UCSF
UC San Francisco Previously Published Works

Title

Opportunities and challenges in using real-world data for health care

Permalink

https://escholarship.org/uc/item/4m59x7ms

Journal

Journal of Clinical Investigation, 130(2)

ISSN

0021-9738

Authors

Rudrapatna, Vivek A
Butte, Atul J

Publication Date

2020-02-03

DOI

10.1172/jci129197

Peer reviewed

Opportunities and challenges in using real-world data for health care

Vivek A. Rudrapatna[1,2] and Atul J. Butte[1,3,4]

[1]Bakar Computational Health Sciences Institute, [2]Division of Gastroenterology, Department of Medicine, and [3]Department of Pediatrics, UCSF, San Francisco, California, USA. [4]Center for Data-Driven Insights and Innovation, University of California Health, Oakland, California, USA.

Real-world data (RWD) continue to emerge as a new source of clinical evidence. Although the best-known use case of RWD has been in drug regulation, RWD are being generated and used by many other parties, including biopharmaceutical companies, payors, clinical researchers, providers, and patients. In this Review, we describe 21 potential uses for RWD across the spectrum of health care. We also discuss important challenges and limitations relevant to the translation of these data into evidence.

Introduction: the what and why of real-world data

We are now practicing medicine in a world immersed with data. Advances in computing and health information technology have given rise to new sources and types of biomedical data. The growing availability of these data and their potential for novel uses have stimulated interest from many parties across the health care delivery spectrum. The United States Food and Drug Administration (FDA) has been using the phrases "real-world data" (RWD) to mean "data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources" (1). They define "real-world evidence" (RWE) as "clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD" (1).

While RWD include case reports or retrospective observational data, this definition actually covers a broader variety of data sources. These include electronic health records (EHRs) (Table 1), administrative and claims data, registries, patient-generated data from websites and wearable sensors, measures of social determinants of health, and environmental exposures (2). Similarly, although evidence derived from observational data can be construed as uncontrolled and low-quality evidence, RWE covers a broader range of analytic designs for causal inference. These

include natural experiments, in which the assignment to the exposure of interest is made by arbitrary forces resembling a randomized trial (3).

Although the best-known uses of RWD have been for the regulation of drug safety, RWD have attracted the attention of many other participants in the health care ecosystem: biopharmaceutical companies, payors, providers, policy makers, and patients (Figure 1 and Table 2). In this Review, we evaluate their potential utility and present limitations. We proceed by highlighting seven broad categories and 21 specific applications of RWD—both existing and emerging ones. We then turn to a detailed discussion of ongoing challenges in their use. It is our belief that a broader awareness of these data can only serve to maximize their potential to improve human health at all levels.

RWD for post-approval safety

Updating side effect rates. As phase III clinical trials may not be sufficiently powered to detect clinically significant adverse events, regulatory bodies and biopharmaceutical sponsors have relied on alternative approaches to study the safety of drugs after approval. The primary method in use by the FDA has been phase IV studies: open-label and noninterventional studies that assess a larger population and longer time period than are typically studied in phase III trials. By contrast, biopharmaceutical companies have primarily relied on national registries to facilitate post-marketing studies of safety and efficacy. This has in large part been because these registries collect other data of commercial interest, such as patient/physician experience, compliance, access/utilization, reimbursement, and competitive intelligence.

One of the earliest examples of regulatory adoption of RWD has been the FDA Sentinel Initiative (4). Sentinel is a federated network established in 2008 that integrates claims, EHR, and registry data nationwide to monitor product safety. Although the FDA is the primary consumer of the Sentinel system, it is increasingly being used by other parties, including biopharmaceutical companies as well as researchers developing methods for event detection. Over time, passive surveillance systems such as Sentinel and other regulatory platforms such as MedWatch (5) may grow in
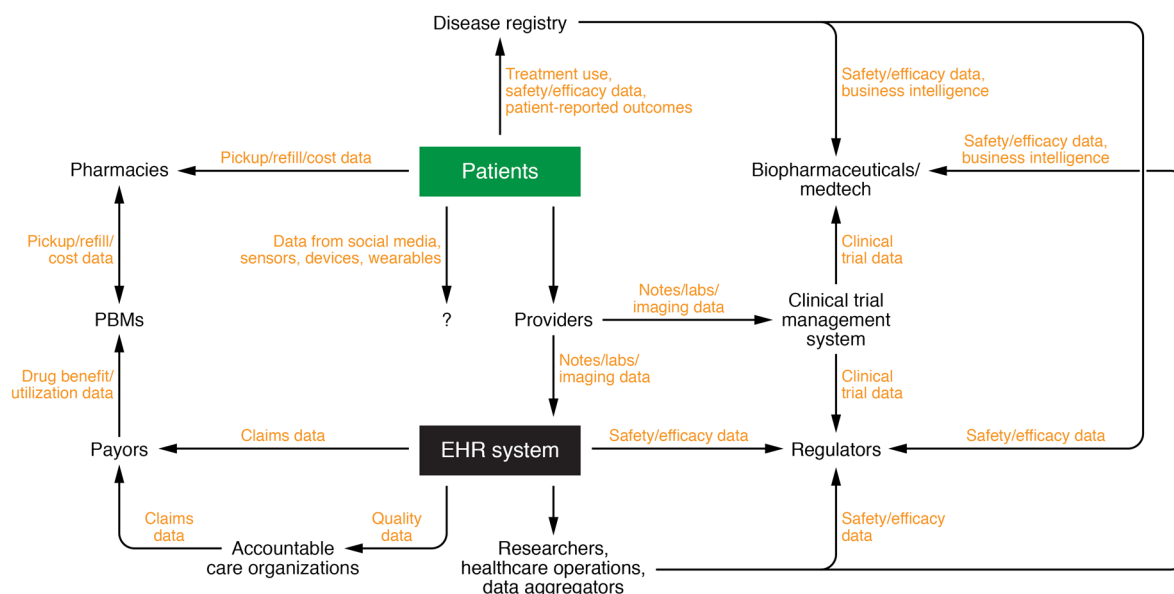
Figure 1. Participants in the health care ecosystem that generate and consume health care data. Patients (and the communities they constitute) are the fundamental source of all clinical data. Much of the clinical data they generate emanates from clinic visits with health care providers. They also generate data from the pharmacies they purchase treatments from, the registries they participate in, and their use of modern/evolving technologies such as social media and wearables. In the setting of a traditional or telehealth-based encounter, clinical data in the form of laboratory test results, imaging, and notes are all generated and housed with an EHR system. These data may be repackaged and sent to managed care organizations and health care payors to facilitate reimbursement. These payors also transmit data relevant to drug benefits to pharmacy benefit managers (PBMs), who negotiate payment for drugs dispensed in pharmacies. Quality data from the EHR are also used by accountable care organizations (ACOs) to support certain quality-based reimbursement schemes. EHR data are also consumed by clinical researchers, individuals who oversee health care operations, and data aggregators. The latter deidentify and repackage these data for consumption by a variety of parties, including biopharmaceutical companies and regulators, as relevant for monitoring of treatment safety and efficacy, among other uses. Although most patient data represent a form of RWD, some patient data are collected in the setting of controlled trials. Although data from cardiac devices (e.g., pacemakers/cardioverter-defibrillators) and glucose meters occasionally end up in the EHR, data from consumer wearables and sensors and social media are currently not integrated into EHR systems. However, these data are increasingly being studied for their potential utility in health care, and may be integrated in the future. Adapted with permission from Datavant (58).

importance as a more efficient, cost-effective, and real-time way of capturing and confirming important safety signals.

Discovering novel side effects. Many of the most valuable data residing in EHRs and many patient-generated sources (e.g., social media) exist in the form of unstructured, free-text data. These data have been more challenging to computationally manipulate than structured data fields (e.g., ICD codes for diagnoses), and as such have been left out from most RWD databases such as Sentinel (Table 1) (6). However, free-text data have several advantages over traditional, protocolized collection of structured data. First, they are more expressive and less encumbered by structured fields with rigidly defined categories. Second, they are less filtered, because this information capture may have been initiated by the patient. As a consequence, these data have the potential to convey a richer survey of unanticipated side effects. Advances in natural language processing are making free text increasingly tractable for this type of analysis.

Beyond EHR data, there may be benefits of analyzing other sources of free text, including social media, as a platform for pharmacovigilance (7). For example, one recent study demonstrated how cutaneous adverse drug reactions resulting from cancer drugs can be identified from these sources about 7 months before their reporting in the literature; it also identified new side effects not previously reported (8).

Data mining techniques on raw, unfiltered RWD also have the potential to discover other kinds of side effects. These include beneficial ones, such as the anti-TNF effect of bupropion in the setting of Crohn's disease (9, 10). They also have the potential to identify unanticipated side effects resulting from combinations of two or more drugs that are difficult to predict during drug discovery and development, yet may be seen in larger cohorts through EHR data (11).

Although observational data play an essential role in the detection of adverse events, confirmatory randomized controlled trials (RCTs) are still necessary, especially in scenarios where prior studies are conflicting. For example, trials of widely used antidiabetic medications were needed to confirm serious risks such as those of bone fractures (12, 13) and amputations (14).

RWD to support regulatory approval

Single-arm experimental trials. Although placebo-controlled and double-blinded RCTs unambiguously represent the gold standard for clinical evidence, this ideal is frequently impractical. As one example, for rare and/or deadly diseases, it can be difficult to ethically justify randomizing some subjects to a no-treatment arm. Trials can also prove too expensive, are difficult to recruit for, and can extend the regulatory timeline.

The challenges associated with obtaining these gold-standard sources of evidence have prompted many to explore expeditious

Table 1. Acronyms and terms used in this Review

Term	Definition
Randomized controlled trial (RCT)	A study design whereby subjects are randomly assigned to exposures of interest (including controls) and assessed for outcomes of interest. This is the gold-standard clinical study design for the inference of causal effects, because it controls common sources of bias in observational data such as confounding and mismeasurement. However, these can be expensive and/or impractical to do for all clinical questions of interest; moreover they can sometimes fail to generalize to common clinical populations or situations, prompting interest in alternative approaches.
Pragmatic clinical trial (PCT)	A clinical trial designed to measure the strength of associations between exposures and outcomes in real-world settings, rather than to prove causal relationships. To achieve this, they can incorporate far less restrictive participant eligibility criteria, use a less complex and more cost-effective design to achieve adequate statistical power, be performed in the target settings where inference is desired, and relax other common features of RCTs, including the use of placebos and blinding. These studies can complement RWE studies to support high-quality clinical evidence in health systems.
Electronic health record (EHR)	Computerized system that captures a large variety and volume of data relevant to patient-provider encounters. These include prescriptions, diagnoses, and procedures, as well as test orders and results, clinical notes, and provider-patient messaging. EHR systems are increasingly including patient-reported outcomes and imaging.
Diagnosis code (e.g., ICD code)	Providers commonly assign diagnosis codes at the time of a clinical encounter in order to facilitate medical billing and reimbursement. These commonly originate from a list of predetermined codes such as the International Statistical Classification of Diseases and Related Health Problems (the 10th revision contains over 70,000 such diagnostic codes).
Structured data	Data elements that encode information in a predefined and consistently organized fashion. These include the use of diagnosis codes, procedure codes, and other elements that place limits on how the information is represented. These types of data are typically easier to analyze but can lose important information and nuance owing to their restrictive nature and inconsistent application.
Unstructured data	In contrast to structured data, these data are not associated with any specifically imposed structure. One common example is the free-text notes associated with the documentation of clinical visits. Unstructured data are more difficult to analyze but contain richer and potentially more accurate content.
Natural language processing	A research field concerned with the computational processing and analysis of human language. Techniques developed in this field are increasingly being used for the analysis of medically related free text such as from EHRs, social networks, and other sources.

and less expensive alternatives. These include single-arm experimental studies, in which data pertinent to the control arm are supplied by historical sources or otherwise derived from sources of RWD such as EHRs. It is important to point out that this study design has been — and continues to be — controversial among authorities. This is in large part because the placebo effect is both substantial and unpredictable in many scenarios. As a result, recent regulatory applications using this approach have been limited to diseases for which the treatment effect is expected to be rapid and substantial, and the natural history of untreated disease is thought to be well understood (15). Nevertheless, some have argued that changes in regulatory attitudes have largely been the result of undue political pressure (via the 21st Century Cures Act as well as lobbying from patient-advocacy groups and biopharmaceutical companies), rather than true advances in clinical science (16).

Beyond the regulatory realm, historical and synthetic controls have also been used to support payor coverage decisions. For example, alectinib, an ALK inhibitor for advanced ALK+ non–small cell lung cancer (17), was approved in both the US and Europe on the basis of two single-arm phase II trials, but European payors requested additional evidence of alectinib's efficacy against the standard of care, ceritinib. The product sponsor Roche collaborated with Flatiron Health to generate a synthetic control cohort of 77 patients that satisfied the requested coverage requirements. A follow-up RCT confirmed similar efficacy between the propensity-matched synthetic control and the ceritinib-treated arm (18, 19).

Digital approvals. Although physicians are permitted to prescribe treatments off-label in regulatory environments such as the US, the regulatory label of a given treatment impacts payor coverage decisions and ultimately how many patients will receive treatment. Regulatory agencies like the FDA are increasingly approving label expansions on the basis of RWD. For instance, the CDK4/6 inhibitor palbociclib was approved only for women with ER+/HER2– breast cancer, because clinical trials only studied women (who constitute >99% of breast cancer cases). In April 2019, the sponsor Pfizer was able to secure approval for its use in males on the basis of outcomes data captured in EHRs related to its off-label use in that population (20). Similarly, in the setting of medical devices, the FDA recently approved the expanded use of the SAPIEN 3 transcatheter heart valve as a valve-in-valve procedure for patients with a failing bioprosthetic aortic or mitral valve with significant surgical risk. The approval for this additional indication was made on the basis of an evaluation of prospectively collected registry data (21).

Biosimilar development. Biologics are a class of medications derived from living sources that have transformed the course of many chronic diseases, such as rheumatologic conditions; however, they are complicated to manufacture and expensive to produce. To facilitate the approval of cheaper, nonbranded biologics at the time of patent expiration, the FDA has established expedited pathways for biosimilars, biologics that must be shown to have substantial similarity to the originator drug in terms of safety, purity, and potency. Many in the clinical community initially regarded biosimilars with substantial concern, in part because expedited regulatory pathways allowed for the extrapolation of efficacy from one drug indication to all indications of the reference product. However, multiple real-world studies have addressed these concerns by repeatedly demonstrating their safety, efficacy, and noninferiority to the reference product across indications and cohorts (22, 23). These studies supporting substantial similarity in the real-world setting have led to increased acceptance of these lower-cost drugs by the clinical community (24).

RWD to inform clinical trial design
Better patient selection. In 2017, a survey of life science companies found that 54% of survey participants were investing in RWE capa-

Table 2. Producers and consumers of RWD in the health care ecosystem

Participant	Role
Regulatory bodies (FDA, European Medicines Agency [EMA])	Promote public health through the control and supervision of medical products, including drugs and medical devices. As discussed in this Review, regulatory bodies are increasingly using RWD to assess the safety and efficacy of medical treatments.
Pharmaceutical and medical device industry	Companies manufacturing medical drugs and devices. These companies commonly use registries to track efficacy, safety, and use of their products (and those of their competitors); they are increasingly interested in other sources of RWD for these purposes.
Providers	A term encompassing all individuals who professionally provide a health care service. This includes physicians, surgeons, nurses, dentists, physical and behavioral therapists, and other allied health professionals. Providers increasingly capture clinical data in EHR systems during clinical encounters. Many disease registries are run by and updated by providers and the individuals employed by them.
Patients and communities	The recipients of health care from providers. They are typically the subjects of RWD, and generate these data through clinical encounters, wearables, social networks, etc. Increasingly, they are also becoming administrators of their own RWD, as well as consumers of it (as discussed in the penultimate section of this Review).
Payors	The entities that finance the provision of health care. In the US, these include the US government, employer-subsidized health insurance, and individually paid health insurance. Payors collect claims data on their enrollees in order to meet financial objectives. Both government- and private-payor claims data have been made available for research purposes.
Managed care organization	A health insurance organization that provisions services in exchange for a predetermined fee. These organizations control the utilization of health care in order to reduce costs while attempting to improve quality. These organizations can be independent of the payor. They analyze RWD submitted by provider groups in order to assess patterns of health care utilization and determine reimbursement.
Facilities	Places where health care is delivered. These include hospitals, clinics, urgent care facilities, surgical centers, pharmacies, and long-term care and rehabilitation facilities. Increasingly, health care is being delivered independent of physical facilities via telemedicine. These facilities are commonly associated with systems that capture health care data, including EHRs, researchers who maintain disease registries, etc.
Accountable care organization (ACO)	An organization of US health care providers whose reimbursements are tied to meeting quality targets and reducing the cost of care. These organizations typically submit data on health care quality during the reimbursement process to Medicare and other payors. ACOs are increasingly interested in using RWD to predict future declines in health.
Other third-party businesses in health care	There are a wide variety of companies in the health care space that produce and consume RWD. These include health care analytics firms and consultancies, operators in the mobile health space, telehealth firms, and direct-to-consumer companies (e.g., microbiome trackers, digital pedometers, heart rate monitors).

bilities in order to support clinical trial design and patient recruitment (25). In the early clinical phase, RWD (e.g., from the EHR) may be used to identify clinical cohorts with unmet clinical needs and a greater likelihood of benefiting from new therapies. These data may help refine trial inclusion/exclusion criteria to improve capture of target patients. Moreover, RWD could be used to identify the best study sites and enable more efficient recruitment and retention within the clinic setting. These interventions have the potential to decrease trial length while increasing both statistical power and generalizability. Much of this promise has already led to the formation of research networks with hopes of optimizing recruitment and data interchange. Examples include the public NIH Clinical and Translational Science Accrual to Clinical Trials (CTSA ACT) program (26) and the private TriNetX network (27).

EHR data are particularly relevant to pragmatic clinical trials (Table 1). These are a variant of RCTs that are designed to answer practical questions faced by decision makers in the routine clinical setting rather than to establish causal relationships. By elucidating typical practice patterns for the disease of interest (e.g., frequency of visits, laboratory/imaging studies, etc.), EHR data may enable the design of more efficient trial protocols. For such trials embedded within the clinic, future trial designs could take greater advantage of randomization schemes built within the practice workflow (e.g., consent at the time of check-in, randomization during the clinic visit, etc.).

Trimming the trials: more efficient data collection. Excessive data collection has been blamed for clinical trials' contributions to substantial expense, complexity, and delay in the drug development process (28). RWD may help reduce the complexity and costs of this data collection process by informing trial designers as to what variables are most often used clinically, which are informative, and which might be redundant. Moreover, new trial designs in the clinical setting, such as adaptive platform trials, which allow for the dynamic evaluation of multiple interventions, may represent a valuable source of RWD with the potential to further increase trial efficiency and reduce costs (29). Of course, trial sponsors would have to synchronize any new data collection expectations with regulators.

RWD to continually establish efficacy

Assessing the efficacy-effectiveness gap. The efficacy-effectiveness gap refers to systematic differences between rates of efficacy reported in RCTs and effectiveness in routine clinical settings (30). Multiple reasons have been proposed for the existence of this gap, including differences between patient populations; differences in endpoints, time under observation, analytic methods, and treatment adherence; and confounding and measurement bias. In part to address questions of RCT generalizability, RWE studies have received considerable attention, especially those that aim to benchmark real-world studies against RCTs (31, 32). Better understanding of these differences has the potential to inform both local clinical practice and future clinical trial design.

Searching for efficacy in specific populations. In particular, RCTs have been criticized for their exclusive eligibility criteria. Women, especially pregnant women and those of childbearing potential, can be excluded from many trials (33). Patients with chronic kidney disease (CKD) have also been excluded from many cardiovascular trials, even though they constitute a large and important

fraction of this disease cohort (34). The desire to control undesired outcome variability and maximize trial efficiency reflects a fundamental tradeoff between the internal validity and external utility of these studies.

To the extent that RCTs do not capture these vulnerable populations well, their exclusion leaves open a critical evidence gap that must be filled by RWD. For example, recent studies have used 10 years' worth of EHR data to study the safety of lower endoscopy in pregnant patients who report alarm symptoms (e.g., rectal bleeding) that might indicate inflammatory bowel disease or cancer (35). Similarly, RWD have illuminated the safety and efficacy of anticoagulants for patients with CKD — a patient population prone to conditions needing this treatment but commonly excluded from controlled studies (36).

Effect modifiers and precision medicine. The specific study of real-world efficacy in subgroups opens the possibility of research into effect modifiers (e.g., treatment by group interactions) and precision medicine. Treatment effect modifiers can take a variety of forms: inherited factors, concurrent medications, comorbidities, surgical history, diet, and other lifestyle habits (e.g., exercise, smoking). The use of RWD to identify treatment effect modifiers can help guide patient selection (e.g., picking the patients most likely to respond to treatment) and tailored behavioral modification (e.g., exercise to augment insulin sensitivity [ref. 37]; NSAIDs and smoking cessation [ref. 38] for Crohn's disease). More importantly, however, these studies can shed important light on pathophysiology and fundamental mechanisms of disease (39). But it is important to note that many of these components, even well-known social determinants of health, are still not captured well in clinical EHRs (40). Overall, it should be noted that the identification of precision medicine subgroups is particularly difficult because it is susceptible to false positives from multiple-hypothesis testing; positive results will still need to be confirmed with independent data sets and/or study designs.

Long-term, post-trial outcomes. Another critique of controlled trials is that they are too short, especially in comparison with chronic disease time scales. In addition to the substantial expense associated with continual and considerable data collection, longer-term trials increase the burden on trial participants, who are then increasingly prone to dropout.

The analysis of long-term outcomes from post-trial data has been a valuable source of information related to treatment efficacy and safety. For instance, 10-year and 30-year follow-up data from the UK Prospective Diabetes Study and the Diabetes Control and Complications Trial, of type 2 and type 1 diabetes mellitus, respectively, have demonstrated a long-term reduction in disease complications following a strategy of intensive glycemic control (41, 42). Sources of RWD, such as the EHR, may be able to identify former study participants who are not already being tracked by post-trial registries and further track their outcomes from an efficacy and safety standpoint.

In addition, with the advent of individual-participant data sharing from clinical trials, linkage methods may facilitate a more complete understanding of important long-term outcomes. But for this to happen, the representation of clinical trial participation in the EHR would need to improve (e.g., how to later record which investigational drug or placebo a patient was treated with, after

the later unblinding of the study), as would the capture of patient-reported outcomes (e.g., HealthMeasures, PROMIS).

RWD for comparative effectiveness

Integrating costs with comparative effectiveness. In the overall effort to bring rational and cost-conscious decision making to the clinic, comparative effectiveness studies represent the next logical step after real-world effectiveness studies. Comparative effectiveness studies are particularly relevant to chronic diseases with multiple medication classes and multiple agents within a class. Head-to-head comparisons of drugs in the setting of clinical trials are rare, in large part because of their expense and a lack of funding incentives from industry. When comparisons do exist, they are more commonly found in the setting of noninferiority studies, rather than assessments of superiority.

To promote the study of comparative effectiveness, the US Congress, as a part of the Affordable Care Act (ACA), established the Patient-Centered Outcomes Research Institute (PCORI) in 2010 to investigate relative effectiveness and inform decision making for Medicare coverage. Although PCORI has funded many important population studies, in part through the use of its data network, PCORnet, it is still prohibited by law from funding cost-effectiveness research using the traditional metric of quality-adjusted life years (43). The reasons for this restriction at the time of bill passage were complicated, including concerns that this would stifle innovation and lead to a slippery slope in coverage decisions and that wide cost variation in the US would make such research misleading, as well as political reasons related to passage of the ACA. This, along with prohibitions on Medicare to centrally negotiate drug prices, is likely contributing to worsening US health care costs.

These limitations highlight the potential for RWD to address this critical evidence gap and place it in a cost-aware framework. Specifically, charge data may be captured in the EHR or in matched claims data. Integrating costs with measures of performance that incorporate average patient use/compliance using RWD can help beneficiaries and payors afford high-quality care (as a simplified example, ranking diabetes drugs by their drop in hemoglobin A_{1c} per dollar). With larger collections of EHR data available, the efficiency in running more of these studies is likely to improve, enabling comparative cost-effectiveness studies that PCORI itself has been challenged to support (44).

Understanding effects of pharmacy practices on health care utilization. Understanding and curbing high health care spending is of obvious importance. One such practice currently being considered for potential regulation is the use of rebates passed between drug manufacturers and pharmacy benefit managers, intermediaries that negotiate and administer prescription benefits on behalf of payors. Although drug rebates are commonly promoted as reducing drug costs, critics have been unsatisfied with the lack of transparency in this practice, including how much of the rebate is actually passed on to beneficiaries (45). Equally unclear is how these practices affect the "list price" of medications (e.g., whether or not they may actually increase the list price and/or paradoxically decrease beneficiary access to affordable medications) (45).

RWD — whether as medical claims or other administrative billing data — may offer important transparency to this otherwise

opaque practice. In particular, by uncovering systematic differences in the prescribing patterns between patients of different payors, RWD may clarify the effects of these market drivers on health care utilization and outcomes.

Studying novel on-label pharmaceuticals versus older off-label drugs. Off-label use of older and cheaper pharmaceuticals can often offer efficacy and safety similar to those of their on-label counterparts. However, because of the lack of formal regulatory evaluation for a given indication, they can often be denied coverage by payors. For instance, ocrelizumab was recently approved for the treatment of multiple sclerosis, although an older but highly similar agent, rituximab, had been used off-label for this condition. Genentech manufactures both drugs, and, to our knowledge, did not pursue FDA approval for rituximab, possibly given its nearing patent expiration. Although there is at least one ongoing clinical trial comparing these agents head to head (NCT02980042, ClinicalTrials.gov; ref. 46), this question may also represent an illustrative opportunity for RWD to compare these agents by cost, efficacy, safety, and other endpoints, such as drug immunogenicity. As pharmaceutical sponsors might not be incentivized to pursue these types of studies, this remains an opportunity for others, such as payors or health care institutions, to use RWD to study cost-effectiveness to inform future coverage decisions.

RWD to study the practice of medicine

Quality of practice and medical errors. Analytics on RWD can help measure the quality of medical practice at the practitioner level. Medical groups associated with medical procedures can use RWD to specifically identify both underperforming and overperforming providers and use this as the basis of a strategy to disseminate best practices. RWD also have the potential to critically assess the equity of health care delivery across race, sex, and other socioeconomic strata. While these types of questions are still answered manually with sampled record reviews, automated systems could enable a more comprehensive and consistent evaluation of quality. Regulatory agencies and payors commonly use clinical data to evaluate the quality of delivered care, and we predict more of these reports will be generated using EHR data over claims data, owing to the higher degree of detail provided.

Although internal data-driven dashboards and physician "scorecards" have historically been heavily guarded from payors and patients, changes to regulations and data-interoperability practices may change over time as these data become increasingly integrated with existing physician-rating platforms on the Internet.

Standardizing care and care delivery. In health care systems such as in the US, excessive practice variation has been implicated as a major contributor to excess health care spending and poorer outcomes (47). One important first step toward reducing unwanted variation (e.g., variation that deviates from evidence-based practice) is the accurate capture and modeling of current practices and identification of actionable changes — whether at the level of the system, community, practice, or provider. Analytic platforms that measure the current state of variation and the response to intervention from a cost and outcomes standpoint represent another use case for RWD — for both payors and accountable care organizations.

The effect of payors on medical care. In the US, payors wield enormous influence on multiple aspects of health care — including

access, costs, and outcomes. Understanding the variation of payor practices can not only yield important insights into their effect on important health care outcomes, but also help disseminate knowledge relevant to marketplace regulation and best practices.

Payor decisions, including preapprovals and denials, are captured in the EHR and might be analyzed using causal inference techniques such as regression discontinuity and instrumental variable analysis in order to understand their effects. Many expensive drugs and devices may only be used after prior authorization by a payor. Physicians (or their staff) may go through cycles of authorization requests and denials or acceptances. The data related to these transactions are increasingly captured in EHR systems, including the actual denial letters. A systematic review of payors and their acceptance/denial rates per medication may be illuminating, especially if such data are published or made open. For example, patients who would have been prescribed a given medication but were shunted to a different treatment as a consequence of payor denial could be analyzed in the setting of a matched-cohort study with similar patients who were able to be treated as intended.

Are new-generation diagnostics improving outcomes? Although the first major clinical appearance of genetic data was in the setting of oncology and rare diseases, genomic testing is growing, with newer polygenic risk score tests being developed and proposed even for the primary care setting (48). But the proliferation of these expensive tests has raised the important question: are they worth it? Early studies are beginning to address the aspect of cost and the practicalities of implementation (49, 50). However, the greater question that remains largely unaddressed is whether genomic data are positively impacting health care outcomes. Until payors demand high-quality evidence of value and regulators tighten control over direct-to-consumer marketing of diagnostics, RWD platforms that capture the presence of testing, the test results, clinical outcomes, and costs may be in the best position to begin to answer this question.

RWD for data-driven decision support

Clinical decision support: the provider perspective. Most clinicians might have seen only a single case of acute porphyria or mesenteric panniculitis over the course of an entire career. However, the odds are far greater that a health system with a thousand providers has collectively encountered these conditions dozens of times. This idea is at the heart of one of the most exciting prospects for RWD and the future of clinical decision support, dubbed the "green button" (51). This concept proposes to strengthen clinical decision making by allowing clinicians to be more data- and experience-driven.

The most straightforward implementation of the green button concept involves providing data on similar patient outcomes, or connecting requesting clinicians to prior providers who can speak to their experience and lessons learned. Another exciting version of this idea is informatics as a clinical consultation service — one that makes EHR-powered data and recommendations accessible to providers within the setting of existing clinical reimbursement frameworks (52). A more sophisticated version might harness machine learning methods such as dimensionality reduction/ clustering, supervised learning, and reinforcement learning to form an embedded recommender system.

The potential impacts of this concept are tremendous. First, it may enable truly personalized medicine by accounting for local factors such as demographics, surgical expertise, etc. Second, clinical data sharing protocols (e.g., FHIR-enabled application programming interfaces) and/or federated-learning methods may ultimately enable the possibility of leveraging insights from millions of patient-years across systems. Third, a careful study of differences in recommendations for the same patient across systems may yield important new insights to help health systems learn from each other.

Clinical decision support: the patient perspective. Although advances in health information technology increasingly enable patients to integrate their clinical data across systems for the purposes of care coordination, it has not been as easy for patients to download their own data and use it as they see fit. The advent of tools such as the "blue button" are increasingly giving patients more access to their health care data (53).

While the average user may take advantage of blue button tools for the purpose of understanding their health status, planning for future health care expenditure, and sharing health information with family and caregivers, the liberalization of health care data could empower the future patient to do much more. If given more control of their data, patients could add to and correct their own health record (54). They may be in the best position to use apps that can answer the question: Given your data, what do you think we should do next?

These future tools could also help clinical researchers reach previously untapped participants. For instance, patients may be able to lend their clinical data to crowdsourced clinical research endeavors such as deep learning on mammograms (55).

Clinical decision support: the community perspective. Given the broad scope of RWD sources extending beyond just the hospital-clinic setting (e.g., air and water pollutants, drug abuse, gun violence, occupational exposures, socioeconomic status, climate and weather patterns), there is a potential for more community-level engagement in directing health care efforts. Some data are already available through governmental or otherwise publicly available platforms; others may become available if compatible with community preferences and local legislation. Once this occurs, these data may become resources for the community itself to do its own data audit and advocate for its own health care priorities, or to invite the global data science community to participate in the effort. Data-driven efforts may be usable to precisely target the best inventions to the right homes in the right cities at the right time, an effort termed precision public health (56, 57).

Challenges in the use of RWD

In this Review, we have briefly highlighted multiple use cases for RWD — some ongoing and many still to be seen. While there are many reasons to be excited about the potential of RWD, many challenges also lie ahead. These generally fall into two categories: epidemiologic challenges and biomedical informatic challenges.

Epidemiologic challenges primarily concern problems of data quality and bias (Table 3). These are issues that generally result from ad hoc data collection and from lack of the quality control that would ordinarily follow from a well-designed and controlled experiment. Although the central assumption underlying most RWD analyses is that these biases can be identified and mitigated

in the analysis phase, doing so requires substantial expertise, including epidemiology, knowledge of the clinical domain and the health system itself, biostatistics, and clinical informatics.

Even in the presence of such expertise to "de-noise" the data and "unbias" the analysis, the success of the process cannot be assessed from within the confines of the data themselves. It requires external validation from independent sources of evidence. Thus our view is that, for the most part, the enterprise of RWE cannot be relied upon in isolation, nor can it be understood as a replacement for controlled trials. To the contrary, it interfaces deeply with both health system stakeholders and pragmatic clinical trialists to enable a learning health system.

There are also many challenges at the level of biomedical informatics. Data collection in real-world settings is frequently haphazard and unstandardized. Although the analysis of free text using natural language processing is evolving with advances in methods and computation, the use of these technologies remains mostly at the level of methodologic development and far beyond the reach of the average investigator. Structured data are the workhorse of most analyses, but come with their own set of challenges, including the lack of standardization and harmonization across data sources.

Data access is another major issue. Much of the promise of RWE follows from its potential to achieve statistical significance by amalgamating population-level data sets. In the US, where the culture of privacy protection is strong and public scrutiny over digital privacy continues to grow with increasing awareness, the sharing of clinical data to unlock important insights remains difficult. The risks of data theft, manipulation, and other malignant use are only becoming more apparent with every news cycle. Deidentification strategies are being tried, but fundamentally may not be possible for every type of data. Strategies to avoid the transfer of actual data, such as synthetic cohorts and federated learning, are being explored, but are largely in their infancy. Tiered access to data seems unavoidable, with many stages of permissioned access likely to come between completely private and widely open access.

The nature of the competitive US health care system also hampers data exchange. We predict that partnered health systems situated across the country will be more likely to share clinical data with each other than with neighboring (and thus competing) systems. We may likely end up with a hopscotch-like set of intersecting regions of "friendly" noncompeting systems sharing data in their own circles. But consistently convincing business reasons to share data at this scale have yet to materialize. Any clinical data interoperability work takes resources, and those with budgets will likely need to see a financial reason to interoperate, beyond helping researchers get better papers published and grants funded. Some of the uses listed above may help make the case.

Despite the substantial hurdles, we remain optimistic about the potential for RWD to transform health care at every level. We believe that the very human ingenuity that led to these data being captured also has the same potential to overcome these challenges, safeguard human rights, and unlock the insights that can help people everywhere lead healthier and more productive lives.

Author contributions

VAR researched and drafted this article. AJB conceived, researched, and critically edited this article.

Table 3. Epidemiologic and biostatistical challenges to the use of RWD

Problem	Description	Potential solutions
Missing data and selection bias	RWD by definition involve routinely collected and nonprotocolized data collection. Thus, the presence of data is typically associated with the outcome, resulting in bias if handled improperly. For example: i. Patients may not follow up with their clinician because they are unwell and/or did not benefit from the first visit. In this case, the patients found within in an EHR database would have better outcomes than a prospective study tracking all patients. ii. Patients with poor health may seek more care. In this case, patients within an EHR database would have poorer outcomes than a prospective study tracking all subjects. iii. Patients who use wearables and health-related social networks may be systematically healthier and less prone to adverse outcomes than all patients with a given condition.	i. Careful study design incorporating epidemiologic principles. Cohorts must be carefully defined to maximize their representativeness of the target population for which inference is desired. ii. Biostatistical methods for bias estimation and correction, sensitivity analyses
Measurement biases	Many RWD analyses involve the "repurposing" of data collected for one purpose to be used for other purposes (e.g., research, business decision making, policy). Because the data collection process was not explicitly designed with the secondary use in mind, it may not accurately capture exposures and outcomes of interest, leading to bias. For example: i. Patients with a disease may be more likely to recall potential exposures than patients without. ii. Clinicians may be more likely to selectively document patient-reported risks that have established associations with disease. iii. Clinicians may make diagnoses on the basis of different tests with different sensitivities and specificities. iv. Many clinical diagnoses are subjective and require blinded adjudication committees in clinical trials.	i. Consultation and/or collaboration with the individuals who made the original measurements to be repurposed ii. Biostatistical methods such as errors-in-variables models, mixed-effects models, quantitative bias analysis, etc. iii. Use of preregistered protocols to clearly define variables and limit subjectivity (e.g., ref. 59)
Confounding	Clinicians may select into treatment only those most likely to respond. Thus, treatments may appear to perform better than they would appear in controlled trials.	i. Consultation and/or collaboration with the individuals who made treatment assignments ii. Restriction, regression-based stratification, and other biostatistical methods
Absence of controls (e.g., placebos)	Because RWD are primarily generated in the course of routine clinical care, the decision to give no treatment in situations in which treatment is needed is rare and is commonly related to the outcome of interest. The prescription of sham treatments (i.e., placebos) is typically prohibited. Thus it is difficult to use RWD to estimate treatment effectiveness and/or confirm that treatments are actually better than no treatment at all.	i. Access to historical data may permit the evaluation of outcomes during periods when treatment was unavailable ii. Use of natural experiments (e.g., a new treatment happened to be used first in one clinic site before other comparable clinics, or was covered by one health care payor before others) iii. Use of quasi-experiments (e.g., the laboratory arbitrarily determined a diagnostic cut point; patients falling on either side of the cut point are highly comparable but assigned to different treatments)
Multiple hypothesis testing/inflated type I error rate	Because most RWD analysis takes place on already collected data, these data are susceptible to being used to perform many post hoc statistical tests and only report positive results. This results in a well-known statistical problem, multiple hypothesis testing, which culminates in an increased rate of false-positive discoveries	i. Preregistration of planned studies with associated protocols ii. Internal and external validation

Address correspondence to: Atul J. Butte, Bakar Computational Health Sciences Institute, University of California, 550 16th Street, 4th Floor Box 0110, San Francisco, California 94158-2549, USA. Phone: 415.514.0511; Email: atul.butte@ucsf.edu.

1. FDA. Framework for FDA's Real-World Evidence Program. 2018. http://www.fda.gov/media/120060/download. Accessed December 5, 2019.

2. Berger M, et al. A framework for regulatory use of real-world evidence. http://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf. Updated September 13, 2017. Accessed December 5, 2019.

3. Craig P, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. *J Epidemiol Community Health*. 2012;66(12):1182–1186.

4. FDA. FDA's Sentinel Initiative. http://www.fda.gov/safety/fdas-sentinel-initiative. Updated October 18, 2019. Accessed December 5, 2019.

5. FDA. MedWatch: The FDA Safety Information and Adverse Event Reporting Program. http://www.fda.gov/safety/medwatch-fda-safety-information-and-adverse-event-reporting-program. Updated December 2, 2019. Accessed December 5, 2019.

6. Sentinel Initiative. Sentinel Common Data Model. http://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model. Updated October 31, 2018. Accessed December 5, 2019.

7. Sarker A, et al. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform*. 2015;54:202–212.

8. Nikfarjam A, et al. Early detection of adverse drug reactions in social health networks: a natural language processing pipeline for signal detection. *JMIR Public Health Surveill*. 2019;5(2):e11264.

9. Altschuler E, Kast R. Methods of modulating TNF using bupropion. US patent 6,5656,005. August 3, 2006.

10. Kast RE, Altschuler EL. Remission of Crohn's disease on bupropion. *Gastroenterology*. 2001;121(5):1260–1261.

11. Lorberbaum T, et al. Coupling data mining and laboratory experiments to discover drug interactions causing QT prolongation. *J Am Coll Cardiol*. 2016;68(16):1756–1764.

12. Kahn SE, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. *N Engl J Med*. 2006;355(23):2427–2443.

13. Home PD, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet*. 2009;373(9681):2125–2135.

14. Neal B, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med*. 2017;377(7):644–657.

15. FDA. Center for Drug Evaluation and Research. 2017 New Drug Therapy Approvals. http://www.fda.gov/files/about%20fda/published/2017-New-Drug-Therapy-Approvals-Report.pdf. Accessed December 5, 2019.

16. Honig N. Will new "real world evidence" standard hurt drug safety? January 30, 2017. *Dome: Law, Legislation & Policy*. http://sites.bu.edu/dome/2017/06/30/will-new-real-world-evidence-standard-hurt-drug-safety. Accessed December 5, 2019.

17. Chatterjee A, Chilukuri S, Fleming E, Knepp A, Rathore S, Zabinski J. Real-world evidence: driving a new drug development paradigm in oncology. McKinsey & Co. http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/real-world-evidence-driving-a-new-drug-development-paradigm-in-oncology. Accessed December 5, 2019.

18. Davies J, Martinec M, Martina R. Retrospective indirect comparison of alectinib phase II data vs ceritinib real-world data in ALK+ NSCLC after progression on crizotinib. *Ann Oncol*. 2017;28(suppl 2):mdx091.018.

19. Mok T, et al. ASCEND-2: a single-arm, open-label, multicenter phase II study of ceritinib in adult patients (pts) with ALK-rearranged (ALK+) non-small cell lung cancer (NSCLC) previously treated with chemotherapy and crizotinib (CRZ). *J Clin Oncol*. 2015;33(15 suppl):8059.

20. FDA. FDA expands approved use of metastatic breast cancer treatment to include male patients. http://www.oncnet.com/news/fda-expands-metastatic-breast-cancer-drug-indication-include-men. Updated April 4, 2019. Accessed December 5, 2019.

21. FDA. FDA expands use of Sapien 3 artificial heart valve for high-risk patients. http://www.fda.gov/news-events/press-announcements/fda-expands-use-sapien-3-artificial-heart-valve-high-risk-patients. Updated June 5, 2017. Accessed December 5, 2019.

22. Meyer A, Rudant J, Drouin J, Weill A, Carbonnel F, Coste J. Effectiveness and safety of reference infliximab and biosimilar in Crohn disease: a French equivalence study. *Ann Intern Med*. 2019;170(2):99–107.

23. De Cock D, Watson K, Hyrich KL. Biosimilars in the UK: early real world data from the british society for rheumatology biologics registers for rheumatoid arthritis. *Ann Rheum Dis*. 2017;76:555–556.

24. Rudrapatna VA, Velayos F. Biosimilars for the treatment of inflammatory bowel disease. *Pract Gastroenterol*. 2019;43(4):84–91.

25. Deloitte. Getting real with real-world evidence (RWE). 2017 RWE Benchwork Survey. http://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/real-world-evidence-benchmarking-survey.html. Accessed December 5, 2019.

26. Visweswaran S, et al. Accrual to Clinical Trials (ACT): a Clinical and Translational Science Award Consortium network. *JAMIA Open*. 2018;1(2):147–152.

27. TriNetX. InSite: The largest European live clinical data network. http://www.trinetx.com/insite. Accessed December 5, 2019.

28. Sargent DJ, George SL. Clinical trials data collection: when less is more. *J Clin Oncol*. 2010;28(34):5019–5021.

29. Saville BR, Berry SM. Efficiencies of platform clinical trials: a vision of the future. *Clin Trials*. 2016;13(3):358–366.

30. Nordon C, et al. The "efficacy-effectiveness gap": historical background and current conceptualization. *Value Health*. 2016;19(1):75–81.

31. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. *BMJ*. 2016;352:i493.

32. Franklin JM, Dejene S, Huybrechts KF, Wang SV, Kulldorff M, Rothman KJ. A bias in the evaluation of bias comparing randomized trials with nonexperimental studies. *Epidemiol Methods*. 2017;6(1):20160018.

33. Liu KA, Mager NA. Women's involvement in clinical trials: historical perspective and future implications. *Pharm Pract (Granada)*. 2016;14(1):708.

34. Zoccali C, et al. Children of a lesser god: exclusion of chronic kidney disease patients from clinical trials. *Nephrol Dial Transplant*. 2019;34(7):1112–1114.

35. Ko MS, Rudrapatna V, Avila P, Mahadevan U. Safety of flexible sigmoidoscopy in pregnant patients with inflammatory bowel disease. *Gastroenterology*. 2019;156(6):S-18–S-19.

36. Godino C, et al. Real-world 2-year outcome of atrial fibrillation treatment with dabigatran, apixaban, and rivaroxaban in patients with and without chronic kidney disease. *Intern Emerg Med*. 2019;14(8):1259–1270.

37. Borghouts LB, Keizer HA. Exercise and insulin sensitivity: a review. *Int J Sports Med*. 2000;21(1):1–12.

38. Alexakis C, Saxena S, Chhaya V, Cecil E, Majeed A, Pollok R. Smoking status at diagnosis and subsequent smoking cessation: associations with corticosteroid use and intestinal resection in Crohn's disease. *Am J Gastroenterol*. 2018;113(11):1689–1700.

39. Kuenzig ME, et al. The NOD2-smoking interaction in Crohn's disease is likely specific to the 1007fs mutation and may be explained by age at diagnosis: a meta-analysis and case-only study. *EBioMedicine*. 2017;21:188–196.

40. US Department of Health and Human Services. Office of the Assistant Secretary for Planning and Evaluation. Incorporating social determinants of health in electronic health records: a qualitative study of perspectives on current practices among top vendors. http://aspe.hhs.gov/pdf-report/incorporating-social-determinants-health-electronic-health-records-qualitative-study-perspectives-current-practices-among-top-vendors. Updated October 15, 2018. Accessed December 5, 2019.

41. Nathan DM, DCCT/EDIC Research Group. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study at 30 years: overview. *Diabetes Care*. 2014;37(1):9–16.

42. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008;359(15):1577–1589.

43. Pearson SD. Cost, coverage, and comparative effectiveness research: the critical issues for oncology. *J Clin Oncol*. 2012;30(34):4275–4281.

44. Rosenthal E. The soaring cost of a simple breath. *New York Times*. October 12, 2013. http://www.nytimes.com/2013/10/13/us/the-soaring-cost-of-a-simple-breath.html. Accessed December 5, 2019.

45. Arnold J. Are pharmacy benefit managers the good guys or bad guys of drug pricing? *STAT*. August 27, 2018. http://www.statnews.com/2018/08/27/pharmacy-benefit-managers-good-or-bad. Accessed December 5, 2019.

46. Tolerability and safety of switching from rituximab to ocrelizumab in patients with relapsing forms of multiple sclerosis. NCT02980042. http://clinical-trials.gov. Accessed December 5, 2019.

47. Krumholz HM. Variations in health care, patient preferences, and high-quality decision making. *JAMA*. 2013;310(2):151–152.

48. Ashley EA, et al. Clinical assessment incorporating a personal genome. *Lancet*. 2010;375(9725):1525–1535.

49. Christensen KD, Phillips KA, Green RC, Dukhovny D. Cost analyses of genomic sequencing: lessons learned from the MedSeq Project. *Value Health*. 2018;21(9):1054–1061.

50. Stark Z, et al. Integrating genomics into healthcare: a global responsibility. *Am J Hum Genet*. 2019;104(1):13–20.

51. Longhurst CA, Harrington RA, Shah NH. A 'green button' for using aggregate patient data at the point of care. *Health Aff (Millwood)*. 2014;33(7):1229–1235.

52. Schuler A, Callahan A, Jung K, Shah NH. Performing an informatics consult: methods and challenges. *J Am Coll Radiol*. 2018;15(3 pt B):563–568.

53. HealthIT.gov. Health IT in Health Care Settings. Blue Button. http://www.healthit.gov/topic/health-it-initiatives/blue-button. Updated April 8, 2019. Accessed December 5, 2019.

54. OpenNotes. http://www.opennotes.org. Accessed December 5, 2019.

55. Maxmen A. AI researchers embrace Bitcoin technology to share medical data. *Nature*. 2018;555(7696):293–294.

56. Khoury MJ, Iademarco MF, Riley WT. Precision public health for the era of precision medicine. *Am J Prev Med*. 2016;50(3):398–401.

57. Kuo AK, Summers NM, Vohra S, Kahn RS, Bibbins-Domingo K. The promise of precision population health: reducing health disparities through a community partnership framework. *Adv Pediatr*. 2019;66:1–13.

58. May T. The fragmentation of health data. *Datavant*. July 31, 2018. https://datavant.com/2018/08/01/the-fragmentation-of-health-data/. Accessed January 8, 2020.

59. Rudrapatna VA, Butte AJ. Robust measurement of the real world effectiveness of Tofacitinib for the treatment of Ulcerative Colitis using electronic health records: a protocol and statistical analysis plan. https://www.protocols.io/view/robust-measurement-of-the-real-world-effectiveness-2bqgamw. Updated May 22, 2019. Accessed December 5, 2019.