Generating the Right Evidence at the Right Time: Principles of a New Class of Flexible Augmented Clinical Trial Designs

Cornelia Dunger-Baldauf, Rob Hemmings, Frank Bretz, Byron Jones, Anja Schiel and Chris Holmes

To support informed decision making, clear descriptions of the beneficial and harmful effects of a treatment are needed by various stakeholders. The current paradigm is to generate evidence sequentially through different experiments. However, data generated later, perhaps through observational studies, can be difficult to compare with earlier randomized trial data, resulting in confusion in understanding and interpretation of treatment effects. Moreover, the scientific questions these later experiments can serve to answer often remain vague. We propose Flexible Augmented Clinical Trial for Improved Evidence Generation (FACTIVE), a new class of study designs enabling flexible augmentation of confirmatory randomized controlled trials with concurrent and close-to-real-world elements. Our starting point is to use clearly defined objectives for evidence generation, which are formulated through early discussion with health technology assessment (HTA) bodies and are additional to regulatory requirements for authorization of a new treatment. These enabling designs facilitate estimation of certain well-defined treatment effects in the confirmatory part and other complementary treatment effects in a concurrent real-world part. Each stakeholder should use the evidence that is relevant within their own decision-making framework. High quality data are generated under one single protocol and the use of randomization ensures rigorous statistical inference and interpretation within and between the different parts of the experiment. Evidence for the decision making of HTA bodies could be available earlier than is currently the case.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
- Studies aimed at broadening the evidence beyond that generated to support regulatory decision making are usually conducted independently from, and subsequent to, initial randomized controlled trials (RCTs), often using observational study designs.

WHAT QUESTION DID THIS STUDY ADDRESS?
- We consider the questions of the optimal design and timeline for broader, high-quality evidence generation in drug development. We address whether the conventional sequential approach to evidence generation can be improved through an integrated design that provides concurrent evidence on drug effects under clinical trial conditions and conditions that are closer to real-world.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
- Active consideration, whereas planning confirmatory clinical trials, of the right time to generate the right evidence can lead to improved evidence gathering. We present a new class of augmented RCT designs that improve the quality, interpretability, and timeliness of generating high-quality evidence.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
- Active consideration of the optimal timeline for evidence gathering and consideration of augmented confirmatory phase III designs (such as the Flexible Augmented Clinical Trial for Improved evidence generation (FACTIVE) design that we propose) will benefit discussions with, and the quality of evidence provided to, regulatory and health technology assessment bodies.

To support informed decision making by pharmaceutical companies, regulators, health technology assessment (HTA) bodies, payers, patients, and physicians, clear descriptions of the benefits and risks of a treatment for a given medical condition should be made available in a timely fashion. The current paradigm is to generate evidence in a sequential manner where at each stage the focus

Received October 24, 2022; accepted January 17, 2023. doi:10.1002/cpt.2869
is on one stakeholder and the information they need in order to progress to the next stage. The consequence of this is not only a delay to all important evidence being available, but also that different parts of the evidence are generated under different experimental conditions considering, for example, experimental design (randomized controlled trial (RCT) vs. observational study, each conducted over a different calendar time), patients recruited, interventions received, or outcomes assessed. We are motivated to explore a more efficient framework where information for regulatory agencies and HTA bodies is generated concurrently, in a way which renders the overall evidence base more interpretable.

In this paper, we present Flexible Augmented Clinical Trial for Improved eVidence gEneration (FACTIVE), via a new class of augmented clinical trial designs. FACTIVE designs bridge conventional phase III RCTs to a broader population, or different experimental conditions, that can be tailored to address a particular HTA question or reflect a particular healthcare system. The proposed framework is different from existing approaches (as we explain later) and allows the generation of the right evidence at the right time, such that key decisions made after Marketing Authorization (MA) can be made sooner than would otherwise be the case.

**METHODS**

We assess the scope and timing of evidence on a new treatment made available to regulatory agencies or HTA bodies in the current clinical development process, and explore scenarios in which important questions remain unanswered even after justified extrapolation from the phase III RCT(s); specifically those relating to potential differences of treatment effects between the setting of the clinical trial and clinical practice. For cases where gaps remain, we consider collecting missing clinical practice elements for the new treatment through augmentation of the conventional clinical trial. We develop a new class of flexible clinical trial designs which can be tailored to meet specific regulatory and HTA requirements and allows to study the impact of factors for treatment effect modification in clinical practice.

**RESULTS**

**Current status**

**Figure 1** (upper panel) summarizes the current evidence generation and decision-making process: following confirmatory phase III trials, an application for MA of a new treatment is submitted to a regulatory agency. The MA is accompanied by a period of discussion and agreement with payers (e.g., HTA bodies, government agencies, and medical insurance companies), who will reimburse the cost of the treatment and influence the price at which the treatment should be marketed. The treatment is then placed on the market (❶ in **Figure 1**). Physicians, healthcare providers, and patients are subsequently informed how the new treatment is positioned in the landscape of already available treatment options (❷) and at some time later (❸) the maximum uptake of its use is achieved. Alongside this, post-authorization trials are conducted to learn more about how the treatment performs in normal clinical practice.

The sequential nature of generating the evidence for different stakeholders is immediately visible. The current main driver when designing confirmatory RCTs is to provide sufficient evidence to regulatory agencies of the efficacy and safety of a new treatment in order that it may be granted MA. Additional post-authorization trials provide further evidence of the treatment’s effectiveness in a broader patient population under clinical practice conditions. Evidence from such trials, in addition to that provided by the confirmatory trials, is then used to inform further market access and pricing discussions with HTA bodies, taking into account the therapeutic landscape. The post-authorization trials can be RCTs, or open-label extension phases to RCTs, but are

---

**Figure 1.** Current (upper panel) and proposed (lower panel) timing of cRW data element collection in phase III treatment development. ❶ First availability of new treatment to patients; ❷ Increasing uptake of new treatment use; ❸ Maximum uptake of new treatment use. cRW, close-to-real-world.
commonly undertaken as observational studies. The pre- and post-authorization experiments are conducted independently of each other, such that if estimates of a particular treatment effect differ between experiments, the reasons for that difference cannot be determined with certainty.

Multiple initiatives have attempted to streamline the evidence generation. Efforts in this direction must address the fact that different stakeholders have different questions to address, including the benefit–risk of an intervention within a specific target population vs. the cost-effectiveness of an intervention, including societal perspectives and a specific healthcare budget. Decision making by the European Medicines Agency (EMA) is centralized on behalf of the European Union (EU) whereas decisions made by one or more country-specific HTA bodies are national or local. Different stakeholders will also identify different sources of uncertainty and evidence gaps they want to see addressed, preferably during the clinical evidence generation phase, or at least post-authorization.

This paper focuses on evidence generation to meet the needs of regulatory agencies and HTA bodies, as they make the two initial and most critical public-sector decisions to determine patient access to medicinal products. The EU is taken as a jurisdiction for illustration, although the benefits of the approach described apply more broadly.

Discussions on streamlining evidence generation have often focused on the design and conduct of RCTs. All stakeholders recognize the high internal validity of this experimental design: the fact that reliable treatment effect estimates for well-defined research questions can be provided through a design where experimental conditions are controlled and well-understood. Indeed, deriving a reliable estimate and being able to interpret the magnitude of treatment effects in the context of experimental conditions that are documented and understood make RCTs the “gold-standard.” In addition, all stakeholders understand that there can be a scientific basis to generalize, or extrapolate, inferences from a clinical trial dataset to a broader patient population or clinical context, although the basis for extrapolation (e.g., other clinical trial data, pharmacological understanding of the mechanism of action, and pharmacological modeling), the extent of the extrapolation (to what proportion of the target population does the extrapolation apply) and the type and strength of evidence needed to support extrapolation is not documented and hence not unified for benefit–risk vs. cost-effectiveness decisions. Importantly, an RCT design in which the experimental conditions are controlled too tightly can leave all stakeholders questioning its external validity (i.e., the applicability of the trial results to the intended patient population and therapeutic use in clinical practice).

External validity of a trial is assessed in relation to its inclusion and exclusion criteria (vs. the population indicated for clinical practice) and its experimental conditions (vs. the therapeutic use expected in clinical practice), such as the outcome variable or comparator, permitted concomitant medications, or combination of treatments. The different mandates for regulators and health technology (HT) assessors can also lead to different clinical outcomes being prioritized for the assessment of efficacy or effectiveness with a consequence for the periods of treatment and follow-up that are of interest, and potentially different treatment effects of interest (i.e., estimands). Importantly, whereas a given clinical development program will be targeted toward a centralized regulatory approval, each HT assessment of the applicability of the trial results to their specific national or local jurisdiction might differ, for example, in relation to other products that are/are not reimbursed and used locally, or the precise target population for which cost-effectiveness can be justified. Concerns over the external validity of a particular RCT does not represent a fundamental flaw in that study design and conduct, only that the specific trial design cannot directly address the needs of a specific HTA body. Section 1 of the Supplementary Material gives additional discussion on external validity and extrapolation.

The result of these dynamics is that a regulator might authorize a product, perhaps with post-authorization evidence generation to address identified uncertainties, whereas an HT assessor might not feel fully informed about how the product will impact their specific healthcare system and budget, and whether a positive decision on cost-effectiveness can be justified. To address a broader set of stakeholder needs, RCTs might be made larger and/or longer and/or less well controlled in respect of patient population and experimental conditions. In addition, different end points or multiple comparators might be used. In reality, however, complementary sources of evidence generation are more efficiently used to provide answers to general and specific questions from HTA bodies. An often-overlooked fact is that the information required to strengthen the external validity can be generated concurrently with the trial data. Additional evidence is not necessarily generated to replicate trial results, rather additional data can explore the effects of treatment beyond the patient population and experimental conditions of the RCT. This paper discusses an experimental design that provides these additional data and seeks to deliver information to all stakeholders in a timely manner, preserving efficient evidence generation for each stakeholder and promoting a methodologically robust approach in an experimental design where different parts are no longer independent.

FLEXIBLE AUGMENTATION TO GATHER THE RIGHT EVIDENCE AT THE RIGHT TIME

We argue that the understanding of the relative effectiveness and time to peak uptake of a new treatment can be enhanced, without compromising safety, by generating additional rigorous evidence throughout the confirmatory development process. To do so, we propose FACTIVE, a new class of augmented RCT designs aimed at widening the evidence base of traditional RCTs. The lower panel of Figure 1 summarizes the potential impacts of using such an augmented RCT design (which we describe below): discussions with HTA bodies are better informed and shortened, along with a potentially greater maximum uptake of treatment use.

A key feature of the new paradigm is that augmentation is wrapped around a conventional RCT that is designed in the usual manner to focus on treatment efficacy and safety in a controlled experimental environment. The consequence is that the core RCT(s), which form the pivotal evidence for regulatory approval, is ring-fenced. A cross-disciplinary team can consider the specific objectives, subsequent design criteria, and the timing for augmentation in view of market value and patient heterogeneity as well...
as early evidence of efficacy and safety in the core RCT. The augmentation can resolve uncertainties that could not be achieved by simple improvements to the RCT. For example, providing insights into multiple, different combinations of active comparators are classic examples of HTA requirements that might dramatically increase the size, duration, and cost of a confirmatory RCT.

We set no limitations to the scope of research questions that can be addressed through augmentation. The questions to be answered by augmentation may be general: to provide estimates of treatment effects in the target population reflecting routine clinical care and under conditions reflecting clinical practice; to facilitate data integration with an existing external data source, by augmenting the RCT with subject eligibility criteria and conditions matched to the external resource; or, alternatively, targeted to obtain complementary information on a specific relaxation of an inclusion/exclusion criteria or different methods of outcome assessment. As an example of a specific question, consider a sponsor needing to address differences in national treatment guidelines regarding a background therapy. Instead of including patients on various background therapies, the core RCT could be conducted on one background therapy. Information as add-on to various background therapies (including the one in the core RCT) could be generated in the augmented part, whereas the patient population and experimental conditions remain otherwise similar, to establish that there is no impact of the background therapy or to characterize the impact that changing background therapy might have on the treatment effect that was observed in the RCT.

FACTIVE supplements the evidence provided by the core RCT for MA through an increased sample size with additional information from close-to-real-world (cRW) elements carefully selected according to safety, feasibility, and, critically, the outstanding questions to be answered. The descriptor “close-to-real-world” reflects an intent to broaden the experiment from the core RCT, whereas acknowledging that the experiment remains a randomized trial with some elements of patient selection, a protocol defining the experimental conditions, and patients being aware that they are participating in a clinical trial. There are established mechanisms for sponsors to interact with regulators (e.g., Scientific Advice procedures in the EU) to understand preferences and standards for a future application. Early dialogue with HTA bodies can provide valuable input about the context in which a treatment might be assessed, and important considerations that are not addressed in the core RCT. For example, Population, Intervention, Comparator(s), Outcomes (PICO) provides a framework to compare the evidence being generated to the question of interest for the HTA and, whereas the PICO can change over time, it can be used to explore whether limitations to the core RCT evidence should be anticipated due to, for example, missing subpopulations, differences in treatment algorithms, or preferences for other comparators. Identifying potential evidence gaps can then inform the purpose, and consequently the design, of the augmentation. Again, the augmentation is not designed to serve as confirmation of the RCT evidence but represents a basis to provide data, or to bridge the RCT data, to the evidence requirements of other stakeholders.

THE FACTIVE DESIGN

Figure 2 visualizes the FACTIVE design intended to generate evidence rigorously and contemporaneously with high-quality information obtained through randomization. Two types of patients are identified in FACTIVE: those eligible for the core RCT (green) and those who are from a broader population (blue). In addition, two experimental settings are identified, the RCT conditions and the alternate cRW experimental settings (e.g., those closer to clinical practice). Both types of patients are first randomized to be studied under RCT treatment conditions or under cRW treatment conditions. The RCT-eligible patients who are studied under RCT conditions form part A of the design, the core RCT used for regulatory submission. Part B is comprised of additional

---

**Figure 2** FACTIVE. Patients screened for the augmented phase III trial are recruited from the target population expected to be treated post-authorization. These are either RCT-eligible (green) or from a broader patient population (blue), excluding those patients with, for example, a safety risk. RCT-eligible patients are randomized either to part A (the core RCT) or to part B, where they are treated under cRW conditions. In both parts, they are additionally randomized to experimental treatment or control. Patients from the broader population are assigned to part B only and are randomized to be treated under cRW conditions or under RCT conditions. In both situations, they are then randomized to experimental treatment or control. In part B, the proportions of green and blue patients treated under cRW conditions need to be agreed beforehand (e.g., to match epidemiology). RCT/cRW conditions define RCT/cRW design elements, such as visit schedule, administration of treatment, monitoring, but exclude specifics about the patient population. The time to initiate part B could be made dependent on accumulating evidence from part A at an interim analysis (IA) as illustrated by the vertical offset and the arrow in yellow between parts A and B, whereas ensuring this does not compromise the integrity of the core RCT. cRW, close-to-real-world; FACTIVE, Flexible Augmented Clinical Trial for improved eVidence eGeneration; RCT, randomized controlled trial; RW, real-world.
RCT-eligible patients (green) and those from the broader population (blue) randomized to cRW treatment conditions and patients from the broader population randomized to RCT treatment conditions. Within each part of the design, patients are randomized to either the experimental treatment or control. This allows all conventional RCT analyses for authorization purposes to be conducted with the evidence generated in part A, supplemented by other analyses addressing specific questions in part B.

The nested structure facilitates rigorous statistical analyses for causal effects of interest; see Section 2 of the Supplementary Material. The augmented design makes it possible to estimate and compare treatment effects, and effect changes, across the four combinations of subject eligibility (RCT eligible patients vs. broader population) and treatment conditions (RCT vs. cRW treatment conditions).

The augmented design is complementary to existing trial designs which look to combine RCT and cRW elements, such as seamless phase III/IV designs and clinical trials using external control information. The distinguishing feature of FACTIVE is the collection of data on randomized cRW elements under the new treatment, before regulatory approval, and concurrently with corresponding RCT data, thus enhancing the available evidence. Note that the left-hand side of part B in Figure 2 could be implemented as a pragmatic trial. It fulfills the pragmatic study criteria for randomized studies under clinical practice treatment conditions of patients expected in routine clinical care, and of assessments meaningful to patients/physicians. However, FACTIVE is broader than conventional pragmatic trial designs: whereas concurrency reduces the sources of time-related bias, the core RCT in part A together with this left-hand pragmatic part does not allow a rigorous comparison of cRW and RCT treatment effects beyond the RCT-eligible patients. The starting point for FACTIVE follows the design of the core RCT, which is subsequently assessed for cRW augmentation. The additional part on the right-hand side, where patients from the broader population are treated under RCT treatment conditions, then provides a more comprehensive understanding of treatment effects.

There can be flexibility also in the design of part B. For example, if few inclusion/exclusion criteria are used in the RCT, the concept of a broader population might be redundant. Likewise, if the experimental conditions used in the RCT are close to clinical practice, it might not be necessary to examine the treatment under alternative conditions. It is not even necessary for the same control arm to be used in both parts of the experiment, although then some of the benefits of the design are lost. If different controls are used in part A and part B, network meta-analysis is one method that can be used to bridge from one control to the other using relevant external data. To justify the proportion of patients from the RCT-eligible and the broader patient population along with the respective sample sizes, whereas one approach might be to adequately power a comparison of treatment vs. control in cRW conditions, other approaches are conceivable. Returning to the example above of investigating the effects of an experimental treatment on different background treatments, the amount of information to be generated might be thought of as similar to generating evidence across subgroups to enable an assessment of consistency. As stated above, the objective is not to replicate findings from the core RCT. In particular, when creating evidence to bridge from the RCT, or to give confidence in the external validity of the RCT results, precision of estimates might be more important and a better basis for planning than tests of statistical significance. Note that removing the cRW components from the augmented design simply returns the original RCT.

DISCUSSION

The development of new treatments is a continuous learning process where evidence collected in early phases contributes to decisions made in later phases and where techniques commonly used in exploratory development can continue to be used on confirmatory data (e.g., clinical pharmacology modeling). Even in confirmatory phase III trials, modifications can be made as knowledge grows during their execution as, for example, in an adaptive trial with dose selection or relaxation of an inclusion criterion after preliminary safety data have been reviewed. Augmenting confirmatory RCTs (part A in Figure 2) with tailored cRW elements (part B in Figure 2) can be considered a natural extension of this process. Although FACTIVE will not be a suitable approach to evidence generation in every program, we think all programs can benefit from a discussion on the merits of structured and concurrent evidence generation beyond the core RCT. The key is not to implement a fixed design or to follow a checklist, but to carefully consider uncertainties or evidence gaps that are critical to address through evidence generation beyond a well-designed confirmatory RCT.

In contrast to the current practice, FACTIVE offers more timely evidence generation and an increased potential to investigate and quantify different modifiers of the treatment effect. Under the current approach, an estimated treatment effect might differ between pre- and post-authorization experiments for reasons that are often not fully understood. This might be due to changes in care over time, changes in patients recruited, perhaps due to lack of or different choice of control arm or changes in experimental conditions, including methods or timings of assessments applied, adherence to treatment or use of concomitant treatments, the choice of investigational sites, etc. Alternatively, estimands may differ (intentionally or unintentionally), such that the experiments address different clinical questions of interest. Compared with uncontrolled observational studies, however, the collection of randomized data based on cRW elements concurrently with the core RCT data allows for a unified statistical analysis with nested models, without the biases inherent when comparing or integrating data from different experiments (see Supplementary Material). If cRW data are collected from observational studies after the RCT, then assumptions on the impact of potential confounders (untestable with the data to hand) are needed for a joint statistical analysis to proceed.

The time to initiate part B would depend on having clarity on the research questions to be addressed and perhaps on accumulating evidence from part A, for example, into the safety of exposing a broader patient population or relaxing the experimental conditions. However, an obvious disadvantage of staggering the start of part B would be the reduction or elimination of the overlap in time between the two parts. Only with the two parts of the experiment
conducted concurrently, enabling randomization of patients between the different parts of the experiment and between experimental conditions, can the full strength of the design and insights into effect modification be leveraged.

Of course, there may be barriers to the adoption of FACTIVE. There are additional up-front costs in planning, designing, and executing this design, which should be weighed against the potential benefits on a case-by-case basis. Interest might be less where, for example, the probability of success for part A is not high, or where a key market has less interest in the outcome of part B but its implementation risks slowing the conduct of the core RCT. The impact of including part B on the time to complete part A is a relevant consideration, but is not easy to predict. Although there is some competition for patients between the parts of the experiment, trial sites might be more motivated to participate because, for example, of the overall lower screening failure rate. We acknowledge our ideas have the potential to be disruptive and may not be taken up easily by all stakeholders. However, in order to break the current linear thinking, changes in mindset are needed, which will take concerted efforts by all stakeholders. The most important change is the move, during confirmatory trial design, toward an active consideration of the optimal time for evidence generation taking a holistic perspective. The optimal timeline will not always be the conventional sequential one (Figure 1, upper panel). A paradigm shift is needed toward understanding that complementary evidence can, and where possible, should be generated in parallel but without the intention to change the framework for regulatory or HT assessment. FACTIVE aligns with initiatives, such as the EMA/HTA parallel Scientific Advice procedure that have promoted early consideration of the different stakeholder needs and post-authorization evidence generation. Maximizing the benefits of this requires full engagement of drug developers bringing all relevant disciplines into discussions from an early stage.

Having the RCT and cRW data available simultaneously to both regulatory agencies and HTA bodies is a new paradigm. How much weight each party should give to the two types of evidence will be context dependent and specific to the designs and results of the different experiments in the context of the totality of evidence generated. It is important that stakeholders are well versed in critically appraising the strengths and weaknesses of different experimental approaches. Optimal stakeholder decision making cannot be based on the rhetoric that RCTs have inadequate external validity and real-world experiments have inadequate internal validity. Each experiment should be judged on its own merits with an understanding that it is possible to generalize results from an RCT if potential effect modification is understood and that it is possible to interpret evidence of benefit from “real-world” experiments when well-conducted and reported. Randomization is a strength in the cRW experiment even if the experimental conditions are less well controlled.

As with any design, there is the potential for misuse. The potential to generate data in a broader patient population under cRW conditions should not be used as a reason for the sponsor to tighten the conditions and lessen the external validity of a core RCT. On the other hand, the potential to generate additional evidence in a timely manner should not mean that the demands of regulators and HTAs increase. We reflect further on the EMA/HTA parallel Scientific Advice procedure, whereby despite a tri-partite conversation that ranges wider than only a regulator’s or an HTA’s individual needs, each stakeholder gives advice to drug developers in a way that reflects and is confined to their own respective legislative basis and mandate. The parallel to FACTIVE is the design of fit-for-purpose RCTs with a vehicle to generate evidence on additional research questions so that stakeholders can make timely decisions, but without altering the established decision-making frameworks. Having clear objectives for the concurrent evidence generation will help to clarify the relevance of the evidence to the different stakeholders.

In conclusion, through FACTIVE we propose a class of augmented RCT designs involving the concurrent collection of RCT and cRW data based on cRW elements in a way that is optimally and sequentially organized by generating the right evidence at the right time. This approach retains the distinctiveness of the core RCT and the questions it seeks to answer from additional research questions aimed at investigating the performance of the treatment in a broader population under cRW conditions or to investigate one or more critical, specific limitations to the RCT design.

Through randomization, FACTIVE ensures that the data on cRW elements are of similar high quality to conventional RCT data and available in closer proximity to the time of MA and able to inform initial HTA discussions, accelerating the journey of the novel treatment from discovery to patient. Discussions with regulators and HTAs can take place either simultaneously or in rapid sequence after completion of the core RCT with the aim that treatments can be made available to patients earlier. In addition, information available to physicians, healthcare providers, and patients will be enhanced by the evidence provided by the augmented design. This could lead to a greater awareness of the benefits of the treatment, leading to its greater use in the community. Post-authorization trials would still be required to collect data on post-authorization aspects of the use of the new treatment, but these are likely to be fewer and smaller in number given the high quality cRW evidence already available.

SUPPORTING INFORMATION
Supplementary information accompanies this paper on the Clinical Pharmacology & Therapeutics website (www.cpt-journal.com).

CONFLICT OF INTEREST
The authors declare no competing interests for this work.

AUTHOR CONTRIBUTIONS
All authors wrote the manuscript, designed the research, performed the research, analyzed the data, and contributed new analytical tools.

© 2023 The Authors. Clinical Pharmacology & Therapeutics published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
1. Eichler, H.G. et al. “Threshold-crossing”: a useful way to establish the counterfactual in clinical trials? Clin. Pharmacol. Therap. 100, 699–712 (2016).
2. Califf, R.M. et al. Transforming evidence generation to support health and health care decisions. N. Engl. J. Med. 375, 2395–2400 (2016).
3. Ray, R., Locke, T., Hendricks-Sturrup, R. Aligning shared evidentiary needs among payers and regulators for a real-world data ecosystem <https://healthpolicy.duke.edu/> (2022).
4. ICH. International Council for Harmonisation Topic E9(R1): Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials <https://ich.org/> (2019).
5. Remiro-Azócar, A. Target estimands for population-adjusted indirect comparisons (with discussion). Stat. Med. (in press) 41, 5558–5569 (2022).
6. Eichler, H.G., Bloechl-Daum, B., Abadie, E., Barnett, D., König, F. & Pearson, S. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. Nat. Rev. Drug Discov. 9, 277–291 (2010).
7. Schmidli, H. Beyond randomized clinical trials: use of external controls. Clin. Pharmacol. Therap. 107, 806–816 (2020).
8. Zuidgeest, M.G.P. et al. Pragmatic trials and real world evidence. J. Clin. Epidemiol. 88, 7–13 (2017).
9. Dias, S., Ades, A.E., Welton, N.J., Jansen, J.P. & Sutton, A.J. Network Meta Analysis for Decision Making (John Wiley & Sons, Chichester, UK, 2018).
10. CHMP. Guideline on the investigation of subgroups in confirmatory clinical trials <www.ema.europa.eu> (2019).
11. Sheiner, L.B. Learning versus confirming in clinical drug development. Clin. Pharmacol. Therap. 61, 275–291 (1997).
12. CHMP. Reflection Paper on ‘Methodological issues in confirmatory clinical trials planned with an adaptive design’ <www.ema.europa.eu> (2007).
13. FDA. Adaptive designs for clinical trials of drugs and biologics <https://www.fda.gov/> (2022).