Pragmatic Randomized Trials
Considerations for Design and Implementation

Andrew Bevan MRes, MRSB, CBiol
Senior Director Project Management, Peri- and Post-Approval Studies, Evidera

Paul Biedenbach, BA
Executive Director Operations, Medical Communications, PPD

Ariel Berger, MPH
Executive Director, Integrated Solutions, Real-World Evidence Evidera

Introduction

The 21st Century Cures Act (Cures Act),[1] which became law in the United States December 13, 2016, has highlighted the need for robust real-world data to demonstrate effectiveness and safety of healthcare innovations that meet the requirements of regulators and payers alike. Included in the Act is an agreement to fund and accelerate cancer research and overall medical product development and delivery, as well as increase choice in, access to, and quality of American healthcare. One result of the Act has been to increase interest from both industry and regulatory authorities, such as the US Food and Drug Administration (FDA), in pragmatic randomized trials (PRTs). In December 2018, the FDA's Framework for Real-World Evidence Program[2] was published, which *"created a framework for evaluating the potential use of real-world evidence (RWE) to help support the approval of a new indication for a drug already approved under section 505(c) of the FD&C Act or to help support or satisfy drug post-approval study requirements."*
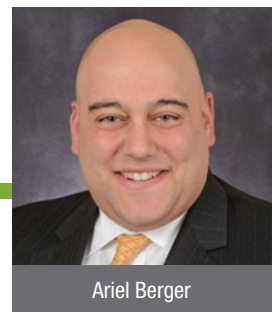
The industry has yet to feel the impact of this paradigm shift, not least because the traditional explanatory randomized controlled trial (RCT) utilizing surrogate
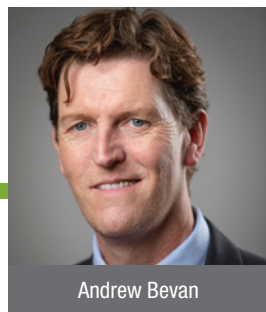
Andrew Bevan
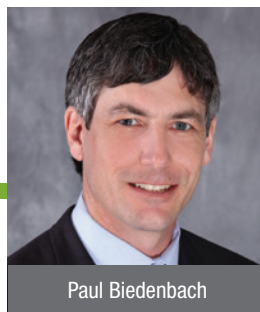
Paul Biedenbach

Ariel Berger

endpoints to establish efficacy and safety in a highly selected population under optimal conditions has been the gold standard for researchers and regulators alike for many decades. These trials, which generate high-quality robust data with high intrinsic validity upon which to base conclusions about causal relationships, answer the explanatory question "is the intervention efficacious and safe in tightly controlled, artificial conditions?" However, they ignore the more pragmatic question "is this an effective and safe option for my patient?" The pragmatic trial design was developed to answer the latter, which is a key question that can inform potentially life-altering decisions required by payers, clinicians, and even patients themselves.

Unlike randomized trials, pragmatic studies use "typical" clinical settings to examine real-world outcomes such as survival, utilization of healthcare services and/or pharmacotherapy, and overall cost of care. Most of these outcomes can be obtained from electronic health records (EHRs), which can shorten study timelines and reduce budget while still providing high-quality information. Pragmatic studies aim to generate evidence and conclusions based on real-world practice that are highly relevant to payers, healthcare providers (HCPs), and ultimately policy makers as they look to gather information to make treatment-related decisions. However, these studies have suffered from concerns of relatively low internal validity due to issues of outcome misclassification and other forms of bias (e.g., selection bias, confounding by indication). These issues can increase the risk of spurious associations between treatment(s) and associated outcomes related to effectiveness and/or safety, thereby reducing the reliability of the conclusions that can be drawn regarding cause and effect, and subsequently limiting their value to regulators. Given the "low intensity" of investigator oversight during the conduct of pragmatic studies, there also are concerns around the quality of information collected – particularly, key outcome measures.

Pragmatic Randomized Trials

PRTs represent a hybrid between traditional randomized controlled clinical trials that have been the gold standard for regulatory decision making, and pragmatic, observational research studies that are often used to generate real-world evidence to support health technology assessment (HTA) and payer decision making. A well-designed PRT that maximizes external validity, but also controls for confounding (including but not limited to selection bias) in order to maintain high levels of internal validity, could theoretically be used to generate evidence that would meet both regulatory and payer requirements. A number of tools have been developed to help researchers design pragmatic trials.[3-5] One such validated tool is the PRagmatic Explanatory Continuum Indicator Summary Version 2 (PRECIS-2),[5] which is a 9-spoked wheel, with each spoke representing a domain that denotes a key element of trial design (Figure 1). Each domain is scored on a 5-point Likert scale ranging from explanatory to pragmatic (i.e., 1=very

Figure 1. The PRagmatic-Explanatory Continuum Indicator Summary 2 (PRECIS-2) Wheel (Adapted from Louden, et al.[5])
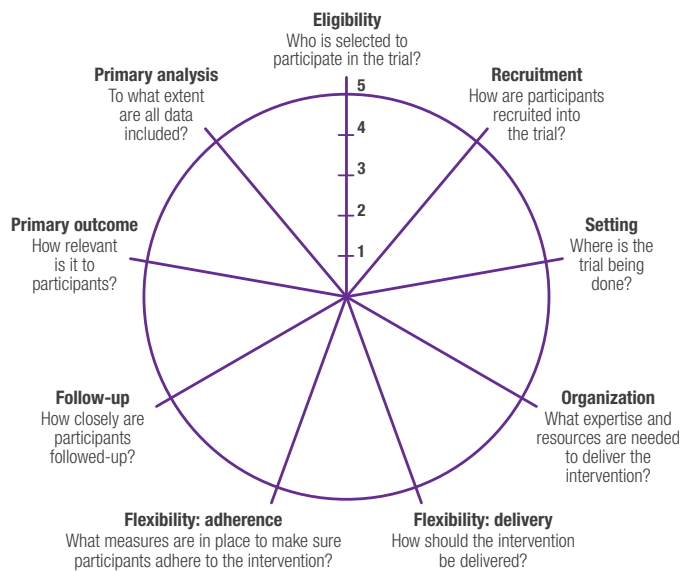
explanatory, 2=rather explanatory, 3=equally pragmatic/explanatory, 4=rather pragmatic, 5=very pragmatic). Trials that are predominantly explanatory in their design generate spoke and wheel diagrams that are close to the hub, whereas those with a more pragmatic approach produce diagrams that are closer to the rim. In reality, few trials are purely explanatory or purely pragmatic, and for the most part, a well-designed PRT that maintains high levels of external *and* internal validity will seek to strike an optimal balance between the two study types, thereby producing a diagram that would be somewhere in the middle of the wheel (and potentially an "uneven" wheel, with aspects more pragmatic pulling the circle closer to the rim and those more explanatory drawing the corresponding point closer to the hub). Representative diagrams for these designs are shown in Figure 2. Design choices should be based primarily on the research question(s) being posed; for example, in a pragmatic cardiovascular outcomes trial, more importance may be placed on high scores on the Eligibility, Primary Outcome, Setting, and Follow-up domains, whereas a trial investigating an intervention in a post-surgical intensive care setting may alternatively preferentially weight the Recruitment, Flexibility-delivery, and Primary Analysis domains.

## Real-World Outcomes and Endpoints

One of the challenges of designing effective PRTs is the choice of outcomes and endpoints. The term outcome is used here to mean a measured variable or event of interest (e.g., time to first occurrence of a composite outcome such as myocardial infarction [MI], stroke, or cardiovascular death, which collectively are referred to as major adverse cardiovascular events [MACE]), whereas an endpoint refers to an analyzed parameter that is expected to change over

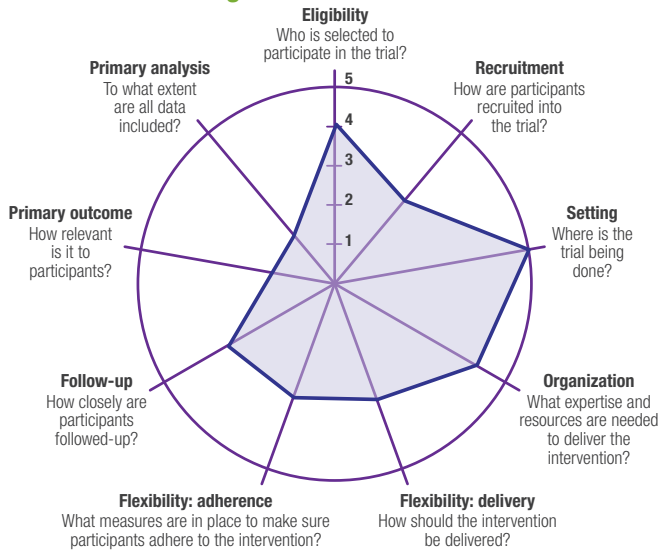**Figure 2.** Hypothetical PRECIS-2 Spoke and Wheel Diagrams

A Retrospective Observational Study
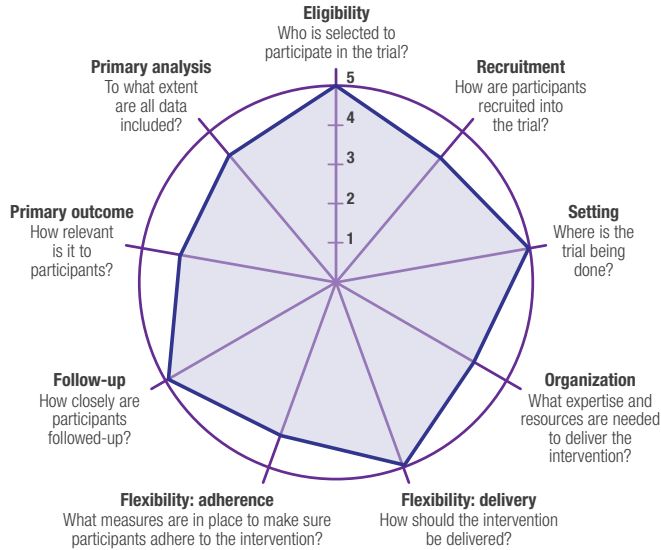
A Prospective Randomized, Double Blind Clinical Trial
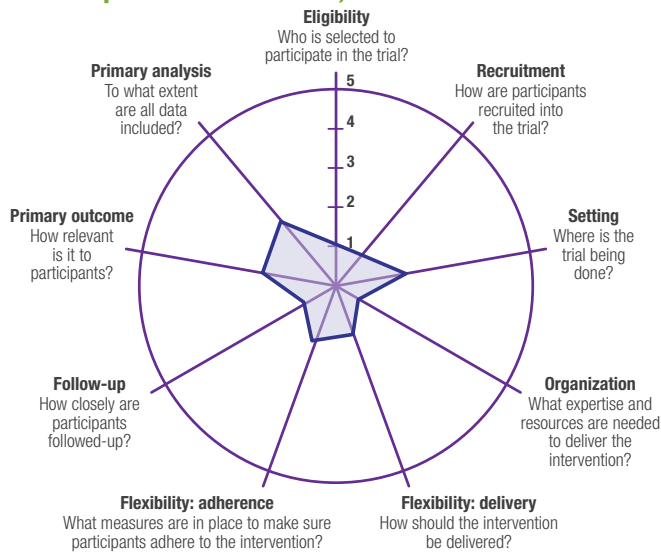
A Pragmatic Randomized Trial

time as a result of an intervention (e.g., change in LDL-C from baseline). For a PRT to meet the requirements of both regulators and payers, it is important that the selected endpoints and outcome measures resonate with key stakeholders (patients, payers, regulators, and healthcare providers), and be defined with sufficient sensitivity (typically more important for safety) and specificity (typically more important for effectiveness estimates) to translate the trial objectives into precise definitions of treatment effect. In addition, endpoints and outcome measures that are routinely available from EHR will render the study more pragmatic as it reduces the need to interact with those running the study (each such interaction moves the patient further from typical care and more towards protocol-mandated care).

In the routine clinical setting, intercurrent events can occur following an intervention – including treatment discontinuation or switching, or use of alternative or contraindicated medications – that can result in treatment effects being misinterpreted. Selecting endpoints and outcome measures without first considering the impact of these intercurrent events will result in uncertainty over the treatment effect, and potentially place a study at risk of not meeting its objectives. The impact of intercurrent events can be controlled for by randomization; however, it may not always be practical or even possible to randomize on an individual subject level, but instead other methods may need to be employed (e.g., cluster randomization [randomizing at the site level] or crossover designs) which can add to complexity of the trial design and analysis. Bias can also be introduced into PRTs through lack of ability to mask treatments. This can be addressed to a certain extent by selecting clinically objective outcomes (e.g., stroke, hospitalization due to non-fatal MI, tumor size), but this may not always be possible (e.g., in studies of Alzheimer's disease, where clinical outcome assessments [COAs] are subject to human interpretation); moreover, structural changes (e.g., items measured using surrogate imaging endpoints) may not translate into clinically meaningful change. Ultimately the selection of a primary endpoint or outcome will be driven by the research question(s) and how to best define the effect(s) of the treatment under study while controlling through design choices for the presence of varied intercurrent events. This topic is addressed in ICH-E9-R1, which introduces the estimand framework to link the trial objectives, the study population, and the variable (or endpoint) of interest to intercurrent events reflected in the research question to more effectively translate the trial objective(s) into a precise definition of the treatment effect(s) under investigation.[6] This revision to ICH guidance will undoubtably shape the approach to the design of randomized clinical trials, especially PRTs, in the future.

The Rise of Health Informatics and Big Data

The increasing availability of rich EHR potentially linkable to medical claims data, and our ability to mine those data using artificial intelligence (AI) and other advanced analytic

methods has expanded the possibilities for implementing embedded PRTs that reduce operational complexity, timelines, and cost, while still allowing for valid comparisons between treatments. AI has enabled computable phenotypes with the precise clinical characteristics that comprise the relevant study population, and clinical and economic outcomes of interest, *all from the same data source(s)*. One example of a real-world data source that has been widely used in the post-marketing evaluation of medicines is the Swedish Healthcare Quality Registries, which collect nationwide clinical data, encompassing a specific disease, intervention, or patient group that is highly relevant to regulators and HTAs.[7] One particular advantage of the Swedish Quality Registries is the ability to link data on specific patient phenotypes with treatments and outcomes. The VALIDATE-SWEDEHEART Trial is an example of a PRT that utilized a Swedish registry platform to compare bivalirudin versus heparin in ST-segment myocardial infarction (STEMI) and non-ST segment myocardial infarction (non-STEMI) patients undergoing percutaneous coronary intervention (PCI) on the composite endpoint of MI, all-cause mortality, and major bleeding.[8] It is included here as a case study to demonstrate the potential of big data within which PRTs can be conducted.

The use of coding algorithms to extract clinical outcomes from EHRs has been gaining traction in pharmacological studies over the past decade. For example, and specific to cardiovascular research, International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes for acute MI have been shown to have a positive predictive value (PPV) of ≥95% compared to manual chart review.[9,10] However, it may not always be clear which code on a particular record to use (each record can have multiple codes), or whether a particular diagnosis relates to the principal discharge diagnosis (i.e., the diagnosis that best describes the reason for the admission), and therefore algorithms that rely on ICD-9-CM diagnoses alone may not translate to a broad clinical research setting. This raises the need for more advanced techniques to identify MACE that may include medication and laboratory data. One such approach used diagnosis codes, procedure codes (in Current Physician's Terminology, 4th Edition [CPT-4] format), and laboratory test results, resulting in a more accurate algorithm to identify MACE (i.e., PPVs between 90% to 97% compared to manual review)[11] that could be readily adapted for use in other pragmatic cardiovascular trials (assuming access to comparable data types).

EHR and claims data have been used to generate RWE in a number of therapeutic areas including but not limited to cardiovascular studies and oncology. One example of the former is a real-world counterpart to the COMPASS pivotal, randomized, multicenter, randomized clinical trial, in which patients with existing and stable coronary artery disease (CAD) or peripheral artery disease (PAD) were randomized to receive low-dose rivaroxaban plus aspirin versus aspirin only[12]; findings from COMPASS, which was stopped early

Pragmatic RCT Case Study

VALIDATE-SWEDEHEART-Trial (Bivalirudin versus Heparin in ST-Segment and Non-ST-Segment Elevation Myocardial Infarction in Patients on Modern Antiplatelet Therapy in the Swedish Web System for Enhancement and Development of Evidence-based Care in Heart Disease Evaluated According to Recommended Therapies Registry Trial)

- **Multicenter, prospective, randomized, open-label design**

- **Utilized a national registry platform to identify eligible patients and extract clinical outcomes using ICD9 and ICD10 diagnosis codes**

- **The registry includes patients admitted to the hospital because of symptoms suggestive of acute coronary syndrome (ACS), and patients undergoing coronary angiography/angioplasty or heart surgery**

- **The study consented and enrolled 6,006 patients (3,005 STEMI, 3,001 NSTEMI) undergoing PCI and treated with P_2Y_{12} inhibitors at 22 centers**

- **Patients randomized 1:1 to receive bivalirudin or heparin**

- **Primary outcome was composite endpoint of MI, all-cause mortality, and major bleeding at 6 months**

- **The study concluded that the rate of composite of death from any cause (MI or major bleeding) was not lower among patients who received bivalirudin compared to heparin**

due to "overwhelming efficacy," provided the evidence needed to expand existing indications for rivaroxaban to include secondary prevention of MACE and major adverse limb events (MALE).[13] In the real-world study, which was run in parallel to COMPASS to demonstrate the burden of MACE and MALE in clinical practice prior to this expanded indication, key trial outcomes (including MACE, MALE, and the incidence of major bleeds), were estimated based on relevant diagnosis codes associated with claims submitted by providers in relevant settings of care (e.g., MI required a relevant principal diagnosis resulting from a visit to an emergency room or admission to hospital).

While access to RWE through EHR and claims data is fairly robust in key markets, the same cannot be said for emerging markets where access has been limited. With greater attention on the Asia Pacific market, demand for access to RWD is growing and, luckily, so is access. For example, PPD and Happy Life Tech (HLT), a Chinese medical AI company with an established relationship with and access to EHR from more than 100 leading hospitals across over 20 provinces in China, entered into an exclusive and unique collaboration.[14] With data representing the health experience of over 300 million patients, HLT data will allow more RWE studies to be done using Chinese patient data, and this should open up the potential to perform embedded global PRTs in this important emerging market. Using statistical methods of meta analyses, information from HLT could be aggregated with comparable data from other countries of interest, thereby potentially extending the power of this "hybrid" study design globally.

Challenges of Interoperability

In addition to ensuring algorithmic approaches to real-world data are generalizable across different sources (e.g., across EHR types, healthcare claims from various insurance payers) and different pragmatic research settings, the ability for one software system and associated data formats to interact with others (i.e., interoperability) represents a challenge to conducting multicenter/multi-country PRTs. Until recently, the mainstay for tackling this obstacle has been to implement common data models (CDMs) such as the FDA's Sentinel Initiative[15,16] to standardize data across multiple sources for research purposes. However, even though sites may format data according to a pre-defined CDM, CDMs require data to be mapped which can result in loss of detail, as information not common to all participating sites/systems tends to be omitted from the final CDM-driven data set. Another answer could be HL7's Fast Healthcare Interoperability Resources or FHIR[17] (pronounced "fire"). FHIR is a draft data standard and Application Programming Interface (API), which is quickly becoming the industry standard for exchanging healthcare data between disparate software systems, including wearable devices,[18] and has great potential

to be an application-based solution to the challenges of interoperability. FHIR aims to provide developers with a user-friendly solution to build applications that enable healthcare data to be accessed irrespective of the EHR system being used and is the data standard that has been adopted by federal agencies and healthcare providers in the US, including the Department of Health and Human Services (HHS), the Veterans Administration, and the Department of Defense; the National Health Service (NHS) in the UK also has adopted FHIR. Recently the Centers for Medicare and Medicaid Services (CMS) announced the launch of a pilot program that leverages FHIR to enable clinicians to directly access Medicare claims data, which according to CMS will "*fill in information gaps for clinicians, giving them a more structured and complete patient history with information like previous diagnoses, past procedures, and medication lists.*"[19] Evidera is already using FHIR to build bespoke data integration solutions to support both retrospective and prospective (including pragmatic) research for our clients that incorporate data from multiple diverse EHR data sources into a single cloud-based platform to support real-world evidence generation and address the challenges of interoperability.

Challenges of Missing Data

Further challenges encountered when performing PRTs are variation in intervals to disease status/check-in and the ability to capture outcomes over time to avoid missing events and incomplete data. The importance of this as a trial design consideration is obviously dependent on

Figure 3. Supporting Remote Patient-Reported Outcomes Data Collection via Global Contact Centers
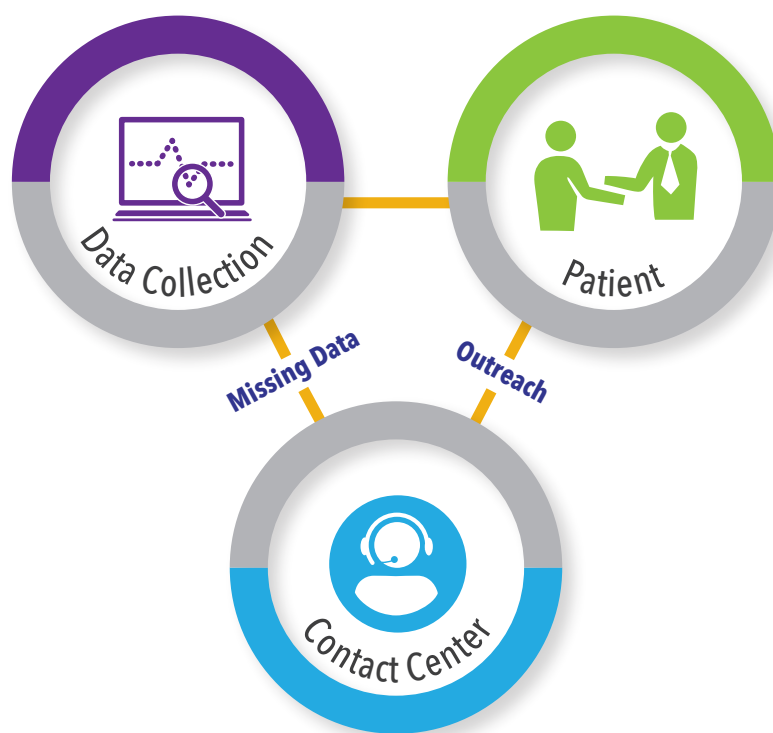
the nature of the condition and the treatment effect(s) of interest (e.g., in oncology, the timing of assessments may be very heterogenous in the real-world setting and this must be taken into account in the trial design). EHR data are a rich source that captures encounters that occur within specific care settings. However, due to the fragmented nature of the US healthcare system, encounters that occur outside of participating settings are not likely to appear in EHR or may be incomplete. One way to address this "missingness" potential is to supplement EHR sources with information from other sources, including but not limited to claims data, patient-reported outcomes, and direct-to-patient follow-up. This approach is being taken with the ground-breaking ADAPTABLE trial (Aspirin Dosing: A Patient-centric Trial Assessing Benefits and Long-Term Effectiveness) to compare the effectiveness of two different widely used doses of aspirin to prevent MI and stroke in patients with heart disease.[20] This trial, which is being conducted in the US, is a collaboration between the National Patient Centered Research Network (PCORnet) and Duke Clinical Research Institute (DCRI) and has recently completed enrollment of the planned 15,000 patients. The trial utilizes a combination of routine querying of EHR via the PCORnet CDM; surveillance data and medical claims data from CMS; and patient-reported outcomes confirmed through contact with DCRI personnel via a centralized call center, to capture endpoints for hospitalizations for MI, stroke, and death events.

This ability to maintain direct-to-patient contact in a long-term, follow-up study is important to ensure that outcomes are not missed and that patients are retained in follow-up. To facilitate comprehensive capture of relevant outcomes, access to global contact center capabilities to support pragmatic clinical trials is important (See Figure 3).

These types of call centers should have a comprehensive understanding of the regulatory, cultural, and logistical complexities associated with providing clinical trial support services across the globe, ideally with 24/7/365 coverage to address patient inquiries and needs during study conduct.

Conclusion

Randomized clinical trials, while traditionally the gold standard of evidence, have several limitations, chief of which is their lack of external validity and consequently a limited ability to impact real-world decision making. Due to recent changes in laws and regulations, including the realization that real-world evidence has an important role to play in informing medical decision making, the pragmatic study design has become an attractive alternative that can address both regulatory and payer needs. Implementing PRTs that a) meet the requirements of regulators and payers, with the ultimate goal of bringing new health technologies to patients quicker and more efficiently, and b) provide the evidence to persuade decision makers to change policies to enable access to those treatments by healthcare providers and their patients is undoubtably a challenge. However, understanding those challenges and how to overcome them through optimizing study design, leveraging existing and comprehensive electronic data stores and technology, and applying data science and operational expertise to generate robust data that demonstrates causal relationship treatment and effect, collectively represent a big step towards making that a reality. n

REFERENCES

1. Government Publishing Office. 21st Century Cures Act. H.R. 34, 114th Congress. 2016. Available at: https://www.gpo.gov/fdsys/pkg/BILLS-114hr34enr/pdf/BILLS-114hr34enr.pdf. Accessed September 5, 2019.

2. US Food and Drug Administration. Framework for FDA's Real-World Evidence Program. December 2018. Available at: https://www.fda.gov/media/120060/download. Accessed September 5, 2019.

3. RE-AIM. Available at: www.RE-AIM.org. Accessed September 5, 2019.

4. Glasgow RE, Vogt TM, Boles SM. Evaluating the Public Health Impact of Health Promotion Interventions: The RE-AIM Framework. Am J Public Health. 1999 Sep;89(9):1322–1327. doi:10.2105/ajph.89.9.1322.

5. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 Tool: Designing Trials That Are Fit for Purpose. BMJ. 2015 May 8;350:h2147. doi:10.1136/bmj.h2147.

6. ICH. ICH Harmonised Guideline. Estimands and Sensitivity Analysis in Clinical Trials , E9(R1). Current Step 2 Version, Dated 16 June 2017. Available at: www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9-R1EWG_Step2_Guideline_2017_0616.pdf. Accessed September 5, 2019.

7. Feltelius N, Gedeborg R, Holm L, Zethelius B. Utility of Registries for Post-Marketing Evaluation of Medicines. A Survey of Swedish Health Care Quality Registries from a Regulatory Perspective. Ups J Med Sci. 2017 Jun;122(2):136-147. doi: 10.1080/03009734.2017.1285837. Epub 2017 Mar 3.

8. Erlinge D, Omerovic E, Fröbert O, et al. Bivalirudin versus Heparin Monotherapy in Myocardial Infarction. N Engl J Med. 2017 Sep 21;377(12):1132-1142. doi: 10.1056/NEJMoa1706443. Epub 2017 Aug 27.

9. Varas-Lorenzo C, Castellsague J, Stang MR, Tomas L, Aguado J, Perez-Gutthann S. Positive Predictive Value of ICD-9 Codes 410 and 411 in the Identification of Cases of Acute Coronary Syndromes in the Saskatchewan Hospital Automated Database. Pharmacoepidemiol Drug Saf. 2008 Aug; 17(8):842-852. doi: 10.1002/pds.1619.

10. Coloma PM, Valkhoff VE, Mazzaglia G, et al. Identification of Acute Myocardial Infarction from Electronic Healthcare Records Using Different Disease Coding Systems: A Validation Study in Three European Countries. BMJ Open. 2013 Jun 20;3(6). pii: e002862. doi: 10.1136/bmjopen-2013-002862.

11. Wei WQ, Feng Q, Weeke P, Bush W, Waitara MS, Iwuchukwu OF, Roden DM, Wilke RA, Stein CM, Denny JC. Creation and Validation of an EMR-Based Algorithm for Identifying Major Adverse Cardiac Events while on Statins. AMIA Jt Summits Transl Sci Proc. 2014 Apr 7; 2014:112-9. eCollection 2014.

12. Eikelboom JW, Connolly SJ, Bosch J, et al. Rivaroxaban with or without Aspirin in Stable Cardiovascular Disease. N Engl J Med. 2017 Oct 5; 377(14):1319-1330. doi: 10.1056/NEJMoa1709118. Epub 2017 Aug 27.

13. Berger A, Simpson A, Bhagnani T, Leeper NJ, Murphy B, Nordstrom B, Ting W, Zhao Q, Berger JS. Incidence and Cost of Major Adverse Cardiovascular Events and Major Adverse Limb Events in Patients with Chronic Coronary Artery Disease or Peripheral Artery Disease. Am J Cardiol. 2019 Jun 15;123(12):1893-1899. doi: 10.1016/j.amjcard.2019.03.022. Epub 2019 Mar 16.

14. Clinical Trials Arena. PPD and China's HLT to Offer Data Science-Driven Clinical Trial Solutions. 2019 Feb 22. Available at: https://www.clinicaltrialsarena.com/news/ppd-hlt-data-science-driven-clinical-trial-solutions/. Accessed September 5, 2019.

15. US Food and Drug Administration. FDA's Sentinel Initiative. Transforming How We Monitor the Safety of FDA-Regulated Products. Available at: https://www.fda.gov/safety/fdas-sentinel-initiative. Accessed September 5, 2019.

16. Sentinel. Sentinel Common Data Model. Available at: https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model. Accessed September 5, 2019.

17. HL7 FHIR. Available at: www.hl7.org/fhir/. Accessed September 5, 2019.

18. Walinjkar A, Woods J. FHIR Tools for Healthcare Interoperability. Biomed J Sci Tech Res. 2018; 9(5).

19. HealthData Management. Slabodkin G. CMS Pilot Taps FHIR to Give Clinicians Access to Claims Data. 2019 Aug 3. Available at: www.healthdatamanagement.com/news/cms-pilot-taps-fhir-to-give-clinicians-access-to-claims-data. Accessed September 5, 2019.

20. Adaptable, The Aspirin Study – A Patient-Centered Trial. Available at: https://theaspirinstudy.org/. Accessed September 5, 2019.