Comparative Effectiveness Research 2

Generating comparative evidence on new drugs and devices after approval

Professor Andrea Cipriani, PhD[1]
Professor John P.A. Ioannidis, MD[2]
Professor Peter M. Rothwell, FMedSci[3]
Professor Paul Glasziou, FRACGP[4]
Tianjing Li, PhD[5]
Professor Adrian F. Hernandez, MD[6]
Anneka Tomlinson, PhD [1]
Professor John Simes, MD[7]
Huseyin Naci, PhD[8]

1. Department of Psychiatry, University of Oxford, Oxford, UK; Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK
2. Meta-Research Innovation Center at Stanford (METRICS), and Departments of Medicine, of Health Research and Policy, of Biomedical Data Science, and of Statistics, Stanford University, Palo Alto, CA, USA
3. Center for the Prevention of Stroke and Dementia, University of Oxford, Oxford, UK
4. Centre for Research in Evidence-Based Practice, University of Bond, Queensland, Australia
5. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
6. Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, USA
7. NHMRC Clinical Trials Centre, University of Sydney, Sydney, Australia
8. Department of Health Policy, London School of Economics and Political Science, London, UK

Correspondence to

Professor Andrea Cipriani
Department of Psychiatry
University of Oxford
Warneford Hospital
OX3 7JX Oxford
UK
Email: andrea.cipriani@psych.ox.ac.uk

Summary

Certain limitations of evidence available on drugs and devices at the time of market approval often persist in the post-marketing period. Too often, post-marketing research landscape is fragmented. When regulatory agencies require pharmaceutical and device manufacturers to conduct studies in the post-marketing period, these studies may remain incomplete many years after approval. Even when completed, many post-marketing studies lack meaningful active comparators, have observational designs, and may not collect patient-relevant outcomes. It is crucial for regulators, in collaboration with the industry and patients, to ensure that the important questions that are unanswered at the time of drug and device approval are resolved in a timely fashion during the post-marketing phase. We propose a set of seven key guiding principles that we believe will provide the necessary incentives for pharmaceutical and device manufacturers to generate comparative data in the post-marketing period. First, regulators and pharmaceutical companies (for drugs), notified bodies and manufacturers (for devices) should develop customised evidence generation plans, ensuring that future post-approval studies address any limitations of the data available at the time of market entry that would influence the benefit-risk profiles of drugs and devices. Second, post-marketing studies should be designed hierarchically: priority should be given to efforts aimed at evaluating a product's net clinical benefit in randomised trials compared with current known effective therapy, whenever possible, to address common decisional dilemmas. Third, post-marketing studies should incorporate active comparators as appropriate. Fourth, use of non-randomised studies for the evaluation of clinical benefit in the post-marketing period should be limited to instances when the magnitude of effect is deemed to be very large or when it is possible to reasonably infer the comparative benefits or risks in settings where doing a randomised trial is not feasible. Fifth, efficiency of randomised trials should be improved by streamlining patient recruitment and data collection through innovative design elements. Sixth, governments should directly support and facilitate the production of comparative post-marketing data by investing in the development of collaborative research networks and data systems that reduce the complexity, cost, and waste of rigorous post-marketing research efforts. Seventh, financial incentives and penalties should be developed or more actively reinforced.

The turn of 21st century marked a period when a number of high-profile safety concerns for commonly-used treatments brought significant attention to the role of regulatory agencies in protecting public health.[1,2] For example, rofecoxib, a nonsteroidal anti-inflammatory drug that was approved by the FDA in 1999, was withdrawn from the market in 2004 after a series of studies found that it increased the risk of major cardiovascular events.[3,4] The rise and fall of rofecoxib brought into sharp focus the limitations of the post-marketing research landscape that had until then relied on ad-hoc efforts to generate data on newly-approved drugs and devices.[5]

Acknowledging the need to monitor and evaluate drugs not only prior to their approval but throughout their life span, regulators in Europe and the US have since adopted a "lifecycle" approach. There has been significant progress on the post-marketing safety evaluation of drugs both in Europe and the US, as represented by the Sentinel initiative in the US,[6] and the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) in the European Union[7] (**see Online Appendix**). Although similar efforts are currently underway for devices, such as the National Evaluation System for health Technology in the US, NEST, these are still in their infancy.[8] Post-marketing safety surveillance of medical devices remains decentralised in Europe.[9]

Together with safety, post-market evaluation of clinical benefit for drugs and devices is important for two reasons (**Figure 1**). First, an increasing proportion of approvals have recently benefited from regulatory programs aimed at expediting the development and review of new drugs.[10] Regulators created expedited programs to address unmet patient need in certain serious and debilitating conditions. Approvals in such programs typically rely on earlier-stage data than what is traditionally required for market entry.[11] Second, regulatory agencies have recently articulated their vision for a future where the line separating pre-approval and post-approval periods is blurred. Instead of making binary decisions as to whether a new treatment should be approved or rejected on the basis of available data, regulators are adopting so-called "adaptive" approaches to iterative data collection and evaluation throughout the life-span of therapies.[12] Historically, evidence standards for medical device approvals have been substantially lower than those for drugs (even more so in Europe); post-approval evaluation is therefore essential.[13]

There are significant challenges associated with relying on post-marketing research to address the limitations of data generated on clinical benefit prior to approval.[14] The relatively little investment on post-approval data needs has led to a fragmented research environment.[15] Consequently, the key limitations of the data available on the clinical benefit of drugs and devices at the time of market approval have largely persisted in the post-marketing period.

In this second article of the *Series*, we focus on the potential for generation of comparative effectiveness evidence in the post-marketing period and its coordination with pre-approval research efforts. Our focus is on drugs and devices (implantable and high-risk devices), however the issues and principles covered in this article apply more broadly to other interventions, such as surgery or even health policy interventions. We first review some of the current key challenges of post-marketing research and its three important methodological features: study designs, endpoints and types of comparators. We then propose strategies to improve the future availability of comparative data on new drugs and devices after market entry.

Current post-marketing research landscape

Once drugs are approved by regulatory agencies, research activity on their clinical benefits is primarily influenced by regulatory and market forces.[16] Regulatory agencies in both Europe and the US frequently recommend the completion of post-marketing studies to address the uncertainties that remain at the time of drug approval. For drugs approved through some expedited programs (accelerated approval in the US and conditional marketing authorisation in Europe), regulators may also have post-marketing study *requirements*. In fact, continued market availability of certain expedited drugs may be conditional on the timely completion of such mandatory post-marketing studies. Although the FDA can require post-approval studies for high-risk devices, the lack of a centralised regulatory agency for medical devices in Europe means that post-approval evaluation of benefit remains ad-hoc under the discretion of notified bodies.[17] In terms of market forces, following marketing authorisation, pharmaceutical manufacturers have a limited period of time (usually 10-12 years) during which they have market protections on their approved products. During this period, companies naturally have incentives to invest in research to broaden the approved indications of their products.

Regulatory agency-driven research in the post-marketing period

According to a recent evaluation of FDA approvals from 2009 to 2012, the vast majority of post-marketing commitments, which are not required by any statute or regulation, were for non-clinical studies.[18] Often, post-marketing studies required by regulatory agencies are insufficiently described and do not contain enough information to characterise important study design features such as comparators, randomisation, and endpoints.[19] This is partly because post-marketing studies are rarely underway (or even designed) at the time of market entry. In a recent systematic review, median times permitted by FDA for pharmaceutical companies to submit protocols for their required post-marketing studies ranged from 3 to 15 months after approval.[20]

Post-marketing commitments and requirements may remain incomplete many years after approval.[14] Pharmaceutical companies seldom meet regulatory deadlines in the post-marketing period: only half the studies started in 2009 and 2010 had been completed by the end of 2015, and some companies failed to submit required annual status reports, with the FDA rarely imposing penalties for lack of due diligence.[21] For drugs that received FDA's accelerated approval from 2009 to 2013, almost half of incomplete studies were either terminated or delayed by more than one year.[22] Of the 93 new cancer indications that received FDA's accelerated approval between 1992 and 2017, 51 (55%) fulfilled their post-marketing requirements and verified clinical benefit, 37 (40%) indications did not complete confirmatory trials or verified benefit, and 5 indications (5%) were withdrawn from the market, as they did not show clinical benefit when confirmatory post-approval trials were completed.[23] Perhaps even more critical than the timeliness of these trials is that they generate sufficient reliable evidence on proven effectiveness of therapies to guide future practice long term. For instance, the recently reported results of ANNOUNCE, a large RCT of olaratumab in patients with advanced or metastatic soft-tissue sarcoma, did not confirm an apparent survival benefit of olaratumab in combination with doxorubicin as compared to doxorubicin alone, a standard-of-care treatment and its FDA approval has now been withdrawn.[24]

In Europe, EMA implemented 69 obligations for 26 conditionally-authorised medicines between 2006 and 2016. Over a third of these obligations were subsequently changed and more than half had delays in data submission.[25] Two of the 26 drugs were ultimately withdrawn from the market for commercial reasons, ten were switched to regular approval, and 14 were still under conditional approval, oftentimes several years after market entry.[26,27]

Even when required confirmatory studies are completed, they resemble the design features of pre-marketing studies. Studies about drugs targeting rare conditions have similar designs as those investigating drugs treating non-rare conditions in the post-marketing period.[28] Among novel therapeutic agents that received accelerated approval between 2000 and 2013, clinical benefit was often confirmed in post-marketing trials which had similar design elements to preapproval trials, including reliance on non-randomised designs, and surrogate endpoints.[22] Cancer drugs approved by the FDA based on the surrogate endpoint of response rate were often tested in post-marketing studies that captured other similar surrogate endpoints.[29]

Among high-risk therapeutic devices approved via FDA's most stringent pathway for medical devices, implementation of post-approval studies has been challenging.[30] According to one review, only approximately 13% of initiated post-marketing studies were completed between three and five years after FDA approval.[31] No corresponding figures are available from Europe;

historically, any relevant post-marketing requirements by notified bodies have not been publicly disclosed. The revised Medical Device Regulations, which will come into effect in May 2020 will require public disclosure of such information in the European Union Database for Medical Devices (EUDAMED).[32]

Industry-initiated research in the post-marketing period

Most new drugs have industry-initiated post-marketing studies; however, the majority of these are conducted in therapeutic areas outside of the approved indication (or including participants that extend beyond the indicated population).[28] Such studies could be useful if they produce unbiased evidence on clinically relevant outcomes for the original approved indication and beyond. Instead, companies conduct post-marketing studies to seek approvals in new indications or expand their already-approved indications.[33,34] In addition, most post-approval studies are small and many are not designed to directly evaluate the clinical benefits of newly-approved drugs.[35] In a large systematic evaluation, the quantity and quality of post-approval clinical evidence varied substantially for novel drugs first approved by the FDA on the basis of limited evidence, with few controlled studies published after approval that confirmed clinical benefit using clinical outcomes for the original FDA approved indication.[36] Post-approval evaluation of high-risk devices is sparse.[31]

Evidence to date suggests that valid data confirming the clinical benefits of drugs and devices on the basis of patient-centred and clinically-relevant outcomes may not routinely emerge in the post-marketing period.[37] According to a recent study, only one-fifth of required post-marketing studies of cancer drug indications approved via the FDA's accelerated approval pathway over the past quarter century demonstrated improvements in overall survival in randomised controlled trials, though patients may occasionally derive quality of life benefits in some limited cases without a survival gain.[38]

Coordination of evidence generation between before and after approval

The nature of the current post-marketing research contributes to the well-known problem of research waste.[39] To produce medical knowledge that is clinically informative and satisfies the goals of different stakeholders, increased coordination over the life-course of a product is required. It is, therefore, crucial for regulators, in collaboration with patient groups, health technology assessment organisations, payers, pharmaceutical and device manufacturers, and public funders, to ensure that the important questions that are unanswered at the time of approval are resolved in a timely fashion during the post-marketing phase.

While some regulatory flexibility in approval standards is important in therapeutic areas with significant unmet need, such cases warrant a careful examination of the gap between the existing (what is available) and the optimal (what is needed) evidence that is required for decision making in clinical practice and health policy. If planned carefully, post-marketing studies on drugs and devices can generate timely evidence across the lifecycle of a medical product to reduce the substantial residual uncertainties at the time of regulatory approval.

What is the "optimal" quantity and quality of evidence to inform decision-making in the post-marketing period?

Although it may be difficult to develop universal evidence standards for all therapeutic areas, there are a number of important principles that determine the internal validity and generalisability of research findings.[40] These principles are summarised by the PICOTS (populations, interventions, comparators, outcomes, time periods, and study designs) framework.[41] Clinical studies supporting the regulatory approval of new drugs and high-risk devices often include highly-selective and narrowly-defined patient populations (P); adopt a strict definition of the intervention implemented in protocol-driven settings (I); examine the clinical benefit of the new product against a placebo or no treatment (C); evaluate surrogate measures of effect rather than clinical outcomes (O); have short follow-up durations (T); and lack important study design elements that are required to establish internal validity, i.e., attribute observed effects to the treatment rather than other factors (S).

An important dimension of comparative effectiveness research in the post-marketing period should be to extend the evidence base to patients for whom the current evidence is considered not applicable over a longer period of time and across a broader definition of the intervention. For example, the mean age of patients included in most trials of antiplatelet drugs in secondary prevention of stroke was about 60 years compared to over 75 years in a population-based study.[42] Although the risk of bleeding complications at age <65 years in the population-based cohort was reassuringly similar to that in the previous trials, both the risk and severity of bleeding complications in patients aged over 75 years was several-fold greater and outcomes were substantially worse.[42]

While it is desirable that post-marketing research efforts address the limitations of the evidence base across the full spectrum of the PICOTS framework, priority should be given to research efforts that are aimed at confirming clinical benefits (new and long-term outcomes) of a new product before setting out to examine its generalisability (expanded patients groups) (**Figure 2**). Our primary focus in this article is on the three key methodological features of post-marketing studies – choice of comparators (C), study outcomes (O), and study designs (S). If

data limitations persist on these three features after approval, it remains difficult to establish whether a new drug or device works, and whether it works any better or worse than existing alternatives.

Choice of comparators

Less than a third of studies in the published clinical literature adopt active comparators[43] and only 22% of studies registered in clinicaltrials.gov have active comparators with the remainder employing either placebo or no control.[44] Clinical trials with active comparators are more likely to be sponsored by non-commercial funders, including governments.[43,44] Some of the largest, and most influential, comparative effectiveness trials in the post-marketing period have been publicly funded. For example, one of the landmark comparative effectiveness trials in psychiatry, the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study, which was funded by the US National Institutes Health, compared in a head-to-head fashion the relative effectiveness of second-generation antipsychotic drugs with perphenazine, an older agent, for the treatment of patients with chronic schizophrenia and found that they were not significantly different in overall effectiveness.[45] Another publicly-funded comparative effectiveness trial, the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), showed that inexpensive thiazide-type diuretics were more effective than some of the newer treatment classes.[46]

Comparative effectiveness studies need not always be undertaken as head-to-head comparisons, especially when the addition of therapies to standard care is being considered. A factorial (or partial factorial design) may be preferred in some instances. For instance, the Second International Study of Infarct Survival (ISIS-2) trial showed that the addition of aspirin or streptokinase provided added benefit over not giving either treatment.[47]

Most comparative effectiveness research sponsored by manufacturers focus on their own products. Previous examinations of the geometry of treatment networks in different therapeutic areas have revealed key insights about the preferences of industry sponsors regarding comparators when designing their research studies.[48] Industry-sponsored studies are not necessarily of lower methodological rigour;[49,50] however, many such studies are designed in a way to produce conclusions in favour of the sponsored intervention by selecting comparators with an inferior benefit or harm profile.[51] The vast literature on antidepressants for depression illustrates this phenomenon.[52] Therefore, the choice of comparators is one of the primary mechanisms through which trial sponsors shape the cumulative evidence available to guide treatment decisions in the post-marketing period.[53] Such practices have long-lasting implications on the

relevance of the evidence base for decision-making and highlight the need for regulatory input on the design of post-marketing studies and the conduct of more such studies that are entirely independent of industry sponsors.

For some truly innovative treatments, active comparator may not exist. In other cases, identification of active comparators may be difficult. What is essential is that the new therapy is compared with the current best standard of care (which may be in addition to or as an alternative to such standard of care). Since most therapies that benefit from expedited regulatory programs are for conditions with an unmet need and sometimes without a recognised established therapy, the choice of comparator in the post-marketing period may include the best supportive care (such as for patients with advanced cancer having failed all lines of effective therapy). Physician's choice as comparator may also be considered in areas when choosing the appropriate comparator proves difficult.[54]

Choice of study outcomes

Study outcomes can be broadly divided into two categories: *clinical outcomes* and *surrogate measures*. Clinical outcomes (such as mortality, morbidity, or health-related quality of life) represent direct clinical benefits that are meaningful to patients and clinicians. Surrogate measures (such as laboratory tests, radiographic images, or other biomarkers that correlate with clinical outcomes), on the other hand, are substitutes for clinical outcomes and typically do not represent direct clinical benefit. An observed correlation between intermediate measures and clinical outcomes – however strong – is not adequate to establish surrogacy; changes in a surrogate measure should also reliably predict changes in the clinical outcome, both at the individual and aggregate levels.[55] Usually, it is easier to demonstrate the surrogacy of measure at the aggregate level. For instance, low-density lipoprotein (LDL) cholesterol is associated with coronary heart disease, and, on average, lowering LDL-cholesterol reduces the risk of coronary heart disease.[56] However, specific individuals who suffer coronary events may not always be those with the worst LDL-cholesterol response.

Non-validated surrogate measures may fail to predict treatment effects on clinical outcomes. For instance, despite the (inverse) association between high-density lipoprotein (HDL) cholesterol and coronary heart disease, RCTs of investigational therapeutic agents have failed to demonstrate a reduction in coronary heart disease risk by increasing HDL cholesterol levels.[57,58] Anti-diabetic agents that effectively lower baseline HbA1c levels do not lower the risk of all-cause mortality or deaths due to cardiovascular causes.[59] In the Cardiac Arrythmia

Suppression Trial, use of encainide and flecainide was associated with excess mortality compared with placebo, despite their effect on a surrogate measure, suppression of ventricular ectopy.[60]

Regulatory agencies in both Europe and the US have a long history of approving new treatments on the basis of their effects on surrogate measures alone. Between 2005 and 2012, approximately half of pivotal clinical studies that supported the FDA approval of new drugs used surrogate measures as primary endpoints.[61] Most high-risk device approvals in the US are supported by surrogate measures alone. Surrogate measures have feasibility advantages over clinical outcomes in drug and device development. Surrogate measures typically require smaller sample sizes and shorter study durations to achieve a statistically significant improvement, thereby substantially reducing the cost and complexity of studies, thus possibly allowing faster patient access to new treatments. According to a recent evaluation, using progression-free survival and response rate in cancer trials was associated with an average 11-month and 19-month, respectively, shorter clinical development period compared with using overall survival.[62]

Although certain surrogate measures are well-validated, many of the surrogate measures used for approval decisions are not comprehensively validated, highlighting the need to confirm clinical benefit in the post-marketing period. Surrogate measures are particularly common in cancer trials. More than four fifths of pivotal studies that supported the approval of cancer drugs in the US relied on surrogate measures alone.[61] According to systematic reviews, the relationship between surrogate measures (such as tumour response or progression-free survival) and clinical outcomes (such as overall survival or quality-of-life) is often poor.[63–66]

Surrogate measures used for approval decisions have important implications for clinical practice and health policy. In some cases, improvements observed on surrogate measures may be false positives. According to a large meta-epidemiological review, clinical studies using surrogate measures produced substantially exaggerated results compared with those using clinical outcomes (with relative odds ratios ranging between 1.28 and 1.48).[67] In addition, clinical studies using surrogate measures were twice as likely to find "positive" results compared with studies that captured clinical outcomes.[67]

For example, bevacizumab was approved for the treatment of metastatic breast cancer on the basis of its effect on progression-free survival. In a subsequent trial, however, there was no evidence that bevacizumab improved overall survival among women with this condition.[68] In some cases, drugs approved on the basis of surrogate measures alone may turn out to be harmful. In the recent BELLINI trial, patients with relapsed, refractory multiple myeloma who received venetoclax had worse overall survival than those who received the control treatment

even though venetoclax appeared superior in terms of its effect on surrogate measures of progression-free survival and response rate.[69]

Choice of study designs

The best way to establish the clinical effectiveness of a new treatments to perform a RCT.[70] In non-randomised studies, treatment assignment is influenced by the patient, the provider, or even the setting, resulting in differences in distribution of prognostic factors in patient groups receiving different treatments. Such confounding by indication (or treatment selection bias) is a material threat to the internal validity of non-randomised studies and explains why clinicians, researchers, and policymakers are often reluctant to use observational studies to reach conclusions about the comparative effectiveness of treatments.[71]

Despite the intractable problems of confounding, there is growing enthusiasm for expanding the use of non-randomised studies in the regulatory setting, driven in part by the increased availability of routinely collected data, such as electronic health records, and methods to process and analyse these data.[72,73] The US 21st Century Cures Act, passed in 2016, allows the use of non-randomised studies when approving new indications for already-approved drugs.[74] While non-randomised studies are helpful in monitoring the safety profiles of treatments, they have well-known validity limitations when determining the clinical effect of treatments (of either benefit or harm) with small-to-moderate effect sizes.[75]

Many methods exist that try to control for confounding in non-randomised studies. Some of them (such as propensity score adjustment and instrumental variables) have become more popular over time,[43] but it cannot be secured that such approaches control confounding effectively.[76] When instrumental variables were used, non-randomised studies failed to control for one or more potentially major confounders, which could lead to over-estimation, under-estimation or complete reversal of the effect estimate.[77] While researchers and regulatory agencies continue to develop approaches to address confounding and bias in non-randomised studies, it will still be important to understand fitness of use and ensure validity for a given context.

Generating comparative effectiveness in the post-marketing period

We recommend seven strategies which may promote and facilitate the generation of post-marketing comparative effectiveness research aimed at addressing the limitations of the evidence available on new drugs and devices at the time of approval (**Table 1**).

1. Ensure post-marketing studies address clinically important evidence gaps

In 2012, the Institute of Medicine in the US recommended implementing a Benefit and Risk Assessment and Management Plan to capture in a single "living" document the FDA's evaluation of the known benefits and risks during the entire life cycle of the product.[78,79] This recommendation has not been adopted, highlighting the challenges associated with establishing and continuously monitoring the fast-evolving evidence base on approved products. Without such a document or living library, however, it is not possible for regulators, health technology assessment organisations, payers, clinicians, and patients to stay abreast of the evolving research on new products. When a new product enters the market, it remains difficult for stakeholders in the health system to characterise and quantify the remaining uncertainties on its benefits, especially in relation to the optimal amount of evidence needed to inform decisions. Regulatory agencies are uniquely positioned to summarise what *is* and *is not* known about the comparative benefits and harms of new products when they enter the market. One exception is the notified bodies in Europe, as they do not conduct the evaluation of evidence submitted by device manufacturers.

With input from patient groups, health technology assessment organisations in Europe and the federal and state-level payers in the US, FDA and EMA should develop a customised plan to guide subsequent post-marketing research efforts and ensure that future studies correspond directly to the limitations of the data available at the time of market entry. In Europe, the recently published guidance on the Summary of Safety and Clinical Performance for high-risk and implantable medical devices will require manufacturers to summarise "if there are any unanswered questions relating to the use of the device."[80] Although health technology assessment organisations and payers differ in their evidence requirements, post-marketing research plans could focus on a minimum set of core principles that are shared among different stakeholders,[81] namely the choice of comparators, study outcomes, and study designs.

2. Design post-marketing studies hierarchically

In recent years, an increasing proportion of new products have entered the market on the basis of non-randomised studies that lack active comparators and include only surrogate measures.[82] When data on drugs and devices deviate from the optimal quantity and quality of evidence, manufacturers should be required by regulators to confirm their clinical benefit in a timely manner.[14] The industry's drug and device development plans should include a detailed, feasible, and timely research plan for generating this evidence. Even though post-marketing studies aimed at extending the approved indication could generate useful evidence on the

effectiveness and safety of products, such studies should not commence before the studies set out to demonstrate clinical benefit within the original indication are well underway.[36]

Drugs and devices approved on the basis of earlier-stage data (i.e., without active comparators, using only surrogate measures as study outcomes, in non-randomised studies) should be required by regulators in the post-marketing period to demonstrate their benefits in randomised trials with active and clinically-meaningful comparators that measure patient-centred clinical outcomes that belong to the set of core outcomes for the disease of interest. Although the regulatory agencies currently lack the statutory authority to require such studies outside of certain programs (e.g., accelerated approval pathway in the US), legislative change should be sought to enable such requirements.

Requiring additional studies in the post-marketing period need not adversely affect investment in drug and device development. In 2008, the FDA issued guidance on the need for outcome trials to assess the cardiovascular safety of new diabetes drugs. Since the FDA's guidance, the research and development landscape for diabetes has significantly improved, with several products demonstrating a positive effect on cardiovascular outcomes.[83] Evidence to date suggests that FDA's regulatory action has not negatively affected drug development.[84]

3. Consider a range of active comparators: alternative drugs, devices, and non-drug treatments

In therapeutic areas with an established standard of therapy, post-marketing studies should adopt active comparators. Post-marketing studies may need to keep pace, in a more adaptive fashion, with evolution in usual care. It is sometimes challenging to choose the most appropriate comparator. Network meta-analysis could help identify the best active comparator and address uncertainties in the available evidence base.[85] Industry sponsors have an obligation to test the comparative benefits and harms of their new products against existing alternatives. However, given the importance of this research agenda and the evidence so far suggesting that industry has not always fulfilled this responsibility, independent organisations should play a greater role in designing and running post-marketing trials, ideally leveraging funding from industry (*see recommendation 7*).

Public and non-governmental funders such as the Patient-Centred Outcomes Research Institute in the US and the National Institute for Health Research in the UK should prioritise sponsoring research studies comparing different treatments (e.g., medical therapy vs. device; drug vs. exercise intervention), including alternative service packages, care pathways, and digital treatments.[86] Informative studies may pit one treatment strategy against another (e.g., psychotherapy vs drug treatment; digital therapeutic options vs. traditional therapeutic options).

Yet, such ground-breaking comparative effectiveness studies are too rare, in part due to the difficulty in designing and conducting such studies. Identifying the appropriate types of outcomes, comparisons, and follow-up durations is difficult in trials that compare different treatment categories, and patients and public should be actively and routinely involved in this process.

Until relatively recently, only about a tenth of comparative effectiveness studies published in high-impact general medical journals compared pharmacological and non-pharmacological interventions.[43] In a recent systematic review of almost 400 randomised trials, there were no direct head-to-head comparisons of antihypertensive drugs and structured exercise interventions in terms of their blood pressure-lowering effects.[87] Similarly, in a previous meta-epidemiological review, there was a paucity of randomised trials that directly compared the mortality benefits of drug and non-drug interventions in major chronic conditions.[88] In the absence of such evidence, clinical practice guidelines typically focus on different categories of interventions in isolation and important public health questions are still unanswered.

4. Use non-randomised study designs more selectively

Non-randomised study designs have a clear role for the post-market evaluation of safety, especially for rare or uncommon adverse effects.[89] However, their role in the evaluation of more common effects of either harm or clinical benefit is contested. For example, the evidence of harm associated with rofecoxib was only fully realised after an analysis of ongoing RCTs. When evaluating clinical benefit, we recommend limiting the use of non-randomised studies in the post-marketing period to settings when the evidence of benefit is very large.[90] According to previous theoretical and simulation studies, very large effects are those when a treatment appears at least 5 or 10 times more effective than its comparator.[90–92] Validity of non-randomised studies can be strengthened by mandatory centralised pre-registration of analytical protocols and public availability of collected datasets.[72]

Non-randomised studies could also be used to evaluate whether drugs and devices with an optimal evidence package can be extended to populations outside of those included in RCTs. In the non-randomised EXPRESS study,[93] urgent treatment of transient ischaemic attack and minor stroke with aspirin, blood pressure-lowering medication, and statin resulted in a reduction in 90-day recurrent stroke risk of 80% (adjusted hazard ratio: 0.20, 95% CI: 0.08-0.49). Although the magnitude of this effect was substantially larger than that obtained from previous RCTs,[94] these findings triggered a re-analysis and time-course evaluation of individual participant data from RCTs of aspirin versus placebo. This re-analysis confirmed the dramatic treatment effect

observed in the non-randomised study, which was due to an acute benefit of aspirin on the 90-day risk of recurrent stroke that had not been detected in the previous analyses of RCTs.[95] Notably, EXPRESS was nested in a population-based incidence study of all transient ischaemic attack and stroke with near-complete ascertainment of all patients and outcomes before and after the change in treatment practice, thereby reducing the selection biases inherent in non-randomised studies, as well as maximising external validity – the study included all patients in the underlying population with the condition.

5. Improve the efficiency of randomised trials

While RCTs in the post-marketing period can adopt simpler "pragmatic" designs (as they do not need to comply with strict regulatory agency requirements), they may also require complex design features to capture diverse clinical outcomes that may develop over long time horizons. Therefore, designing studies that are useful in the post-marketing period cannot happen unless there are drastic improvements in the efficiency of RCTs. Costs for clinical trials are very high, especially in the US. The median cost of pivotal regulatory trials was estimated at $19 million for drugs that received FDA approval between 2015 and 2016.[96] However, there is significant variation in reported estimates, with cost per recruited patient ranging from $41 to $6,990 in different studies.[97]

RCTs could benefit from innovative methodological designs (i.e. adaptive design trials, basket trials, registry trials, umbrella protocols), which have their own strengths and weaknesses (**Panel 1**). A key driver of clinical trial expenses is the complexity of patient enrolment, trial procedures and data analysis.[98] To overcome these problems, a new framework is needed which reduces the amount of transactions needed to get the data from a patient into a database for analysis.

For example, the registry-based RCT design leverages data sampling from high-quality registries to facilitate high participant inclusion rates at relatively low costs and, therefore, may offer a robust mechanism by which relevant clinical questions are answered in the post-marketing period. In such trials, online registration identifies patients eligible for inclusion, random allocation occurs in the registry, and study set-up is part of clinical care, including the informed consent process.[99] Ensuring seamless integration of such trials into routine clinical practice may require buy-in from care providers and substantial investment from governments. Another necessary precondition for registry-based trials is the existence of a high-quality registry covering the population to be studied, as the quality of the study data is bound by the quality of

the data in the registry.[100] Registries (as those in the Nordic countries and in the UK) offer a potential source of relevant data.[101,102]

6. Invest in data infrastructure for comparative effectiveness research

Electronic health records, administrative data, and clinical registries currently exist in silos in health care systems. Efforts are underway to build collaborative data infrastructures by linking and leveraging information obtained from separate sources. In compliance with existing regulations to protect the confidentiality of personal data (such as the European General Data Protection Regulation), we recommend accelerating these efforts to facilitate comparative effectiveness research in the post-marketing period,[103] particularly for facilitating pragmatic RCTs. There are already examples of integrated partnerships involving clinical researchers in academia and industry, patients and institutions, also for medical devices.[104,105] In 2015, the Patient-Centered Outcomes Research Institute developed a network including patient-powered research networks and clinical data research networks and launched the randomised ADAPTABLE (Aspirin Dosing: A Patient-centric Trial Assessing Benefits and Long-Term Effectiveness) trial, which is currently underway, comparing two different aspirin doses in high-risk patients with a history of heart disease.[106] ADAPTABLE reflects a pragmatic design by embedding the RCT within usual care, recruiting a diverse patient population with minimal eligibility criteria, promoting the continuation of usual care without standardised treatment protocols, and relying on electronic data collection with reduced need for costly primary data collection.[107]

Another US-based initiative, the National Evaluation System for Health Technology (NEST) focuses on medical devices and coordinates the participation of institutions in a data network to develop data quality and methods standards.[108] The first NEST studies involve multiple health systems answering key clinical and safety questions on a range of medical devices from cardiac and orthopaedic implants to catheters used for soft-tissue ablation, intervertebral body fusion devices, and craniomaxillofacial distractors.[104]

However, progress has been slow, mainly for concerns about the quality and interoperability of underlying data in such systems, as they are not collected for research purposes, and ethical issues regarding data availability and data sharing in non-randomised settings. The future post-marketing research agenda could greatly benefit from the direct engagement of patients by consenting to sharing their electronic data from multiple sources through mobile health apps and electronic platforms, with researchers, regulators and other stakeholders. Concerns about data sharing may pose challenges to such patient-powered research

efforts in the post-marketing period.

7. Create a new set of incentives and reinforce accountability

Pharmaceutical and device manufacturers should be held accountable for demonstrating and confirming the clinical benefit of their products in approved indications. Several guiding principles should be considered to reinforce such accountability. First, the level of payment for drugs and devices should correspond to their added benefit according to robust comparative effectiveness studies. Second, longer marketing protections should be considered for products that convincingly demonstrate their superiority to established standards of care.[109] Third, public reporting of best research practices in the post-marketing period may incentivise companies to invest in comparative studies.[110] Fourth, regulatory approval may be more formally linked to payer policies such as coverage with evidence development whereby the treatment is only available within the context of an ongoing post-marketing clinical trial.[111] However, such strategies should be used very selectively and designed carefully so that they do not place undue administrative burden on public payers.

In terms of penalty mechanisms, regulatory agencies should more actively consider license suspensions, indication restrictions, monetary fines, or even market withdrawal on a case by case basis. FDA and EMA already have the statutory authority to impose monetary penalties for not completing some required studies in a timely manner in expedited programs for drugs. However, regulators currently lack the administrative capacity and financial resources to exercise these powers.[112] Therefore, regulators have yet to penalise pharmaceutical manufacturers for not generating post-marketing data with due diligence. In Europe, the proposal to implement a conditional marketing authorisation pathway for high-risk and implantable devices was rejected, severely restricting attempts at enforcing accountability. This should be the focus of future legislative change. Experience to date suggests that sizeable penalties may be effective to change industry behaviour. Some of the largest corporate fines for criminal offences (imposed by the US Department of Justice and not by regulators) have been for pharmaceutical companies for off-label promotion of their products.[113] Such financial penalties and the media coverage associated with them affected subsequent marketing practices and use.[114,115]

Conclusions

Comparative evidence on the benefits and harms of new and existing drugs and devices rarely emerges in the post-marketing period. There is an opportunity to coordinate research efforts between before and after approval. Policymakers and regulators can incentivise the

generation of comparative data in the post-marketing period by ensuring that post-marketing studies directly correspond to the limitations of pre-approval studies; designing post-marketing studies hierarchically (first to confirm clinical benefit and then to examine generalisability); limiting the use of non-randomised study designs when evaluating clinical benefit; improving the efficiency of randomised trials; investing in data infrastructure; and creating new incentive and penalty mechanisms.

Acknowledgments

This study was funded by the National Institute for Health Research (NIHR) Oxford Health Biomedical Research Centre (BRC-1215-20005). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health.

Contributors

HN and AC conceived and designed the study. AC wrote the first draft of the manuscript. All other authors contributed to the writing of the final version of the manuscript, and agreed with the results and conclusions of this Article.

Declaration of interest

AC is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility, by an NIHR Research Professorship (grant RP-2017-08-ST2-006) and by the NIHR Oxford Health Biomedical Research Centre (grant BRC-1215-20005). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health. All other authors declare no competing interests.

Role of the funding source

The funder of this study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit for publication. AC and HN had full access to all the data, and AC was responsible for the decision to submit for publication.

Panel 1. Innovative (or non-conventional) study designs for randomised trials. This panel aims to outline key features of selected innovative trial designs. Adaptive trials use information generated during trial conduct to alter subsequent operations in a pre-specified way. In 2018 the FDA provided a draft guidance on "master protocols", which refer to a master (or core) protocol, upon which multiple questions can be asked about the effectiveness of interventions for a particular disease or condition. Novel trial designs that use master protocols include basket, umbrella and adaptive platform trials. Both elements (master protocol and adaptive design features) add complexity, but with the intent of improving the efficiency of knowledge generation. Registry-based trials include a randomisation module in a large inclusive clinical registry with unselected consecutive enrolment, to combine the advantages of a prospective randomised trial with the strengths of a large-scale all-comers clinical registry.

Adaptive trials

Adaptive trials are designed to maximise flexibility, without compromising the validity and integrity of the trial. Modifications ("adaptations") to aspects of the ongoing trial can be pre-specified and prospectively planned, including adding or dropping treatment arms, changing dosages, sample size re-estimations and alterations. Adaptive trials aim to identify the patients who are most likely to benefit from a treatment*:*

- When single or multiple different disease populations are studied
- When single or multiple interventions are studied; adaptive trials can utilise multiple therapies
    - The sample size can vary significantly from very large to small depending on the study sub-design, interim sample size reassessments are also utilised
- *Use:* Both exploratory and confirmatory clinical trials
- *Advantage:* Reduces the use of resources, decreases the time of trial completion, limits the number of participants allocated to inferior interventions and improves the probability of success of the trial
- *Disadvantage:* Subject to operational bias, due to the leakage of interim results and the potential to influence investigator behaviour

Basket trials

Basket trials are used to test the effect of a single drug, or a combination of drugs on single mutation (a single target) in multiple diseases ('baskets'):

- When multiple disease populations are being studied

- Including different histology types or different tumour types, often referred to as 'histology *independent*'

- When a single intervention is studied, which is targeted, matched or is biomarker specific
  - The sample size is relatively large but typically smaller than umbrella trial sample sizes and are generally single-arm trials
- *Use:* Commonly discovery-based trials used in early stages of development
- *Advantage*: An efficient way of identifying if a drug targeting a specific genetic mutation in one site of the body may be effective in treating that same genetic mutation found in tumour located in a different part of the body
- *Disadvantage:* The use of the underlying assumption that molecular profiling is a sufficient replacement of histological tumour typing

Platform trials (or adaptive platform trials)

(Adaptive) platform trials are able to study multiple interventions in a disease or condition in a perpetual manner, with interventions entering and leaving the platform on the basis of a predefined decision algorithm.

- When a single disease population is studied, usually limited to a single disease or single histology/tumour type
  - A broad cohort of participants are enrolled, and later stratified into different subtypes based on clinical or biomarker criteria
- When multiple interventions are studied; they utilise multiple therapies in a perpetual trial design
  - Large sample sizes are often required as platform trials have the capacity to add and drop trial arms as futility or efficacy are demonstrated, often using a decision algorithm
- *Use:* Can range from proof of concept studies through to confirmation of application trials
- *Advantage*: Platform enables characterisation of the safety and efficacy of novel treatment combinations, potentially across diseases, mechanisms and sponsors, that would otherwise not be feasible in one trial
- *Disadvantage*: Potential complexity of the trial implementation and planning, often requiring complex collaborations across sponsors and participating sites

Registry-based trials

Registry trials are pragmatic trials that use registries as platforms for health records, data collections, randomisation and follow-up. The advancement of electronic data collection systems has led to the increasing number of developed registries used for research, policy, and administrative purposes. A clinical registry can be used for collection of baseline variables and to identify eligible patients for a study:

- When single or multiple different disease populations are studied
- When single or multiple interventions are studied
    - Typically use large sample sizes as large observational cohorts of patients.
- *Use*: Often later on in drug development, and not suitable for trials that need comprehensive safety reporting or intense pharmacokinetic or pharmacodynamic profiling
- *Advantage:* Low cost, enhanced generalisability of findings (real-world setting), rapid consecutive enrolment and follow-up
- *Disadvantage*: Variable data quality, potentially poorly defined variables, limited facility to collect detailed safety reporting

Umbrella trials

Umbrella trials are designed to evaluate the impact of different drugs on different mutations in a single type of cancer:

- When a single disease population is studied: trials are limited to a single disease or single histology/tumour type
    - Multiple biomarker matched subgroups of patients are used, patients are assigned to biomarker subgroups using a biomarker allocation algorithm
- When multiple interventions are studied; umbrella trials utilise multiple therapies
    - Large sample sizes are often required, patients with multiple biomarkers can be included in more than one trial arm
- *Use*: Can range from proof of concept studies to confirmation of application trials
- *Advantage*: Capacity to draw meaningful conclusions specific to a tumour type
- *Disadvantage*: Flexibility is limited, due to use of a single tumour type, particularly with rare diseases, where further subclassification may lead to poor accrual.

Key References

1. Antoniou M et al. Biomarker-guided adaptive trial designs in phase II and phase III: a methodological review. *PLoS ONE* 2016;11:e0149803.
2. Berry SM, Connor JT, Lewis RJ. The Platform Trial: An Efficient Strategy for Evaluating Multiple Treatments. *JAMA.* 2015;313(16):1619–1620.
3. Bhatt D, Mehta C. Adaptive Designs for Clinical Trials. *N Engl J Med* 2016;375:65-74.
4. Bothwell LE et al. Assessing the Gold Standard--Lessons from the History of RCTs. *N Engl J Med.* 2016 Jun 2;374(22):2175-81.
5. James S et al. Registry-based randomized clinical trials--a new clinical trial paradigm. *Nat Rev Cardiol.* 2015;12(5):312-6.
6. Janiaud P et al. New clinical trial designs in the era of precision medicine: An overview of definitions, strengths, weaknesses, and current use in oncology. *Cancer Treat Rev.* 2019;73:20-30.
7. Renfro LA, Sargent DJ. Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. *Ann Oncol.* 2017;28(1):34-43.
8. The Adaptive Platform Trials Coalition. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat Rev Drug Discov.* 2019 Oct;18(10):797-807.
9. US FDA. Master protocols: efficient clinical trial design strategies to expedite development of oncology drugs and biologics guidance for industry. https://www.fda.gov/media/120721/download. Accessed December 10, 2019
10. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med.* 2017;377(1):62-70.

694 **Table 1.** Strategies aimed at incentivising pharmaceutical and device manufacturers to generate comparative data on drugs and high-risk and
695 implantable devices in the post-marketing period, with a summary of the key recommendations and the target stakeholders.
696

Overall strategies	Key recommendations	Target stakeholders (in alphabetical order)
(1) Ensuring post-marketing studies address evidence gaps	• Develop a customised plan to guide subsequent post-marketing research efforts • Ensure that future studies correspond directly to the limitations of the data available at the time of market entry	• Health technology assessment organisations • Public payers • Regulatory agencies (FDA and EMA)
(2) Designing post-marketing studies hierarchically	• Products approved on the basis of studies without active comparators should be required to have active-comparators (or controlled trials compared with an known effective control)*. • Products approved on the basis of surrogate measures alone should be required to confirm their clinical benefit on patient-relevant outcomes. • Products approved on the basis of non-randomised study designs should be required to have randomised trials. • Once randomised trials with active-comparators and clinical outcomes are available, post-marketing research efforts should pivot to evaluating the applicability of evidence to broader patient populations over longer time horizons.	• Device manufacturers • Pharmaceutical manufacturers • Regulatory agencies (FDA and EMA) • Government research funding bodies
(3) Considering active comparators beyond drugs and devices	• Network meta-analyses should be used to choose appropriate active comparators. • Active comparators should include non-pharmacological interventions.	• Device manufacturers • Government research funding bodies • Non-governmental research funders • Pharmaceutical manufacturers
(4) More selective use of non-randomised study designs	• When evaluating clinical benefit, non-randomised studies should be used only when the evidence of benefit is overwhelmingly positive. • When randomised trials are available, non-randomised studies should be used to evaluate applicability to populations outside of those included in trials.	• Device manufacturers • Government research funding bodies • Non-governmental research funders • Pharmaceutical manufacturers • Regulatory agencies (FDA and EMA)

(5)	Improving the efficiency of randomised trials	• Randomised trials should adopt innovative design elements to improve efficiency of patient recruitment and data collection.	• Device manufacturers • Government research funding bodies • Non-governmental research funders • Pharmaceutical manufacturers
(6)	Investing in data infrastructure for comparative effectiveness research	• Efforts aimed at establishing collaborative research networks and data systems should be accelerated.	• Governments
(7)	Creating a new set of incentives and reinforcing accountability	• Payment for new drugs and devices should correspond to their added benefit over existing alternatives. • Penalties should be invoked for not generating comparative data in the post-marketing period.	• Governments • Health technology assessment organisations • Public payers • Regulatory agencies (FDA and EMA)

697
698
699
700
701 * Active comparator trials may include head to head comparisons of a new product with a known effective treatment; the addition of the product to
702 current effective treatment versus that same treatment used alone; and placebo-controlled trials of the new product when given in addition to best
703 standard care. The essential requirement is that the new product (alone or in combination) is being compared with a known effective therapy.

References

1. Horton R. Vioxx, the implosion of Merck, and aftershocks at the FDA. *The Lancet*. 2004;364(9450):1995-1996.

2. Topol EJ. Failing the Public Health — Rofecoxib, Merck, and the FDA. *N Engl J Med*. 2004;351(17):1707-1709. doi:10.1056/NEJMp048286

3. Mukherjee D, Nissen SE, Topol EJ. Risk of Cardiovascular Events Associated With Selective COX-2 Inhibitors. *JAMA*. 2001;286(8):954-959. doi:10.1001/jama.286.8.954

4. Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *The lancet*. 2004;364(9450):2021-2029.

5. Fontanarosa PB, Rennie D, DeAngelis CD. Postmarketing surveillance—lack of vigilance, lack of trust. *Jama*. 2004;292(21):2647-2650.

6. Platt R, Brown JS, Robb M, et al. The FDA Sentinel Initiative - An Evolving National Resource. *N Engl J Med*. 2018;379(22):2091-2093. doi:10.1056/NEJMp1809643

7. Kurz X, Perez-Gutthann S, ENCePP Steering Group. Strengthening standards, transparency, and collaboration to support medicine evaluation: Ten years of the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). *Pharmacoepidemiol Drug Saf*. 2018;27(3):245-252. doi:10.1002/pds.4381

8. Fleurence RL, Shuren J. Advances in the Use of Real-World Evidence for Medical Devices: An Update From the National Evaluation System for Health Technology. *Clin Pharmacol Ther*. March 2019. doi:10.1002/cpt.1380

9. Fraser AG, Butchart EG, Szymański P, et al. The need for transparency of clinical evidence for medical devices in Europe. *The Lancet*. 2018;392(10146):521-530. doi:10.1016/S0140-6736(18)31270-4

10. Kesselheim AS, Wang B, Franklin JM, Darrow JJ. Trends in utilization of FDA expedited drug development and approval programs, 1987-2014: cohort study. *BMJ*. 2015;351.

11. Wallach JD, Ross JS, Naci H. The US Food and Drug Administration's expedited approval programs: Evidentiary standards, regulatory trade-offs, and potential improvements. *Clin Trials*. 2018;15(3):219-229.

12. Eichler H-G, Oye K, Baird LG, et al. Adaptive licensing: taking the next step in the evolution of drug approval. *Clin Pharmacol Ther*. 2012;91(3):426-437.

13. Kramer DB, Xu S, Kesselheim AS. Regulation of Medical Devices in the United States and European Union. *N Engl J Med*. 2012;366(9):848-855. doi:10.1056/NEJMhle1113918

14. Wallach JD, Ross JS, Naci H. The US Food and Drug Administration's expedited approval programs: Addressing premarket flexibility with enhanced postmarket evidence generation. *Clin Trials*. 2018;15(3):243-246.

15. Ridley DB, Kramer JM, Tilson HH, Grabowski HG, Schulman KA. Spending On Postapproval Drug Safety. *Health Aff (Millwood)*. 2006;25(2):429-436. doi:10.1377/hlthaff.25.2.429

16. Zeitoun J-D, Ross JS, Atal I, et al. Factors Associated With Postmarketing Research for Approved Indications for Novel Medicines Approved by Both the FDA and EMA Between 2005 and 2010: A Multivariable Analysis. *Clin Pharmacol Ther*. 2018;104(5):1000-1007. doi:10.1002/cpt.1038

17. Kramer DB, Xu S, Kesselheim AS. How Does Medical Device Regulation Perform in the United States and the European Union? A Systematic Review. *PLOS Med*. 2012;9(7):e1001276. doi:10.1371/journal.pmed.1001276

18. Wallach JD, Luxkaranayagam AT, Dhruva SS, Miller JE, Ross JS. Postmarketing commitments for novel drugs and biologics approved by the US Food and Drug Administration: a cross-sectional analysis. *BMC Med*. 2019;17(1):117. doi:10.1186/s12916-019-1344-3

19. Wallach JD, Egilman AC, Dhruva SS, et al. Postmarket studies required by the US Food and Drug Administration for new drugs and biologics approved between 2009 and 2012: cross sectional analysis. *BMJ*. 2018;361. doi:10.1136/bmj.k2031

20. Wallach JD, Egilman AC, Ross JS, Woloshin S, Schwartz LM. Timeliness of Postmarket Studies for New Pharmaceuticals Approved Between 2009 and 2012: a Cross-Sectional Analysis. *J Gen Intern Med*. 2019;34(4):492-495.

21. Woloshin S, Schwartz LM, White B, Moore TJ. The Fate of FDA Postapproval Studies. *N Engl J Med*. 2017;377(12):1114-1117.

22. Naci H, Smalley KR, Kesselheim AS. Characteristics of Preapproval and Postapproval Studies for Drugs Granted Accelerated Approval by the US Food and Drug Administration. *JAMA*. 2017;318(7):626-636.

23. Beaver JA, Howie LJ, Pelosof L, et al. A 25-Year Experience of US Food and Drug Administration Accelerated Approval of Malignant Hematology and Oncology Drugs and Biologics: A Review. *JAMA Oncol*. 2018;4(6):849-856.

24. Tap WD, Wagner AJ, Papai Z, et al. ANNOUNCE: A randomized, placebo (PBO)-controlled, double-blind, phase (Ph) III trial of doxorubicin (dox) + olaratumab versus dox + PBO in patients (pts) with advanced soft tissue sarcomas (STS). *J Clin Oncol*. 2019;37(18_suppl):LBA3-LBA3.

25. Bloem LT, Mantel-Teeuwisse AK, Leufkens HGM, De Bruin ML, Klungel OH, Hoekman J. Postauthorization Changes to Specific Obligations of Conditionally Authorized Medicines in the European Union: A Retrospective Cohort Study. *Clin Pharmacol Ther*. 2019;105(2):426-435.

26. Banzi R, Gerardi C, Garattini S. Approvals of drugs with uncertain benefit–risk profiles in Europe. *Eur J Intern Med*. 2015;26(8):572-584.

27. Banzi R, Gerardi C, Garattini S. Conditional approval of medicines by the EMA. *BMJ Br Med J Online*. 2017;357.

28. Skydel JJ, Luxkaranayagam AT, Dhruva SS, Ross JS, Wallach JD. Analysis of Postapproval Clinical Trials of Therapeutics Approved by the US Food and Drug Administration Without Clinical Postmarketing Requirements or Commitments. *JAMA Netw Open*. 2019;2(5):e193410-e193410.

29. Chen EY, Raghunathan V, Prasad V. An Overview of Cancer Drugs Approved by the US Food and Drug Administration Based on the Surrogate End Point of Response Rate. *JAMA Intern Med*. May 2019.

30. Reynolds IS, Rising JP, Coukell AJ, Paulson KH, Redberg RF. Assessing the Safety and Effectiveness of Devices After US Food and Drug Administration Approval: FDA-Mandated Postapproval Studies. *JAMA Intern Med*. 2014;174(11):1773-1779.

31. Rathi VK, Krumholz HM, Masoudi FA, Ross JS. Characteristics of Clinical Studies Conducted Over the Total Product Life Cycle of High-Risk Therapeutic Medical Devices Receiving FDA Premarket Approval in 2010 and 2011. *JAMA*. 2015;314(6):604-612.

32. Melvin T, Torre M. New medical device regulations: the regulator's view. *EFORT Open Rev*. 2019;4(6):351-356.

33. DiMasi JA. Innovating by developing new uses of already-approved drugs: trends in the marketing approval of supplemental indications. *Clin Ther*. 2013;35(6):808-818.

34. Wang B, Kesselheim AS. Characteristics of efficacy evidence supporting approval of supplemental indications for prescription drugs in United States, 2005-14: systematic review. *BMJ*. 2015;351:h4679.

35. Naci H, Wouters OJ, Gupta R, Ioannidis JP. Timing and characteristics of cumulative evidence available on novel therapeutic agents receiving Food and Drug Administration accelerated approval. *Milbank Q*. 2017;95(2):261-290.

36. Pease AM, Krumholz HM, Downing NS, Aminawung JA, Shah ND, Ross JS. Postapproval studies of drugs initially approved by the FDA on the basis of limited evidence: systematic review. *BMJ*. 2017;357:j1680.

37. Federico CA, Wang T, Doussau A, Mogil JS, Fergusson D, Kimmelman J. Assessment of Pregabalin Postapproval Trials and the Suggestion of Efficacy for New Indications: A Systematic Review. *JAMA Intern Med*. 2019;179(1):90-97.

38. Gyawali B, Hey SP, Kesselheim AS. Assessment of the Clinical Benefit of Cancer Drugs Receiving Accelerated Approval. *JAMA Intern Med*. May 2019.

39. Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet*. 2014;383(9912):166-175.

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

41. Guyatt G, Drummond R, Meade M, Cook D. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd Ed.*

42. Li L, Geraghty OC, Mehta Z, Rothwell PM. Age-specific risks, severity, time course, and outcome of bleeding on long-term antiplatelet treatment after vascular events: a

population-based cohort study. *Lancet Lond Engl*. 2017;390(10093):490-499. doi:10.1016/S0140-6736(17)30770-5

43. Hochman M, McCormick D. Characteristics of Published Comparative Effectiveness Studies of Medications. *JAMA*. 2010;303(10):951-958.

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

45. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of Antipsychotic Drugs in Patients with Chronic Schizophrenia. *N Engl J Med*. 2005;353(12):1209-1223.

46. 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):2981-2997.

47. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. *Lancet Lond Engl*. 1988;2(8607):349-360.

48. Salanti G, Kavvoura FK, Ioannidis JPA. Exploring the geometry of treatment networks. *Ann Intern Med*. 2008;148(7):544-553.

49. Dias S, Welton NJ, Ades AE. Study designs to detect sponsorship and other biases in systematic reviews. *J Clin Epidemiol*. 2010;63(6):587-588.

50. Naci H, Dias S, Ades AE. Industry sponsorship bias in research findings: a network meta-analysis of LDL cholesterol reduction in randomised trials of statins. *BMJ*. 2014;349:g5741.

51. Lathyris DN, Patsopoulos NA, Salanti G, Ioannidis JP. Industry sponsorship and selection of comparators in randomized clinical trials. *Eur J Clin Invest*. 2010;40(2):172-182.

52. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*. 2018;391(10128):1357-1366.

53. Fabbri A, Lai A, Grundy Q, Bero LA. The influence of industry sponsorship on the research agenda: a scoping review. *Am J Public Health*. 2018;108(11):e9-e16.

54. Donoghue M, Lemery SJ, Yuan W, et al. Eribulin mesylate for the treatment of patients with refractory metastatic breast cancer: use of a "physician's choice" control arm in a randomized approval trial. *Clin Cancer Res*. 2012;18(6):1496-1505.

55. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996;125(7):605-613.

56. Ference BA, Ginsberg HN, Graham I, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J*. 2017;38(32):2459-2472.

57. Niacin in Patients with Low HDL Cholesterol Levels Receiving Intensive Statin Therapy. *N Engl J Med*. 2011;365(24):2255-2267.

58. Lincoff AM, Nicholls SJ, Riesmeyer JS, et al. Evacetrapib and Cardiovascular Outcomes in High-Risk Vascular Disease. *N Engl J Med*. 2017;376(20):1933-1942.

59. Boussageon R, Bejan-Angoulvant T, Saadatian-Elahi M, et al. Effect of intensive glucose lowering treatment on all cause mortality, cardiovascular death, and microvascular events in type 2 diabetes: meta-analysis of randomised controlled trials. *BMJ*. 2011;343.

60. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and Morbidity in Patients Receiving Encainide, Flecainide, or Placebo. *N Engl J Med*. 1991;324(12):781-788.

61. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting fda approval of novel therapeutic agents, 2005-2012. *JAMA*. 2014;311(4):368-377.

62. Chen EY, Joshi SK, Tran A, Prasad V. Estimation of Study Time Reduction Using Surrogate End Points Rather Than Overall Survival in Oncology Clinical Trials. 2019;179(5):642-647.

63. Prasad V, Kim C, Burotto M, Vandross A. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. *JAMA Intern Med*. 2015;175(8):1389-1398.

64. Haslam A, Hey SP, Gill J, Prasad V. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. *Eur J Cancer*. 2019;106:196-211.

65. Hwang TJ, Gyawali B. Association between progression-free survival and patients' quality of life in cancer clinical trials. *Int J Cancer*. 2018;0(0).

66. Kovic B, Jin X, Kennedy SA, et al. Evaluating Progression-Free Survival as a Surrogate Outcome for Health-Related Quality of Life in Oncology: A Systematic Review and Quantitative Analysis. *JAMA Intern Med*. 2018;178(12):1586-1596.

67. Ciani O, Buyse M, Garside R, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *BMJ*. 2013;346:f457.

68. Carpenter D. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton University Press; 2010. http://www.jstor.org/stable/j.ctt7t5st.

69. Kumar S, Rajkumar SV. Surrogate endpoints in randomised controlled trials: a reality check. *Lancet Lond Engl*. 2019;394(10195):281-283.

70. Garattini S, Jakobsen JC, Wetterslev J, et al. Evidence-based clinical practice: overview of threats to the validity of evidence and how to minimise them. *Eur J Intern Med*. 2016;32:13-21.

71. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *The Lancet*. 2008;372(9656):2152-2161.

72. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clin Pharmacol Ther*. 2019;105(4):867-877.

73. Corrigan-Curay J, Sacks L, Woodcock J. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. *JAMA*. 2018;320(9):867-868.

74. Avorn J, Kesselheim AS. The 21st Century Cures Act—will it take us back in time? *N Engl J Med*. 2015;372(26):2473-2475.

75. Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox P, Gliklich R. Why observational studies should be among the tools used in comparative effectiveness research. *Health Aff (Millwood)*. 2010;29(10):1818-1825.

76. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. 2006;59(5):437. e1-437. e24.

77. Garabedian LF, Chu P, Toh S, Zaslavsky AM, Soumerai SB. Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Ann Intern Med*. 2014;161(2):131-138.

78. Mello MM, Goodman SN, Faden RR. *Ethical Considerations in Studying Drug Safety—the Institute of Medicine Report*. Mass Medical Soc; 2012.

79. Psaty BM, Meslin EM, Breckenridge A. A lifecycle approach to the evaluation of FDA approval methods and regulatory actions: opportunities provided by a new IOM report. *Jama*. 2012;307(23):2491-2492.

80. European Commission. *MDCG 2019-9 Summary of Safety and Clinical Performance A Guide for Manufacturers and Notified Bodies*.; 2019. https://ec.europa.eu/docsroom/documents/37323. Accessed October 30, 2019.

81. Kleijnen S, George E, Goulden S, et al. Relative Effectiveness Assessment of Pharmaceuticals: Similarities and Differences in 29 Jurisdictions. *Value Health*. 2012;15(6):954-960.

82. Zhang AD, Puthumana J, Downing NS, Shah ND, Krumholz H, Ross JS. Clinical Trial Evidence Supporting FDA Approval of Novel Therapeutic Agents Over Three Decades, 1995-2017: Cross-Sectional Analysis. *medRxiv*. January 2019:19007047. doi:10.1101/19007047

83. Menon V, Lincoff AM. Cardiovascular Safety Evaluation in the Development of New Drugs for Diabetes Mellitus. *Circulation*. 2014;129(25):2705-2713.

84. Hwang TJ, Franklin JM, Kesselheim AS. Effect of US Food and Drug Administration's Cardiovascular Safety Guidance on Diabetes Drug Development. *Clin Pharmacol Ther*. 2017;102(2):290-296.

85. Salanti G, Nikolakopoulou A, Sutton AJ, et al. Planning a future randomized clinical trial based on a network of relevant past trials. *Trials*. 2018;19(1):365.

86. Peden CJ, Stephens T, Martin G, et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. *The Lancet*. 2019;393(10187):2213-2221.

87. Naci H, Salcher-Konrad M, Dias S, et al. How does exercise treatment compare with antihypertensive medications? A network meta-analysis of 391 randomised controlled trials assessing exercise and medication effects on systolic blood pressure. *Br J Sports Med*. December 2018:bjsports-2018-099921.

88. Naci H, Ioannidis JPA. Comparative effectiveness of exercise and drug interventions on mortality outcomes: metaepidemiological study. *BMJ*. 2013;347:f5577.

89. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet Lond Engl*. 2004;363(9422):1728-1731.

90. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ*. 2007;334(7589):349-351.

91. Razavi M, Glasziou P, Klocksieben FA, Ioannidis JPA, Chalmers I, Djulbegovic B. US Food and Drug Administration Approvals of Drugs and Devices Based on Nonrandomized Clinical Trials: A Systematic Review and Meta-analysis. *JAMA Netw Open*. 2019;2(9):e1911111-e1911111.

92. Djulbegovic B, Glasziou P, Klocksieben FA, et al. Larger effect sizes in nonrandomized studies are associated with higher rates of EMA licensing approval. *J Clin Epidemiol*. 2018;98:24-32.

93. Rothwell PM, Giles MF, Chandratheva A, et al. Effect of urgent treatment of transient ischaemic attack and minor stroke on early recurrent stroke (EXPRESS study): a prospective population-based sequential comparison. *Lancet Lond Engl*. 2007;370(9596):1432-1442.

94. Rothwell PM, Algra A, Amarenco P. Medical treatment in acute and long-term secondary prevention after transient ischaemic attack and ischaemic stroke. *The Lancet*. 2011;377(9778):1681-1692.

95. Rothwell PM, Algra A, Chen Z, Diener H-C, Norrving B, Mehta Z. Effects of aspirin on risk and severity of early recurrent stroke after transient ischaemic attack and ischaemic stroke: time-course analysis of randomised trials. *The Lancet*. 2016;388(10042):365-375.

96. Moore TJ, Zhang H, Anderson G, Alexander GC. Estimated Costs of Pivotal Trials for Novel Therapeutic Agents Approved by the US Food and Drug Administration, 2015-2016. *JAMA Intern Med*. 2018;178(11):1451-1457.

97. Speich B, von Niederhäusern B, Schur N, et al. Systematic review on costs and resource use of randomized clinical trials shows a lack of transparent and comprehensive data. *J Clin Epidemiol*. 2018;96:1-11.

98. Reith C, Landray M, Devereaux PJ, et al. Randomized Clinical Trials — Removing Unnecessary Obstacles. *N Engl J Med*. 2013;369(11):1061-1065.

99. Yndigegn T, Hofmann R, Jernberg T, Gale CP. Registry-based randomised clinical trial: efficient evaluation of generic pharmacotherapies in the contemporary era. *Heart Br Card Soc*. 2018;104(19):1562-1567.

100. Li G, Sajobi TT, Menon BK, et al. Registry-based randomized controlled trials- what are the advantages, challenges, and areas for future research? *J Clin Epidemiol*. 2016;80:16-24.

101. Stewart R, Soremekun M, Perera G, et al. The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: development and descriptive data. *BMC Psychiatry*. 2009;9(1):51.

102. Nyberg K, Hedman P. Swedish guidelines for registry-based randomized clinical trials. *Ups J Med Sci*. 2019;124(1):33-36.

103. Fitzpatrick T, Perrier L, Shakik S, et al. Assessment of Long-term Follow-up of Randomized Trial Participants by Linkage to Routinely Collected Data: A Scoping Review and Analysis. *JAMA Netw Open*. 2018;1(8):e186019-e186019.

104. Fleurence RL, Blake K, Shuren J. The future of registries in the era of real-world evidence for medical devices. *JAMA Cardiol*. 2019.

105. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc*. 2014;21(4):578-582.

106. Johnston A, Jones WS, Hernandez AF. The ADAPTABLE trial and aspirin dosing in secondary prevention for patients with coronary artery disease. *Curr Cardiol Rep*. 2016;18(8):81.

107. Hernandez AF, Fleurence RL, Rothman RL. The ADAPTABLE Trial and PCORnet: shining light on a new research paradigm. *Ann Intern Med*. 2015;163(8):635-636.

108. Shuren J, Califf RM. Need for a national evaluation system for health technology. *Jama*. 2016;316(11):1153-1154.

109. Kesselheim AS. Using market-exclusivity incentives to promote pharmaceutical innovation. *N Engl J Med*. 2010;363(19):1855-1862.

110. Miller J, Ross JS, Wilenzick M, Mello MM. Sharing of clinical trial data and results reporting practices among large pharmaceutical companies: cross sectional descriptive study and pilot of a tool to improve company practices. *BMJ*. 2019;366:l4217.

111. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. *Value Health*. 2012;15(3):570-579.

112. Herder M. Pharmaceutical Drugs of Uncertain Value, Lifecycle Regulation at the US Food and Drug Administration, and Institutional Incumbency. *Milbank Q*. 2019.

113. Kmietowicz Z. Eli Lilly pays record $1.4bn for promoting off-label use of olanzapine. *BMJ*. 2009;338.

114. Kesselheim AS, Darby D, Studdert DM, Glynn R, Levin R, Avorn J. False Claims Act prosecution did not deter off-label drug use in the case of Neurontin. *Health Aff (Millwood)*. 2011;30(12):2318-2327.

115. Chace MJ, Zhang F, Fullerton CA, Huskamp HA, Gilden D, Soumerai SB. Intended and unintended consequences of the gabapentin off-label marketing lawsuit among patients with bipolar disorder. *J Clin Psychiatry*. 2012;73(11):1388-1394.