

Bridging the divide between clinical research and clinical care in oncology: An integrated real-world evidence generation platform

Digital Health
Volume 7: 1–5
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076211059975
journals.sagepub.com/home/dhj



Ariel B. Bourla¹  and Neal J. Meropol¹

Abstract

Real world data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources; real-world evidence (RWE) generated by RWD analyses can become an important component of drug development programs and, potentially, regulatory decision-making. As a RWD source, electronic health records (EHRs) can now provide patient-level data at unparalleled depth and granularity. We propose a RWE generation framework that could maximize the synergy between RWD and prospective clinical trials by capitalizing on an emerging data curation infrastructure that may be applied to both retrospective and prospective research. In this platform, centralized data collection and monitoring could be enabled via routine EHR use, and seamlessly integrated with select intentional data capture during prospective study periods. By bridging the divide between routine care and clinical research, this integrated platform aggregates retrospective and prospective data, collected both routinely and intentionally. This approach makes clinical trial participation more available to patients, increasing the potential depth of data, representativeness and efficiency of clinical research.

Keywords

cancer, electronic health records, technology, medicine, oncology, studies, mixed methods, EHR, real-world data, RWD, real-world evidence, RWE

Submission date: 19 November 2020; Acceptance date: 27 October 2021

Introduction

Epidemiologic studies based upon real-world data (RWD) have been common since the 1950s, but recently, they are becoming more than descriptive and hypothesis-generating. RWD, related to patient health status and/or the delivery of health care, are routinely collected from a variety of sources, such as electronic health records (EHRs) or registries (Figure 1A).^{1,2} Real-world evidence (RWE), generated by analyzing RWD, provides insights into a disease's natural history and the potential benefits or risks of a medical intervention. RWD are now an important component of drug development programs and regulatory decision-making^{3–7} via the application of new research designs, technologic and analytic innovations, and the acknowledgment of societal norms governing research use of patient data. This article focuses on a therapeutic area of high unmet need, oncology, as case in point to propose an RWE generation platform that recognizes a

continuum between retrospective and prospective research, challenging the historical distinctions between the two, in an effort to accelerate patient-oriented research. This platform could enable centralized data collection and monitoring, intentional data capture outside of routine clinical care (e.g. genomic data), and permit prospective data collection during an active study period followed by seamless and light-weight long-term follow up, with low burden for clinical trial sites and investigators. Bridging the divide between routine clinical care and clinical research would increase the efficiency of the evidence generation process and the generalizability of resulting learnings, and would expand clinical trial

¹Flatiron Health, Inc., New York, NY, USA

Corresponding author:

Ariel B. Bourla, Flatiron Health, Inc., 233 Spring St, New York, NY 10013, USA.

Email: abourla@flatiron.com



A. Key concepts and taxonomy in RWE research

Real-World Data (RWD). Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources including but not limited to EHRs, claims data, data from clinical registries, patient-reported data (including home-based), and data gathered from other sources, such as mobile devices.

Real-World Evidence (RWE). The clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD. RWE can be generated by different study designs or analyses, including retrospective and prospective observational studies.

Retrospective Study (Observational). A research output where the exposure/treatment of interest are not implemented for the purpose of the study, but have instead occurred in the past and the impact of the exposure/treatment are available through historical data. The data themselves are generated prior to the initiation of the research for purposes outside of the study and are curated to meet a study-specific data model.

Prospective Study (Observational). A study where the population of interest is identified before initiation of the study, and exposure/treatment and outcome data are collected from that point forward. The start of the study is defined as the time at which the research protocol for the specific study question is initiated.

Intentionally Collected Data. Data whose collection is above and beyond what would otherwise be collected as part of routine practice. Intentionally collected variables are critical for analyses or for protocol integrity, and as a result require particularly high completeness and quality. Sites or patients participating in a study are likely to be informed about these requirements and queried when there are concerns or inconsistencies with these data points. Intentionally collected data are either routinely collected data that are queried to improve quality or are the result of a mandated schedule of assessments where the data reflect an impact on patient care (for example, prespecified assessment timing, non-standard diagnostic testing, or additional collection of patient reported outcomes).

Routinely Collected Data. Data that are generated during a physician's encounter with a patient and are commonly documented. These variables are not queried when there are concerns or inconsistencies with the data.

B. Integrated RWE generation platform.

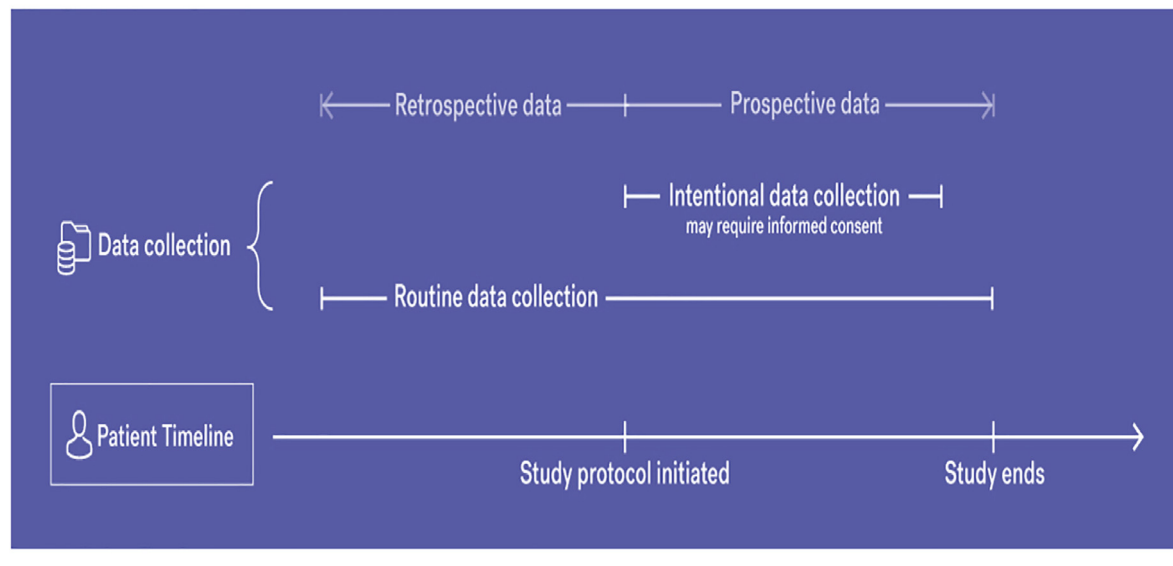


Figure 1. In the proposed platform and for each individual patient, longitudinal research data streams can overlay routinely collected and intentionally collected data, obtained retrospectively and prospectively (A), and the access to those two layers of data may be governed by informed consent and privacy boundaries (B).

participation opportunities for investigators and patients. While we use oncology as our example, this approach is broadly applicable to other disciplines.

In the current oncology landscape, the exponential growth of scientific knowledge and new therapies has led to redefinition of diseases, and information-intensive clinical decision-making, research questions and designs. While prospective trials are the research gold standard, randomized controlled trials (RCTs), with their narrow scope,

may not provide the best estimates of the real-world effectiveness of an intervention.⁸ RWE can unlock information about routine care in both broad and hard-to-accrue rare populations.⁹

EHR-derived RWE can be used to complement traditional clinical trial-based research. EHR infrastructure facilitates both patient management and the capture of comprehensive longitudinal data with depth and quality not previously available. Upon this infrastructure, retrospective and prospective

collection of data fit-for-purpose may be integrated into a single platform for evidence generation, capturing information on representative real-world populations with increased data depth and completeness. This integration requires a new RWD taxonomy that introduces the concept of intentional versus routine data collection (Figure 1A). Whereas observational research typically focuses on collection of routine care data, intentional collection of additional data elements and clinical activities may enable investigations not otherwise feasible with routinely-collected RWD alone. This approach may encompass more complete capture of routinely collected data elements (i.e. diagnosis date, functional status), capture of non-routine elements (i.e. adverse event grading, patient reported outcomes, genomic data), and prespecification of activities (i.e. timing of assessments), all achieved through efforts including site training, new workflows, and ability to query sites to address incomplete or missing data. As discussed below, a different approach to informed-consent may also be required.

Seamless retrospective-prospective data collection recognizes that many clinical parameters collected in clinical trials are components of routine care, thus enabling a single infrastructure for data collection and processing (Figure 1B). Research data elements and activities outside of routine care would be collected intentionally in “routine/intentional” data collection designs.

For example, a dataset to analyze genomic correlates of response to therapy would require, at a minimum, information on disease and treatments, clinical outcomes, and genomic information. While retrospective routinely collected data could provide a great deal of evidence, the availability of this information for all eligible patients may be limited. With prospective intentional data capture, missing or unclear data could be queried as needed, and samples could be collected for study-specific genomic analyses. In addition, the timing of disease assessments, e.g. CT scans, could be prespecified, thereby mitigating surveillance bias.^{10,11} Depending on the research goal, additional data sources could be layered on, including collection of patient-reported outcomes. In this way, one can dial up or down the depth of information on a fit-for-purpose basis. In addition, structured data prompts could be integrated into the EHR workflow to optimize data completeness.¹²

Key features of a new RWE generation platform

Closing the gap between retrospective RWE and the ‘controlled clinical trial’ approach requires a research infrastructure that incorporates the following critical features.

Common data model

A common data model would accelerate integration across data sources and ultimately the research process itself. Defined by shared language and a standardized set of schemas, this would eliminate the inefficiency of having

to transform data for each use across sources. Common data model features include data element names and attributes, and metadata features. It will be important to create a structural framework that maps both structured and unstructured data to the same standardized source-agnostic format. In order to address the challenges of consolidating data models, it may be wise to focus initially on discrete therapeutic disciplines. For instance, efforts to develop and promote adoption of a common oncology data model are underway.^{13,14} But, ultimately, this type of platform could be functional across disciplines and therapeutic settings, with the capability to answer multidisciplinary and holistic clinical research questions.

Data quality

Quality standards will be necessary in the integration of heterogeneous sources. All data sources, both individually and aggregated together, must be deemed relevant and reliable. Regulatory-grade RWE checklists have been proposed for characterizing and communicating about real-world data quality;¹⁵ the continued development of a data quality ‘dictionary’ will enable greater characterization of RWD relevance and quality. It will also be important to track how changing standards of care and technologies, or healthcare system disruptions (e.g. the COVID-19 pandemic), affect the quality of the data over time. This type of data stewardship entails centralization once data are collected, for quality control and curation purposes.

Analytic guidance

RWD is generally collected with research as a secondary goal, therefore analytic challenges related to data quality (including missingness) and bias are to be expected¹⁶ and appropriate methodological approaches are needed to mitigate their impact. Strategies such as simulation studies or sensitivity analyses can be included in instances where external factors may threaten the internal validity of a dataset. Scientific questions and hypotheses must be transparent, with pre-specification as a guiding principle for cohort selection and analytical plans.¹⁷ Throughout this process, clinical perspective is key to maintain research relevance, configure analytic guidances, and contextualize results, both for their impact and their potential limitations.

Regulatory framework

Combining routine and intentional data collection will mandate careful consideration of the statutes, regulations and ethical principles governing the utilization of personal health information for research purposes. This type of RWE generation platform merges into a single continuum disparate scenarios, where data access for research purposes is regulated differently. Data collected intentionally under pre-specified

protocols is typically accessed under ‘informed consent’, where patients agree to the research use of fully identifiable data for project-specific purposes. On the other hand, more general privacy regulations (such as the US Health Insurance Portability and Accountability Act [HIPAA]¹⁸ and the EU General Data Protection Regulation [GDPR]¹⁹) usually dictate access terms for identifiable information collected during routine medical care. For individual studies, these considerations will be orchestrated by the institutional review board (IRB) or ethics committee approval process. Therefore, our proposed integrated RWE platform will have to (a) navigate potential transition points between routine privacy and informed-consent settings, and (b) devise technical means to differentiate data collected intentionally and routinely, possibly via metadata tags, in order to guarantee appropriate research use and interpretation. The suitability of data collection and research access from a regulatory and ethical perspective provides an additional input into whether a RWD set is appropriate for a specific investigation.

Conclusion

Digital capture of health information and technologies that enable efficient curation have created an opportunity to develop a comprehensive data platform that can draw from the strengths of both retrospective and prospective data collection -- merging representativeness and scale with clinical depth and standardization. The seamless integration of clinical data from routine clinical care, plus intentionally-collected incremental information related to specific research questions, will help bring clinical practice and research closer together and enhance the ability of clinicians and patients to participate in research. From a logistical vantage point, the development of new capabilities to process, analyze, and optimize quality control of RWD is critical to achieve this goal. A number of multidisciplinary challenges will have to be addressed by clinicians, public health authorities, biomedical researchers, technology companies, the pharmaceutical industry, and patients. Maintaining clinical, scientific, and ethical safeguards that ensure optimal patient care and privacy protection is paramount, and will require proper research and regulatory standards. This new integrated RWE generation platform has the potential to increase the availability, depth, representativeness, and the efficiency of clinical research, and ultimately improve patient care.

Acknowledgements: The authors wish to acknowledge Julia Saiz from Flatiron Health, for editorial support.

Declaration of conflicting interests: The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: All authors report employment at Flatiron Health, Inc., which is an independent subsidiary of the Roche group, equity interest in Flatiron Health, Inc., and stock ownership in Roche.

Funding: The author(s) received no financial support for the research, authorship and/or publication of this article.

Contributorship: All authors participated in the conception of the article, in writing and reviewing and gave final approval for the submitted manuscript.

Ethical approval: NA.

Guarantor: AB.

Informed consent: Not applicable, because this article does not contain any studies with human or animal subjects.

ORCID ID: Ariel B. Bourla  <https://orcid.org/0000-0002-9838-0544>

Trial registration: Not applicable, because this article does not contain any clinical trials.

References

1. US Food & Drug Administration. Framework for FDA’s real-world evidence program, <https://www.fda.gov/media/120060/download> (2020, accessed 29 January 2020).
2. US Food & Drug Administration. Real world evidence, <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence> (accessed 29 January 2020).
3. Khozin S, Blumenthal GM and Pazdur R. Real-world data for clinical evidence generation in oncology. *J Natl Cancer Inst* 2017; 109: djx18. DOI: 10.1093/jnci/djx187
4. Wedam S, et al. FDA Approval summary: palbociclib for male patients with metastatic breast cancer. *Clin Cancer Res* 2019; 26: 1208–1212.
5. Booth CM, Karim S and Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. *Nat Rev Clin Oncol* 2019; 16: 312–325.
6. US Food & Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices, <https://www.fda.gov/media/99447/download> (accessed 20 May 2020).
7. Duke Margolis Center for Health Policy. Characterizing RWD quality and relevancy for regulatory purposes, https://healthpolicy.duke.edu/sites/default/files/atoms/files/characterizing_rwd.pdf (accessed 25 May 2020).
8. Rothwell PM. External validity of randomised controlled trials: “to whom do the results of this trial apply?”. *Lancet* 2005; 365: 82–93.
9. Miksad RA, Samant MK, Sarkar S, et al. Small but mighty: the use of real-world evidence to inform precision medicine. *Clin Pharmacol Ther* 2019; 106: 87–90.
10. Kapetanakis V, Prawitz T, Schlichting M, et al. Assessment-schedule matching in unanchored indirect treatment comparisons of progression-free survival in cancer studies. *Pharmacoeconomics* 2019; 37: 1537–1551.
11. Adamson BJ, Ma X, Griffith SD, et al. Differential frequency in imaging-based outcome measurement: bias in real-world oncology

- comparative-effectiveness studies. *Pharmacoepidemiol Drug Saf* 2021; 1–9. <https://doi.org/10.1002/pds.5323>
12. Bertagnolli MM, Anderson B, Quina A, et al. The electronic health record as a clinical trials tool: opportunities and challenges. *Clin Trials* 2020; 17: 237–242.
 13. Snyder JM, Pawloski JA and Poisson LM. Developing real-world evidence-ready datasets: time for clinician engagement. *Curr Oncol Rep* 2020; 22: 45.
 14. mCODE™: Minimal Common Oncology Data Elements, <https://mcodeinitiative.org/> (accessed 29 January 2020).
 15. Miksad RA and Abernethy AP. Harnessing the power of real-world evidence (RWE): a checklist to ensure regulatory-grade data quality. *Clin Pharmacol Ther* 2018; 103: 202–205.
 16. Collins R, Bowman L, Landray M, et al. The magic of randomization versus the myth of real-world evidence. *N Engl J Med* 2020; 38: 674–678.
 17. Girman CJ, et al. Pre-study feasibility and identifying sensitivity analyses for protocol pre-specification in comparative effectiveness research. *J Comp Eff Res* 2014; 3: 259–270.
 18. U.S. Department of Health & Human Services, Office of the Assistant Secretary for Planning and Evaluation. Health Insurance Portability and Accountability Act of 1996, <https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996> (accessed 29 January 2020).
 19. Official Journal of the European Union. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=E> (2016, accessed 29 January 2020).
-