at improving the quality of research reports (90, 125), adequate information about study methods is available in only about 65% of published reports (90, 126). Distorted presentation of results is also common, particularly when nonsignificant findings are obtained (127).

PROBLEMS IN THE SYNTHESIS OF THE EVIDENCE AND IN THE FORMATION OF RECOMMENDATIONS FROM STUDIES OF DRUG EFFECTS

Clinicians increasingly turn to evidence summaries to stay abreast of the quickly expanding deluge of research findings. Systematic reviews, meta-analyses, and clinical practice guidelines allow medical providers to interpret the evidence, to learn what is known, and to describe the extent to which the existing evidence is applicable to individual patients. In the past two decades, there has been a surge of interest in the use of evidence syntheses to inform clinical decision making. Yet flaws in the development of systematic reviews, meta-analyses, and clinical practice guidelines remain, impeding their relevance in translating evidence to clinical decisions.

Problems with Reliability of Systematic Reviews and Meta-Analyses

Conducted in a transparent way, systematic reviews and meta-analyses offer a less biased alternative to narrative reviews, which lack an explicit description of systematic methods of searching for, identifying, and including clinical studies. Although systematic reviews and meta-analyses originally emerged as superior to narrative reviews, recent evidence suggests that they too perpetuate biases seen in primary clinical studies. For example, meta-analyses integrating early evidence tend to exaggerate treatment effects by approximately 15–30% (3, 16, 128). Such reviews also distort the evidence base in favor of certain products, as can be seen in the greatly overlapping meta-analyses on "hot" topics (129).

Systematic reviews and meta-analyses are very influential, i.e., receive more citations than other types of study designs (130); therefore, their conclusions have a major impact on clinical practice. Notably, although industry-sponsored meta-analyses produce findings numerically similar to those of non-industry-sponsored ones, they consistently report conclusions that are favorable to industry products (44, 45, 131). Industry-sponsored authors of systematic reviews and meta-analyses may "spin" the phrasing of the conclusions in favor of industry products (132, 133).

Too many meta-analyses adopt a narrow focus, providing an incomplete summary of the existing evidence. Almost half of meta-analyses focus only on specific agents, with 15% comparing one agent and placebo/control (134). A typical such paper shows pooled analyses done by the industry combined with data from several sponsored trials on a single product and almost invariably show that the product is effective and/or safe. Such pooled analyses might be viewed as marketing advertisements rather than scientific papers. Systematic reviews rarely incorporate evidence from

multiple data sources, such as information in publicly available FDA review reports, which would offer valuable data considered in regulatory decisions. Moreover, failure to consider the wider clinical research agenda in systematic reviews and meta-analyses—especially when some drugs are evaluated in hundreds of clinical studies for many different indications—inevitably results in the identification of spurious "successes" owing to multiple testing (23).

Problems with Reliability of Clinical Practice Guidelines

Clinical practice guidelines are at the interface between evidence and practice; they aim to translate research findings into real-world settings (135). Viewed as a key solution to improving health-care quality, guidelines have benefitted from considerable effort and resources spent on their development and dissemination (136), including by several major medical organizations.

However, recent evaluations demonstrate that guidelines fall considerably short of expectations (137). For example, financial conflicts of interest appear pervasive among developers of clinical practice guidelines (138–140). More than 80% of guideline recommendations are developed from lower levels of evidence, with almost half of all recommendations based on expert opinion (141). Despite their purportedly objective assessments of the existing evidence, many recent guidelines have become marketing and opinion-based pieces, delivering directive rather than assistive statements (142). Over the past few decades, clinical practice guidelines have progressively lowered the threshold for drug treatment of conditions such as heart disease (143) despite compelling evidence that nonpharmacological treatments, including exercise-based interventions, work equally well for numerous chronic conditions (144).

LACK OF SYSTEMATIC ASSESSMENTS OF CLINICAL UTILITY

In the current research environment, much of what we know about the clinical utility of drugs emerges haphazardly. There is a paucity of implementation research that seeks to understand what, why, and how drugs work in real-world settings and to test approaches that measure and improve their utility (145). As discussed above, contrary to what the end users of research expect, many clinical studies are not designed, conducted, analyzed, reported, and synthesized to evaluate the real-world effectiveness of drugs, but rather to meet regulatory demands and commercial interests. Large and simple pragmatic trials (146, 147), which impose minimal restrictions on patient populations and clinical decision making, are exceedingly rare. When evaluating head-to-head comparisons of relevant drugs using patient-centered outcomes, such studies are the gold standard of assessing clinical utility (**Table 1**).

Instead of pressing for more investment in large and simple pragmatic trials, drug regulators, clinicians, patients, and other health-care decision makers often settle for lower standards of

Table 1 Design features, advantages, and potential limitations of randomized study designs to evaluate clinical utility

Design features	Advantages	Potential challenges
Randomization can occur at either the individual level (pragmatic, practical trial) or the group level, e.g., clinic or hospital (cluster randomized trial).	Patient populations and care patterns that are representative of routine practice can be included.	Extensive patient crossover and contamination of comparator groups may complicate interpretability of findings.
Flexible inclusion criteria allow variations among patients and providers.	Data collection can be integrated into electronic health record systems, reducing the burden of primary data collection.	Effective blinding of participants and study personnel is often not possible.
Minimal restrictions are imposed on clinical decision making.	Registry-based patient recruitment and data collection can introduce significant efficiencies.	

evidence. Following market entry, drug regulators encourage pharmaceutical companies to conduct Phase IV studies to gather evidence on the safety of their products in real-world settings. Generally nonrandomized in nature, these studies include large patient populations to detect safety signals that may have been missed in premarketing studies. Regulators have limited influence over the design and conduct of Phase IV studies relative to their influence over earlier clinical studies; more than half of postmarketing commitments remain unfulfilled many years after market entry (148). Even when they are fulfilled—the situation is improving (149)—pharmaceutical companies may use postmarketing studies to boost prescribers' familiarity with their products; indeed, postmarketing studies have been referred to as seeding trials (150, 151). In the case of the postmarketing studies of diabetes drugs, for example, a recent analysis demonstrated that the studies had limited scientific value and that they were designed as marketing tools to promote wider use of more expensive products (152). Such practices highlight the inadequacy of the current clinical research enterprise in evaluating the clinical utility of drugs.

POTENTIAL IMPROVEMENTS

Closing the gap between the initially promising effects of drugs and their later clinical utility will require improvements to each aspect of clinical studies—their design, conduct, analysis, reporting, and synthesis (i.e., their incorporation into systematic reviews and meta-analyses; see above). Historically, progress in clinical research has been fragmented and has occurred primarily through self-regulatory mechanisms. Perhaps we need a new emphasis on clinical utility that takes into account the interplay among the producers (industry and researchers), regulators, and end users (clinicians, patients, other health-care decision makers) of clinical studies of drug effects. We make some recommendations to improve the relevance of clinical evidence in predicting the real-world effectiveness and safety of drugs (**Table 2**).

Study Registration and Analysis Prespecification

Until the recent institution of policies mandating the registration of clinical studies (153), it was essentially impossible to know whether (and to what extent) the published literature represented a biased subset of all studies that had been completed (154). The International Committee of Medical Journal Editors now requires the registration of clinical studies as a prerequisite for consideration for publication (155); this requirement helps identify clinical studies that have been completed or are under way. Coupled with the FDA's mandate to report studies at ClinicalTrials.gov, recent journal policies resulted in the registration of more than 160,000 randomized studies by early 2014 (156). Since 2009, federal legislation in the United States has also required the reporting of summary results within one year of study completion (157).

We believe that a similar strategy is needed to encourage the registration and publication of complete clinical study protocols. The clinical study protocol plays a key role in study planning, conduct, interpretation, oversight, and external review (158). Particularly important is the inclusion of information on the definitions of outcome and the analytical procedures that will be used (24). Such efforts improve the transparency of clinical studies and enable scrutiny and comparison of reported analyses with what was planned at the study outset (159).

Provision of Raw Data for Independent Analysis

Condensed summaries of clinical studies published in peer-reviewed reports represent unavoidably incomplete synopses of actual studies—resulting in loss of valuable information. This loss

Table 2 Potential ways to improve the clinical utility of drugs

	What is needed?	Recent developments	What should happen next?
Study registration and analysis prespecification	Registration and publication of complete clinical study protocols, including information on the standardized outcome definitions and planned analytical procedures	Federal legislation requires the registration of clinical studies and their summary results on ClinicalTrials.gov.	Widespread adoption of reporting guidelines to improve protocol development, implementation, and registration
Provision of raw data	Provision of access to raw data from clinical studies of drug effects	The European Medicines Agency and numerous pharmaceutical companies have committed to making raw data on drug effects available in a timely fashion.	Agreement on a feasible timeline to resolve any challenges (e.g., need for deidentification, informed consent stipulations, legal trade secret claims)
Independent sponsorship of studies, systematic reviews, and clinical practice guidelines	Elimination of conflicts of interest (financial or otherwise) from studies of drug effects, clinical practice guidelines, and systematic reviews Greater alignment of incentives and objectives of industry, regulators, researchers, clinicians, and patients	Despite lack of real progress in ridding clinical studies of financial conflicts of interest, numerous guidelines offer recommendations for managing and minimizing such conflicts in clinical practice guidelines.	Increased investment from governments and noncommercial sponsors for independent clinical studies of drug effects Guidance from regulatory entities to establish standards for clinical studies of drug effects Development of clinical practice guidelines and systematic reviews by evidence review and synthesis methodologists rather than content area experts
Head-to-head comparisons	Requirement of comparative data at the time of new drug approvals	The European Medicines Agency has stated its preference for head-to-head comparisons (over placebo-controlled studies) of new and existing drugs.	Investment by industry sponsors in active-comparator studies to compare new drugs to existing alternatives Routine use of network meta-analyses by regulators to compare the benefit-harm profiles of new and existing drugs
Continuous feedback from implementation research	Integration of implementation research into actual clinical practice	Patient-Centered Outcomes Research Institute has recently launched the National Patient-Centered Clinical Research Network to conduct clinical outcomes research.	Collaboration among industry sponsors and independent, nonconflicted entities to implement large and simple megatrials in clinical practice

is particularly problematic when study-level aggregation of results does not accurately reflect the underlying participant-level data. Such data from clinical studies of drug effects—included in comprehensive clinical study reports (160)—are currently considered commercial, confidential information. Clinical study reports may be the most detailed, complete, and integrated source of information about study design, conduct, analysis, and results (160). Raw data (containing information on individual study participants) from clinical studies are rarely made available to independent researchers (161). Such data may exist in different formats (either as collected or deidentified), and access can be either open or restricted (e.g., tailored data to address specific research questions) (160, 162, 163).

Clinical study sponsors, investigators, and regulators should commit to making raw data on drug effects available in a timely fashion. Sharing raw data would have many benefits, including facilitation of the independent reanalysis of results to detect errors and selective reporting and an ability to complement meta-analyses with more granular information to explore patient variability (14, 162). By increasing the efficiency of drug development and reducing duplication of efforts, industry also stands to benefit from the wider provision of raw data (164). Acknowledging this potential, drug regulators such as the European Medicines Agency are seeking collaboration between study sponsors and researchers that respects the commercial interests of the industry yet embraces the reciprocal gains from the public availability of raw data (165). Firmer action is needed to push for greater transparency and to ensure that raw data from clinical studies are available for independent scrutiny and that any obstacles (e.g., need for deidentification, informed consent stipulations, legal trade secret claims) are dealt with efficiently and transparently. Although adopting data-sharing requirements would be relatively feasible for new, future trials, considerable challenges exist for obtaining access to raw data from past trials. These data may sometimes be lost even to their sponsors, but improvements that promote access to raw data could be made on that front as well.

Sponsoring of Clinical Research by Nonconflicted Entities

Much of clinical research is plagued by conflicts of interest—financial or otherwise—with the pharmaceutical industry footing most of the bill for sponsored clinical studies. In the current environment, "we get what we pay for": Research and development activities of pharmaceutical companies are driven by marketing and sales motives, creating asymmetries in the evidence base (15). The current system is inefficient and contains redundancy and wasted resources, with too many similar, small studies of limited value. Improvements in the design of the overall research agenda may lead to cost savings across the research and development enterprise.

There is a need to align the incentives and objectives of producers and end users of research. Independently funded clinical studies should address research questions that are truly important and needed; include head-to-head comparisons of drugs with other drugs or nonpharmacological interventions; evaluate patient populations that closely resemble those in clinical practice; and measure outcomes that matter to clinicians, patients, and their caregivers. In our view, there is a need to increase funding for independent clinical studies of drug effects. One potential solution would be to supplement current government funding of comparative effectiveness research with industry investments that are allocated to an independent trial fund tasked with designing and conducting clinical studies of drug effects (166).

In addition, nonconflicted entities should urge pharmaceutical sponsors to invest in more rigorous evaluations of drug effects. In an attempt to develop a feasible approach for the industry, regulatory advice should guide pharmaceutical companies to invest in underresearched areas that will encourage more efficient allocation of investment and potentially lead to an increase in the

number of innovative drugs that offer greater clinical benefit than do existing treatment options. Pharmaceutical companies stand to benefit from such regulatory guidance. Indeed, sponsors of promising products should favor transformative changes in the financing, design, and conduct of the research agenda. Such changes should be accompanied by the institution of transparent reward mechanisms that favor genuine innovations.

Guidance from regulatory entities is also important for establishing standards for clinical studies of drug effects—particularly standards with a new emphasis on clinical utility. Whenever possible, studies should conform to such standards in terms of patient populations, outcomes, assessment techniques, follow-up studies, dosing regimens, and choice of comparators.

Requirement of Comparative Data at the Time of Drug Approvals

In our view, regulators should raise the bar for market authorization of new drugs and require comparative data at the time of new drug approval, especially in therapeutic areas that have competing drugs or other treatment regimens. Such evidence is needed by a range of decision makers upon market entry (167). Comparative evidence at market entry will help equip clinicians, patients, and caregivers with information as to how new therapies fare against those currently approved; help drug regulators protect the public from inferior and harmful products; and ensure that third-party payers engage in pricing and coverage strategies on the basis of therapeutic value. Recent proposals call for requiring head-to-head comparisons of new drugs with viable alternatives at the time of market authorization decisions (167, 168). In addition, network meta-analyses that compare the benefit-harm profiles of new drugs with those of existing alternatives should become a more standard tool in the assessment of new drugs (169).

Nonconflicted Conduct of Systematic Reviews and Clinical Practice Guidelines

Conflicts of interest among those who write systematic reviews and prepare clinical practice guidelines compromise the objectivity of the conclusions and recommendations, biasing the interpretation of existing clinical evidence. The Institute of Medicine suggests that individuals with a conflict of interest should not partake in developing clinical practice guidelines. Such suggestions should extend to those who write systematic reviews. Minimizing potential bias, in our view, requires that those who develop clinical practice guidelines and write systematic reviews should primarily be systematic review methodologists rather than content area experts (132). Clinical practice guidelines should also be subject to independent scrutiny before their recommendations are finalized (139).

Continuous Feedback from Implementation Research

Efforts are needed to integrate implementation research into clinical practice to assess the real-world performance of drugs (170). Nonconflicted public or nongovernmental entities, such as the Patient-Centered Outcomes Research Institute in the United States, should establish research networks within health-care systems to improve the scientific scrutiny of drugs after market entry by streamlining the conduct of large, simple megatrials in clinical practice (166, 171).

CONCLUSIONS

Existing clinical evidence can be a shaky foundation for defining the clinical utility of drugs. Problems in clinical studies are an indication of missed opportunities to successfully define the

real-world effectiveness and safety of drugs. Driven largely by commercial interests, many clinical studies generate more noise than meaningful evidence to guide clinical decision making. Greater involvement of nonconflicted bodies is needed in the design and conduct of clinical studies, along with more head-to-head comparisons, representative patient populations, hard clinical outcomes, and appropriate analytical approaches. Documenting, registering, and publishing study protocols at the outset and sharing participant-level data at study completion would help ensure transparency and enhance public trust in the clinical research enterprise. Such an approach is needed to generate evidence that is better suited to the tasks of predicting the clinical utility of drugs and providing the information needed by patients and clinicians. Future efforts should focus on engaging the industry, researchers, regulators, clinicians, patients, and other decision makers in discussions to develop transformative ideas with the aim of tackling the numerous defects in the current research environment. Emerging ideas should be piloted and subjected to scientific scrutiny before they are widely implemented and touted as solutions.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

METRICS is funded by a grant from the Laura and John Arnold Foundation. The work of J.P.A.I. is also supported by an unrestricted gift from Sue and Robert O'Donnell. H.N. is funded by a research grant from the European Commission's Seventh Framework Program.

LITERATURE CITED

1. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, et al. 2014. Biomedical research: increasing value, reducing waste. *Lancet* 383(9912):101–4
2. Eichler HG, Abadie E, Breckenridge A, Flamion B, Gustafsson LL, et al. 2011. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. *Nat. Rev. Drug Discov.* 10(7):495–506
3. Pereira TV, Ioannidis JPA. 2011. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *J. Clin. Epidemiol.* 64(10):1060–69
4. Ioannidis JPA, Trikalinos TA. 2005. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J. Clin. Epidemiol.* 58(6):543–49
5. Pereira TV, Horwitz RI, Ioannidis JPA. 2012. Empirical evaluation of very large treatment effects of medical interventions. *JAMA* 308(16):1676–84
6. Prasad V, Gall V, Cifu A. 2011. The frequency of medical reversal. *Arch. Intern. Med.* 171(18):1675–76
7. Ioannidis JPA. 2005. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 294(2):218–28
8. Roses AD. 2008. Pharmacogenetics in drug discovery and development: a translational perspective. *Nat. Rev. Drug Discov.* 7(10):807–17
9. Topol EJ. 2004. Failing the public health—rofecoxib, Merck, and the FDA. *N. Engl. J. Med.* 351(17):1707–9
10. Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. 2004. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet* 364(9450):2021–29
11. Singh S, Loke YK, Spangler JG, Furberg CD. 2011. Risk of serious adverse cardiovascular events associated with varenicline: a systematic review and meta-analysis. *CMAJ* 183(12):1359–66

12. Fullerton CA, Busch AB, Frank RG. 2010. The rise and fall of gabapentin for bipolar disorder: a case study on off-label pharmaceutical diffusion. *Med. Care* 48(4):372–79
13. Chalmers I. 1990. Underreporting research is scientific misconduct. *JAMA* 263(10):1405–8
14. Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, et al. 2014. Increasing value and reducing waste: addressing inaccessible research. *Lancet* 383(9913):257–66
15. Light DW, Lexchin JR. 2012. Pharmaceutical research and development: What do we get for all that money? *BMJ* 345:e4348
16. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, et al. 2014. How to increase value and reduce waste when research priorities are set. *Lancet* 383(9912):156–65
17. Robinson KA, Goodman SN. 2011. A systematic examination of the citation of prior research in reports of randomized, controlled trials. *Ann. Intern. Med.* 154(1):50–55
18. Ioannidis JPA. 2008. Perfect study, poor evidence: interpretation of biases preceding study design. *Semin. Hematol.* 45(3):160–66
19. Cooper NJ, Jones DR, Sutton AJ. 2005. The use of systematic reviews when designing studies. *Clin. Trials* 2(3):260–64
20. Tatsioni A, Bonitsis NG, Ioannidis JPA. 2007. Persistence of contradicted claims in the literature. *JAMA* 298(21):2517–26
21. Gøtzsche PC. 1987. Reference bias in reports of drug trials. *BMJ* 295(6599):654–56
22. Goudie AC, Sutton AJ, Jones DR, Donald A. 2010. Empirical assessment suggests that existing evidence could be used more fully in designing randomized controlled trials. *J. Clin. Epidemiol.* 63(9):983–91
23. Ioannidis JPA, Karassa FB. 2010. The need to consider the wider agenda in systematic reviews and meta-analyses: breadth, timing, and depth of the evidence. *BMJ* 341:c4875
24. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, et al. 2014. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 383(9912):166–75
25. Sackett DL, Haynes RB, Tugwell P. 1985. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. London: Little, Brown. 2nd ed.
26. Ioannidis JPA, Haidich AB, Pappa M, Pantazis N, Kokori SI, et al. 2001. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 286(7):821–30
27. Ioannidis JPA. 2005. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 294(2):218–28
28. Djulbegovic B, Hozo I, Ioannidis JPA. 2014. Improving the drug development process: more not less randomized trials. *JAMA* 311(4):355–56
29. Hochman M, McCormick D. 2010. Characteristics of published comparative effectiveness studies of medications. *JAMA* 303(10):951–58
30. Prasad V, Jorgenson J, Ioannidis JPA, Cifu A. 2013. Observational studies often make clinical practice recommendations: an empirical evaluation of authors' attitudes. *J. Clin. Epidemiol.* 66(4):361–66.e4
31. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, et al. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343:d5928
32. Gluud LL. 2006. Bias in clinical intervention research. *Am. J. Epidemiol.* 163(6):493–501
33. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, et al. 1998. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 352(9128):609–13
34. Kjaergard LL, Villumsen J, Gluud C. 2001. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann. Intern. Med.* 135(11):982–89
35. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JPA, et al. 2002. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 287(22):2973–82
36. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, et al. 2012. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann. Intern. Med.* 157(6):429–38
37. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, et al. 2008. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 336(7644):601–5

38. Dorsey ER, de Roulet J, Thompson JP, Reminick JI, Thai A, et al. 2010. Funding of US biomedical research, 2003–2008. *JAMA* 303(2):137–43
39. Patsopoulos NA, Ioannidis JPA, Analatos AA. 2006. Origin and funding of the most frequently cited papers in medicine: database analysis. *BMJ* 332(7549):1061–64
40. Bekelman JE, Li Y, Gross CP. 2003. Scope and impact of financial conflicts of interest in biomedical research. *JAMA* 289(4):454–65
41. Lexchin J, Bero LA, Djulbegovic B, Clark O. 2003. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 326(7400):1167–70
42. Bero L. 2013. Industry sponsorship and research outcome: a Cochrane review. *JAMA Intern. Med.* 173(7):580–81
43. Sismondo S. 2008. Pharmaceutical company funding and its consequences: a qualitative systematic review. *Contemp. Clin. Trials* 29(2):109–13
44. Jorgensen AW, Hilden J, Gotzsche PC. 2006. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *BMJ* 333(7572):782
45. Yank V, Rennie D, Bero LA. 2007. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. *BMJ* 335(7631):1202–5
46. Lathyris D, Patsopoulos N, Salanti G, Ioannidis JPA. 2010. Industry sponsorship and selection of comparators in randomized clinical trials. *Eur. J. Clin. Investig.* 40(2):172–82
47. Smith R. 2005. Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Med.* 2(5):e138
48. Sackett DL, Oxman AD. 2003. HARLOT plc: an amalgamation of the world's two oldest professions. *BMJ* 327(7429):1442–45
49. Munos B. 2009. Lessons from 60 years of pharmaceutical innovation. *Nat. Rev. Drug Discov.* 8(12):959–68
50. Garattini S. 1997. Are me-too drugs justified? *J. Nephrol.* 10(6):283–94
51. Hollis A. 2005. Comment on "The economics of follow-on drug research and development: trends in entry rates and the timing of development." *Pharmacoeconomics* 23(12):1187–92; discussion 1193–202
52. Stafford RS, Wagner TH, Lavori PW. 2009. New, but not improved? Incorporating comparative-effectiveness information into FDA labeling. *N. Engl. J. Med.* 361:1230–33
53. Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G. 2014. Characteristics of networks of interventions: a description of a database of 186 published networks. *PLoS ONE* 9(1):e86754
54. Salanti G, Kavvoura FK, Ioannidis JPA. 2008. Exploring the geometry of treatment networks. *Ann. Intern. Med.* 148(7):544–53
55. Ioannidis JPA, Karassa FB, Druyts E, Thorlund K, Mills EJ. 2013. Biologic agents in rheumatology: unmet issues after 200 trials and $200 billion sales. *Nat. Rev. Rheumatol.* 9(11):665–73
56. Goldberg NH, Schneeweiss S, Kowal MK, Gagne JJ. 2011. Availability of comparative efficacy data at the time of drug approval in the United States. *JAMA* 305(17):1786–89
57. van Luijn JC, van Loenen AC, Gribnau FW, Leufkens HG. 2008. Choice of comparator in active control trials of new drugs. *Ann. Pharmacother.* 42(11):1605–12
58. van Luijn JC, Gribnau FW, Leufkens HG. 2007. Availability of comparative trials for the assessment of new medicines in the European Union at the moment of market authorization. *Br. J. Clin. Pharmacol.* 63(2):159–62
59. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. 2012. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 380(9836):37–43
60. Liberopoulos G, Trikalinos NA, Ioannidis JPA. 2009. The elderly were under-represented in osteoarthritis clinical trials. *J. Clin. Epidemiol.* 62(11):1218–23
61. Dai C, Stafford RS, Alexander GC. 2005. National trends in cyclooxygenase-2 inhibitor use since market release: nonselective diffusion of a selectively cost-effective innovation. *Arch. Intern. Med.* 165(2):171–77
62. Glasgow RE, Lichtenstein E, Marcus AC. 2003. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *Am. J. Public Health* 93(8):1261–67

63. Van Spall HC, Toren A, Kiss A, Fowler RA. 2007. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA* 297(11):1233–40
64. Konrat C, Boutron I, Trinquart L, Auleley GR, Ricordeau P, Ravaud P. 2012. Underrepresentation of elderly people in randomised controlled trials: the example of trials of 4 widely prescribed drugs. *PLoS ONE* 7(3):e33559
65. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. 1999. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N. Engl. J. Med.* 341(27):2061–67
66. Rehman HU. 2005. Under-representation of the elderly in clinical trials. *Eur. J. Intern. Med.* 16(6):385–86
67. Heiat A, Gross CP, Krumholz HM. 2002. Representation of the elderly, women, and minorities in heart failure clinical trials. *Arch. Intern. Med.* 162(15):1682–88
68. Lee PY, Alexander KP, Hammill BG, Pasquali SK, Peterson ED. 2001. Representation of elderly persons and women in published randomized trials of acute coronary syndromes. *JAMA* 286(6):708–13
69. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, et al. 2000. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. *JAMA* 284(10):1290–96
70. Ocana A, Tannock IF. 2011. When are "positive" clinical trials in oncology truly positive? *J. Natl. Cancer Inst.* 103(1):16–20
71. Kazi DS, Hlatky MA. 2012. Repeat revascularization is a faulty end point for clinical trials. *Circ. Cardiovasc. Qual. Outcomes* 5(3):249–50
72. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, et al. 2007. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 334(7597):786
73. Fleming TR, Powers JH. 2012. Biomarkers and surrogate endpoints in clinical trials. *Stat. Med.* 31(25):2973–84
74. Svensson S, Menkes DB, Lexchin J. 2013. Surrogate outcomes in clinical trials: a cautionary tale. *JAMA Intern. Med.* 173(8):611–12
75. Fleming TR, DeMets DL. 1996. Surrogate end points in clinical trials: Are we being misled? *Ann. Intern. Med.* 125(7):605–13
76. Yudkin JS, Lipska KJ, Montori VM. 2011. The idolatry of the surrogate. *BMJ* 343:d7995
77. Pandor A, Ara RM, Tumur I, Wilkinson AJ, Paisley S, et al. 2009. Ezetimibe monotherapy for cholesterol lowering in 2,722 people: systematic review and meta-analysis of randomized controlled trials. *J. Intern. Med.* 265(5):568–80
78. Brown BG, Taylor AJ. 2008. Does ENHANCE diminish confidence in lowering LDL or in ezetimibe? *N. Engl. J. Med.* 358(14):1504–7
79. Doggrell SA. 2012. The ezetimibe controversy—can this be resolved by comparing the clinical trials with simvastatin and ezetimibe alone and together? *Expert Opin. Pharmacother.* 13(10):1469–80
80. Nissen SE. 2008. ENHANCE and ACCORD: controversy over surrogate end points. *Curr. Cardiol. Rep.* 10(3):159–61
81. Tzoulaki I, Siontis KC, Evangelou E, Ioannidis JPA. 2013. Bias in associations of emerging biomarkers with cardiovascular disease. *JAMA Intern. Med.* 173(8):664–71
82. Ciani O, Buyse M, Garside R, Pavey T, Stein K, et al. 2013. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *BMJ* 346:f457
83. Ioannidis JPA, Panagiotou OA. 2011. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA* 305(21):2200–10
84. Moore TJ, Furberg CD. 2014. Development times, clinical testing, postmarket follow-up, and safety risks for the new drugs approved by the US Food and Drug Administration: the class of 2008. *JAMA Intern. Med.* 174(1):90–95
85. Cohen D. 2010. Rosiglitazone: what went wrong. *BMJ* 341:c4848
86. Sismondo S. 2008. How pharmaceutical industry funding affects trial outcomes: causal structures and responses. *Soc. Sci. Med.* 66(9):1909–14

87. Song F, Parekh S, Hooper L, Loke YK, Ryder J, et al. 2010. Dissemination and publication of research findings: an updated review of related biases. *Health Technol. Assess.* 14(8):iii, ix–xi, 1–193
88. Ioannidis JPA. 1998. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 279(4):281–86
89. Stern JM, Simes RJ. 1997. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 315(7109):640–45
90. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, et al. 2014. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 383(9913):267–76
91. Kjaergard LL, Gluud C. 2002. Citation bias of hepato-biliary randomized clinical trials. *J. Clin. Epidemiol.* 55(4):407–10
92. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. 2008. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med.* 5(2):e45
93. Eyding D, Lelgemann M, Grouven U, Harter M, Kromp M, et al. 2010. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ* 341:c4737
94. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. 2004. Empirical evidence for selective reporting of outcomes in randomized trials. *JAMA* 291(20):2457–65
95. Chan AW, Altman DG. 2005. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 330(7494):753
96. Dwan K, Gamble C, Williamson PR, Kirkham JJ. 2013. Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. *PLoS ONE* 8(7):e66844
97. Wieseler B, Wolfram N, McGauran N, Kerekes MF, Vervolgyi V, et al. 2013. Completeness of reporting of patient-relevant clinical trial outcomes: comparison of unpublished clinical study reports with publicly available data. *PLoS Med.* 10(10):e1001526
98. Vedula SS, Bero L, Scherer RW, Dickersin K. 2009. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *N. Engl. J. Med.* 361(20):1963–71
99. Rising K, Bacchetti P, Bero L. 2008. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med.* 5(11):e217
100. Vera-Badillo FE, Shapiro R, Ocana A, Amir E, Tannock IF. 2013. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. *Ann. Oncol.* 24(5):1238–44
101. Hart B, Lundh A, Bero L. 2012. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. *BMJ* 344:d7202
102. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, et al. 2010. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 340:c365
103. Ioannidis JPA, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, et al. 2004. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann. Intern. Med.* 141(10):781–88
104. Papanikolaou PN, Churchill R, Wahlbeck K, Ioannidis JPA. 2004. Safety reporting in randomized trials of mental health interventions. *Am. J. Psychiatry* 161(9):1692–97
105. Yazici Y, Yazici H. 2007. A survey of inclusion of the time element when reporting adverse effects in randomised controlled trials of cyclo-oxygenase-2 and tumour necrosis factor α inhibitors. *Ann. Rheum. Dis.* 66(1):124–27
106. Ioannidis JPA. 2009. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. *Arch. Intern. Med.* 169(19):1737–39
107. Pitrou I, Boutron I, Ahmad N, Ravaud P. 2009. Reporting of safety results in published reports of randomized controlled trials. *Arch. Intern. Med.* 169(19):1756–61
108. Lee PE, Fischer HD, Rochon PA, Gill SS, Herrmann N, et al. 2008. Published randomized controlled trials of drug therapy for dementia often lack complete data on harm. *J. Clin. Epidemiol.* 61(11):1152–60
109. Ioannidis JPA, Lau J. 2001. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 285(4):437–43
110. Hopewell S, Wolfenden L, Clarke M. 2008. Reporting of adverse events in systematic reviews can be improved: survey results. *J. Clin. Epidemiol.* 61(6):597–602

111. Ernst E, Pittler MH. 2001. Assessment of therapeutic safety in systematic reviews: literature review. *BMJ* 323(7312):546
112. Zorzela L, Golder S, Liu Y, Pilkington K, Hartling L, et al. 2014. Quality of reporting in systematic reviews of adverse events: systematic review. *BMJ* 348:f7668
113. Assmann SF, Pocock SJ, Enos LE, Kasten LE. 2000. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 355(9209):1064–69
114. Sun X, Ioannidis JPA, Agoritsas T, Alba AC, Guyatt G. 2014. How to use a subgroup analysis: users' guide to the medical literature. *JAMA* 311(4):405–11
115. Sun X, Briel M, Busse JW, You JJ, Akl EA, et al. 2012. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ* 344:e1553
116. Rothwell PM. 2005. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 365(9454):176–86
117. Chalmers I, Matthews R. 2006. What are the implications of optimism bias in clinical research? *Lancet* 367(9509):449–50
118. Saquib N, Saquib J, Ioannidis JPA. 2013. Practices and impact of primary outcome adjustment in randomized controlled trials: meta-epidemiologic study. *BMJ* 347:f4313
119. Siontis KC, Evangelou E, Ioannidis JPA. 2011. Magnitude of effects in clinical trials published in high-impact general medical journals. *Int. J. Epidemiol.* 40(5):1280–91
120. Chen YL, Yang KH. 2009. Avoidable waste in the production and reporting of evidence. *Lancet* 374(9692):786
121. Gotzsche PC, Hrobjartsson A, Maric K, Tendal B. 2007. Data extraction errors in meta-analyses that use standardized mean differences. *JAMA* 298(4):430–37
122. Bassler D, Briel M, Montori VM, Lane M, Glasziou P, et al. 2010. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* 303(12):1180–87
123. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, et al. 2005. Randomized trials stopped early for benefit: a systematic review. *JAMA* 294(17):2203–9
124. Bassler D, Montori VM, Briel M, Glasziou P, Guyatt G. 2008. Early stopping of randomized clinical trials for overt efficacy is problematic. *J. Clin. Epidemiol.* 61(3):241–46
125. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. 2012. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst. Rev.* 1:60
126. Glasziou P, Meats E, Heneghan C, Shepperd S. 2008. What is missing from descriptions of treatment in trials and reviews? *BMJ* 336(7659):1472–74
127. Boutron I, Dutton S, Ravaud P, Altman DG. 2010. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 303(20):2058–64
128. Doshi P, Jones M, Jefferson T. 2012. Rethinking credible evidence synthesis. *BMJ* 344:d7898
129. Siontis KC, Hernandez-Boussard T, Ioannidis JPA. 2013. Overlapping meta-analyses on the same topic: survey of published studies. *BMJ* 347:f4501
130. Patsopoulos NA, Analatos AA, Ioannidis JPA. 2005. Relative citation impact of various study designs in the health sciences. *JAMA* 293(19):2362–66
131. Jorgensen AW, Maric KL, Tendal B, Faurschou A, Gotzsche PC. 2008. Industry-supported meta-analyses compared with meta-analyses with non-profit or no support: differences in methodological quality and conclusions. *BMC Med. Res. Methodol.* 8:60
132. Gotzsche PC, Ioannidis JPA. 2012. Content area experts as authors: helpful or harmful for systematic reviews and meta-analyses? *BMJ* 345:e7031
133. Stamatakis E, Weiler R, Ioannidis JPA. 2013. Undue industry influences that distort healthcare research, strategy, expenditure and practice: a review. *Eur. J. Clin. Investig.* 43(5):469–75
134. Haidich AB, Pilalas D, Contopoulos-Ioannidis DG, Ioannidis JPA. 2013. Most meta-analyses of drug interventions have narrow scopes and many focus on specific agents. *J. Clin. Epidemiol.* 66(4):371–78
135. Grimshaw JM, Russell IT. 1993. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 342(8883):1317–22

136. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. 1999. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ* 318(7182):527–30
137. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. 1999. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 281(20):1900–5
138. Neuman J, Korenstein D, Ross JS, Keyhani S. 2011. Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. *BMJ* 343:d5621
139. Lenzer J, Hoffman JR, Furberg CD, Ioannidis JPA. 2013. Ensuring the integrity of clinical practice guidelines: a tool for protecting patients. *BMJ* 347:f5535
140. Mendelson TB, Meltzer M, Campbell EG, Caplan AL, Kirkpatrick JN. 2011. Conflicts of interest in cardiovascular clinical practice guidelines. *Arch. Intern. Med.* 171(6):577–84
141. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. 2009. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 301(8):831–41
142. Shaneyfelt TM, Centor RM. 2009. Reassessment of clinical practice guidelines: Go gently into that good night. *JAMA* 301(8):868–69
143. Ioannidis JPA. 2014. More than a billion people taking statins? Potential implications of the new cardiovascular guidelines. *JAMA* 311(5):463–64
144. Naci H, Ioannidis JPA. 2013. Comparative effectiveness of exercise and drug interventions on mortality outcomes: metaepidemiological study. *BMJ* 347:f5577
145. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. 2013. Implementation research: what it is and how to do it. *BMJ* 347:f6753
146. Roland M, Torgerson DJ. 1998. Understanding controlled trials: What are pragmatic trials? *BMJ* 316(7127):285
147. Tunis SR, Stryer DB, Clancy CM. 2003. Practical clinical trials. *JAMA* 290(12):1624–32
148. Carpenter D. 2010. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton, NJ: Princeton Univ. Press
149. Fain K, Daubresse M, Alexander G. 2013. The Food and Drug Administration Amendments Act and postmarketing commitments. *JAMA* 310(2):202–4
150. Steinman MA, Bero LA, Chren M-M, Landefeld CS. 2006. Narrative review: The promotion of gabapentin: an analysis of internal industry documents. *Ann. Intern. Med.* 145(4):284–93
151. Hill KP, Ross JS, Egilman DS, Krumholz HM. 2008. The ADVANTAGE seeding trial: a review of internal documents. *Ann. Intern. Med.* 149(4):251–58
152. Gale EAM. 2012. Post-marketing studies of new insulins: sales or science? *BMJ* 344:e3974
153. Dickersin K, Davis BR, Dixon DO, George SL, Hawkinse BS, et al. 2004. The Society for Clinical Trials supports United States legislation mandating trials registration. *Clin. Trials* 1(5):417–20
154. Zarin DA. 2013. Participant-level data and the new frontier in trial transparency. *N. Engl. J. Med.* 369(5):468–69
155. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, et al. 2004. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N. Engl. J. Med.* 351(12):1250–51
156. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. 2011. The ClinicalTrials.gov results database—update and key issueS. *N. Engl. J. Med.* 364(9):852–60
157. Prayle AP, Hurley MN, Smyth AR. 2012. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. *BMJ* 344:d7373
158. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gotzsche PC, et al. 2013. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann. Intern. Med.* 158(3):200–7
159. Fleming TR. 2005. Surrogate endpoints and FDA's accelerated approval process. *Health Aff.* 24(1):67–78
160. Doshi P, Jefferson T. 2013. Clinical study reports of randomised controlled trials: an exploratory review of previously confidential industry reports. *BMJ Open* 3(2):e002496
161. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA. 2011. Public availability of published research data in high-impact journals. *PLoS ONE* 6(9):e24357
162. Doshi P, Goodman SN, Ioannidis JPA. 2013. Raw data from clinical trials: within reach? *Trends Pharmacol. Sci.* 34(12):645–47

163. Doshi P, Jefferson T, Del Mar C. 2012. The imperative to share clinical study reports: recommendations from the Tamiflu experience. *PLoS Med.* 9(4):e1001201
164. Eichler H-G, Pétavy F, Pignatti F, Rasi G. 2013. Access to patient-level trial data—a boon to drug developers. *N. Engl. J. Med.* 369(17):1577–79
165. Gotzsche PC, Jorgensen AW. 2011. Opening up data at the European Medicines Agency. *BMJ* 342:d2686
166. Ioannidis JPA. 2013. Mega-trials for blockbusters. *JAMA* 309(3):239–40
167. Sorenson C, Naci H, Cylus J, Mossialos E. 2011. Evidence of comparative efficacy should have a formal role in European drug approvals. *BMJ* 343:d4849
168. Naci H, Cylus J, Vandoros S, Sato A, Perampaladas K. 2012. Raising the bar for market authorisation of new drugs. *BMJ* 345:e4261
169. Naci H, O'Connor AB. 2013. Assessing comparative effectiveness of new drugs before approval using prospective network meta-analyses. *J. Clin. Epidemiol.* 66(8):812–16
170. Tunis SR, Stryer DB, Clancy CM. 2003. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 290(12):1624–32
171. Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, et al. 2014. Increasing value and reducing waste in biomedical research regulation and management. *Lancet* 383(9912):176–85