JNCI J Natl Cancer Inst (2017) 109(11): djx187

doi: 10.1093/jnci/djx187
First published online September 13, 2017
Commentary

COMMENTARY

Real-world Data for Clinical Evidence Generation in Oncology

Sean Khozin, Gideon M. Blumenthal, Richard Pazdur

Affiliations of authors: Oncology Center of Excellence, Food and Drug Administration, Silver Spring, MD

Correspondence to: Sean Khozin, MD, MPH, Food and Drug Administration, Oncology Center of Excellence, 10903 New Hampshire Ave., Building 22 RM2322 Silver Spring, MD 20993 (e-mail: sean.khozin@fda.hhs.gov).

Abstract

Conventional cancer clinical trials can be slow and costly, often produce results with limited external validity, and are difficult for patients to participate in. Recent technological advances and a dynamic policy landscape in the United States have created a fertile ground for the use of real-world data (RWD) to improve current methods of clinical evidence generation. Sources of RWD include electronic health records, insurance claims, patient registries, and digital health solutions outside of conventional clinical trials. A definition focused on the original intent of data collected at the point of care can distinguish RWD from conventional clinical trial data. When the intent of data collection at the point of care is research, RWD can be generated using experimental designs similar to those employed in conventional clinical trials, but with several advantages that include gains in efficient execution of studies with an appropriate balance between internal and external validity. RWD can support active pharmacovigilance, insights into the natural history of disease, and the development of external control arms. Prospective collection of RWD can enable evidence generation based on pragmatic clinical trials (PCTs) that support randomized study designs and expand clinical research to the point of care. PCTs may help address the growing demands for access to experimental therapies while increasing patient participation in cancer clinical trials. Conducting valid real-world studies requires data quality assurance through auditable data abstraction methods and new incentives to drive electronic capture of clinically relevant data at the point of care.

Real-world data (RWD) is a general term that can be described as data generated or obtained outside of conventional clinical trials. A wide range of data elements can be captured in the real-world setting, including variables on the individual (eg, patient demographics, physical, and physiologic parameters), the environment, and clinical outcomes (eg, survival and tumor dynamics such as response rate). Sources of real-world data include insurance claims, patient registries, electronic health records (EHRs), patient health records, and digital health solutions such as mobile applications and devices, including those with sensor capabilities (eg, gyroscopic accelerometers). A definition that focuses on the original intent behind collection of clinical data can be used to characterize the core characteristics of RWD (Table 1). Using this framework, purposeful collection of data at the point of care for research, rather than routine delivery of health care services, narrows the gap between real-world and conventional clinical trial data, providing a foundation for optimal experimental designs that include randomization at the point of care.

The major contributing factor to the emergence of RWD as a viable source of clinical evidence has been the recent acceleration in the use of EHRs. As direct conduits into point of care activities and transactions, EHRs are practical and scalable tools for data collection. The driving force behind increased adoption of EHRs in the United States was the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act, enacted as part of the American Recovery and Reinvestment Act to promote the adoption and meaningful use of interoperable health information technology systems. HITECH facilitated adoption of EHRs by making investments in incentive programs for health care providers and hospitals (1). By 2014, adoption of EHRs by office-based physicians nearly doubled. Similarly, the

Received: April 24, 2017; **Revised:** June 28, 2017; **Accepted:** August 8, 2017

Table 1. The intended use of point-of-care data at the time of collection is the primary feature informing potential use cases of real-world data for clinical evidence generation*

Intended use of data at the time of collection	Primary sources of data	Potential use cases	Challenges
Delivery of routine health care services	EHRs and PHRs Insurance claims Patient registries Digital health solutions	Development of external control Studying the natural history of disease Postmarket pharmacovigilance Hypothesis generation to support design of prospective clinical trials	Can primarily support retrospective analyses Limited availability of clinically relevant structured data elements in EHRs Extraction of data from unstructured content (eg, physician notes and diagnostic reports) is resource intensive Requires special procedures for assurance of data quality
Research	EHRs and PHRs Digital health solutions	All of the above plus: Prospective pragmatic clinical trials that support randomization and other experimental design principles employed in conventional clinical trials	Creation of new incentives for capturing clinically relevant structured data elements at the point of care Providing appropriate training for community oncologists to ensure adherence to ethical, regulatory, and legal standards in conducting clinical research

*EHR = electronic health record; PHR = patient health record.

proportion of nonfederal acute care hospitals reporting EHR use increased from 13% in 2009 to 76% in 2014 (2).

Leveraging RWD has been of great interest to the US Food and Drug Administration (FDA). In 2008, the Agency launched the Sentinel Initiative in response to the FDA Amendments Act (FDAAA) calling for creation of an active surveillance system for monitoring the safety of approved drugs and medical products. Working with several data partners (including insurers, universities, and hospitals), important surveillance reports have been generated using data contained in Sentinel's network of claims and EHR-related content (3). More recently, the FDA's Office of Oncology and Hematology Products launched the Information Exchange and Data Transformation (INFORMED) initiative, a multidisciplinary effort that focuses on building technical and organizational infrastructure in several key areas of big data analytics to explore new pipelines of data from sources such as EHRs and digital health solutions in making regulatory decisions (4). The enactment of the 21st Century Cures Act in 2016 has paved a new path for the use of RWD to support new product indications and postapproval requirements (5).

Opportunities for Using Real-world Evidence in Cancer Drug Development

Existing FDA regulations offer adequate flexibility for the use of emerging sources of clinical evidence in making regulatory decisions (6). As a result, efforts for using real-world data should focus on the development of appropriate study designs and strategies for acquisition of high-quality data from EHRs. The experimental principles that have shaped the basic tenets of biomedical research can be used to assess the role of point-of-care data in clinical evidence generation and guide the design of valid real-world studies (7).

Real-world Evidence for Digital Pharmacovigilance

As defined by the World Health Organization, pharmacovigilance encompasses the science and activities related to the detection, assessment, understanding, and prevention of adverse effects and other drug-related problems (8). The primary mechanism employed by regulatory agencies to support postmarket pharmacovigilance is based on passive surveillance through analysis of voluntary reports of adverse events by health care professionals and patients, in addition to required reporting by pharmaceutical companies (9). Passive reporting has several limitations and can be influenced by factors unrelated to the intrinsic effects of a product such as media attention and the length of time the product has been on the market. Establishment of the Sentinel Initiative by the FDA was driven by the need to create an active surveillance system where RWD can be proactively interrogated for detection of new safety signals (10,11). The rapid growth in information technology solutions in recent years provides an opportunity for the development of an integrated approach based on RWD from EHRs and patient-generated sources such as mobile applications and internet search logs for bringing pharmacovigilance into the digital age (12,13). In randomized clinical trials, web-based tools designed to capture self-reported symptoms have been shown to improve survival in patients with advanced cancers undergoing treatment, highlighting the value of a digital framework in identification of adverse events and triggering mitigation strategies to prevent downstream adverse consequences (14,15). A digital pharmacovigilance system linking RWD (generated by health care professionals through EHRs and by individual patients through web and mobile applications) to clinical investigators and pharmacovigilance scientists in industry and regulatory agencies can support an active surveillance system capable of rapid deployment of mitigation strategies once a valid safety signal is detected. The validity of safety signals in a digital pharmacovigilance system can be determined using data mining techniques, such as proportional reporting ratios and empirical Bayesian geometric mean scores already employed by the FDA and other regulatory agencies (16,17). A digital framework that includes heterogeneous pipelines of real-world data can also take advantage of deep learning approaches based on artificial intelligence and natural language processing to supplement insights gained from traditional data mining methods with computational reasoning to improve safety signal detection practices (18,19,20).

Real-world Evidence as Means of Studying the Natural History of Disease and Development of External Controls

Applications supporting FDA approval of a new drug or biologic product are required to contain data from adequate and well-controlled clinical investigations (21). Appropriate design of such studies relies on making accurate assumptions about the natural history of the disease under investigation. The natural history of disease depicts the course of a disease from the time right before its inception to the presymptomatic phase, the different clinically symptomatic stages, and the point where the patient is cured, chronically disabled, or dead without external intervention (22). One way of describing this continuum is by identifying the impact of two types of covariates that influence the risk of developing asymptomatic disease from a healthy state (type 1) and progression to the symptomatic stage of disease (type 2) (23). Our current understanding of the natural history of disease is largely informed by studies at academic referral centers, despite the fact that most patients are treated in community-based medical practices (24). RWD provides an opportunity to study the covariates that influence the natural history of disease in settings where the majority of the population is under routine monitoring and treatment. For example, real-world analyses of type 1 covariates responsible for increasing the risk of cancer in healthy individuals can be done through retrospective extraction of EHR data to characterize patient and environmental factors influencing disease occurrence. Retrospective examination of EHR data for characterization of type 2 covariates can similarly aid in identifying factors influencing progression of cancer into the symptomatic stage. These types of studies can inform the design of prospective clinical trials assessing the impact of cancer screening and early intervention on patient outcomes.

In clinical study designs such as single-arm trials, the use of external control data can potentially support the development of comparative benchmarks for regulatory decision-making, especially for serious conditions of high unmet medical need such as advanced malignancies (25). If preliminary clinical evidence for an anticancer agent in a single-arm trial suggests a substantial treatment effect (for example, tumor response rates of large magnitude and duration), evaluating outcomes in similar groups of patients using RWD may provide a reliable assessment of the safety and effectiveness of available therapies for comparison. Anticancer therapies under the breakthrough therapy designation program are especially appropriate candidates for considering alternative trial designs that may allow the use of real-world-derived external control data as primary or supportive evidence for regulatory decisions (26,27).

Progress in genomic sequencing and computational proteomics is uncovering increasing numbers of rare tumor variants based on somatic mutations, proteomic signatures, and cell signaling pathway alterations of oncogenic potential (28,29,30,31). When clinical development goals require delineation of disease outcomes in these rare subsets, retrospective RWD analyses may be the most practical way of understanding the prognostic implications of the biomarkers of interest. Therefore, linking clinical outcomes in RWD repositories to genomic and proteomic profiles for the prognostication and development of external control benchmarks is of critical importance and calls for increasing capacity in big data analytics for optimal interpretation of rare and complex signatures defined by the computational outputs of multiomic pipelines.

Observational Real-world Studies

In cancer drug development, randomized clinical trials (RCTs) are the gold standard for establishing causal relationships and evaluating the efficacy and safety of new therapies, primarily as a result of widely implemented procedures governing the design and conduct of clinical research. Emerging evidence suggests similarities between the results of RCTs and well-designed observational studies, signaling an opportunity for building robust methodologies in support of EHR-based observational research (32,33,34). The results of observational studies can generate new hypotheses that can be tested in randomized clinical trials or used as supportive evidence in regulatory decision-making. Observational real-world studies can accommodate assessment of safety and effectiveness of therapies in patients that are excluded in conventional cancer trials, such as those with poor performance status, history of prior malignancies, organ dysfunction, or brain metastases. New regulatory incentives for drug developers for submission of RWD on patients that are typically excluded in conventional clinical trials can improve the generalizability of the information on FDA labels, helping prescribers better tailor their treatment decisions (35).

Pragmatic Clinical Trials

Prospective real-world studies are similar to pragmatic (also known as practical) clinical trials (PCTs), clinical studies that are designed to produce results that uniquely support clinical decision-making at the point of care (36,37). As the primary instruments for conducting prospective PCTs, EHRs are widely available conduits into the health care delivery system. The chain of technological innovations such as structured documentation and practice management tools in modern EHRs can power a clinical trial enterprise anchored at the point of care and digitally connected to patients through emerging technological solutions such as sensors and mobile applications. With EHRs supporting the purposeful collection of clinically relevant data reflecting the true diversity of cancer patients, we can bring the real-world evidence base to drug development while driving the focus on improving quality, patient safety, and value in cancer care delivery (38).

In cancer drug development, PCTs have potential advantages over conventional clinical trials that are typically confined to specialized centers with adequate resources and economy of scale to maintain research programs. The low rates of participation in cancer trials (<5%), especially for minorities, the elderly, low-income individuals, and those living in rural areas, are clear indications of the barriers posed by segmentation of clinical research to geographically dispersed sites (39,40,41,42). Indeed, the recent right-to-try debates do suggest that barriers to gaining convenient access to experimental therapies, rather than patient preferences, are the prohibitive force behind low participation in cancer trials (43,44). PCTs allow community oncologists to assume an active role in clinical research, especially for late-phase studies where highly controlled experimental conditions for dose finding or drug mechanism explorations are not needed. PCTs can maintain existing standards in methodological, ethical, legal, and regulatory oversight of clinical research while increasing access to experimental therapies in a safe and efficient manner. For community oncologists, the incentives provided by sponsors of PCTs can balance the demands of the additional time devoted to EHR-based structured data entry.

Table 2. Potential sources of bias in real-world studies threatening internal validity

Sources of bias	Individual	Technology	System
Arising from	Patient-provider dynamics and patient characteristics	EHRs	Trends and influences on the health care system
Primary type(s)	Information bias* influencing accuracy of data collection (recall, observer/interviewer, and reporting bias)	Information bias* due to variations in EHR interfaces, data entry procedures, or data retrieval methods leading to compromising data quality	Selection bias due to variation in access to care affecting sampling frame
	Confounding bias† due to patient characteristics and comorbidities	Selection bias§ arising from selection of patients using EHR diagnostic and therapeutic codes	Confounding bias† due to regional variations in standards of care or available therapies due to third-party formularies
	Compliance bias‡ due to patient nonadherence to treatment		

*Information bias: erroneous or inaccurate capture of patient variables. EHR = electronic health record.
†Confounding bias: association between treatment and outcome being influenced by the presence of extraneous variables.
‡Compliance bias: variations in patient adherence to planned treatment affecting study outcomes.
§Selection bias: study population not representative of the true distributions in the overall population.

COMMENTARY

Assessing Threats to Internal Validity of Real-world Clinical Studies

The existing controls that govern the conduct of clinical research, in particular RCTs, focus on creating ideal conditions supporting assumptions and operations that reduce bias and augment the internal validity of the studies in order to minimize alternative explanations of treatment effects. However, factors such as narrow eligibility criteria and differences between protocol-specified patient care and routine medical practice have led to deficits in the external validity of conventional clinical trials (45,46). A balanced approach based on prospective collection of RWD can protect against common threats to internal validity while increasing external validity of clinical research to facilitate evidence-based decisions at the point of care with greater precision (46,48).

Establishing internal validity requires careful assessment of the design and conduct of clinical studies. This assessment is done logically, as opposed to statistically, through detailed examination of the study design and gauging adherence to the execution of study procedures (9). In a typical clinical study report, the materials and methods section outlines the technical and methodological components of the study design aimed at protecting internal validity, while the discussion section draws conclusions on the potential sources of bias influencing interpretation of the study results (49). For clinical studies intended to support regulatory decisions, quality in execution of study procedures is explicitly stated by the study sponsor's attestation of adherence to good clinical practice guidelines (50). Additionally, the FDA's regulatory review process contains several measures directed at validating appropriate conduct of clinical research, including site inspections for source document verification and random audits on data sets to verify accuracy. These well-established procedures contribute to the FDA's logical framework for evaluating threats to internal validity and confirming the integrity of clinical research studies. For real-world studies intended to support regulatory decisions, a similar thread of information can accommodate a comparable approach for assessing internal validity and estimating the influence of extraneous factors on treatment effects. Ensuring the internal validity of real-world studies, in particular in nonrandomized designs, requires controlling for the sources of bias arising from provider-patient dynamics, data collection and processing techniques, and variations in practice patterns in regional health care systems (Table 2).

Discussion

The conventional clinical trial enterprise leans heavily toward producing internally valid clinical studies, often at the expense of compromising external validity. This has resulted in uncertainties in making evidence-based individualized treatment decisions and excessive reliance on clinical judgment, which can introduce variations in practice patterns and cancer care quality. Furthermore, fragmentation of clinical research to geographically dispersed sites poses a significant barrier to cancer clinical trial participation, despite growing demands for patient access to experimental therapies.

Advances in health information technologies and a dramatic increase in the adoption of EHRs have created new opportunities for optimizing and streamlining clinical evidence generation through collection of RWD. Applying the principles that have shaped the theoretical foundation of biomedical research to the experimental design of real-world studies can result in an appropriate balance between the internal and external validity of clinical research, enabling more individualized treatment decisions at the point of care. Prospective PCTs conducted at the point of care are particularly beneficial for increasing patient participation in clinical research, bringing new efficiencies to cancer drug development, and improving the evidentiary standards used for making decisions regarding the safety and effectiveness of therapies.

The shortfalls of existing EHRs have been widely discussed because most systems were primarily designed to support billing and practice management activities as opposed to clinical research (51,52). Clinically relevant information is often hard to retrieve, buried deep in unstructured content such as physician notes and diagnostic reports. Despite these deficits, methods based on natural language processing and technology-enabled abstraction are producing reliable EHR data sets for clinical investigation that can be used today to support evidence generation (53,54).

As we carve a path toward the use of real-world evidence as a means of modernizing clinical research and evidence

generation, more focus should be placed on identifying and addressing organizational barriers. For example, purely technical solutions aimed at increasing interoperability among EHRs are critical for efficient health information exchange, but they are not sufficient for solving the challenges of data capture at the point of care or leveraging patient-generated data on mobile applications and devices. Building a scalable framework for broadening the use of real-world evidence requires organizational support driven by appropriate incentives for busy oncologists to provide clinically relevant data at the point of care and for patients to benefit from sharing their data generated on digital health platforms outside of the traditionally defined boundaries of the health care system.

References

1. Blumenthal D. Launching HITECH. *N Engl J Med*. 2010;362(5):382–385.
2. Charles D, Gabriel M, Searcy T. Adoption of electronic health record systems among U.S. non-federal acute care hospitals: 2008-2014. Office of the National Coordinator for Health Information Technology. 2016. https://www.healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf. Accessed April 24, 2017.
3. Food and Drug Administration. Sentinel initiative, a national product monitoring system. 2017. http://www.mini-sentinel.org/Reports. Accessed April 24, 2017.
4. Khozin S, Kim G, Pazdur R. From big data to smart data: FDA's INFORMED initiative. *Nat Rev Drug Discov*. 2017;16(5):306.
5. United Stated Congress. 21st Century Cures Act. 2016. https://www.congress.gov/bill/114th-congress/house-bill/34/text#tocHBC0DF02539024F65AEF0DD6A88063188. Accessed April 24, 2017.
6. Sherman RE, Davies KM, Robb MA, Hunter NL, Califf RM. Accelerating development of scientific evidence for medical products within the existing US regulatory framework. *Nat Rev Drug Discov*. 2017;16(5):297–298.
7. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. 2nd ed. Boston, MA. Houghton Mifflin Comp; 1967.
8. Härmark L, van Grootheest AC. Pharmacovigilance: Methods, recent developments and future perspectives. *Eur J Clin Pharmacol*. 2008;64(8):743–752.
9. Edwards IR. Pharmacovigilance. *Br J Clin Pharmacol*. 2012;73(6):979–982.
10. Food and Drug Administration. Sentinel initiative. 2017. http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm149340.htm. Accessed April 24, 2017.
11. Food and Drug Administration. Sentinel initiative: A national strategy for monitoring medical product safety. 2008. http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm089474.htm. Accessed April 24, 2017.
12. White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clin Pharmacol Ther*. 2014;96(2):239–246.
13. Salathé M. Digital pharmacovigilance and disease surveillance: Combining traditional and big-data systems for better public health. *J Infect Dis*. 2016;214(suppl 4):S399–S403.
14. Denis F, Lethrosne C, Pourel N, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. *J Natl Cancer Inst*. 2017;109(9):djx029.
15. Basch E, Deal AM2, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA*. 2017;318(2):197–198.
16. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10(6):483–486.
17. Dumouchel W. Bayesian data mining in large frequency tables, with an application to the FDA Spontaneous Reporting System. *Am Stat*. 1999;53(3):177–190.
18. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *JAMIA*. 2015;22(3):671–681.
19. Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: Recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *J Am Med Inform Assoc*. 2017; in press.
20. Graves A, Wayne G, Reynolds M, et al. Hybrid computing using a neural network with dynamic external memory. *Nature*. 2016;538(7626):471–476.
21. Food and Drug Administration. Code of federal regulations. 2016. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=314.126. Accessed April 24, 2017.
22. de la Paz MP, Villaverde-Hueso A, Alonso V. Rare diseases epidemiology research. *Adv Exp Med Biol*. 2010;686:17–39.
23. Brookmeyer R. Statistical problems in epidemiologic studies of the natural history of disease. *Environ Health Perspect*. 1990;87:43–49.
24. Ellenberg JH, Nelson KB. Sample selection and the natural history of disease studies of febrile seizures. *JAMA*. 1980;243(13):1337–1340.
25. Simon R, Blumenthal G, Rothenberg M, et al. The role of nonrandomized trials in the evaluation of oncology drugs. *Clin Pharmacol Ther*. 2015;97(5):502–507.
26. Friends of Cancer Research Real World Data Working Group. A blueprint for breakthrough: Exploring utility of real world evidence (RWE). 2016. http://www.focr.org/events/blueprint-breakthrough-exploring-utility-real-world-evidence-rwe. Accessed April 24, 2017.
27. Food and Drug Administration. FDA guidance for industry. Expedited programs for serious conditions. 2014. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm358301.pdf. Accessed April 24, 2017.
28. Leiserson MD, Vandin F, Wu HT, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet*. 2015;47(2):106–114.
29. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science*. 2013;339(6127):1546–1558.
30. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511(7511):543–550.
31. Kim Y, Jeon J, Mejia S, et al. Targeted proteomics identifies liquid-biopsy signatures for extracapsular prostate cancer. *Nat Commun*. 2016;7:11906.
32. Konnerup M, Kongsted HC. Are more observational studies being included in Cochrane Reviews? *BMC Res Notes*. 2012;5:570.
33. Vázquez GH, Holtzman JN, Lolich M, Ketter TA, Baldessarini RJ. Recurrence rates in bipolar disorder: Systematic comparison of long-term prospective, naturalistic studies versus randomized controlled trials. *Eur Neuropsychopharmacol*. 2015;25(10):1501–1512.
34. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with non-experimental designs compared with those assessed in randomised trials. *Cochrane Database Syst Rev*. 2014;(4):MR000034.
35. Beaver JA, Ison G, Pazdur R. Reevaluating eligibility criteria—balancing patient protection and participation in oncology trials. *N Engl J Med*. 2017;376(16):1504–1505.
36. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *J Chronic Dis*. 1967;20:637–648
37. Califf RM. Pragmatic clinical trials: Emerging challenges and new roles for statisticians. *Clin Trials*. 2016;13(5):471–477.
38. Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. *J Clin Oncol*. 2010;28(27):4268–4274.
39. Al-Refaie WB, Vickers SM, Zhong W, Parsons H, Rothenberger D, Habermann EB. Cancer trials versus the real world in the United States. *Ann Surg*. 2011;254(3):438–443.
40. Nass SJ, Moses HL, Mendelsohn J. *Physician and Patient Participation in Cancer Clinical Trials*. Washington, DC: National Academies Press; 2010.
41. Meropol NJ. Health policy: Overcoming cost barriers to clinical trial participation. *Nat Rev Clin Oncol*. 2016;13(6):333–334.
42. Meropol NJ, Buzaglo JS, Millard J, et al. Barriers to clinical trial participation as perceived by oncologists and patients. *J Natl Compr Canc Netw*. 2007;5(8):655–664.
43. Yang YT, Chen B, Bennett C. "Right-to-try" legislation: Progress or peril? *J Clin Oncol*. 2015;33(24):2597–2599.
44. Cohen-Kurzrock BA, Cohen PR, Kurzrock R. Health policy: The right to try is embodied in the right to die. *Nat Rev Clin Oncol*. 2016;13(7):399–400.
45. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials*. 2006; in press.
46. Summerskill W. Evidence-based practice and the individual. *Lancet*. 2005;365(9453):13–14.
47. Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004;58(8):635–641.
48. Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: Systematic review of recent trials. *BMJ*. 2008;336(7649):876–880.
49. Slack MK, Draugalis JR. Establishing the internal and external validity of experimental studies. *Am J Health Syst Pharm*. 2001;58(22):2173–2181.
50. The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH guidelines for good clinical practice. 1996. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf. *International Harmonization for Better Health*. Accessed April 24, 2017
51. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care*. 2013;51(8)(suppl 3):S30–S37.
52. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research. *JAMIA*. 2013;20(1):144–151.
53. Liu F, Weng F, Yu H. Natural language processing, electronic health records, and clinical research. In: RL Richesson, JE Andrews, eds. *Clinical Research Informatics*. New York, NY: Springer; 2012:293–312.
54. Ohno-Machado L. Realizing the full potential of electronic health records: The role of natural language processing. *J Am Med Inform Assoc*. 2011;18(5):539.