

# ADVANCES IN DRUG DEVELOPMENT

Current Developments in Oncology Drug Research

Section Editor: Mark J. Ratain, MD

## Opportunities for Using Big Data to Advance Cancer Care



Neal J. Meropol, MD  
Vice President of Research Oncology  
Flatiron Health  
New York, New York

### H&O Could you please define big data?

**NM** In cancer care, big data commonly refers to a large patient-level data source collected for some other purpose, such as medical records or insurance claims. The data elements are pooled or processed to glean new insights. Big data derived from electronic health records is a source of “real-world evidence.” Real-world evidence is clinical evidence generated from de-identified real-world data sets collected as a part of routine care rather than through a prospective randomized clinical trial. A legislative mandate, driven by the 21st Century Cures Act, requires the US Food and Drug Administration (FDA) to consider how to incorporate real-world evidence into their decision-making. This mandate, along with the proliferation of electronic health records, has accelerated progress in the development of ways to generate, analyze, and apply real-world evidence.

There is a huge and growing demand for evidence in cancer drug development today. Although clinical trials are vital to the drug development process, they have certain limitations. For example, they are expensive and lengthy. Few patients enroll, and those who do are not necessarily representative of the broader population. Molecular characteristics are now being used to categorize cancers into subgroups too small to sufficiently populate a randomized study. Furthermore, rapid FDA approvals of exciting new treatments are associated with mounting postapproval commitments. The enormous evidence gap cannot be met via the historical approach to evidence development, which focused solely on prospective clinical trials. Analysis of big data can generate evidence to address many of the questions now arising in cancer care and therapeutic development.

With real-world evidence, it is possible to learn

from every patient by using information obtained during routine care. This approach requires a way to make sense of both structured and unstructured data that reside in electronic health records. Broadly speaking, the care of nearly all cancer patients is documented in an electronic health record. However, these electronic health records were not designed for research purposes, and much of the crucial information needed for research is found in physician narratives, radiology reports, biomarker test reports, and other unstructured formats that are not easily pooled and processed. In order to make sense of information in electronic health records, it is necessary to use both the structured data—such as height, weight, and chemotherapy regimens—and the unstructured data that are locked in the electronic health record. Historically, the need for researchers to read and interpret these records has presented a challenge in terms of scalability.

At Flatiron Health, we use a technology-enabled abstraction process to generate real-world evidence. Technology-enabled abstraction can efficiently locate particular data in an electronic medical record, and then tee it up to human abstractors, including nurses and certified tumor registrars, who extract the information. With this approach, it is possible to process both structured and unstructured data. The structured data must also be processed through a harmonization and normalization process. We provide centralized training for our abstraction workforce, and employ a formal quality oversight process. They populate models with clean data elements used to answer research questions. Our approach toward turning electronic health records into an evidence source can be thought of as a manufacturing process: we manufacture research-grade (and ultimately, what we call “regulatory-grade”) de-identified real-world data sets from electronic health records to test hypotheses and generate evidence.

## H&O What kinds of technologies are allowing the consolidation and analysis of real-world evidence?

**NM** The key element in this process is the ability to perform high-quality chart abstraction at scale on hundreds of thousands of patients. This requires technology to improve efficiency and to provide oversight and monitoring. Machine-learning technology can be used to help identify patients and designate cohorts for research. Machine-learning approaches help us efficiently build and populate data models for conducting research. Another technological advance that enables the use of real-world evidence is the development of statistical methodologies to account for potential biases and unmeasured variables that might skew the results of observational research. This possibility has always raised concerns about basing clinical decisions on retrospective real-world evidence. Careful planning and validation of the analytic approach can justify greater confidence in the meaning of the results.

## H&O How can real-world evidence be used for clinical trial design?

**NM** Real-world evidence can assist the development of eligibility criteria that reflect patients in clinical practice, thereby enhancing accrual and generating results that are applicable to the overall population. It is well-known that patients who enter clinical trials are systematically different from those treated in the real world. Real-world data confirm that patients treated in clinical trials are younger and have fewer comorbidities than patients in the real world. This discrepancy hinders generalizability of trial data to patients treated in the clinic.

Real-world evidence can also inform the statistical design for a clinical trial of a new therapy by showing typical outcomes for patients treated with standard therapy. By modeling the outcomes in such a “control” population, retrospective real-world data could inform sample size and power calculations.

Real-world evidence can also be used to help design a clinical trial’s follow-up plan and assessment schedule so that it is more aligned with routine clinical practice. For example, some trials mandate a schedule of computed tomography scans that does not match the intervals followed in clinical practice. If the trial’s design reflects clinical practice, then it will be easier to apply the results to the real-world care of patients. Using de-identified real-world data, we have been able to model the impact of different assessment schedules on the outcomes of randomized clinical trials. We are also exploring scenarios in which contemporaneous real-world evidence might be used to supplement or potentially even replace the control arm of a clinical trial that is studying a rare population of patients, where randomization may be infeasible.

## H&O Can real-world evidence be used in other aspects of drug development?

**NM** Real-world evidence is already being used to provide supplemental information to regulatory bodies to support expanded indications for approved agents. In addition, there is potential for fulfilling postmarketing commitments related to safety or efficacy of new drugs. Regulatory bodies are increasingly recognizing the potential for real-world evidence to accelerate drug development. There are various examples in which the FDA and regulatory authorities worldwide have used real-world data to support approval decisions.

## H&O What are some of those examples?

**NM** In the United States, alectinib (Alecensa, Genentech) is approved by the FDA for patients with anaplastic lymphoma kinase–positive metastatic non–small cell lung cancer. Overseas, health technology authorities required supportive evidence of the benefit of this therapy for reimbursement purposes. Flatiron Health generated data for outcomes of patients treated with the standard therapy at the time to use as a comparator against data reported in single-arm clinical trials of alectinib. The standard-treatment real-world evidence supported the benefit of alectinib. These data led to expanded access to alectinib in more than 20 countries across Europe almost a year earlier than they had anticipated, while a randomized study was maturing.

Flatiron de-identified real-world data have also been used to provide information on populations of patients who were excluded from clinical trials, such as those with organ dysfunction, to show that a treatment would be safe and effective in clinical use.

## H&O How can the accuracy of real-world evidence be improved?

**NM** An intense focus on quality control is necessary to ensure the validity of insights derived from real-world data. This includes policies and procedures that govern data collection, processing (in the case of data derived from electronic health records), and oversight, as well as carefully designed a priori analytic plans.

As another use case, the combination of molecular data with clinical outcomes data enables the identification of resistance mechanisms as well as the discovery or validation of predictive biomarkers for response to therapy. This requires the accurate measurement of patient outcomes in the real world. Typically, real-world data sources do not provide information on computed tomography scans or changes in tumor burden as assessed by the Response Evaluation Criteria in Solid Tumors (RECIST), as might

be available in a clinical trial. We are therefore developing so-called real-world endpoints, such as tumor progression, based on physician documentation. This requires careful validation in a context-specific way to ensure that real-world endpoints are closely associated with clinical outcomes of importance to patients.

One example concerns mortality, which is the gold standard for clinical benefit. Unfortunately, the date of death is frequently missing from electronic health records. We therefore pursued a detailed analysis of the mortality endpoint in real-world data. By supplementing information from electronic health records with mortality data obtained from external databases, we improved sensitivity and specificity of the mortality variable to greater than 90%. This type of validation is required to have confidence in the evidence derived from real-world data.

### H&O What is the role of the physician in the accumulation and analysis of real-world evidence?

**NM** The quality of real-world data is dependent, in part, on the quality of physician documentation. The more information that is contained as structured data elements, the better. The challenge is that it is not possible—nor beneficial—to require physicians to enter structured data into the medical record if it interferes with the workflow of caring for patients. Therefore, there is a need for new methods to improve the completeness and accuracy of data entered into the electronic health record in a way that is standardized across platforms and does not interfere with physician workflow.

### H&O What are some challenges in determining how to best use real-world evidence?

**NM** As I noted earlier, the credibility of real-world evidence requires an intense focus on data quality. Critical features include completeness and accuracy of the data elements, representativeness of the population of interest, recency of the data, and completeness of clinical follow-up. Validation of each real-world clinical endpoint is also necessary to ensure that outcomes data are credible.

### H&O Can physicians use real-world evidence in clinical practice?

**NM** At the macro level, as results from real-world evidence studies are presented and published, they will provide new evidence to assist physicians with treatment decisions. There is also a lot of interest in the use of real-world evidence to provide decision support at the point

of care. However, it is necessary to proceed with caution when developing point-of-care tools that aggregate real-world evidence. The quality of the output must be ensured before it is used to guide treatment for individual patients.

### H&O What are some other potential opportunities for real-world evidence in oncology?

**NM** Real-world evidence can be used to conduct analyses of the value of diagnostic tools and treatment interventions in routine practice. In the future, real-world evidence could help predict outcomes, such as short-term survival or likelihood of hospitalization; identify the likelihood of benefit and adverse events with specific therapies; and select treatments most likely to benefit individual patients.

### H&O Are there any concerns regarding privacy?

**NM** This issue should be top of mind for anyone conducting research leveraging patient data. The right to privacy is a fundamental human right. That is why it is so critical that any use of patient information for research purposes is undertaken with great care to protect patient privacy, and that steps are taken to earn and retain patient trust. Engaging patients and the patient advocacy community in our work is critical.

#### Disclosure

*Dr Meropol is an employee of Flatiron Health, an independent subsidiary of the Roche Group.*

#### Suggested Readings

Agarwala V, Khozin S, Singal G, et al. Real-world evidence in support of precision medicine: clinico-genomic cancer data as a case study. *Health Aff (Millwood)*. 2018;37(5):765-772.

Curtis MD, Griffith SD, Tucker M, et al. Development and validation of a high-quality composite real-world mortality endpoint [published online May 14, 2018]. *Health Serv Res*. doi:10.1111/1475-6773.12872.

Khozin S, Abernethy AP, Nussbaum NC, et al. Characteristics of real-world metastatic non-small cell lung cancer patients treated with nivolumab and pembrolizumab during the year following approval. *Oncologist*. 2018;23(3):328-336.

Miksad RA, Abernethy AP. Harnessing the power of real-world evidence (RWE): a checklist to ensure regulatory-grade data quality. *Clin Pharmacol Ther*. 2018;103(2):202-205.

O'Connor JM, Fessele KL, Steiner J, et al. Speed of adoption of immune checkpoint inhibitors of programmed cell death 1 protein and comparison of patient ages in clinical practice vs pivotal clinical trials. *JAMA Oncol*. 2018;4(8):c180798. doi:10.1001/jamaoncol.2018.0798.

Presley CJ, Tang D, Soulos PR, et al. Association of broad-based genomic sequencing with survival among patients with advanced non-small cell lung cancer in the community setting. *JAMA*. 2018;320(5):469-477.