The Evaluation of Effectiveness, or How Do We Know Whether a Treatment Works in Real Life? Part 2

Thomas M. Schindler[1] **and Helen Bridge**[2] / [1]Innovation Medical Writing, Boehringer Ingelheim Pharma, Biberach a.d. Riss, Germany; [2]Medical Communications, AstraZeneca, Cambridge, United Kingdom

Efficacy and Effectiveness

In our previous article (Schindler TM, Bridge H. The evaluation of efficacy, or how do we know whether a treatment works? Part 1. *AMWA J.* 2020;35(2):82-86), we described how randomized clinical trials (RCTs) determine efficacy of a treatment. Efficacy is generally understood as the ability of a drug or other intervention to reliably produce a positive effect in patients with a defined disease under controlled conditions. In other words, clinical trials are good tools to evaluate efficacy. Clinical trials provide a well-controlled framework that is characterized by the following:

- Selecting participants according to clearly defined eligibility criteria
- Randomly allocating participants to treatment groups
- Concealing study treatments to participants and study doctors (blinding)
- Clearly defining endpoint measures before study start

Provided a drug consistently demonstrates efficacy in several clinical studies and appears safe, regulators will approve it and permit its marketing for the treatment of a disease. However, once a drug is on the market, it will be used for the entire spectrum of patients and under circumstances that might not have been tested in the clinical studies. For example, the new drug will be used in patients who have several comorbidities in addition to the approved indication. The evaluation of a drug's effects under everyday conditions is called effectiveness research.

Why Does a Treatment's Effectiveness Differ From Its Efficacy?

There are many reasons why the effects of a drug in everyday clinical practice are different from those seen in clinical trials. In short, the entire societal-medical context contributes to the effectiveness of a drug or intervention. Before we look at some of the key factors, we need to remind ourselves that both "efficacy" and "effectiveness" are determined in groups of people and not individuals. While efficacy pertains to groups of participants in clinical trials, effectiveness pertains to all people with a disease who could be treated with the drug.

Issue 1: Physical Availability and Accessibility

Above all else, to be effective, a drug needs to be available to patients, at best to all patients with the disease it was developed to treat. There are national, regional, and local aspects of availability of a treatment. Approval of a drug by regulators permits its marketing and selling in a certain country. For example, a drug that is approved in the United States may not be available in Canada or Mexico.

Before the new drug reaches a patient, a number of obstacles in the supply chain need to be overcome. The first one is reimbursement, ie, the negotiation of price for the new drug with government agencies, insurance companies, and pharma wholesalers. Discussions on price may have a drastic effect on availability. If a pharmaceutical company considers the proposed price too low to ever regain its investments, it may decide not to provide the drug at all.

Regional differences in availability may arise when regional pharma wholesale companies decide on stocking of medicines. They may decide not to stock a certain drug in certain locations because they do not believe that it will be widely used there. Even if a drug is available via a wholesaler and the physician is willing to prescribe it, a patient's health insurance plan may not cover the entire cost for the new treatment. Likewise, if a patient is in hospital, she might not get the drug because the company that runs the hospital may not see its benefits and may therefore not include it in their treatment plans and offer an alternative treatment. Furthermore, health care providers may not believe in the benefits of a certain drug and may prescribe treatments that they consider to have superior therapeutic effects instead.

Issue 2: Medical Tradition

Medical doctors, ie, the prescribers of medicines, undergo many years of training, and they apply this knowledge in their clinical practice, thereby maintaining a certain medical tradition. Such traditions are tenacious, and it may take a long time before new therapeutic options are widely accepted and offered to all patients. Conversely, health care providers and hospitals may continue to offer treatments to their patients that have long been shown to be less effective because this is their established practice, offers financial advantages, or is expected by patients.

Medical practitioners follow the insights they gain in their medical practice. If they have the impression that a particular medicine works for their patients, they will continue to prescribe it. They will stick to what they believe is helpful and may not use a new drug with which they have little experience. This is particularly likely when a new drug has only been tested against placebo and not against established treatments. A new drug may not be given to a patient because the physician has not heard about it or does not believe it is superior to the drugs that she usually prescribes.

Issue 3: Treatment Adherence

In clinical trials, participants are closely monitored with regard to how and when and at which dose they take their medication. For example, study participants are reminded to take their dose at the same time every day, with or without food, to achieve the optimal effect. In normal life, things tend to be different, and patients might occasionally forget to take their medication. On days when their disease is particularly discomforting, they may be inclined to increase the dose of their medicine. If their lives are busy, patients may forget to renew their prescription in time and may therefore not take any medication for a while. Thus, although the drug is available to the patients, its effects may be smaller than those observed in the clinical studies because of limited adherence to medication plans. Furthermore, patients might be taking additional drugs to treat other conditions, and these drugs may influence the effects of the new treatment.

Real-World Data and Real-World Evidence to Determine Effectiveness

Many different factors affect the use of a drug in everyday medical practice. It is therefore very difficult to determine how effective a drug is "out there." Post-marketing (Phase 4) studies usually focus on safety rather than effectiveness. However, for many stakeholders in the health care system, it is important to know how well a drug works in the real world:

- patients taking the medicine want to know whether it works,
- payers like health insurance companies need to understand whether they are paying for an effective treatment,
- government agencies want to know how the new medicine affects public health,
- pharmaceutical companies want to understand to what extent available medicines meet patients' needs and whether there is space for additional medicines.

Recently, there has been much discussion about the usefulness of "real-world data" (RWD) and "real-world evidence" (RWE) to evaluate effectiveness of treatments and support health care decisions. RWD and RWE are generally understood to refer to data and evidence from sources other than traditional clinical trials, but it has not always been clear exactly what these categories include. Over a number of years, regulators were pressed to delineate ways to incorporate nonclassical data in their evaluations. In 2018, the US Food and Drug Administration (FDA) released the framework for the Agency's Real-World Evidence Program. There, they give the following definitions:

- RWD is "data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources."
- RWE is "the *clinical* evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD."

Real-World Data: Abundant, Promising, but Messy

Sources of RWD include registries, electronic health records, medical claims databases, mobile devices, and so on. The recent excitement about RWD and RWE has focused on the availability of apparently rich sources of data, particularly those resulting from digitization. Sources such as electronic health records from entire health systems or wearable devices that generate data on a plethora of variables at frequent intervals promise an abundance of easily collected data.

There is a growing tendency to believe that data collected in the course of routine health care, whether via the records kept by health care providers or directly from patients using apps, better reflect how treatments actually perform than data collected in the research settings of clinical trials. Large quantities of RWD can easily be collected from thousands, or even millions, of patients, and such "big data" seem to promise greater representativeness than data from the comparatively small, selected populations of patients included in clinical trials.

There are also clear practical and economic grounds for exploring RWD: vast quantities of such data can be collected and accessed quickly and cheaply by comparison with the laborious and costly collection of data in traditional clinical trials. RWD therefore holds appeal for companies keen to reduce the expense and time taken to bring treatments to market.

Easy and cost-effective access to large quantities of medical data collected from the full range of users of a treatment in real-world conditions may sound too good to be true. Sure enough, there are several major obstacles to using RWD to reach reliable conclusions regarding the effectiveness of treatments.

Obstacle 1: Data Quality

One of the factors behind the expense of running clinical trials is the quest for good data quality and the consequent rigor with which data are collected. The clinical trial protocol specifies precisely which data are to be collected for each patient, the time points at which they are to be collected, and the methods, often including the precise equipment, that are to be used. A trial is designed to answer specific scientific questions, and the data to be collected are those that are required to answer these questions. Data are recorded on a case report form designed specifically for the trial. Strict procedures are followed to ensure data are collected in accordance with Good Clinical Practice principles, and clinical trial monitors verify the data for completeness and accuracy. These methods ensure that clinical trial data are highly standardized from patient to patient and from center to center, with the same variables measured and recorded in the same way and at the same time points. These provisions make clinical trial data reliable and trustworthy.

Data from real-world sources are unlikely to share any of these qualities. Table 1 outlines the main problems with quality of RWD. These deficiencies introduce noise and bias that make it difficult to draw reliable conclusions about the effectiveness of a treatment.
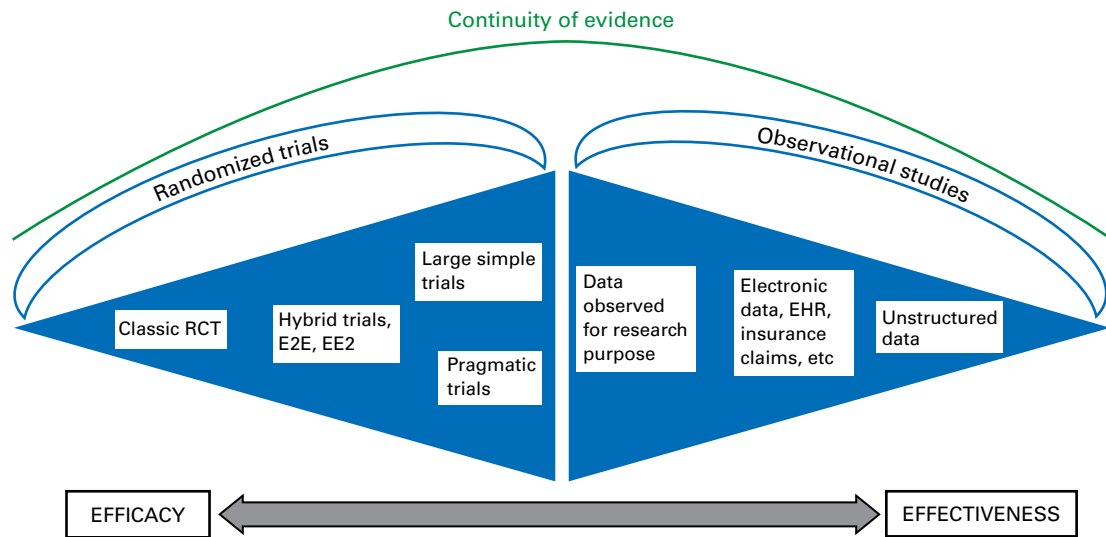
Obstacle 2: Lack of Randomization

Medical research generally falls into 1 of 2 broad categories: clinical trials and observational studies (Figure 1). The key feature that separates these approaches is the presence or absence of randomization. A recent opinion piece in the *New England Journal of Medicine* insightfully contrasts the "magic of randomization" in RCTs with the "myth of real-world evidence" from observational studies. Randomization ensures that there are no systematic differences between treatment groups with regard to patients' characteristics that may affect efficacy outcomes. It is likely impossible to achieve such a balance in an observational study because the groups that are compared did not result from randomization. Whenever the treatment is a choice, whether by a doctor or the patient, groups of patients taking different treatments are likely to differ systematically, often in ways that are difficult to identify. Moreover, the reasons for the choice of a particular treatment are almost never entered into health records or databases. This

Table 1. Real-World Data: Common Problems With Data Quality

Characteristic of Data	Likely Problems With RWD
Relevance to research questions	Data are collected for purposes other than research and will not be optimal for answering the question of interest. There is a risk of allowing the available data to determine the research questions that are asked (rather than defining the question first and then looking for the data).
Reliability	Inaccuracies in the data may result from human error or faulty devices. Data collected cannot be verified by comparison with source data. Random errors and systematic bias are difficult to identify and impossible to correct.
Completeness	RWD are characterized by a high quantity of missing values, with no information as to why data are missing.
Consistency and interoperability	Variables recorded, measurement methods, and data formats and data standards usually vary greatly across patients, health care providers, devices, companies, etc. The lack of common data standards, ie, the way data are structured, stored, and summarized, makes it challenging to collate and analyze the data.

is likely to result in a biased comparison of the treatments. Indeed, there have been well-publicized cases in which observational studies and RCTs have come to opposing conclusions about particular treatments.

An obstacle to using RWD to evaluate effectiveness is that these data originate in routine health care contexts in which patients are not randomized to treatments. Although such data have long been used by regulators to evaluate the safety of treatments, using them to arrive at unbiased evaluations of effectiveness is challenging and requires sophisticated considerations on methodology. In recent years, statisticians have worked on developing new study designs and complex analysis methods, including so-called "causal inference" and machine-learning methods, to help overcome some of the limitations of analyzing observational data. The FDA has committed (in its RWE framework document) to evaluating "the potential role of observational studies in contributing to evidence of drug product effectiveness" and has supported a series of workshops that included discussion of methods for assessing and minimizing bias in observational studies.

Continuity of evidence

Randomized trials

Observational studies

Large simple trials

Classic RCT

Hybrid trials, E2E, EE2

Pragmatic trials

Data observed for research purpose

Electronic data, EHR, insurance claims, etc

Unstructured data

EFFICACY

EFFECTIVENESS

©TMSchindler & HBridge 2021

Figure 1. Idealized theoretical framework for the generation of evidence for efficacy and effectiveness. Hybrid trials are studies that combine efficacy and effectiveness assessments. E2E, efficacy-to-effectiveness trial (sequential assessment); EE2, efficacy-and-effectiveness-too trial (simultaneous assessment); EHR, electronic health records.

Obstacle 3: The Need for Rigorous Research Methodology
No amount of data can, in itself, help us evaluate treatments. Data only become evidence once they are used within a methodological framework or research study to answer specific questions, such as "Is treatment X effective in patients with heart failure?" or "How much more effective is treatment A than treatment B at extending survival in patients with advanced non-small cell lung cancer?" For RWD to provide evidence of effectiveness, they need to be analyzed using an appropriate research methodology. This includes the definition of data formats for datasets, the availability of a comprehensive study protocol, and detailed analysis plans to ensure reproducibility of results. If data are derived from novel data sources such as wearables, these data need to accurately reflect the clinical outcome that is being investigated.

Combining Clinical Trials and Real-World Data for Insight into Effectiveness
Observational studies are not the only way of using RWD to tackle the question of effectiveness. A promising alternative approach advocated in the FDA framework document as a way of generating RWE is to make clinical trials more "real life." This can be done by simplifying the trial design to become a "large, simple trial" or by otherwise incorporating pragmatic elements in the design so that patients' treatment within the trial closely resembles routine clinical practice (Figure 1). Alternatively, hybrid designs can be used whereby health data that are routinely collected are used in the trial, together

with data specified by the trial protocol (for example, efficacy-to-effectiveness or efficacy-and-effectiveness-too trials). Combining clinical trial methodology, notably randomization, with the collection and evaluation of RWD has the benefit of enabling a more unbiased evaluation of treatment effects in settings that are close to real-world clinical practice.

Particularly in rare diseases, it is often impossible to conduct RCTs because of the low number of patients available. Recruiting a sufficient number of patients into a study may take too long to yield useful results. In these instances, single-arm open-label studies may be conducted and the results compared with external controls, ie, RWD collected outside of the study. The control data could come from registries, medical records, scientific literature, or expanded access programs. A recent example of such an approach is the approval of avelumab in metastatic Merkel cell carcinoma. The drug was approved in 2017 based on a single-arm open-label study that compared the study outcomes with historical controls retrieved from electronic health records. In 2019, approval of palbociclib for HR+, HER2- advanced breast cancer in men, a label extension, was based on post-marketing reports and electronic health records.

Responses to the coronavirus disease 2019 (COVID-19) pandemic provide a further, highly topical example of how randomized trials that incorporate real-world elements can generate reliable evidence of effectiveness to guide clinical decision-making about treatments. At the time of writing, a number of large, simple trials to evaluate various potential

treatments for COVID-19 are ongoing. For example, the global Solidarity trial, initiated by the World Health Organization, had recruited over 12,000 patients by October 2, 2020, with 116 countries having joined or expressed an interest in joining the trial. In the United Kingdom, the RECOVERY trial is being conducted at all major hospitals and had enrolled over 20,000 patients by December 2020. These trials have simple protocols and heavily streamlined procedures. The aim is to maximize recruitment and minimize the burden of participation on health care staff. The RECOVERY trial, for example, has minimal eligibility criteria, simple and quick informed consent and randomization processes, and minimal data collection requirements, with follow-up information to be recorded at a single time point. Within 3 months of trial initiation, results were released showing the effectiveness of dexamethasone for reducing mortality in patients on mechanical ventilation or supplemental oxygen. The trial has also shown hydroxychloroquine, lopinavir and ritonavir, and azithromycin to be ineffective at reducing mortality from COVID-19.

These examples show the potential of combining RCTs with RWD to produce robust evidence of effectiveness that can inform decision-making about treatments. They also indicate that the use of RWD, while informative as a supplement to RCTs in certain contexts, is unlikely to replace traditional RCTs. Until a treatment is widely used in clinical practice, there are no RWD relating to its use. Consequently, data to support first-time marketing approvals for new drugs have to come from RCTs. The context of a clinical trial also allows close monitoring of patients, which is essential for their safety with drugs that have not yet received marketing authorization. As the examples given show, the situation is very different when a new drug is to be compared with a widely used drug or when well-established drugs are to be evaluated as potential treatments for a new disease.

Author contacts: Thomas.Schindler@boehringer-ingelheim.com; Helen.Bridge@astrazeneca.com

Resources

Baumfeld Andre E, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: The changing landscape of the regulatory approval process. *Pharmacoepidemiol Drug Saf.* 2020;29(10):1201-1212.

Beaulieu-Jones BK, Finlayson SG, Yuan W, et al. Examining the use of real-world evidence in the regulatory process. *Clin Pharmacol Ther.* 2020;107(4):843-852.

Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. *N Engl J Med.* 2020;382(7):674-678.

Epstein D, ProPublica. When evidence says no, but doctors say yes. The Atlantic website. Published February 22, 2017. Accessed January 8, 2021. https://www.theatlantic.com/health/archive/2017/02/when-evidence-says-no-but-doctors-say-yes/517368/

Ford I, Norrie J. Pragmatic trials. *N Engl J Med.* 2016;375(5):454-463.

Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? *Clin Pharmacol Ther.* 2017;102(6):924-933.

Najafzadeh M, Gagne JJ, Schneeweiss S. Synergies from integrating randomized controlled trials and real-world data analyses. *Clin Pharmacol Ther.* 2017;102(6):914-916.

Randomised evaluation of COVID-19 therapy (RECOVERY) Clinical Study Protocol, V12.1. RECOVERY Trial website. Published December 16, 2020. Accessed January 8, 2021. https://www.recoverytrial.net/files/recovery-protocol-v12-1-2020-12-16.pdf

Reynolds MW, Bourke A, Dreyer NA. Considerations when evaluating real-world data quality in the context of fitness for purpose. *Pharmacoepidemiol Drug Saf.* 2020;29(10):1316-1318.

Saesen R, Huys I. COVID-19 clinical trials: See it big and keep it simple. *BMJ Evid Based Med.* Published online July 8, 2020. doi:10.1136/bmjebm-2020-111447

Schindler TM, Bridge H. The evaluation of efficacy, or how do we know whether a treatment works? Part 1. *AMWA J.* 2020;35(2):82-86.

Scott IA, Attia J. Cautionary tales in the interpretation of observational studies of effects of clinical interventions. *Intern Med J.* 2017;47(2):144-157.

Selker HP, Eichler H-G, Stockbridge NL, et al. Efficacy and effectiveness too trials: Clinical trial designs to generate evidence on efficacy and on effectiveness in wide practice. *Clin Pharmacol Ther.* 2019;105(4):857-866.

Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence – What is it and what can it tell us? *New Engl J Med.* 2016;375(23):2293-2297.

World Health Organization. "Solidarity" clinical trial for COVID-19 treatments. Published 2020. Accessed January 8, 2021. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments

US Food and Drug Administration. Framework for FDA's real-world evidence program. Published December 2018. Accessed January 8, 2021. https://www.fda.gov/media/120060/download

Wapner J. How prescription drugs get their prices, explained. Newsweek website. Published April 14, 2017. Accessed January 8, 2021. https://www.newsweek.com/2017/04/14/prescription-drug-pricing-569444.html#:~:text=Reimbursement%20is%20the%20amount%20the,as%20a%20pharmacy%20benefit%20manager