

Are Novel, Nonrandomized Analytic Methods Fit for Decision Making? The Need for Prospective, Controlled, and Transparent Validation

Hans-Georg Eichler^{1,2,*}, Franz Koenig², Peter Arlett¹, Harald Enzmann^{3,4}, Anthony Humphreys¹, Frank Pétavy¹, Brigitte Schwarzer-Daum^{2,5}, Bruno Sepodes^{4,5,6}, Spiros Vamvakas¹ and Guido Rasi^{1,7}

Real-world data and patient-level data from completed randomized controlled trials are becoming available for secondary analysis on an unprecedented scale. A range of novel methodologies and study designs have been proposed for their analysis or combination. However, to make novel analytical methods acceptable for regulators and other decision makers will require their testing and validation in broadly the same way one would evaluate a new drug: prospectively, well-controlled, and according to a pre-agreed plan. From a European regulators' perspective, the established methods qualification advice procedure with active participation of patient groups and other decision makers is an efficient and transparent platform for the development and validation of novel study designs.

The opportunities for learning fast about drugs' benefits and harms have never been greater. The past decade has seen impressive changes in the generation and availability of health-related data. The majority of patient-provider interactions in developed healthcare environments are now recorded electronically and electronic health records (EHRs) have been made available for secondary use to answer research questions. On another frontier, patient-level data from completed randomized controlled trials (RCTs) are now being shared on an unprecedented and growing scale. This enables cross-trial analyses and, although more challenging, combining RCT data with different types of "real world data" (RWD), including EHRs and insurance claims. Last, new data sources, including the medical internet of things, wearables, social platforms, or smart phone apps, might possibly be mined in the future for healthcare-relevant information.

A number of obstacles will have to be overcome before the full potential of these data sources can be brought to bear on pharmaceutical research and care. The obstacles have been broadly grouped in two domains¹: (i) technical/operational readiness, which relates to factors like extent of EHR coverage, use of structured data, interoperability of databases, and data quality; and (ii) data governance readiness, which addresses legal issues impeding secondary data analysis, including data privacy concerns, level of consent required, and clarity on who has legal access to health data for research purposes. A recent Organisation for Economic Co-operation and Development (OECD) report highlighted that all OECD countries still face challenges in both domains.¹ Applicability of RWD across healthcare systems remains an additional issue.

Yet, we are optimistic. As the ecosystem for e-health develops, so will data quantity. Issues of data quality, including missing data, and differences in terminologies and data formats will be more challenging to resolve.² However, the need for quality assurance and control procedures has been recognized and a range of initiatives are aiming to bring RWD quality to a level of regulatory acceptability. Collaborations among stakeholders and opportunities for data processing and quality improvement are constantly growing.^{3,4} Progress will likely happen in fits and starts but we foresee a future where healthcare data from different sources and of sufficient quantity and, eventually, quality will be available for rapid secondary analysis by researchers. Some secondary use of RWD is well established⁵ (e.g., drug utilization, disease epidemiology, or safety evaluation).⁶ However, if fully exploited, RWD could also contribute to, for example, demonstrating efficacy and treatment stratification to inform regulatory, reimbursement, and personalized treatment⁷ decisions. The broad range of research questions that might be addressed with the help of new data sources have been described elsewhere.^{4,8}

METHODOLOGY AVERSION

Alas, *data*, even of good quality, do not necessarily translate into credible *evidence* in the absence of adequate (statistical) methods to extract, analyze, and interpret them.^{2,9} Addressing this obvious bottleneck, a range of relatively novel methodologies have been proposed or refined over the past decade to enable the analysis of RWD or to combine RWD with RCT data. Unsurprisingly, proponents of these methodologies argue that they can address

¹European Medicines Agency (EMA), Amsterdam, The Netherlands; ²Medical University of Vienna, Vienna, Austria; ³Federal Institute for Drugs and Medical Devices (BfArM), Bonn, Germany; ⁴EMA's Committee for Medicinal Products for Human Use (CHMP), Amsterdam, The Netherlands; ⁵EMA's Committee for Orphan Medicinal Products (COMP), Amsterdam, The Netherlands; ⁶Universidade de Lisboa, Lisbon, Portugal; ⁷University Tor Vergata, Rome, Italy. *Correspondence: Hans-Georg Eichler (Hans-Georg.Eichler@ema.europa.eu)

Received July 16, 2019; accepted September 2, 2019. doi:10.1002/cpt.1638

potential biases and deliver robust evidence. However, many commentators remain unpersuaded and argue that acceptance of non-RCT methodologies is tantamount to lowering the quality of evidence because these methods are prone to a myriad of undetected or undetectable biases. The pros and cons of non-RCT methodologies have been aired extensively and will not be repeated here.

The RCT will, in our view, remain the best available standard and be required in many circumstances, but will need to be complemented by other methodologies to address research questions where a traditional RCT may be unfeasible or unethical.

It is self-evident that uncritical adoption of novel methodologies may lead to false conclusions, poor healthcare decisions, and, ultimately, patient harm. However, the opposite—that is, *not* to use novel, robust methodologies—has equally detrimental consequences: Bauer and König¹⁰ coined the term “methodology aversion in drug regulation” referring to a purported unwillingness of regulators (and presumably other decision makers) to adopt novel statistical or other methods of data analysis. Some part of the unwillingness may stem from a fear that, “without in-depth knowledge, [...] toolboxes may quickly turn into black boxes.” In turn, this fear may be precipitated by a lack of familiarity with new methodologies, partly due to conservatism, partly to lack of resources.¹⁰

We concur that unfounded methodology aversion is a potential roadblock to making the best use of new data sources. Regulators have been accused of methodology aversion (although they are often simultaneously accused of recklessly abandoning the “gold standard” RCT) but we see various degrees of methodology aversion in all stakeholder groups within the pharmaceutical ecosystem.

METHODOLOGY DEVELOPMENT BY DESIGN

How can we overcome methodology aversion without the risk of adopting unreliable study designs? We believe the appropriate course of action is to take a break from the often heated exchanges and start to evaluate and validate novel statistical, epidemiological, or other methodologies² in broadly the same way one would evaluate a new drug: prospectively, well controlled, and according to a pre-agreed plan.

Table 1 lists a number of non-RCT methodologies that have been proposed in the context of the assessment of drugs. The list is neither exhaustive nor definitive but has been compiled on the basis of commonalities that these methodologies share:

- Their concept and theoretical underpinnings have been developed in detail.
- The potential gains from their use for drug development and assessment could be considerable in terms of resource and time savings and in getting more relevant information on the effectiveness and safety of medicines to patients faster.
- Lack of confidence in these non- (or not fully) randomized designs is limiting their impact.
- Sharing of patient-level RCT data and RWD will be key enablers of these methodologies.

- We are not starting from scratch: at least some validation efforts have already been made, if only by individual parties (e.g., individual companies or academic groups), or the method is already in use but has not been thoroughly tested and accepted.
- It is perfectly feasible to validate these methods in a prospective controlled way, often by “bolting-on” the methods validation exercise to a standard drug development plan or RCT.
- Prospective evaluation will not require additional *de novo* data generation, hence, can be relatively low cost.

We emphasize the last two bullet points: there are no unsurmountable obstacles to designing a prospective evaluation plan for any of these methodologies in the context of and in parallel to routine premarketing and postmarketing drug development programs. Compared with the cost and timelines of *de novo* data generation in an interventional trial, the resource requirements appear not prohibitive.

How could these methodologies be prospectively tested and validated in practice? Consider, for example, the first two topics in **Table 1**. Prospective evaluation of these methodologies can be built into the planning of conventional two-arm RCTs. Provided there are relevant RWD or patient-level data from previous RCTs available that conform to the RCT selection criteria and to the control treatment (placebo, no treatment, a defined standard treatment, or best supportive care), an add-on analysis can be planned to compare the level of agreement between the standard intergroup comparison of the randomized groups with the comparison between the experimental group from the RCT and the control group with borrowed data (or the virtual external control group). The prospective analysis plan needs to ensure that the conduct and interpretation of the RCT itself is not compromised, but at the same time needs to ensure credibility of the methods-evaluation exercise (i.e., the exercise should be free from *post hoc* bias; it must not “paint the bull’s eye around the arrow”).

Arguably the most controversial among the methods listed in **Table 1** is the concept of replacing RCTs by RWD analysis. It is tempting to compare the effectiveness of two drug treatments that have been on the market for some time by way of retrospective RWD analysis. The strengths and limitations and the types of research questions that could potentially be addressed by RWD instead of an RCT have been reviewed elsewhere.¹¹ Compared with running a head-to-head RCT, RWD analysis could obviously lead to cost and time savings. However, the track record of nonrandomized comparative studies is not convincing. In some high-profile cases, findings from subsequent RCTs differed not only in effect size but even in direction, resulting in qualitatively different causal conclusions.¹²

Can we improve the track record? Proponents of RWD analysis would argue that we have learned from past mistakes and the field has strengthened in the past decades. Indeed, there are some encouraging examples where RWD studies correctly predicted RCT results *before* the RCT results became available¹² (but we do not know how many unpublished RWD studies did not correctly predict RCT results). The prospective nature of the exercise needs re-emphasizing because analyzing RWD after the RCT is always

Table 1 Examples of (novel) methodologies, in no particular order, for analysis of different types of data that would benefit from prospectively designed validation

Methodology	Potential benefit for drug developers and decision makers	Current limitations	How to validate prospectively
Borrowing of data ²¹⁻²³	“Borrowing” cases from past studies for the control arm of a current RCT could increase the efficiency of decision making with the current study. This may translate into smaller sample size for the current trial and/or unequal randomization.	Relies on the assumption of similarity of historical information to the current control data; may result in bias if assumption is not satisfied. Several methods for historical borrowing have been proposed.	Use conventional RCTs to concurrently analyze results as per usual and with borrowed data according to preplanned protocol and data sources. Compare and assess various methods of borrowing.
Use of external control group, threshold crossing ^{24,25}	May enable causal inferences about drug effects on the basis of external (historical) control groups for products and indications where RCTs are not feasible.	Comparisons with external controls are based on assumptions that often cannot be verified, which may lead to biased conclusions about drug effects. ²⁶ External control groups tend to have worse outcomes than a similar control group in an RCT. ^{27,28}	Use conventional RCTs to concurrently analyze results as single-arm trials with historical comparators; compare results from the randomized and nonrandomized analyses based on pre-agreed plan. ²⁵
Indirect comparisons for relative efficacy ^{29,30}	Allows for estimation of relative efficacy of two (or more) treatments in the absence of any head-to-head RCTs (= direct comparisons) Frequently used by HTA bodies for REA, because many (new) drugs have insufficient RCT information for direct comparisons.	Although indirect comparisons usually rely on randomized data, the treatments of interest have not been randomized against each other (head-to-head), only to a common comparator. A variety of methods exist to mitigate this, but each method rests on a number of assumptions about the data used. Methods are still evolving and sometimes generate discrepant results.	Use the opportunity afforded by the planning of a head-to-head RCT where previous RCTs of the drugs of interest against a common comparator (e.g., placebo) are available to develop a prospective analysis plan for indirect and MTC. The aim is to compare different methods for indirect or MTC, explain discrepancies in results from different methods, and cross-validate methods against each other and against the head-to-head RCT.
Replacing RCT by RWD analysis ¹²	Conceptually, RCTs could, in some situations, be replaced by comparative analyses of RWD. Replacing even a small proportion of postmarketing RCTs with nonrandomized RWD analyses would in many cases translate into faster availability of relevant information using substantially fewer resources.	Major concerns about comparative RWD analyses include lack of ability to tightly control measurements of patient characteristics and health outcomes and susceptibility to bias. A general lack of confidence in nonrandomized RWD analyses has limited their impact.	Prospectively design new RWD studies to match the design of planned RCTs. This is feasible when both drugs have been in routine use for a sufficient time. The concurrent approach avoids bias by matching the RCTs and RWD analyses as closely as possible (e.g., for patient characteristics, dose regimens), while avoiding the temptation to trim RWD analysis to the RCT results once they become available. It also allows for sensitivity analyses to identify whether alternative designs or analyses could have improved agreement between the designs.
Reweighting of RCT results to reflect real life ^{31,32}	Using RWD (e.g., from disease registries), to “reweight” RCT results may improve external validity and generalizability of RCT results.	A demonstration project has shown the feasibility but the concept has not been prospectively validated.	Use results of conventional RCTs of novel drugs to obtain reweighted results and compare with measured outcomes once enough RWD has accumulated; according to preplanned protocol and data sources.
Extrapolation of knowledge to an unstudied population ^{33,34}	In some populations (e.g., neonates or young children), the conduct of clinical trials is fraught with operational or ethical challenges leading to an absence of information on drug effects. “Implicit extrapolation,” although subjective, is often the only basis for treatment or dosing decisions in these populations. A systematic framework for “explicit extrapolation” of relevant information from a source population (e.g., adults), to a target population (e.g., small children) preferably based on quantitative methodology has the potential to improve treatment decisions.	Although some of the methods proposed for the extrapolation exercise are not novel, experience with their use in extrapolation exercises is limited. Few, if any, systematic extrapolation exercises have undergone prospective validation.	As clinical experience grows during the postmarketing phase, the assumptions and predictions made on the basis of extrapolations can be checked against prospectively planned collection of RWD. Apply the concept of extrapolation also in areas where RCTs are possible (e.g., extension of indications in adults where further RCTs are conducted and compare whether the extrapolation concept; requiring different/less data) would have resulted in similar results. Assess various concepts of extrapolation simultaneously. Might require that some additional data are collected in the current RCT (such as PK/PD).

(Continued)

Table 1 (Continued)

Methodology	Potential benefit for drug developers and decision makers	Current limitations	How to validate prospectively
Predictive approaches to heterogeneous treatment effects ^{35–37}	(Positive) RCTs can only help predict that at least some patients similar to those enrolled in the trial will likely benefit from the intervention (“reference class forecasting”). However, determining the best treatment for an individual patient is different from determining the best average treatment, because of heterogeneity of treatment effects. Improved prediction of outcome risk and understanding of heterogeneity of treatment effect could be key enablers of personalized treatment decisions and more successful treatment outcomes.	Conventional subgroup analyses, aiming to describe effect modifiers, often fall short because each patient belongs to multiple different subgroups, each of which may yield different inferences. More elaborate, regression-based approaches have been proposed to address heterogeneity of treatment effect, including risk modeling and treatment effect modeling. However, experience with these methods is limited, especially with externally derived models. There have been few, if any, attempts to systematically evaluate their usefulness in clinical practice.	Develop models concurrently with the design of an RCT. Where possible, incorporate assessment of the use of RWD for predictive analysis of heterogeneity of treatment effect. The ultimate test of a predictive approach is to compare decisions or outcomes in settings that use such predictions with usual care in a prospectively planned experiment. ³⁶

HTA, health technology assessment; MTC, mixed treatment comparison; PK/PD, pharmacokinetic/pharmacodynamic; RCT, randomized controlled trial; REA, relative effectiveness assessment; RWD, real-world data.

fraught with the risk that the design and analysis will be tuned to the known RCT findings. Only the prospective, structured evaluation of RWD studies to match the results of RCT in different clinical settings can avoid potential for *post hoc* adjustments. Projects are currently underway to that intent.¹³ The US Food and Drug Administration (FDA) has developed and is now funding targeted demonstration projects.¹⁴ We are aware that interest in conducting head-to-head RCTs in the postmarketing phase is often limited, but what is needed is at least several RCTs comparing two (or more) treatments that have been on the market for a sufficient period of time to enable simultaneous RWD analysis. This would provide an opportunity to “bolt-on” a methods evaluation exercise by simultaneously developing the RCT protocol and the parameters of the RWD analysis, which is to be conducted concurrently with the RCT but before RCT results become available.

Along similar lines, the conduct of a postmarketing head-to-head RCT¹⁵ can be used to prospectively plan for the comparison of different methods for indirect or mixed treatment comparisons, provided that previous common-comparator RCTs (e.g., against placebo) are available for the drugs to be compared. The exercise could help to explain discrepancies in results from different methods and to cross-validate methods against each other. As above, the benefit of running the exercise during the planning stages of the head-to-head RCT is to avoid that design and analysis choices are tuned to match the known RCT results.

For some methodologies listed in **Table 1**, the validation exercise will have to be done during the postmarketing phase. For example, predictions of efficacy, optimal dosage, or safety based on extrapolation from a source population (e.g., adolescents) to a target population (e.g., younger children or infants) made at the time of marketing authorization can only be assessed at a much later stage (i.e., once sufficient clinical experience in the target population has accumulated). Yet, the validation plan should be formulated proactively and agreed upon at the time when the extrapolation is performed.

The overall goal of the parallel validation exercises described above and in **Table 1** is to gain practical experience with novel analytical methodologies. The technical goals are to:

- See where they can or cannot be used
- Understand why some studies fail while others succeed
- Avoid design or analytic flaws that have plagued much of non-randomized research
- Allow for sensitivity analyses to explore whether alternative designs could increase the level of agreement between the novel and conventional methodologies
- Define if any characteristics can predict with high certainty the validity of a nonconventional study or analysis

As experience grows, the scientific field will mature, decision-makers will have robustness checks in place, and confidence in the reliability of results will grow—if justified.

How much work needs to be done before a given new analytic method can be declared fit-for-decision making? We would caution against the expectation of a simple answer. The potential impact of relying on inappropriate methodology is context dependent. The impact is comparatively small when new methods are merely used to support additional evidence, because regulatory and other decision makers are used to looking at the totality of evidence, based on information coming from different types of data and methods. Examples include scenarios where positive benefit-risk has been established by conventional means on the basis of pivotal RCTs and their pre-defined primary end point(s) and novel methodologies are merely used for analysis of a secondary end point, or for extension of the label to a (biologically similar) indication of a late-comer in a given therapeutic class. It would seem prudent for decision makers to first accept novel analytic methods for such lower risk situations and then gradually expand acceptability as confidence in the method grows.

Returning, for example, to the controversial question of whether RWD can ever replace RCTs (**Table 1** and above), we believe the

answer at this point in time should be neither a categorical no or yes but an open-minded, prospective exploration to identify scenarios where RWD analysis can provide sufficiently robust, decision-relevant supportive evidence.

Most of the methods listed in **Table 1** rely (partly) on the use of RWD. Healthcare systems and healthcare environments are different from one country to another, and it is often argued that RWD cannot easily be extrapolated across regions. We concur but emphasize the difference between data sources and analytic methods, where there are bigger opportunities for improvement. Although the results of a given RWD analysis from, for example, the United States may not be relevant for a healthcare environment in the European Union (or vice versa), the learnings from methods development on how to, say, address a given type of bias, are expected to have global relevance.

THE EUROPEAN MEDICINES AGENCY REGULATORY SCIENCE STRATEGY AND METHODS QUALIFICATION PROCEDURE

Recognizing the fast pace of innovation and its own role in catalyzing and enabling regulatory science and innovation, the European Medicines Agency (EMA) has recently published its newly developed regulatory science strategy (RSS).¹⁶

One of the key goals identified in the RSS is “Driving collaborative evidence generation—improving the scientific quality of evaluations”; the public health aims of this drive are to “provide regulators and [Health Technology Assessment bodies] HTAs/payers with better evidence to underpin regulatory assessment and decision-making” and “advancing patient centred access to medicines.” The strategy lists a number of core recommendations and proposed underlying actions. These include fostering innovation in clinical trials (with a focus on novel trial designs, statistical concepts, end points, or techniques for gathering data) and developing methodology to incorporate clinical care data sources in regulatory decision making. The study designs and methodologies listed in **Table 1** are representative (but not exhaustive) of the novel methodologies that could be explored in the context of the RSS.

The EMA has been offering for some time a scientific advice procedure “to support the qualification of innovative development methods for a specific intended use in the context of research and development into pharmaceuticals.”¹⁷ The process foresees an initial consultation and advice phase with repeated interactions between innovators and regulators to define what studies or other activities will be required to qualify a new methodology “fit for purpose.” At that stage, for methodologies which seem promising, the EMA publishes a letter of support after agreement with the sponsor.¹⁷ The advice phase is followed by a formal opinion on the acceptability of the specific use of a method. Before final adoption of a qualification opinion, after agreement with the sponsor, the evaluations are made available for public consultation by the scientific community; for specific examples, please refer to ref.¹⁸ and ref.¹⁹ This is to ensure that all relevant information is open to scientific scrutiny and discussion.

Several methodology advice procedures have been successfully concluded or have started in collaboration with EU HTA bodies and patient groups. The hoped-for result of such collaborative

efforts is the widest possible acceptance of useful methodologies by key healthcare decision makers, beyond regulators only.

We believe that the qualification advice procedure with active participation of HTA bodies, healthcare payers, and patient groups is an efficient, transparent, and inclusive platform for the development and validation of novel study designs, such as those summarized in **Table 1**. We invite researchers from academia, industry, and public-private consortia to avail themselves of this opportunity to open up their methodology developments to external scrutiny and, in the process, familiarize decision makers with their concepts—which is the best way to enhance their acceptability, if justified.

CONCLUSION

Our goal has been to draw attention to a potential roadblock in the use of new data sources: methodology development, and its flip side, methodology aversion. It will not be sufficient for researchers to elaborate novel study designs and analyses. A necessary and self-evident second step is the testing and validation of any such approaches with a view to avoiding the “black-box trap,” that is to jeopardize the acceptance of even a useful method because it cannot be understood by external stakeholders who were not involved at any stages of the development.⁵ To overcome methodology aversion, the developers of new methods should also prospectively address “data aversion” and new methods should be tested both retrospectively utilizing the open-access to RCT and RWD and prospectively as discussed.

Transparent, collaborative platforms, such as the EMA’s methods qualification procedure or similar procedures offered by the FDA and other public bodies, are likely the best available avenues to achieve the goal.

The ultimate key to achieving credibility is to start with an open but “agnostic” mind-set and submit novel methods to a fair, transparent, and *prospective* validation exercise; this cannot be done only by dry runs with old products. It is understandable that drug developers are wary of jeopardizing the development programs for their valued new assets. However, we emphasize that if developers want trial assessors to accept novel methods, they will have to expose some of their experimental drugs to methodology development exercises. This would need to happen with a clear upfront agreement on a “firewall” between the methods-evaluation and the product-evaluation, with assurances that the methods-evaluation will neither jeopardize nor rescue a product. We are confident that, with proper planning, optimal drug development can be combined with optimal methodology development.

We are also aware that methodology developments are labor-intensive activities that will require the collaboration of methodologists, drug developers, patient representatives, data custodians, and prospective trial assessors (regulators, HTA bodies, and payers), as well as adequate funding streams. We hope that funding bodies, such as the EU Innovative Medicines Initiative,²⁰ will support development of methodologies that require public-private partnerships and represent a paradigm shift that would benefit a range of therapeutic areas or products. Funds will need to be dedicated over the prolonged periods of time needed to see methods development through all stages from exploration to validation and acceptance.

To revisit the analogy with drug development, these are long-term projects that will not come to fruition as a result of short-term efforts by individual players.

The stakes are high—overcoming methodology aversion and ensuring that all stakeholders arrive at a nuanced view between categorical rejection and naïve adoption of novel methods.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

As an Associate Editor for *Clinical Pharmacology & Therapeutics*, Spiros Vamvakas was not involved in the review or decision process for this paper. The authors declared no competing interests for this work.

DISCLAIMER

The views expressed in this article are the personal views of the author(s) and may not be understood or quoted as being made on behalf of or reflecting the position of the regulatory agency/agencies or organizations with which the author(s) is/are employed/affiliated.

© 2019 The Authors. *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals, Inc. on behalf of American Society for Clinical Pharmacology and Therapeutics.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

- Oderkirk, J. Readiness of electronic health record systems to contribute to National Health Information and Research. OECD Health Working Papers, No. 99 (OECD Publishing, Paris, France, 2017). <https://doi.org/10.1787/9e296bf3-en>.
- Cave, A., Kurz, X. & Arlett, P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. *Clin. Pharmacol. Ther.* **106**, 36–39 (2019).
- European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) <<http://www.encepp.eu/>>.
- Eichler, H. et al. Data rich, information poor: can we use electronic health records to create a learning healthcare system for pharmaceuticals? *Clin. Pharmacol. Ther.* **105**, 912–922 (2019).
- Schneeweiss, S. Real-world evidence of treatment effects: the useful and the misleading. *Clin. Pharmacol. Ther.* **106**, 43–44 (2019).
- Dal Pan, G.J. Real-world data, advanced analytics, and the evolution of postmarket drug safety surveillance. *Clin. Pharmacol. Ther.* **106**, 28–30 (2019).
- Miksad, R.A., Samant, M.K., Sarkar, S. & Abernethy, A.P. Small but mighty: the use of real-world evidence to inform precision medicine. *Clin. Pharmacol. Ther.* **106**, 87–90 (2019).
- Koenig, F. et al. Sharing clinical trial data on patient level: opportunities and challenges. *Biom. J.* **57**, 8–26 (2015).
- Ramamoorthy, A. & Huang, S.M. What does it take to transform real-world data into real-world evidence? *Clin. Pharmacol. Ther.* **106**, 10–18 (2019).
- Bauer, P. & König, F. The risks of methodology aversion in drug regulation. *Nat. Rev. Drug Discov.* **13**, 317–318 (2014).
- Franklin, J.M., Glynn, R.J., Martin, D. & Schneeweiss, S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin. Pharmacol. Ther.* **105**, 867–877 (2019).
- Franklin, J.M. & Schneeweiss, S. When and how can real world data analyses substitute for randomized controlled trials? *Clin. Pharmacol. Ther.* **102**, 924–933 (2017).
- Randomized controlled trials duplicated using prospective longitudinal insurance claims: applying techniques of epidemiology <<https://www.rctduplicate.org/>>.
- ElZarrad, M.K. & Corrigan-Curay, J. The US Food and Drug Administration's real-world evidence framework: a commitment for engagement and transparency on real-world evidence. *Clin. Pharmacol. Ther.* **106**, 33–35 (2019).
- European Medicines Agency (EMA). Votrient – pazopanib EPAR <<https://www.ema.europa.eu/en/medicines/human/EPAR/votrient>>.
- European Medicines Agency (EMA). RSS; “EMA Regulatory Science to 2025” <<https://www.ema.europa.eu/en/about-us/how-we-work/regulatory-science-2025>>.
- European Medicines Agency (EMA). Qualification of novel methodologies for medicine development <<https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance/qualification-novel-methodologies-medicine-development>>.
- European Medicines Agency (EMA). Qualification opinion on cellular therapy module of the European Society for Blood & Marrow Transplantation (EBMT) Registry <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/draft-qualification-opinion-cellular-therapy-module-european-society-blood-marrow-transplantation_en.pdf>.
- European Medicines Agency (EMA). Qualification opinion on stride velocity 95th centile as a secondary endpoint in Duchenne Muscular Dystrophy measured by a valid and suitable wearable device <https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-secondary-endpoint-duchenne-muscular-dystrophy_en.pdf>.
- Innovative Medicines Initiative <www.imi.europa.eu>
- Schmidle, H., Gsteiger, S., Roychoudhury, S., O'hagan, A., Spiegelhalter, D. & Neuenschwander, B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* **70**, 1023–1032 (2014).
- Viele, K. et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm. Stat.* **13**, 41–54 (2014).
- Spiegelhalter, D.J., Abrams, K.R. & Myles, J.P. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation* (Wiley, Chichester, NH, 2004).
- Glynn, R., Schneeweiss, S., Wang, P., Levin, R. & Avorn, J. Selective prescribing led to overestimation of the benefits of lipid-lowering drugs. *J. Clin. Epidemiol.* **59**, 819–828 (2006).
- Eichler, H.G. et al. “Threshold-crossing”: a useful way to establish the counterfactual in clinical trials? *Clin. Pharmacol. Ther.* **100**, 699–712 (2016).
- International Council on Harmonization (ICH). ICH Topic E10: Choice of control groups in clinical trials <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf>.
- Sacks, H., Chalmers, T.C. & Smith, H. Randomized versus historical controls for clinical trials. *Am. J. Med.* **72**, 233–240 (1982).
- Schulz, K.F., Chalmers, I., Hayes, R.J. & Altman, D.G. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* **273**, 408–412 (1995).
- Guideline comparators & comparisons: Direct and indirect comparisons, adapted version 2015 <https://www.eunetha.eu/wp-content/uploads/2018/01/Comparators-Comparisons-Direct-and-indirect-comparisons_Amended-JA1-Guideline_Final-Nov-2015.pdf>.
- Naci, H. & O'Connor, A.B. Assessing comparative effectiveness of new drugs before approval using prospective network meta-analyses. *J. Clin. Epidemiol.* **66**, 812–816 (2013).
- Happich, M. et al. Reweighting RCT evidence to better reflect real life: a case study of the innovation medicines initiative. *Value Health* **19**, A711 (2016).
- Stevens, J.W., Fletcher, C., Downey, G. & Sutton, A. A review of methods for comparing treatments evaluated in studies that form disconnected networks of evidence. *Res. Synth. Methods* **2018**(9), 148–162 (2018).
- European Medicines Agency (EMA). EMA Reflection paper on the use of extrapolation in the development of medicines for paediatrics <https://www.ema.europa.eu/en/documents/scientific-guideline/adopted-reflection-paper-use-extrapolation-development-medicines-paediatrics-revision-1_en.pdf>.

34. Ollivier, C. *et al.* Commentary on the EMA Reflection Paper on the use of extrapolation in the development of medicines for paediatrics. *Br. J. Clin. Pharmacol.* **85**, 659–668 (2019).
35. Kent, D.M., Steyerberg, E. & van Klaveren, D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ* **363**, k4245 (2018).
36. Ballarini, N.M., Rosenkranz, G.K., Jaki, T., König, F. & Posch, M. Subgroup identification in clinical trials via the predicted individual treatment effect. *PLoS One* **13**, e0205971 (2018).
37. Lamont, A. *et al.* Identification of predicted individual treatment effects in randomized clinical trials. *Stat. Methods Med. Res.* **27**, 142–157 (2018).