Therapeutic Innovation & Regulatory Science
https://doi.org/10.1007/s43441-022-00492-z

DIA

COMMENTARY

Examining Endpoint Concordance in Clinical Trials and Real-World Clinical Practice to Advance Real-World Evidence Utilization

Debra Schaumberg[1] · Kay Larholt[2] · Elizabeth Apgar[2] · Chris L. Pashos[3] · Gigi Hirsch[2]

Received: 24 June 2022 / Accepted: 24 December 2022

Abstract

Real-world evidence (RWE) is increasingly contributing to more informed decisions regarding the optimal access to and use of therapeutics to improve patient outcomes. However, in many cases, a disconnect between evidence derived from clinical trials and the RWE that follows market approval impedes the potential value and widespread adoption of RWE to optimize patient care. Collaborators with the Learning Ecosystems Accelerator for Patient-centered, Sustainable innovation (LEAPS), a major project of the Tufts Medical Center [formally Massachusetts Institute of Technology (MIT)] NEW Drug Development ParadIGmS (NEWDIGS) initiative, propose assessing the relationship between efficacy endpoints used in randomized controlled trials (RCTs) and effectiveness measures that inform treatment decisions within real-world clinical settings as one way to bridge this divide and further leverage RWE to improve care and patient outcomes. This commentary outlines elements of an endpoint concordance study using Rheumatoid Arthritis as a case study. The authors describe the ways in which better understanding of the relationship between effectiveness and RCT endpoints could improve the confidence in and adoption of RWE by both contextualizing existing RWE as well as identifying ways in which to improve the value of RWE in improving care and outcomes.

Keywords Real-world evidence (RWE) · Clinical trials · Endpoint concordance · Precision medicine

Commentary

Understanding the real-world performance of therapeutics in the evolving clinical context in which they are being used is critical to shaping clinical practice and informing decisions regarding their access and use. While the potential of real-world evidence (RWE) to inform health care decisions and ultimately improve patient outcomes is increasingly recognized [1–8], there is usually a disconnect between the evidence derived from clinical trials and the RWE that follows, due at least in part to differences between the efficacy measures used in randomized controlled trials (RCTs) and the effectiveness measures commonly used in clinical practice. A better understanding of this gap between the RCT efficacy endpoints and the clinical effectiveness measures that inform real-world treatment decisions by key stakeholders (regulators, payers, clinicians, and patients) is critical for optimizing the adoption and value of RWE in improving patient well-being.

RCTs are intended to isolate a medical product's treatment effect in support of regulatory evidence requirements for approval and as such incorporate key design features such as strict inclusion/exclusion criteria, randomization, masking, and standardized procedures and assessments to reduce variability. However, because of the typically controlled settings and narrowly defined patient populations, evidence derived from the RCTs does not adequately inform decisions regarding access to and use of the medical product in real-world settings characterized by heterogeneity of patients and clinical care environments. Augmenting evidence from RCTs with RWE is crucial to optimize treatment regimens and improve patient outcomes.

LoCasale et al. [9] describe the disconnect between RCTs and RWE and propose bridging this divide by incorporating endpoints used in real-world clinical practice into pivotal

✉ Gigi Hirsch
phirsch@tuftsmedicalcenter.org

[1] Evidera, a PPD Company, Bethesda, MD, USA

[2] Center for Biomedical System Design & NEWDIGS, Institute for Clinical Research & Health Policy Studies, Tufts Medical Center, 800 Washington Street, #1013, Boston, MA 02111, USA

[3] Genesis Research, Hoboken, NJ, USA

Published online: 09 January 2023

RCTs when possible. While this approach would generate valuable information, its applicability to real-world treatment decisions remains limited due to the differences between the research environment of RCTs and the clinical practice patterns of routine care. Rather the reverse, assessing the relationship between clinical effectiveness measures and RCT endpoints in the real-world clinical practice setting in which they are used to direct treatment decisions, has the potential to impact the role of RWE-based clinical decision-making more broadly. Evaluating metrics such as the concordance between RCT and RWE assessments within a real-world setting offers the potential to "calibrate" endpoints used in RCTs versus those used in RWE studies. Such an approach may improve confidence in RWE more broadly and enable greater adoption of RWE to inform decisions of regulators, payers, providers, and patients. (Fig. 1) This concept is also an important step towards broader implementation of point-of-care platforms designed to meet the evidentiary needs of precision medicine [10–13].

Collaborators with the Learning Ecosystems Accelerator for Patient-centered, Sustainable innovation (LEAPS), a major project of the Tufts Medical Center [formerly Massachusetts Institute of Technology (MIT)] NEW Drug Development ParadIGmS (NEWDIGS) initiative, propose the design of an endpoint concordance study as one way to bridge this gap and further the broader goal of LEAPS to modernize how we plan, produce, and use RWE to advance the knowledge, practice, and sustainability of precision medicine [14–16]. The endpoint concordance design was developed within the larger concept of an Adaptive Point-of-Care (APoC) platform using Rheumatoid Arthritis (RA) as a case study. Briefly, the RA APoC platform would integrate elements of adaptive clinical trials [11, 17, 18] (e.g., response adaptive randomization, disease-focused platform infrastructure) and point-of-care study designs [12, 13] into a perpetual learning platform embedded in clinical care to assess the comparative effectiveness of therapeutic regimens for RA subpopulations. RA was selected as the initial LEAPS case study in part because of the critical knowledge gaps in RA care that impede regimen optimization. In the 2021 American College of Rheumatology (ACR) Guidelines for the treatment of RA, the evidence level for most of the recommendations was rated as very low or low certainty and the guideline panel identified several key clinical knowledge gaps including comparative effectiveness evidence to inform optimal treatment selection between therapeutic classes, adding versus switching therapeutics, de-escalation or stopping, and predicting response [8]. Indeed, many RA treatment decisions are influenced by providers' prior experience and treatment setting (e.g., academic center versus

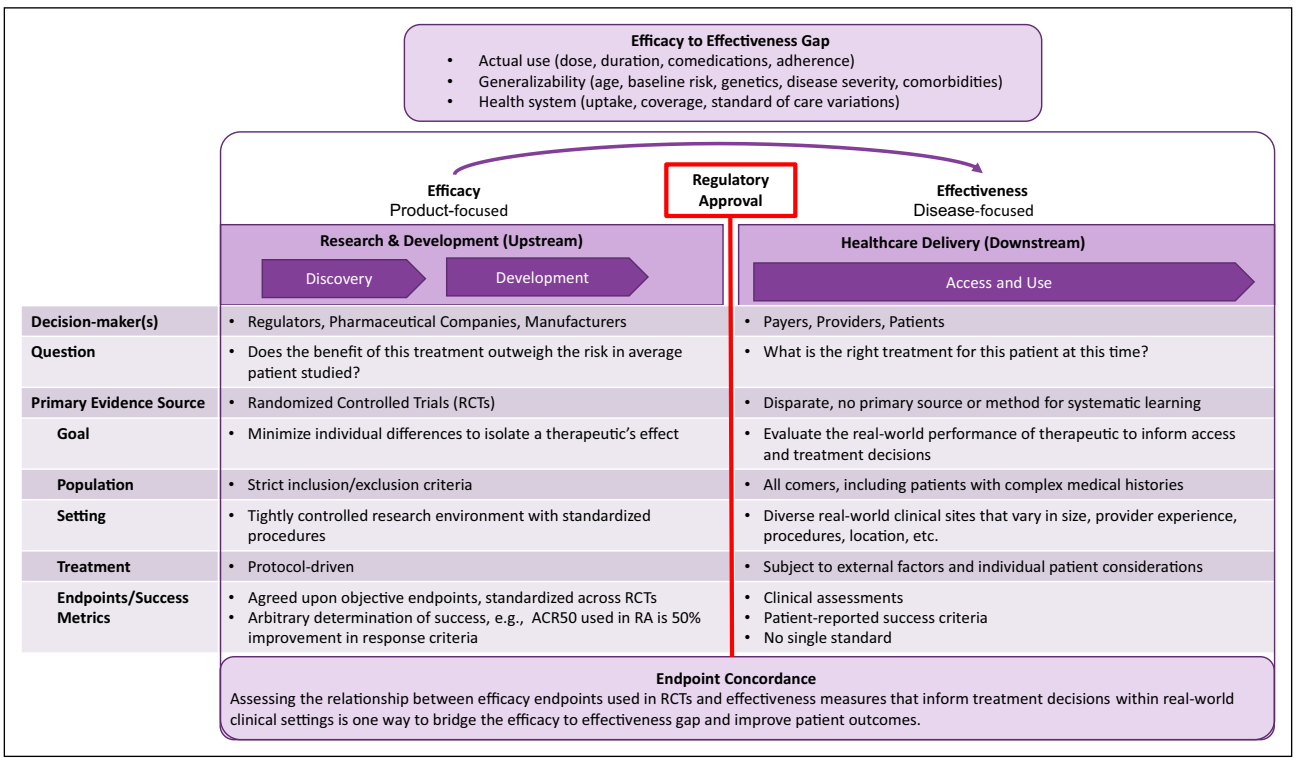

Figure 1. The efficacy to effectiveness gap leaves a disconnect between evidence derived from RCTs and RWE that follows market approval. Endpoint concordance is one way to bridge this divide and further leverage RWE to improve patient care and outcomes. *ACR* American College of Rheumatology, *RA* rheumatoid arthritis, *RCT* randomized controlled trial, *RWE* real-world evidence.

community care) as well as non-clinical factors such as payers' formulary or step therapy requirements [19]. Narrowing the disconnect between RCT and RWE in RA could begin to address these knowledge gaps. Furthermore, RA provides a good illustration for an endpoint concordance study as endpoints in RA are more subjective than in other disease and therapeutic areas, e.g., HbA1c level for diabetes and cardiac events, and therefore more subject to variation in assessment.

The general concept of an endpoint concordance study within a real-world clinical practice setting is to assess all patients in the study for both the routine measure(s) typically employed by the practice clinicians [clinical effectiveness measure(s)] and the endpoint(s) used in pivotal RCTs. To reduce bias, the order of assessment (i.e., effectiveness-RCT vs. RCT-effectiveness) would be randomized by patient. Additional strategies to reduce bias such as having different providers assess the effectiveness versus RCT endpoint could be considered if operationalizable. The concordance of the two different endpoints could then be evaluated using the appropriate statistical methods [20]. For the RA design case, the proposed endpoint concordance study would leverage the APoC platform at the treatment timepoint of conventional synthetic disease-modifying anti-rheumatic drugs (csDMARD) failure and initiation of second-line treatment with biologic DMARDs (bDMARDS). Within the context of this larger study, a subset of provider groups would collect, at baseline and follow-up, the routine measure typically used in their clinical practice, e.g., Routine Assessment of Patient Index Data 3 (Rapid-3), Clinical Disease Activity Index (CDAI), or Disease Activity Score-28 (DAS-28), as well as the RA RCT endpoint. For a dichotomous RCT measure of response such as the ACR20, patients would be categorized as responders at follow-up by the ACR20 criteria as well as by a pre-determined equivalent cut off for the effectiveness endpoint, e.g., $\geq 50\%$ improvement from baseline for CDAI [21]. The concordance of response to bDMARD treatment as assessed by the ACR20 versus the effectiveness measure could then be calculated using Cohen's kappa [20]. To strengthen the value of this study, potential predictors of concordance, such as disease severity and duration, demographics, and provider or health system profiles, would also be analyzed and could lead to the development of enhanced multivariable approaches to endpoint definitions.

We recognize that conducting the concordance study would impose an additional work-flow burden on the clinical care providers, but believe that study findings would ultimately lead to greater efficiencies and effectiveness of patient care. With this goal in mind, we would propose to expand the study, at least in a subset of settings, to include evaluation of performance of other efficacy and effectiveness measures. For example, the proposed concordance study could also be designed to further understand the role of effectiveness endpoints in evaluating the performance of therapeutics in real-world use by exploring the relationships between ACR20 and routine measures of clinical effectiveness with other measures of disease severity and treatment response (e.g., other clinical assessments and patient-reported outcomes) collected as part of the APoC platform. This work would provide insight as to whether the routine clinical assessments adequately reflect real-world performance of therapeutics, thereby demonstrating the value of RWE, or reveal if in some cases it will be necessary to incorporate better measures into clinical practice to improve the value of RWE.

Results of the proposed endpoint concordance study would help contextualize real-world assessment of RA treatment response and evaluate the consistency of evidence derived across different study designs, clinical settings, and populations. Understanding the relationship between RCT and real-world endpoints would bridge one element of the disconnect between RCTs and RWE, thereby increasing confidence in RWE. Greater understanding of this relationship would also identify both how existing RWE is best utilized as well as avenues through which to improve the value of RWE. Additionally, learnings would inform endpoint concordance studies in other diseases, further promoting the value and adoption of RWE to inform optimal access to and use of therapeutics.

The concepts described in this commentary are relevant to all stakeholders in the clinical research ecosystem, especially those planning RCTs or real-world studies. While RA was used as a case study, there is value in applying the concept of endpoint concordance across all therapeutic areas, especially where there are critical differences between typical RCT efficacy and RWE effectiveness endpoints. Comparisons of RCT versus real-world endpoints in any therapeutic area will consider design differences/similarities in factors such as timing, measurement error, misclassification, and endpoint type. However, the relative importance of and focus upon specific items in an endpoint concordance study may vary based on the endpoint, indication, and therapeutic area, as well as the intended use of the resulting evidence.

Author Contributions

All authors contributed to the conception or design of the work, drafting the work, or revising it critically for important intellectual content, final approval of the version to be published, and agreement to be accountable for all aspects of the work ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

This work was conducted and funded by the New Drug Development Paradigms (NEWDIGS) initiative under Massachusetts Institute of Technology's (MIT's) Center for Biomedical Innovation (CBI). Since

completion of the work, NEWDIGS has moved to Tufts Medical Center's Center for Biomedical System Design. NEWDIGS members and collaborator partners are identified on the website (https://newdigs.tuftsmedicalcenter.org/about/members-and-collaboration-partners/). The opinions expressed herein are those of the author(s), and not necessarily those of the institution with whom they are affiliated, or those of NEWDIGS members and collaborators.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed for the described work.

Declarations

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Klonoff DC. The expanding role of real-world evidence trials in health care decision making. J Diabetes Sci Technol. 2020;14(1):174–9.
2. Jarow JP, LaVange L, Woodcock J. Multidimensional evidence generation and FDA regulatory decision making: defining and using "Real-World" data. JAMA. 2017;318(8):703–4.
3. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA. 2018;320(9):867–8.
4. US Food and Drug Administration. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drugs and Biological Products. Guidance for Industry. In: Center for Drug Evaluation and Research, editor. September 2021.
5. Khosla S, White R, Medina J, Ouwens M, Emmas C, Koder T, et al. Real world evidence (RWE) - a disruptive innovation or the quiet evolution of medical evidence generation? F1000Res. 2018. https://doi.org/10.12688/f1000research.13585.2.
6. Chambers JD, Panzer AD, Pope EF, Graff JS, Neumann PJ. Little consistency in evidence cited by commercial plans for specialty drug coverage. Health Aff. 2019;38(11):1882–6.
7. Panzer AD, Margaretos NM, Lai RC, Enright DE, Chambers JD. What types of real-world evidence studies do US commercial health plans cite in their specialty drug coverage decisions? Pharmacoepidemiol Drug Saf. 2020;29(10):1307–11.
8. Fraenkel L, Bathon JM, England BR, St Clair EW, Arayssi T, Carandang K, et al. American college of rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Care Res. 2021;73(7):924–39.
9. LoCasale RJ, Pashos CL, Gutierrez B, Dreyer NA, Collins T, Calleja A, et al. Bridging the gap between RCTs and RWE through endpoint selection. Ther Innov Regul Sci. 2021;55(1):90–6.
10. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. N Engl J Med. 2017;377(1):62–70.
11. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. Clin Pharmacol Ther. 2009;86(1):97–100.
12. Angus DC. Fusing randomized trials with big data: the key to self-learning health care systems? JAMA. 2015;314(8):767–8.
13. Angus DC, Berry S, Lewis RJ, Al-Beidh F, Arabi Y, van Bentum-Puijk W, et al. The REMAP-CAP (Randomized embedded multifactorial adaptive platform for community-acquired pneumonia) study rationale and design. Ann Am Thorac Soc. 2020;17(7):879–91.
14. Hirsch G. Leaping together toward sustainable, patient-centered innovation: the value of a multistakeholder safe haven for accelerating system change. Clin Pharmacol Ther. 2019;105(4):798–801.
15. MIT NEWDIGS LEAPS Project. Downstream Innovation Part I: Strategic Perspectives and the Case for LEAPS. MIT NEWDIGS Research Brief [Internet]. June 10, 2022; 2019L311v045. Available from: https://newdigs.tuftsmedicalcenter.org/wp-content/uploads/2019/11/LEAPS_Research_Brief_2019L311v045-Downstream_Innovation_Part_I.pdf.
16. MIT NEWDIGS LEAPS Project. Downstream Innovation Part II: Real-World Discovery LEAPS to Life. MIT NEWDIGS Research Brief [Internet]. June 10, 2022; 2020L302v46. Available from: https://newdigs.tuftsmedicalcenter.org/downstream-innovation-part-ii/.
17. Alexander BM, Ba S, Berger MS, Berry DA, Cavenee WK, Chang SM, et al. Adaptive global innovative learning environment for glioblastoma: GBM AGILE. Clin Cancer Res. 2018;24(4):737–43.
18. Pancreatic Cancer Action Network. Precision Promise. Available from: https://www.pancan.org/research/precision-promise/.
19. American College of Rheumatology Committee on Rheumatologic Care. Position Statement: Step Therapy [updated February 2019]. Available from: https://www.rheumatology.org/Portals/0/Files/Step%20Therapy.pdf.
20. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: measures of agreement. Perspect Clin Res. 2017;8(4):187–91.
21. Smolen JS, Aletaha D. Scores for all seasons: SDAI and CDAI. Clin Exp Rheumatol. 2014;32(85):75–9.