

Workshop 2019:

***Use of Real World Data/Evidence
to Inform Regulatory Decision Making***

December 3, 2019
Tunney's Pasture, Ottawa, Canada

A joint workshop of:

**Health Canada
&
Canadian Society for Pharmaceutical Sciences**

Conference Chairs:

Cathy Lau, Janssen Canada, Toronto, Canada
Co Pham, Health Canada, Ottawa, Canada

Scientific Chairs:

Robert Platt, McGill University, Montreal, Canada
Patrick Ryan, Janssen R&D
Nirosha Mahendraratnam Lederer, Duke-Margolis Center for Health Policy

Conference Organizing Committee:

Health Canada: Agnes Klein, Co Pham
CSPS: Cathy Lau, Fakhreddin Jamali, Raimar Loebenberg, Arshia Ghani

REMARKS FROM ORGANIZERS

Cathy Lau
Janssen Canada
clau2@its.jnj.com

Agnes Klein
Health Canada
agnesv.klein@canada.ca

Co Pham
Health Canada
co.pham@canada.ca

Real world data (RWD) and real world evidence (RWE) are playing increasing roles in health-care decisions. Real world data are routinely employed to support reimbursement and coverage decisions for drugs and devices. More recently, clinical trials incorporating pragmatic designs and observational studies are considered to supplement traditional clinical trials (e.g., randomized clinical trials). Regulatory agencies and large co-operative groups including academia and industry are exploring whether leveraging big databases such as electronic medical records and claims databases can be used to garner clinical insights extending beyond those gained from randomized controlled studies. Whether RWE can ultimately replace or improve traditional clinical trials is the big question.

The workshop held on December 3, 2019 at Health Canada included presenters from regulatory agencies, industry and academia. Health Canada, US FDA and European Medicine Agency presented current thinking, draft frameworks and guidance available in the public domain. While the three agencies might be at different stages of utilizing RWE for regulatory decision making, the consensus is not whether RWE would be used but when and how it can be incorporated into regulatory decision making while maintaining a high evidentiary bar.

The complexity of data sourcing, curating databases, aligning on common data models, illustrated by high-profile work conducted as part of Sentinel, DSEN, OHDSI and Duke-Margolis initiatives, was presented and discussed during the workshop, creating great learning opportunities for the attendees. The design and analysis of RWE studies were compared and contrasted to those of RCTs. While there are gaps, they are closing quickly as novel analytical methods are employed and innovative ways of curating data, including natural language processing and artificial intelligence, are explored.

This proceeding contains summaries of information presented by the speakers, including current highlights about the use of RWE in regulatory decision making. In the world where the uptake of “big data” in everyday life is happening at unprecedented speed, we can expect RWE to be a fast-moving area and with the potential for big impact in health-care decision making in the years to come.

TABLE OF CONTENTS

SESSION 1: REGULATORY FRAMEWORKS FOR REAL WORLD EVIDENCE

<i>Health Canada's Efforts to Optimize the Use of Real World Evidence for Regulatory Decision Making</i>	5
Gayatri Jayaraman , Health Canada	
<i>FDA's Real World Evidence Program</i>	6
John Concato , Food and Drug Administration	
<i>Real World Evidence: European Medicines Agency Experience</i>	10
Peter Arlett , European Medicines Agency	

SESSION 2: DATA SOURCES AND DATA QUALITY – MEETING THE BAR FOR ‘FIT FOR PURPOSE’ IN REGULATORY DECISION MAKING

<i>Canadian Network for Observational Drug Effect Studies (CNODES): Data Quality Assurance Practices</i>	14
Robert W. Platt and Kristian B. Filion , McGill University	
<i>Medical Product Safety Surveillance: Data Quality in the Sentinel Initiative</i>	18
Judith C. Maro , Harvard Pilgrim Health Care Institute	

SESSION 3: CAN RWE REPLACE TRADITIONAL CLINICAL TRIALS IN REGULATORY DECISION MAKING? WHAT METHODOLOGIES ARE NEEDED, AND IN WHAT CONTEXT CAN RWE BE USED?

<i>Promise and Pitfalls of Replacing Traditional Clinical Trials for Regulatory Decision Making</i>	22
Joseph S. Ross , Yale School of Medicine	
<i>Randomized Trials and Observational Studies: What's in a Name?</i>	25
John Concato , Food and Drug Administration	
<i>Should We Replace or Enhance? Modernizing Clinical Trial Designs Using Real-World Data</i>	30
Elodie Baumfeld Andre , Pfizer Inc.	
<i>An Industry Perspective on Real-World Evidence in Regulatory Decision Making: An Example of Transparency in Pharmacoepidemiologic Research</i>	34
James Weaver , Janssen R&D	

SESSION 4: DEMONSTRATING HOW REAL WORLD DATA/EVIDENCE CAN BE USED IN SUPPORT OF REGULATORY APPROVALS

Demonstrating and Communicating Real World Data (RWD) Reliability to Support Fitness for Use for Regulatory Decision-Making..... 38
Nirosha Mahendraratnam Lederer, Duke-Margolis Center for Health Policy

What's Next? Pathways to Unlock the Value of RWD from Oncology EHRs..... 41
Nicole Mahoney, Flatiron Health Inc.

Real-world data curation to transform medical investigation: Technology to reimagine the economics of evidence gathering and support regulatory decision-making 44
Aaron Leibtag, Pentavere Research Group Inc.

SESSION 1:

Regulatory Frameworks for Real World Evidence

Health Canada's Efforts to Optimize the Use of Real World Evidence for Regulatory Decision Making

Gayatri Jayaraman, MPH, PhD
Health Canada
gayatri.jayaraman@canada.ca

The workshop began with a presentation by Dr. Gayatri Jayaraman on Health Canada's efforts to optimize the use of real world evidence (RWE) for regulatory decision making. Health Canada's RWE efforts have evolved in the context of the Regulatory Review of Drugs and Devices (R2D2) initiative, with the goal of improving the assessment of health product safety and effectiveness across the lifecycle. This initiative is conducted in collaboration with Health Technology Assessment (HTA) agencies such as CADTH and INESSS, and other stakeholders through the Core Action Team comprised of payers, academics, industry, and key data holders such as the Canadian Institute for Health Information (CIHI).

Health Canada already uses RWE to inform regulatory decisions. In the pre-market space, RWD has primarily been used to generate supportive rather than pivotal evidence. In the post-market space, pharmacovigilance through a triangulation of approaches including the implementation of sound risk management plans and the analysis of spontaneous reports and other types of RWD are used to inform regulatory decisions including making changes to indications and product monographs, label revisions, and product recall. Health Canada also undertakes its own research, largely in collaboration with Canada's Drug Safety and Effectiveness Network (DSEN), which includes Canadian Network for Observational Drug Effect Studies (CNODES). Internationally Health Canada is collaborating with international regulators through the International Council on Harmonisation, with the FDA through their Sentinel Network, and with the EMA on executing common protocols.

However, we can do more. The intent is to create an agile 'learning by doing' environment which explores how RWD might be optimised to inform health product safety and effectiveness during pre-market approvals and for ongoing post market surveillance. It is important to recognize, for example, that clinical trials may not include certain segments of the population, like the paediatric population, pregnant women, and seniors. Clinical trials may also be unfeasible or unethical, for example, in the context of rare diseases. Health Canada advises sponsors to refer to the notice posted on their website (<https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/announcements/optimizing-real-world-evidence-regulatory-decisions.html>) for further information and to contact Health Canada to determine whether pre-submission meetings may be required.

The desired state is one where RWE is used appropriately across the entire lifecycle of health products in order to improve timely access to safe and effective health products for all Canadians.

FDA's Real World Evidence Program

John Concato, MD
Food and Drug Administration
John.Concato@fda.hhs.gov

The second presenter was Dr. John Concato from the Office of Medical Policy in the Center for Drug Evaluation and Research, U.S. Food and Drug Administration (FDA), who spoke about FDA's Real-World Evidence Program. In the U.S. context, the real-world evidence program was formalized by Section 3022 of the 21st Century Cures Act,¹ passed by Congress in December 2016. The Act mandated that FDA establish a program to evaluate the potential use of RWE for supporting a new indication for an approved drug, or for supporting or satisfying post-approval study requirements. The Act also instructed the agency to develop a draft framework within two years, and in response the *Framework for FDA's Real-World Evidence Program*² was published in December 2018. An additional requirement instructed the agency to develop draft guidance for industry within five years of the Act's passage, i.e., by December 2021. In parallel, the commitment letter for the Prescription Drug User Fee Act (PDUFA) VI, covering 2018-2022, affirms that the agency will explore using of real-world evidence in regulatory decision-making.

Definitions of real-world evidence (RWE) and real-world data (RWD) can differ, and FDA's *Framework* formally defined RWE as "clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of real-world data."² In turn, real-world data (RWD) was defined as "data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources."² Beyond definitions, three main considerations summarize FDA's approach to RWD/RWE; see Figure 1. When evaluating real-world evidence, one consideration is whether the real-world data involved are fit for use. A second consideration is whether the trial or study design used to generate the evidence can provide adequate scientific support to address the relevant regulatory question. A third consideration is whether the study itself meets regulatory requirements (e.g., for study monitoring and data collection). These three elements also outline the basis for how FDA will evaluate a sponsor's RWE-based application to the agency.

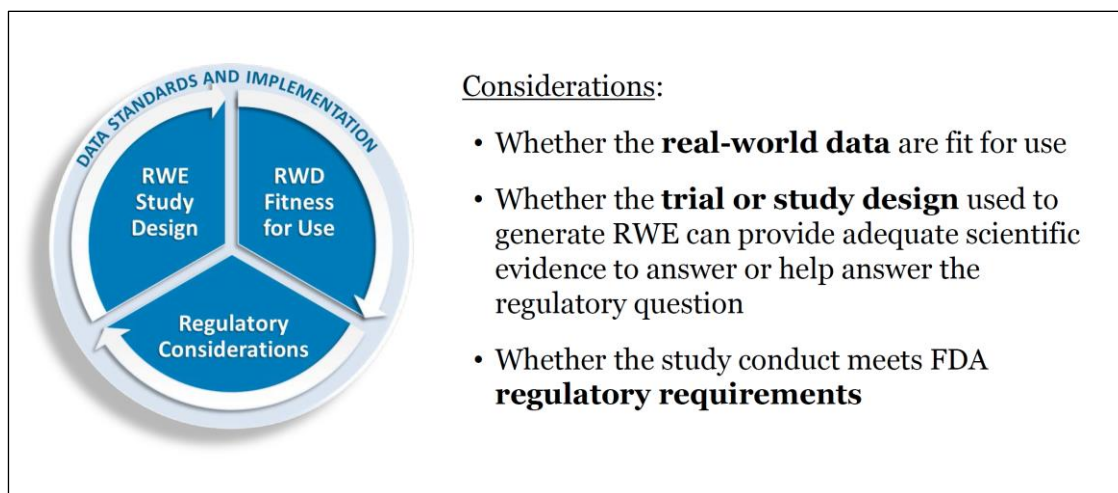


Figure 1. Overview of FDA's conceptual approach for evaluating real-world evidence.

Regarding fit-for-use data, and given that another session of this workshop covers real-world data in more detail, Dr. Concato mainly emphasized that the FDA does not endorse any one type of RWD. Indeed, to address a given research question, a single source may not capture all data elements, and using judgement in applying standards of good scientific practice is needed to select suitable sources. The FDA *Framework*² also discusses the concepts of data reliability and relevance, as well as related aspects such as data standards, digital health tools, and interoperability of data systems.

Figure 2 illustrates the spectrum of study designs that are discussed in the framework, with increasing reliance on RWD moving from the left- to right-side of the graphic. On the left, for example, and by using elements of RWD within traditional randomized trials, enrollment and feasibility can be assessed, and electronic health records or claims data can be used to capture selected data elements. In the middle of Figure 2, randomized controlled trials (RCTs) in clinical practice settings are highlighted, representing so-called pragmatic designs. Also, as a non-randomized approach with an interventional component, a single-arm study can be compared to an external control arm, albeit with caveats and concerns (beyond the scope of this presentation). Furthest to the right are familiar epidemiologic study types, including cohort and case-control studies as non-interventional (“observational”) designs that also have limitations.

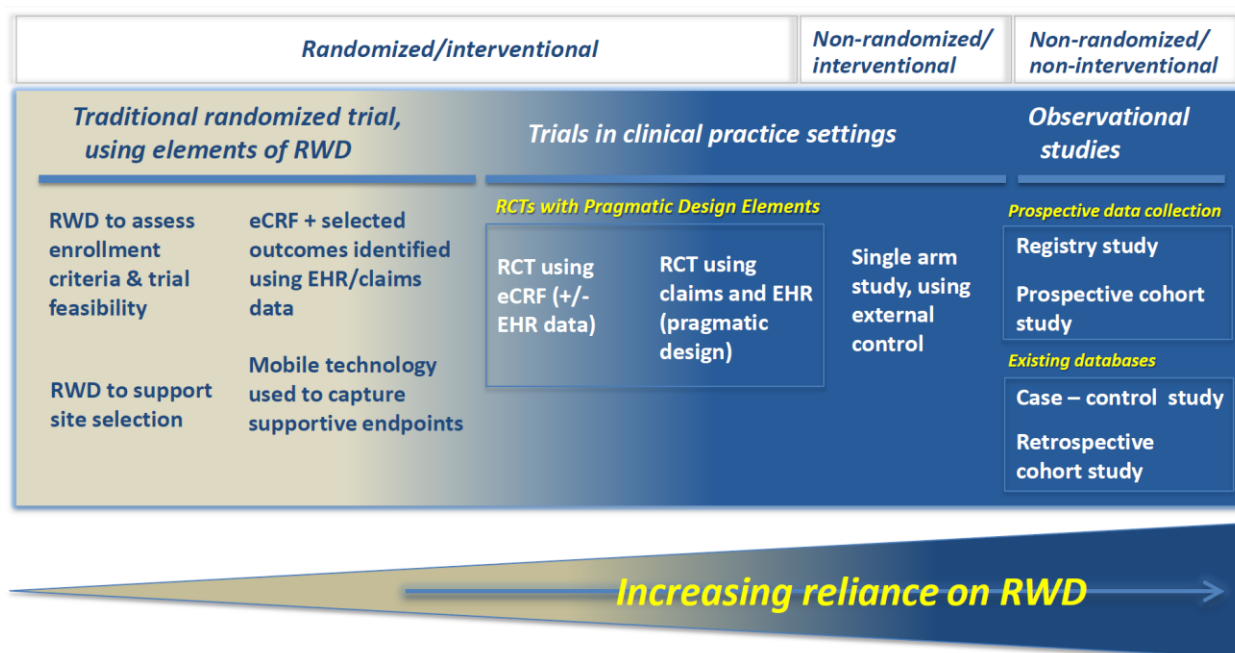


Figure 2. Overview of selected study designs involving real-world evidence.

Regarding regulatory standards for RWE, the FDA and Health Canada can be characterized informally as having a similar perspective. In the U.S., the requirement for substantial evidence derived from adequate and well-controlled studies – based on 21 Code of Federal Regulations 314.126 – remains in effect. Specifically, studies of a cause-effect (treatment-outcome) association should be free of other influences, such as spontaneous changes in diseases course, the placebo effect, or biased observations. In the current era of RWE, a key question is: *When do various studies rise to the level of substantial evidence?* This topic will be covered later in the workshop, but a prominent take-home

message is that a dichotomy of “randomized versus observational studies” is an overly simplistic viewpoint. In brief, counterarguments include a) real-world data being fit-for-use is often a more pressing concern than research design, and b) methods for conducting and analyzing non-interventional (observational) studies have improved over time, albeit not necessarily leading to regulatory acceptance.

The next component of the presentation highlighted representative RWD/RWE demonstration projects involving the FDA and various collaborators. For example, and as an early example of work in this space, an ancillary study linked to the HARMONY-OUTCOMES trial³ is examining how EHRs can be suitable for recruitment of participants, as well as for assessment of baseline measures and endpoints. Another example, the IMPACT-AFib study,⁴ is a proof-of-concept design using the FDA Sentinel infrastructure for conducting interventional effectiveness trials involving RWD. In addition, the FDA has commissioned a formal comparison of observational and randomized evidence done on the same topics, with the goal of determining what can be learned from such comparisons; this effort is reflected in the RCT DUPLICATE project,⁵ with more detail contained in Session 3.

Looking forward with regard to FDA’s guidance development, the agency’s approach is to focus on a number a number of topic areas, including (but not limited to): assessing fitness of RWD sources when used for regulatory decisions; data standards, including criteria for collection and submission to FDA; considerations regarding decentralized clinical trials, registries, and other approaches; potential for observational study designs using RWD to support effectiveness decisions; and regulatory considerations for study designs using RWD/RWE. Dr. Concato noted that draft guidance documents will become available for comment as each work product is completed. Final guidance documents will be developed subsequently, after review of any feedback that is received.

The presentation wrapped-up by returning to pertinent topics such as the potential value of EHR data in research activities, and how different groups and institutions have already been improving data-capture methodologies. Also, and as a coincidence of timing, a recently published article describing the “large-scale assessment of a smartwatch to identify atrial fibrillation”⁶ was mentioned as being representative of how digital technology is expanding the scope of real-world data. Regulatory agencies will have to keep up with, and anticipate future changes in, such technological advances.

In closing, the three main take-home messages from this presentation include: a) FDA’s Real-World Evidence Program is advancing as outlined in the December 2018 RWE *Framework*; b) key considerations for RWE include fit-for-use data, adequate study design, and appropriate regulatory requirements; and c) ongoing activities related to RWD/RWE will inform the preparation of FDA guidance documents, as well as the agency’s practice in terms of medical product evaluation.

Disclaimer: This article reflects the perspective of the author and should not be construed to represent the views or policies of the FDA.

References

1. H.R. 34 — 114th Congress: 21st Century Cures Act. 2016. <https://www.govtrack.us/congress/bills/114/hr34>. Accessed 17 Feb 2020.
2. US Food and Drug Administration (FDA). Framework for FDA's Real-World Evidence Program. 2018. <https://www.fda.gov/media/120060/download>. Accessed 17 Feb 2020.
3. Hernandez AF, Green JB, Janmohamed S, D'Agostino RB, Sr., Granger CB, Jones NP, et al. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. *Lancet*. 2018 Oct 27;392(10157):1519-29.
4. Vinereanu D, Lopes RD, Bahit MC, Xavier D, Jiang J, Al-Khalidi HR, et al. A multifaceted intervention to improve treatment with oral anticoagulants in atrial fibrillation (IMPACT-AF): an international, cluster-randomised trial. *Lancet*. 2017 Oct 14;390(10104):1737-46.
5. Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital. RCT DUPLICATE: Randomized Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology. <https://www.rctduplicate.org/>. Accessed 17 Feb 2020.
6. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. *N Engl J Med*. 2019 Nov 14;381(20):1909-17.

**Real World Evidence:
European Medicines Agency Experience**

Peter Arlett, MD, FRCP
Head of Data Analytics and Methods Task Force
European Medicines Agency
Peter.Arlett@ema.europa.eu

Box 1: Key points

- EMA strives to meet the needs of patients for new and better therapies.
- Already today, RWD forms an important part of the evidence that we use for regulatory decision making, complementing the foundational evidence from clinical trials.
- The type of evidence chosen to support regulatory decisions may vary with the nature of the decision and the types of evidence available.
- To further enable use of RWD in regulatory decision-making we will build on the recommendations of the Big Data Task Force [1] to:
 - obtain greater access to data;
 - develop a data quality framework;
 - ensure robust data governance;
 - invest in analytics technology,
 - build new regulatory processes, and,
 - train people across stakeholder groups.
- The RWE approach of the European regulators is to deliver more through a multi-stakeholder and international approach.

The European Union is made up of about half a billion citizens and we have 24 official languages. This is highly relevant for real world data, because electronic health records, claims and registry data reflect the different languages, different cultures, and different systems of delivering medicine across the EU. Many EU Member States have universal coverage and systems of general practitioners as gate-keepers to healthcare meaning that individual patients can often be followed up over many years with all, or at least most, of their medical interventions recorded.

In 2018 we studied the different electronic healthcare databases that are available in the European Union to look at whether they are accessible and potentially of use for regulatory decision making [2]. We found that only 13 member states had electronic healthcare databases accessible and with potential to support some regulatory decisions. However, the situation is evolving positively and evolving fast. In the last couple of years initiatives have taken place in different Member States to enable research based on real world data. Examples include new laws in Finland enabling research access to data and major investments in France, Denmark and Germany.

To realise the potential of RWD to support regulatory decision-making we need to invest in data quality, in study methods, and in the regulatory framework. In addition, we need to fully understand the reliability of evidence obtained from RWD and how steps of the process used to obtain such evidence can be verified. In the EU, we believe that if we work on all of these elements, we will be able to bridge the gap between data and evidence and we will be able to deliver better medicines to patients faster and to optimize the safe and effective use of those medicines on the market.

Real world data and real world evidence are not new. However, the utility of real world data will be very dependent on the regulatory use. Examples of these potential or current uses are included in Box 2. Real world evidence is best established in the study of diseases and in product safety. The study of product efficacy is the most controversial. That said, real world evidence is already used to demonstrate efficacy in certain very rare diseases or diseases where randomized clinical trial is not ethical [3]. Indeed, if we use a randomized control trial to demonstrate efficacy, i.e. that the drug works, we can then use real world evidence to show how the drug performs in terms of effectiveness when it's used in clinical practice, based on the already-demonstrated efficacy.

Box 2: examples of regulatory uses for real-world evidence

Disease epidemiology

- identifying unmet medical needs;
- understanding the natural course of disease;
- studying disease incidence and prevalence;
- demonstrating differences in clinical practice between different member states;
- comparison of surrogate and clinical outcomes;
- development of clinical predictor models for treatment response;
- measurement of background rates of events, e.g. for the study of vaccine safety;
- characterization and representativeness of patients in disease registries.

Product-specific use cases

- understanding how product is used in clinical practice;
- monitoring for safety;
- monitoring the effectiveness or risk minimization;
- extrapolation of safety and effectiveness from the clinical trial population to different populations such as the very elderly or children;
- extending indications beyond the conditions studied in clinical trials.

EMA has been conducting RWD studies in-house for at least a decade. We analyse pseudo-anonymised electronic health record data held in-house to support our committees' decision making. We currently have in-house electronic health records from the UK, France, and Germany and between 2013 and 2019, we conducted 88 in-house studies

where the results were fed into one of our formal decision-making committees. The majority of this work has been on product safety reflecting the established role for RWE in this area of regulatory work. Examples in hydrochlorothiazide and skin cancer, peripheral neuropathy and fluoroquinolones, autism spectrum disorder and antidepressant use in pregnancy, and psychiatric adverse events following exposure to levonorgestrel-containing intrauterine devices. However, we also have examples where we've gone beyond drug safety and in the future we'll be investigating potential uses in both product benefits and risks on the market, what many would call "product performance."

At EMA, to complement in-house studies we also have contracts in place with academia for studies. The advantage here is that many of the datasets in Europe are not directly accessible by the EMA. By having contracts with academics we can then access data on a far greater number of patients. In 2019, we've commissioned two studies on sodium valproate in pregnancy, two on retinoids in pregnancy, and we have just commissioned a large multinational study on the use of ranitidine to help us manage the current evaluation of nitrosamine contamination of ranitidine products. Commissioned studies are all registered in the EU PAS Register to ensure high levels of transparency [4].

International collaboration between regulators is established and is strengthening. One example is the study comparing DOACs where we agreed a common protocol between Health Canada and EMA, and then commission a study to look at bleeding risk in regions in Canada as well as different member states in the European Union. By doing so, we had the power to look at differences in practice, to study a very large number of patients exposed and compare adverse outcomes between products. On the European side, the results of these studies, as well as a parallel study that was conducted by the FDA, were considered by the EMA's main benefit risk committee (CHMP), to see if there is any need for product labelling changes. This is a really important piece of work showing the power of collaboration.

How are the European regulators facilitating the use of RWD for decision-making? In 2019 the EMA published the OPTIMAL framework for regulatory use of real world evidence. OPTIMAL stands for "Operational, Technical, and Methodological." In each of these areas you can find challenges, and in each of these areas we need to make efforts to address those challenges in order to realize the potential of real world data [3]. Complementary to the OPTIMAL framework, in 2020, three years of work conducted by the Big Data Task Force of the EMA and heads of national medicines agencies concluded with the publication of its final report and ten priority recommendations on Big Data with a particular focus on real world data [1]. Box 3 lists some of the EU initiatives relevant to facilitating use of RWD.

Box 3: examples of EU initiatives to facilitate the use of Real World Data

- ENCePP: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance [5]
- The EHDEN project [6]
- GetReal Initiative [7]
- VAC4EU [8]
- European Medicines Agency Registries Initiative [9]
- Scientific Advice (including qualification advice) [10]

EMA strives to meet the needs of patients. We already use real world data as an important part of our evidence base and the evidence value depends on the regulatory use, products, diseases, and the data itself. There's a need for a lot of work to fully realise the potential of real world data in regulatory decision-making and EMA will work with multiple stakeholders and international partners to deliver through collaboration, guided by the OPTIMAL framework and the priority recommendations of the Big Data Task Force.

***Disclaimer:** The views expressed in this article are the personal views of the author and may not be understood or quoted as being made on behalf of or reflecting the position of the EMA.*

References

1. Big Data Task Force report: https://www.ema.europa.eu/en/documents/other/hma-ema-joint-big-data-taskforce-phase-ii-report-evolving-data-driven-regulation_en.pdf
2. Pacurariu A. et al. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. *BMJ Open* 8(9):e023090. doi:10.1136/bmjopen-2018-023090 (2018).
3. Alison Cave, Xavier Kurz, and Peter Arlett. Real world data for regulatory decision making: challenges and possible solutions for Europe. *Clinical Pharmacology and Therapeutics* 2019. DOI:10.1002/cpt.1426.
4. EU PAS Register: <http://www.encepp.eu/encepp/studiesDatabase.jsp>
5. ENCePP: <http://www.encepp.eu>
6. The EHDEN project: <https://www.ehden.eu/>
7. GetReal Initiative: <https://www.imi-getreal.eu/GetReal-Initiative>
8. VAC4EU: <https://vac4eu.org/>
9. European Medicines Agency Registries Initiative: <https://www.ema.europa.eu/en/human-regulatory/post-authorisation/patient-registries>
10. Scientific Advice (including qualification advice): <https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance>

SESSION 2:

**Data Sources and Data Quality
Meeting the Bar for ‘Fit for Purpose’
in Regulatory Decision Making**

**Canadian Network for Observational
Drug Effect Studies (CNODES):
Data Quality Assurance Practices**

Robert W. Platt, PhD
McGill University
Co-lead, Canadian Network for Observational Drug Effect Studies
robert.platt@mcgill.ca

Kristian B. Fillion, PhD, FAHA
Departments of Medicine and of Epidemiology, Biostatistics, and Occupational Health
McGill University
kristian.fillion@mcgill.ca

The Canadian Network for Observational Drug Effect Studies (CNODES) is a network funded through a partnership between Health Canada and the Canadian Institutes of Health Research (CIHR) run under the Drug Safety and Effectiveness Network office of the CIHR ^{1,2}.

CNODES uses population-based administrative health care data, that is, health care claims data collected through provincial insurers and others, to provide timely responses to queries from Canadian stakeholders regarding drug safety and effectiveness. Typically, a query will come from Health Canada, and CNODES conducts a study in response. It is a national network with representation in almost all of the provinces across the country. The network is distributed, in that investigators are situated in each province (with the exception of the Atlantic provinces, which are overseen by a single site in Nova Scotia). This distributed nature is dictated primarily by the need for data security and the interest of individual sites and their data custodians in keeping data in-house. The key principle behind CNODES is that rather than conduct separate individual studies in each of the sites, CNODES conducts a coordinated set of studies with common protocols and planning with subsequent meta-analysis, with an end result that is both higher power and higher quality.

CNODES is run by a coordinating centre housed at McGill University and the Lady Davis Institute of the Jewish General Hospital. CNODES also manages four cross-cutting teams that work with the coordinating centre. The Database Team manages our databases and ensures access to data on an ongoing basis, as well as working on measurement and on definitions of key variables. The Methods Team works on best practices in statistical and epidemiologic methods, supports query-related studies, and conducts methods research ³. We also have a Knowledge Translation (KT) Team which both conducts active KT with stakeholders and conducts KT research ⁴. Finally, the

Training Team organizes seminars and practical sessions to support professional development within the network, facilitates capacity building among network trainees, and conducts online training programs and other activities.

The typical CNODES query proceeds as follows. A query comes from Health Canada (or other government partner; referred to as a query submitter) via the Drug Safety and Effectiveness Network. They may ask, for example, for a study of the association between incretin-based drugs and pancreatic adverse events⁵⁻⁷. The submission of a query leads to an initial discussion between the query submitter and CNODES to refine the research question and assess feasibility. Once the question is determined and the feasibility confirmed, CNODES scientists develop a scientific protocol. This protocol is typically developed by a team of experts within CNODES, including clinical experts as needed, in collaboration with query submitters to ensure that the study is answering the question they have asked. It is very important to ensure that the scientific question corresponds to the regulatory need. The scientific protocol is registered with clinicaltrials.gov, in the same way that a trial protocol might be, in order to ensure that the study is conducted in a transparent and replicable way. The protocol includes pre-specification of variables and analyses, and allows for site-specific analyses to be conducted and deposited with the Coordinating Centre blind to the results from other sites; results are combined by an independent analyst at the Coordinating Centre.

CNODES undertakes several key steps during study conduct to ensure study and data quality. The first key step to ensure quality and to minimize errors in a CNODES study is to create a detailed statistical analysis plan. Since the analysts conducting the study are in separate sites, they need to have maximum detail on how to conduct the study while respecting their own data structures and limitations. Other networks such as Sentinel⁸ and the Observational Health Data Science Initiative (OHDSI)⁹ use more structured settings with automated or semi-automated analysis of data that has been mapped into a common data model so that coding is identical across sites. CNODES uses a version of this approach in some projects, but data holder restrictions do not permit network-wide implementation. The CNODES analysis plan is designed to be sufficient such that any analyst can use it on any of the CNODES databases and receive replicable results.

Second, CNODES conducts a phased implementation of our studies. A typical phase I is a preliminary look at the data to describe the patients using a specific drug, their characteristics, and the baseline risk of the outcome under study. Phase II is the actual (usually comparative) analysis of exposure and outcome. Phase I is done for two reasons. First, it informs whether the study is worth doing in Canada. If, for example, there are few or no exposures to the drug of interest, or if there are few or no incidences of the outcome in the relevant time window, CNODES stops conduct at the end of phase I. Second, this phase identifies data anomalies. If coding of key outcomes is substantially different in one site vs. another, this phase will often serve to detect it. In a CNODES study of domperidone and sudden cardiac death, we observed that the incidence ranged from 20 per 10,000 person-years to 53 per 10,000 person years across provinces¹⁰. Even taking possible health, social, and system-related reasons for a difference in outcome rate, these differences are not plausible. Phase I helps us eliminate these data anomalies before conducting further analysis¹¹.

Finally, CNODES uses a series of tools that are set up to ensure that they study works systematically. There are a set of tools for data analysts, project and protocol developers, the Coordinating Centre and for KT professionals to ensure consistency across CNODES studies.

Two examples illustrate aspects of the CNODES quality-assurance program. In a study of proton pump inhibitors and the incidence of community-acquired pneumonia, an important anomaly was identified at the analytic stage; while results were null across almost all sites, Nova Scotia data showed a strong apparent association¹². Investigation showed substantial differences in formulary between Nova Scotia and other sites. This led to an important lesson

learned – prior to initiating any study, formulary restrictions must be assessed. In addition to helping identify the most appropriate comparator, such restrictions can be an important source of heterogeneity and need to be considered when checking results for internal consistency across participating sites. In a study of high- vs low-dose statins and new-onset diabetes, a protective association a similar anomaly was observed; in this case, one site showed a substantial protective effect while all others showed a slight increased risk for users of high-dose statins¹³. A thorough investigation was undertaken including an audit of the SAS code used, and inter-site sharing of SAS code (in which an analyst at another site used the anomalous site's SAS code and verified that results were consistent with other results at that second site). These analyses, along with careful investigation of the formularies and other setting-specific differences, did not identify any structural reason for this difference. The heterogeneity observed in this study is consistent with other studies that have shown that unexpected findings can sometimes be explained by differences in data structure or capture, confounding due to different local conditions, and/or chance. This highlights the importance of replication, a key strength of CNODES.

In conclusion, networks like CNODES can put in place structures and procedures that ensure high-quality work. CNODES has adapted both network-level and study-level quality assurance processes. With the use of a distributed protocol and analytic approach, the primary attention has focused on protocol development and internal consistency, using external information where possible. While this level of consistency is very important, our procedures and the above examples also show the need for local expertise; individuals who know the data source and the local health system are best positioned to conduct the study in that environment. Study quality and data quality are, therefore, the responsibility of the entire network.

Acknowledgements

CNODES, a collaborating centre of the Drug Safety and Effectiveness Network (DSEN), is funded by the Canadian Institutes of Health Research (Grant Number DSE-146021). Dr. Platt holds the Albert Boehringer I Chair in Pharmacoepidemiology at McGill University. Dr. Filion is supported by a salary support award from the *Fonds de recherche du Québec – santé* and a William Dawson Scholar award from McGill University.

References

1. Suissa S, Henry D, Caetano P, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med* 2012;6(4):e134–40.
2. Platt RW, Henry DA, Suissa S. The Canadian Network for Observational Drug Effect Studies (CNODES): Reflections on the first eight years, and a look to the future. *Pharmacoepidemiol Drug Saf* 2019;6:1.
3. Platt RW, Dormuth CR, Chateau D, Filion KB. Observational Studies of Drug Safety in Multi-Database Studies: Methodological Challenges and Opportunities. *EGEMS (Washington, DC)* 2016;4(1):1–10.
4. Gamble JM, Traynor RL, Gruzd A, Mai P, Dormuth CR, Sketris IS. Measuring the impact of pharmacoepidemiologic research using altmetrics: A case study of a CNODES drug-safety article. *Pharmacoepidemiol Drug Saf* 2018;102(46):16569–16510.
5. Azoulay L, Filion KB, Platt RW, et al. Association Between Incretin-Based Drugs and the Risk of Acute Pancreatitis. *JAMA Intern Med* 2016;176(10):1464–1410.
6. Azoulay L, Filion KB, Platt RW, et al. Incretin based drugs and the risk of pancreatic cancer: international multicentre cohort study. *Br Med J* 2016;i581–8.

7. Filion KB, Azoulay L, Platt RW, et al. A Multicenter Observational Study of Incretin-based Drugs and Heart Failure. *N Engl J Med* 2016;374(12):1145–54.
8. Platt R, Brown JS, Robb M, et al. The FDA Sentinel Initiative — An Evolving National Resource. *N Engl J Med* 2018;379(22):2091–2.
9. Observational Health Data Sciences and Informatics. *The Book of OHDSI*. Independent; 2019.
10. Renoux C, Dell’Aniello S, Khairy P, et al. Ventricular tachyarrhythmia and sudden cardiac death with domperidone use in Parkinson’s disease. *Br J Clin Pharmacol* 2016;82(2):461–72.
11. Doyle CM, Lix LM, Hemmelgarn BR, Paterson JM, Renoux C. Data variability across Canadian administrative health databases: Differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf* [Internet] 2019; Available from: <http://dx.doi.org/10.1002/pds.4889>
12. Filion KB, Chateau D, Targownik LE, et al. Proton pump inhibitors and the risk of hospitalisation for community-acquired pneumonia: replicated cohort studies with meta-analysis. *Gut* 2014;63(4):552–8.
13. Dormuth CR, Filion KB, Paterson JM, et al. Higher potency statins and the risk of new diabetes: multicentre, observational study of administrative databases. *Br Med J* 2014;348(may29 6):g3244–g3244.

Medical Product Safety Surveillance: Data Quality in the Sentinel Initiative

Judith C. Maro, PhD
Harvard Pilgrim Health Care Institute
jmaro@mit.edu

The next presentation was by Dr. Judith C. Maro of Harvard Medical School and Harvard Pilgrim Health Care Institute about the U.S. Food and Drug Administration's (FDA) Sentinel Initiative.¹⁻³ The Harvard Pilgrim Health Care Institute is the lead institute and coordinating center for the Sentinel Operations Center. The Sentinel Operations Center works with the Sentinel Data Partners, which include US-based claims insurers and integrated delivery system organizations where clinical records and claims transaction records are held by the same organization. All data from the data partners remain on-site within their organizations and are stored in the [Sentinel Common Data Model format](#). This distributed data network structure incentivizes data partner participation as data partners have maximal local control over their data and its uses.^{4,5} In addition, the FDA and the Sentinel Operations Center leverages a network of scientific and technology partners that provide expertise and advice on methodological questions.

Data in the Sentinel Distributed Database include primarily claims-based information, electronic health record information, and information from state-based registries (e.g., death). The data model is designed to be flexible enough to accommodate new data domains (e.g., unstructured clinical text), but prioritizes parsimony and is expanded as needed when data are appropriate to add and deemed fit for purpose.⁶

The data are stored in the most granular level possible with a minimal amount of mapping. By keeping the data as raw as possible, one can use the querying process to do the hard work of combining data elements into medical concepts such as computable phenotypes. Algorithms or computable phenotypes are stored in a library for future use. For example, if the computable phenotype for congestive heart failure consists of a suite of diagnosis codes along with pharmacy codes indicating evidence of treatment, then that same computable phenotype can be retrieved from the library to be reused in a future study.

Sentinel relies on local data partner expertise to guide appropriate use and interpretation of the data. Therefore, not all data partners populate all data tables. This is especially important as the Sentinel Common Data Model is being used by more international partners, like the Canadian Network for Observational Drug Effect Studies (CNODES) and the United Kingdom's Clinical Practice Research Datalink (CPRD) for study implementation.

Sentinel uses a robust data quality review and characterization program to declare data fit for purpose as shown in a schematic in Figure 1. More than 1600 quality checks are performed across the Sentinel Common Data Model. Data quality checks have been divided up into levels for ease of use and understanding. The most basic checks – Level 1 checks – are for data model compliance to ensure that the data are in the expected format. For example, these checks ensure numeric variables are numeric. Level 2 checks ensure cross-table accuracy and consistency for any given patient. For example, these checks ensure that hospital admission dates precede discharge dates; that admission dates are consistent with a patient's active enrollment period; and that diagnoses and procedures that are labeled as inpatient occur during inpatient stays. Level 3 checks relate to data integrity over time. Data partners refresh their dataset quarterly or annually, and it is important to ensure that data stay consistent across each refresh and that the same patient can be tracked across refreshes. Finally, Level 4 checks are plausibility checks to ensure that implausible diagnoses or treatments are not present in the data. These are in development. When a data partner's dataset

successfully passes all these checks, it is deemed fit for purpose and the dataset passes into production and is available for routine querying. The Sentinel Operations Center has approved more than 200 datasets from 18 data partners over the lifetime of the program, with some data partners on their 30th refresh.

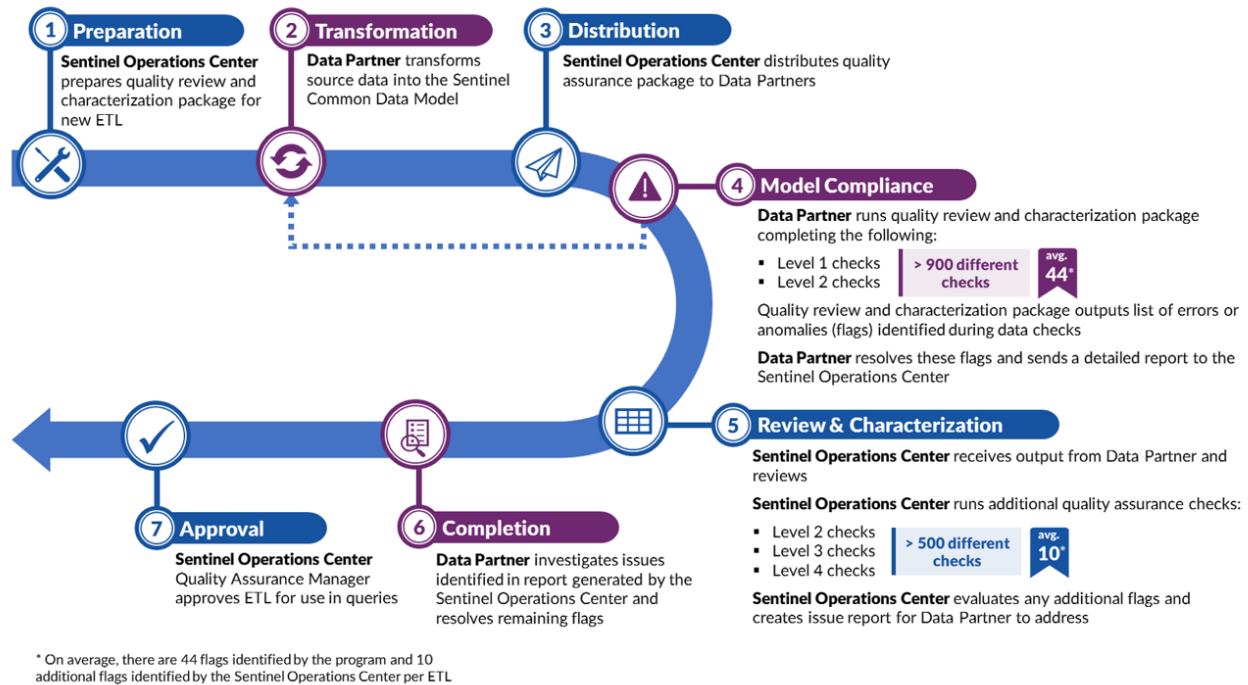


Figure 1. Sentinel’s Data Quality Review and Characterization Process (<https://www.sentinelinitiative.org/sentinel/data-quality-review-and-characterization>)

When the US FDA wants to ask a regulatory question, the Sentinel Operations Center works with the FDA to prepare a query (i.e., a SAS-based computer program). The Sentinel Operations Center distributes the query to the data partners through a secure network. They all receive the exact same copy of the program, download it, and execute it against their production version of the dataset. By completing these steps, the data partners are *opting in* to answer that regulatory question although they are not required or compelled to do so. Data partners review the query’s aggregated data results, and should they choose, they send these results back to the Sentinel Operations Center for aggregation and report creation for the US FDA. FDA and the Sentinel Operations Center strive to ensure that minimum necessary data standards are always followed.

The querying system used in Sentinel is called the Active Risk Identification and Analysis system (ARIA). Several different types of regulatory questions can be asked within ARIA. The concept of continuous and complete capture of patient information over time underlies all querying. That is, for every patient, is important to know that the absence of a particular diagnosis or outcome is because it did not occur, not because it was missing due to incomplete information. In Sentinel, this continuous and complete capture is operationalized as enrollment information on the patient, which is primarily sourced through claims data. As the model is used in more international jurisdictions like the United Kingdom, registration in a primary care catchment area is an analogous construct.

ARIA querying, as summarized in Figure 2 below, includes signal identification querying, descriptive feasibility querying (Level 1), and inferential querying either done once at the end-of-study (Level 2), or prospectively and repeatedly throughout a study period (Level 3). ARIA querying involves structured computer programs that use parameterized input files to tailor the analysis to whatever regulatory question is being asked. These computer programs have all been pre-tested and quality-checked, include standard output, and are available [for download](#).



Figure 2. The Active Risk Identification and Analysis (ARIA) querying system. (<https://www.sentinelinitiative.org/active-risk-identification-and-analysis-aria>)

First, [signal identification querying](#) identifies exposed cohorts of interest and elevated occurrences of anywhere from 8,000 to nearly 100,000 health outcomes of interest, operationalized as individual diagnosis codes.⁷ These signal identification queries use common epidemiological study designs such as propensity-score matching and self-controlled designs.

Level 1 ARIA queries develop thorough descriptive cohorts associated with a particular regulatory question. Level 1 queries are typically looking at medical product utilization, background rates of outcomes of interest, or unadjusted incidence rates. These data would support power calculations for possible inferential analysis. For example, FDA queried the Sentinel Distributed Database to estimate real-world off-label use of sodium-glucose cotransporter 2 inhibitors in patients with type 1 diabetes and the rates of diabetic ketoacidosis among those patients, comparing them with rates observed in sotagliflozin clinical trials.⁸

Level 2 and Level 3 ARIA queries are inferential analyses designed to yield a hazard ratio, a relative risk, or an odds ratio. Level 2 queries perform this estimation process once as in a typical retrospective epidemiologic study. Level 3 queries perform hypothesis tests sequentially and prospectively as data accrue, holding type 1 error control constant over the course of an entire analysis of analysis. Cohort designs and self-controlled designs are included in the ARIA querying suite of programs. An example of the former is a query to evaluate stroke risk among users of typical antipsychotics compared to users of atypical antipsychotics in a non-elderly and non-demented US population.⁹ An example of the latter is a query to evaluate seizure risk immediately following incident ranolazine exposure.¹⁰

New modules recently added to the ARIA querying system include the ability to perform a propensity-score matched analysis among pregnant women to estimate the risk of adverse birth outcomes (e.g., cardiac malformations) and the

ability to perform an interrupted time series analysis. Additionally, FDA and the Sentinel Operations Center have developed a user interface – Sentinel’s Query Builder – to facilitate building [medical utilization queries for organizations with data in the Sentinel Common Data Model format and a SAS license](#).

FDA has a commitment to transparency. Parameterized, inferential ARIA queries (i.e., the quality-checked computer programs) that have been used to make regulatory determinations are [published to enable easy replication by those with data in the Sentinel Common Data Model format](#). Additionally, the results of ARIA analyses [are published to the website to ensure widespread dissemination](#).

In September 2019, FDA announced Sentinel would [expand to three centers: the Sentinel Operations Center, Innovation Center, and Community Building and Outreach Center](#). The Sentinel Operations Center will continue to perform the activities described herein, primarily facilitating FDA’s ability to answer postmarket regulatory questions. The Innovation Center will develop innovative methods, focusing on novel ways to extract and structure information from electronic health records. The Community Building and Outreach Center’s will focus on communication and collaboration as well as deepening stakeholder involvement and broadening awareness, access, and use of Sentinel tools and data infrastructure.

References

1. Ball, R., Robb, M., Anderson, S. A. & Dal Pan, G. The FDA’s sentinel initiative--A comprehensive approach to medical product surveillance. *Clin. Pharmacol. Ther.* **99**, 265–268 (2016).
2. Platt, R. *et al.* The FDA Sentinel Initiative - An Evolving National Resource. *N. Engl. J. Med.* **379**, 2091–2093 (2018).
3. Dal Pan, G. J. Real-World Data, Advanced Analytics, and the Evolution of Postmarket Drug Safety Surveillance. *Clin. Pharmacol. Ther.* **106**, 28–30 (2019).
4. Maro, J. C. *et al.* Design of a National Distributed Health Data Network. *Ann. Intern. Med.* **151**, 341–344 (2009).
5. Brown, J. S. *et al.* Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med. Care* **48**, S45-51 (2010).
6. Curtis, L. H. *et al.* Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol. Drug Saf.* **21 Suppl 1**, 23–31 (2012).
7. Yih, W. K. *et al.* Assessment of Quadrivalent Human Papillomavirus Vaccine Safety Using the Self-Controlled Tree-Temporal Scan Statistic Signal-Detection Method in the Sentinel System. *Am. J. Epidemiol.* doi:10.1093/aje/kwy023
8. Hampp, C. *et al.* Use of Sodium-Glucose Cotransporter 2 Inhibitors in Patients With Type 1 Diabetes and Rates of Diabetic Ketoacidosis. *Diabetes Care* **43**, 90–97 (2020).
9. Taylor, L. G., Panucci, G., Mosholder, A. D., Toh, S. & Huang, T.-Y. Antipsychotic Use and Stroke: A Retrospective Comparative Study in a Non-Elderly Population. *J. Clin. Psychiatry* **80**, (2019).
10. Eworuke, E., Welch, E. C., Tobenkin, A. & Maro, J. C. Use of FDA’s Sentinel System to Quantify Seizure Risk Immediately Following New Ranolazine Exposure. *Drug Saf.* **42**, 897–906 (2019).

SESSION 3:

Can RWE Replace Traditional Clinical Trials in Regulatory Decision Making? What Methodologies are Needed, and in What Context Can RWE be Used?

Promise and Pitfalls of Replacing Traditional Clinical Trials for Regulatory Decision Making

Joseph S. Ross, MD, MHS
Yale School of Medicine
joseph.ross@yale.edu

Every day, patients and clinicians are faced with decisions about whether to use a medicine to treat a disease, weighing whether the potential benefits of that product will exceed its risks. However, the evidence available to inform those treatment decisions is limited. For instance, at the time of original regulatory approval by the U.S. Food and Drug Administration (FDA), the vast majority of clinical uses for new medicines are supported by no more than 2 pivotal randomized controlled trials demonstrating safety and efficacy.¹ Further, these trials are often small, of short duration, and are focused on surrogate measures of benefit: trials in aggregate supporting approval enroll fewer than 800 patients on average; 44% of drugs approved for chronic treatment were supported by a trial that was six months or longer, only 12% of a year or longer; and 45% were supported by trials exclusively focused on surrogate markers.¹ Moreover, even after market approval, most clinical uses for new medicines are not further evaluated in controlled clinical trials.² And there are low completion rates of clinical trials required by the FDA as a condition of marketing approval.³

Therefore, for patient and clinicians, as a medical product becomes available for use, there is still a tremendous amount of uncertainty of both its benefit and safety. This puts an onus on researchers, regulators, and industry to generate evidence to inform clinical decision-making as to which products are best to use to manage their disease and has been characterized as a lifecycle approach to regulation. However, the question remains as to how best to do this. Clearly, first seeking out opportunities to conduct randomized evaluations using pragmatically designed trials is best. These trials need to be practical, inclusive of the diverse patient populations receiving care in the real world, and engaged with health systems, patients, and clinicians. They also need to be relevant and the results need to be designed to directly inform decision making. Such trials are needed because in order to get more evidence-based practice, more practice-based evidence is required.⁴

But not every outstanding clinical question for which there is uncertainty can be answered with a clinical trial. And generating some type of clinical evidence is preferable to having no evidence whatsoever.⁵ Thus, when pragmatic clinical trials cannot be pursued, with the increasing availability of health and clinical data – from electronic health records (EHRs) to personal digital devices tracking sleep, activity and diet – and clear advances in statistical methods

and computing capacity, there are increasing opportunities to use ‘real-world data’ to generate ‘real-world evidence’, evidence that complements clinical trials and increases their efficiency.

Real-world data have been defined by the U.S. Food and Drug Administration (FDA) as “*data related to patient health status and/or the delivery of health care routinely collected from... electronic health records (EHRs), claims and billing data, data from product and disease registries, patient-generated data including home-use settings, and data gathered from other sources that can inform on health status, such as mobile devices.*”⁶ Importantly, this definition does not preclude randomization, and deliberately encourages embedding trials within routine clinical practice and using data available within EHRs for data collection and outcome ascertainment.

However, real-world data have important limitations that must be taken into consideration. For instance, many patients receive care at multiple health systems and from multiple clinicians who use different EHRs, which may not be connected. If one were to embed a trial within routine hospital care and use the EHR for outcome ascertainment, complete outcome data would be available for patients if the follow-up duration is through hospital discharge, but there would be missing data for many patients if the timeframe extends longer, e.g. 3 months or 1 year. Similarly, while EHRs are used for clinical management and documentation, they have been primarily designed for billing. Diagnosis codes and problem lists are not consistently accurate and useful for outcome ascertainment or risk adjustment; more sophisticated algorithms will be needed,⁷ including analysis of patient notes, a complicated endeavor. Even taking a best-case scenario approach, we have shown that only 15% of U.S.-based clinical trials published in high-impact journals in 2017 could be feasibly replicated through analysis of administrative claims or EHR data, a proportion so low because the trials’ intervention, indication, inclusion and exclusion criteria, and primary end points could not be routinely ascertained from these real-world data sources.⁸

Working with both the FDA and the Medical Device Innovation Consortium’s National Evaluation System for health Technology (NEST), along with a patient-centered data sharing platform, Hugo Health, our team has been testing approaches to ascertain multiple sources of real-world data to inform medical product evaluation efforts. We are aggregating multiple types of health data from multiple sources (e.g. EHR data from multiple healthcare systems), leveraging Blue Button technology or Application Programming Interfaces (APIs). This approach addresses the problem of fragmented longitudinal follow-up by aggregating EHR data from multiple healthcare systems, while also allowing ascertainment of patient-reported outcome measure (PROM) and other patient-generated health data. For instance, we conducted a prospective cohort study of 60 patients undergoing bariatric surgery or catheter-based atrial fibrillation ablation at two U.S. tertiary care hospitals and used Hugo to obtain and aggregate patient health data from multiple sources, including EHR data for all patients who underwent the procedures at both hospitals, as well as from 10 additional health systems; pharmacy data obtained for patients using CVS or Walgreens pharmacies; personal digital device data from physical activity monitors, digital weight scales, and single-lead ECGs; and PROM data obtained through surveys to assess post-procedure recovery and disease-specific symptoms.⁹ We are taking this same patient-centered approach to pragmatic clinical study design for additional projects to better understand patients’ experiences and outcomes with the Apple Watch among patients with atrial fibrillation and the Pear Therapeutics prescription digital therapy among patients with depression and insomnia, along with to characterize, among patients experiencing acute pain and receiving care in primary care settings, the emergency department, or dental clinics, patients’ pain severity and persistence, as well as use of prescription and over-the-counter pain medications.

In summary, while randomized controlled trials are generally considered the gold standard for the generation of clinical evidence, there is clear promise in the use of real-world evidence to evaluate the efficacy and safety of medical products. However, there are pitfalls to consider as well and it is unreasonable to expect that observational data can be used to address the same clinical questions being answered by traditional clinical trials. While real-world

data are not likely to obviate the need for traditional clinical trials, these data will soon offer valuable opportunities to generate complementary evidence, along with the potential to embed trials within routine clinical practice, and will advance our understanding of medical products that will ultimately lead to better care for patients.

Acknowledgements

Funding/Support and Role of the Sponsor: This work was not supported by any external grants or funds.

Potential Conflicts of Interest (representing the past 36 months): Dr. Ross declares the following interests: receiving research support through Yale University from Johnson and Johnson, from the Food and Drug Administration (FDA) (U01FD004585), from the Medical Devices Innovation Consortium (MDIC)/National Evaluation System for Health Technology (NEST), from the Centers of Medicare and Medicaid Services (CMS) (HHSM-500-2013-13018I), from the Agency for Healthcare Research and Quality (R01HS022882), from the National Heart, Lung and Blood Institute of the National Institutes of Health (NIH) (R01HS025164), and from the Laura and John Arnold Foundation.

References

1. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutics, 2005-2012. *JAMA*. 2014;311:368-377.
2. Pease AM, Krumholz HM, Downing NS, Aminawung JA, Shah ND, Ross JS. Postapproval studies of drugs initially approved by the FDA on the basis of limited evidence: systematic review. *BMJ*. 2017;357:j1680.
3. Wallach JD, Egilman AC, Dhruva SS, McCarthy M, Miller JE, Woloshin S, Schwartz L, Ross JS. Postmarket studies required by the US Food and Drug Administration for new drugs and biologics approved between 2009 and 2012: a cross-sectional evaluation. *BMJ*. 2018;361:k2031.
4. Green LW. Public health asks of systems science: to advance our evidence-based practice, can you help us get more practice-based evidence? *Am J Public Health*. 2006;96:406-409.
5. Jarow JP, LaVange L, Woodcock J. Multidimensional Evidence Generation and FDA Regulatory Decision Making: Defining and Using "Real-World" Data. *JAMA*. 2017;318(8):703-704.
6. U.S. Food and Drug Administration. Guidance for Industry and Food and Drug Administration Staff: Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf>. Published 2017, August. Accessed.
7. Schulz WL, Young HP, Coppi A, et al. Temporal Relationship of Computed and Structured Diagnoses in Electronic Health Record Data. <https://www.medrxiv.org/content/10.1101/2019.12.28.19015628v1>. Published 2020. Updated January 2, 2020. Accessed January 21, 2020.
8. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence. *JAMA Netw Open*. 2019;2(10):e1912869.
9. Dhruva SS, Ross JS, Akar JG, et al. Aggregating Multiple Real-World Data Sources using a Patient-Centered Health Data Sharing Platform: an 8-week Cohort Study among Patients Undergoing Bariatric Surgery or Catheter Ablation of Atrial Fibrillation. *npj Digital Medicine*. 2020;In press.

Randomized Trials and Observational Studies: What's in a Name?

John Concato, MD
Food and Drug Administration
John.Concato@fda.hhs.gov

Dr. Concato gave his second presentation on the topic of randomized trials and observational studies that use real-world data. As a brief and simplistic overview, observational studies, including cohort and case-control designs, are credited mainly with identifying causes and correlates of disease. In contrast, randomized, controlled trials (RCTs) have become the benchmark of study designs for evaluating therapeutic effectiveness. Perhaps aided by the emergence of evidence-based medicine,[1] in recent years the phrase *randomized trials versus observational studies* is commonly encountered during discussions of real-world evidence.

After briefly reviewing definitions and concepts from his first presentation – including the terms real-world data (RWD) and real-world evidence (RWE), as well as major considerations of FDA's RWE Framework from 2018 – Dr. Concato emphasized that a randomized-vs.-observational conceptualization when describing study designs is incomplete and sometimes problematic. His main point was that such an approach can mistakenly conflate issues of whether the *intervention* was assigned randomly by an investigator or given during clinical care, and whether the analyzed *data* were collected as primary (research-based) or secondary (clinically-based) information. Although these two components often track together in specific patterns, a simple dichotomy does not exist. The underlying issues are not controversial but become somewhat complicated when considered together.

As one example of the issues involved, a *clinical trial* is defined as a research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes.[ref] Observational studies are not clinical trials, but at the same time not all clinical trials are randomized – and causal inference is viewed as largely being supported by randomization itself.

As another perspective on study design, it's important to note that the era of “big data” has enabled access to various types and vast amounts of secondary data. When observational studies are criticized in comparison to RCTs, often the underlying problem is that the data are not fit-for-purpose in terms of the research question being asked. Given that secondary use of data is the most common approach used in observational studies, it's accurate to say the potential for an observational study design, per se, to generate valid results is actually uncertain if the data involved in a given study are inadequate.

Turning his attention to how results of RCTs and observational studies compare when evaluated head-to-head, Dr. Concato initially discussed what can now be considered first-generation studies in this field. Specifically, groups of RCTs and observational studies on the same clinical topic have been compared systematically. Among representative examples, one study published in 2000 found “little evidence that estimates of treatment effects in observational studies [are] consistently larger than or qualitatively different from those obtained in randomized, controlled trials.”[2]

Another publication, in the same issue of the same journal, stated that “results of well-designed observational studies [...] do not systematically overestimate the magnitude of the effects of treatment as compared with those in

randomized, controlled trials on the same topic.”[3] Even with these publications, however, a long-standing viewpoint is reflected in statements such “Only randomized treatment assignment can provide a reliably unbiased estimate of treatment effects”:[4] see Figure 1.

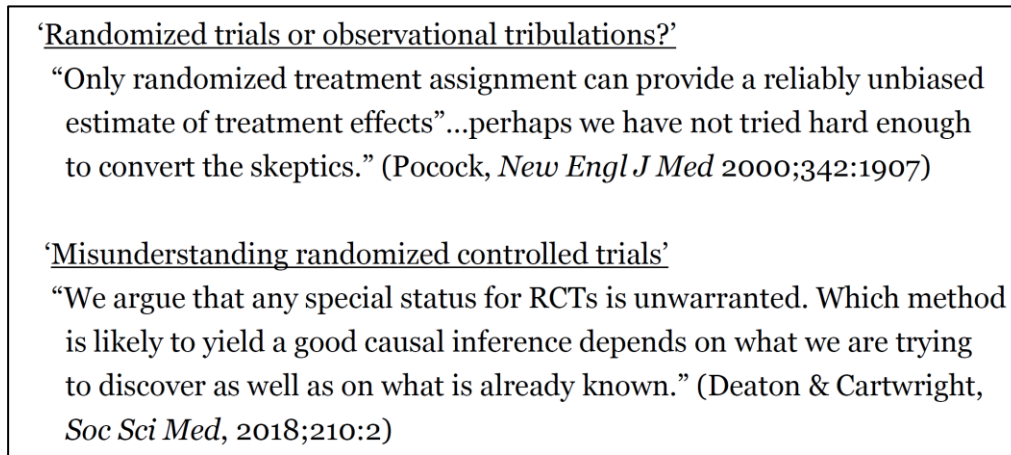


Figure 1. Selected, divergent perspectives on randomized trials and observational studies.

A more recent example of category-based comparisons of RCTs and observational studies was generated by the Cochrane Collaboration, a group best known for conducting meta-analyses of RCTs. After examining systematic reviews on this topic, the Cochrane report found that “on average, there is little evidence for significant effect estimate differences between observational studies and RCTs.”[5] A related statement suggested that “Factors other than study design *per se* need to be considered when exploring reasons for a lack of agreement between results.”[5]

A somewhat different inference may be drawn, however, from the finding that “though not statistically significant [...] the difference in point estimates between pharmacological RCTs and observational pharmacological studies is greater than the difference in point estimates from non-pharmacological studies.”[5] This result was considered to be either a) a reflection of the difficulties involved in removing all potential confounding in observational pharmacological studies, or b) perhaps that positive studies are more likely to be reported, as an artifact of selective reporting bias in industry-sponsored studies.

Focusing on a specific cause-effect association to compare results from different study designs can also provide relevant insights. Dr. Concato referred to the example of hormone replacement therapy (HRT) in postmenopausal women, noting that this topic is often used mistakenly to suggest that observational studies are inherently flawed. Instead, based on a methodological review of available data, a consensus has emerged that observational and randomized studies actually “fall in line.”[6] Specifically, observational studies had evaluated women who were already receiving HRT and who had passed the period of increased cardiovascular risk, whereas the randomized trials examined incident use of HRT that does confer such risk. As such, this example provides a lesson-learned that endorses the use of new-user designs, rather than suggest an internally invalid result for the observational study analysis.[7]

The presentation included other examples that highlight how evidence can be found to support virtually any a-priori viewpoint regarding strength of study designs. For example, two observational studies were published in 2010 examining whether oral bisphosphonates are associated with esophageal cancer. Despite both studies having analyzed data from the same database, one report[8] found that the bisphosphonate use was not significantly associated with

incident esophageal CA, whereas the other report[9] found an approximate doubling of risk. This example gives pause when considering whether a single observational study can contribute to “substantial evidence” of a medical product’s association with clinical outcome(s).

Dr. Concato also pointed out a parallel situation for RCTs, however, based on evidence[10] finding significant discordance among trials involving 36 clinical topics and more than 200 trials. Within each topic, some RCTs showed that a drug improved a particular disease or condition, whereas other trials found the same drug had no (null) effect, and yet others found a paradoxically harmful effect. Thus, results from RCTs can be discordant, often due to different design features or due to sampling considerations and other issues;[11] see also Figure 1.

Returning to the topic of observational studies, the Observational Health Data Sciences and Informatics (OHDSI) group conducted a “systematic evaluation of the impact of analytical choices on effect estimates in observational studies”[12] and found overall that “[results of] clinical studies using observational databases can be sensitive both to study design choices and to specific analytic choices.”[12] The current state-of-science recognizes the strengths of RCTs when evaluating or planning observational studies,[13] and the design of a rigorous observational study can start with the logic of a corresponding randomized trial and emulate that trial as much as possible; see Figure 2.

Observational studies emulating randomized, controlled trials: Key issues

- harmonize design features, including eligibility criteria, treatment strategies, endpoint(s), start/end of follow-up, causal contrast
- use similar strategy for data analysis to estimate the causal effect
- conduct sensitivity analyses to investigate impact of relevant factors

Figure 2. Considerations when evaluating or planning observational studies.

From the U.S. regulatory standpoint – and depending on relevant clinical factors (e.g., heterogeneity of patients, prognosis of the disease) as well as other evidence available on the same treatment-outcome association – meeting the substantial evidence standard[14] requires either two adequate and well-controlled clinical investigations, or one such investigation plus confirmatory evidence. A contemporary question, and the focus of this workshop session, is whether and how observational studies can satisfy that evidentiary standard.

In addressing that question, and as just one example of FDA demonstration projects involving RWE, the RCT DUPLICATE project[15] is an FDA-funded collaboration with investigators at Brigham and Women’s Hospital. In brief, real-world data (from insurance claims) are being analyzed to replicate randomized trials, with a goal of understanding what types of clinical questions can or cannot be addressed through real-world evidence observational studies. Results are expected to be presented in 2020.

The presentation also emphasized the importance of transparency in the planning and conduct of studies involving RWD/RWE. A commonly cited concern is that the low barrier to conducting multiple analyses on an available data set potentially enables investigators to search until finding a favorable association. The onus is on sponsors to be

accountable and to convince regulators that the protocol was developed before conducting the data analysis, and that analytic process involved was trustworthy; regulators have the responsibility of assessing and verifying what was done.

Dr. Concato summarized his presentation with several take-home points: a) clear and consistent use of terminology is important in the current era of real-world evidence, b) although all study designs have strengths and limitations, analyses involving real-world data and real-world evidence have distinctive attributes, c) efforts are ongoing to identify factors and attributes that promote generation of valid “observational” evidence, and d) various (e.g., observational) study designs can potentially support – but not replace – traditional clinical trials for regulatory decision making. Closing remarks reminded the audience that the FDA is pleased to be part of the effort seeking to improve the state of science regarding real-world evidence.

***Disclaimer:** This article reflects the perspective of the author and should not be construed to represent the views or policies of the FDA.*

References

1. Guyatt G, Cairns J, Churchill D, Cook D, Haynes B, Hirsh J, et al. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA*. 1992;268:2420-25.
2. Benson K, Hartz AJ. A Comparison of Observational Studies and Randomized, Controlled Trials. *New England Journal of Medicine*. 2000;342:1878-86.
3. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000 Jun 22;342:1887-92.
4. Pocock SJ, Elbourne DR. Randomized Trials or Observational Tribulations? *New England Journal of Medicine*. 2000;342:1907-09.
5. Anglemeyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database of Systematic Reviews*. 2014 (4).
6. Vandembroucke JP. The HRT controversy: observational studies and RCTs fall in line. *Lancet*. 2009;373(9671):1233-5.
7. Hernan MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008 Nov;19:766-79.
8. Cardwell CR, Abnet CC, Cantwell MM, Murray LJ. Exposure to oral bisphosphonates and risk of esophageal cancer. *Jama*. 2010 Aug 11;304:657-63.
9. Green J, Czanner G, Reeves G, Watson J, Wise L, Beral V. Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort. *BMJ*. 2010;341:c4444.
10. Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med*. 1987;82:498-510.
11. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2-21.

12. Madigan D, Ryan PB, Schuemie M. Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies. *Ther Adv Drug Saf.* 2013;4:53-62.
13. Lodi S, Phillips A, Lundgren J, Logan R, Sharma S, Cole SR, et al. Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *American Journal of Epidemiology.* 2019;188:1569-77.
14. Adequate and well-controlled studies. U.S. Code of Federal Regulations. 21 C.F.R. § 314.126 (2019).
15. RCT DUPLICATE: Randomized Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology, at <https://www.rctduplicate.org/>. Accessed 17 Feb 2020.
16. Gökbuget N, Kelsh M, Chia V, Advani A, Bassan R, Dombret H, et al. Blinatumomab vs historical standard therapy of adult relapsed/refractory acute lymphoblastic leukemia. *Blood Cancer Journal.* 2016;6:e473-e73.

Should We Replace or Enhance? Modernizing Clinical Trial Designs Using Real-World Data

Elodie Baumfeld Andre, PhD
Pfizer Inc.
Elodie.BaumfeldAndre@pfizer.com

Background

Globally, regulatory authorities show an increased willingness to incorporate real-world evidence (RWE) in clinical development to modernize clinical trials. Public workshops, demonstration projects, and white papers aim to improve our understanding of how, where and when it is appropriate to use RWE to support regulatory decision making (e.g., Duke-Margolis RWE Collaborative). The Food and Drug Administration (FDA), the European Medicines Agency (EMA) and Health Canada have developed frameworks and programs to promote the use of high-quality real-world data (RWD) and to support the identification of opportunities where RWE can enhance clinical trials by overcoming clinical trial limitations (1-5). Other geographies, such as China, Japan, Korea, Taiwan, and Brazil, have more recently initiated conversations around RWE to change their regulations to allow adoption of RWE. Regulators, especially in the US, have been increasingly vocal about the need to improve efficiencies in the current clinical trial system and to look at tools such as RWD to facilitate such change.

Hybrid Trials

There is a wide spectrum of potential uses for RWE/RWD in clinical studies (6). In fact, there are many randomized clinical trials (RCTs) that are currently utilizing RWD to optimize them or supplement the RCT data. In instances of a high unmet medical need but a lack of evidence, RWD also offers an opportunity to generate much-needed evidence. Regulators have shown receptivity to certain types of hybrid trials, defined as designs leveraging elements from both RCTs and real-world (RW) studies (7) to answer research questions. Examples of RWE-RCT hybrid trials include external comparators, pragmatic trials, and innovative extensions.

External Comparators

External comparators allow researchers to build a comparator arm using RWD. There are several options for leveraging external comparators to replace or augment RCT control arms. A comparator arm could serve as a RW benchmark to contextualize single-arm trial data. In this case, characteristics of the populations from the RCT experimental arm and the RWD external comparator arm are aligned as closely as possible, but no direct comparison is made (i.e. data from each arm is analyzed separately). Another option (best-case scenario) is to use RWD with matching populations to create a comparator arm to simulate an RCT control arm. External comparators have several advantages such as minimizing the number of patients receiving placebo, an important ethical concern among rare disease patient advocacy groups. Further, the use of RWD to augment the placebo arm would allow for the inclusion of more severe patients or those who are further along in disease progression. It would also enable interpretation of long-term outcomes. Regulatory agencies have been increasingly accepting of external comparators for rare disease or small targeted populations, but it is important to remember that it might be difficult to identify enough patients to achieve statistical power and that there will be key considerations for how to conduct the analysis to address, for instance, potential population heterogeneity.

Pragmatic Trials

Pragmatic trials, a design used as early as 1954, are similar to interventional studies (i.e. patients are randomized) but the differentiating factor is that patients are followed in the RW (i.e. as per standard of care) instead of at the site, based on a prespecified schedule of assessment. Some characteristics of pragmatic trials include the broader inclusion/exclusion criteria which increase the generalizability of findings, a common limitation of RCTs, but a large sample size is often needed to detect an effect. In a pragmatic trial, data can be collected through a combination of primary data and existing data (e.g., electronic health records, claims, or registries). Pragmatic trials substantially reduce patient burden since follow-up is as per standard of care and patients do not have to return to the site as frequently or at all. Furthermore, pragmatic trials are useful in addressing comparative effectiveness research questions since patients are not prescribed placebo. Regulators have been receptive to this type of trial when designed to fulfill an evidence gap. However, since these studies are conducted in the RW, there are obstacles to keeping assessments objective and measure treatment effect when, for example, the trials are generally unblinded and patients might change drugs or doses multiple times. Appropriate statistical analyses need to be deployed in these instances.

Innovative Extension

As a more novel type of hybrid design, innovative extensions are long-term follow-up studies that have emerged due to the recent increase in availability and access to patient health data as well as rapidly evolving technology. In certain instances, for example, with new treatments such as gene therapies, regulatory agencies institute long-term requirements (i.e. 5-, 10-, or 15-year) to understand the durability and safety profile of the treatment. In such instances it is useful to explore new methods of collecting data without requiring patients to return to the clinical site over an extended period of time. The idea is to bring clinical trials to the patient instead of having patients go to the clinical site to reduce patient burden and to reduce loss to follow-up by helping patients adhere to the long-term follow-up schedule. Depending on the endpoints, data could be collected by using wearables, such as a Fitbit, via telephone, or by utilizing RWD sources such as claims databases and registries. Innovative extensions are an encouraging option to achieve regulatory requirements for these long-term follow-up studies. Validation studies on certain endpoints would be expected to ensure that the data collected during an innovative extension study follows similar quality standards as data that would have been collected at the site. Although innovative extensions can often provide efficiencies, it is not a suitable design if the variables of interest are treatment adherence, healthcare resource utilization, and treatment patterns.

When to use hybrid designs?

The research question, clinical development phase, and the context can impact the study design of choice. For example, external controls to support regulatory decision-making are most frequently used for Phase II or III trials and when an RCT is not feasible or when it is not ethical to randomize patients to placebo. Pragmatic trials are most frequently used for Phase IV post-approval studies to address comparative effectiveness and safety or to generate evidence that could not be obtained through RCTs. Innovative extensions typically address regulatory requirements for long-term follow-up of clinical trial patients especially in new disease areas with little existing evidence on long-term durability of effect and/or safety. Innovative extensions are most often planned around the end of Phase III trials, but there have been examples of innovative extensions following Phase I and Phase II.

Regulatory Engagement around Use of RWE

Engaging early with regulatory agencies around RWE approaches is the best way to set up use cases for success. Regulators, such as the FDA RWE Subcommittee in the US and the EMA Innovation Task Force in Europe, are open to dialogue, an opportunity to build partnerships. For transparency with regulatory agencies, it is important to

document the rationale behind using an RWE approach rather than a traditional clinical trial approach. Being able to articulate the ability of RWE to add to the totality of evidence is crucial, since at present time, RWE by itself may not be sufficient to meet the substantial evidence criteria for demonstrating efficacy in most instances. Regulatory agencies look at the whole data package submitted to drive regulatory decisions.

A Word of Caution

Although RWE is currently a buzzword, caution must be exercised when using RWE in clinical development. There are multiple factors to consider when it comes to determining what constitutes “fit-for-purpose” data, such as regulatory objective (i.e. initial approval or label expansion), data availability, and quality of data. Since high-quality data is essential in regulatory submissions, steps must be taken to elevate lower quality data to regulatory grade. It is recommended to include experts, such as pharmacoepidemiologists and RWD biostatisticians, when designing hybrid trials to ensure rigorous and adequate methodology is being implemented. In addition to ensuring robustness of the evidence submitted to regulatory agencies, it is also important to be transparent and to identify potential limitations early on.

Well-designed and executed observational studies have been able to replicate and even predict the results of RCTs (8, 9). Despite this, observational studies cannot necessarily control for potential bias to the extent that RCTs can, and thus, in many contexts, do not replace RCTs, but rather, complement them (10).

There are several examples of successful patient-centric use cases that have effectively harnessed RWE to support regulatory decisions by the FDA, EMA, and Health Canada, such as Invega Sustenna (pragmatic trial), Bavencio (external comparator), and Ibrance or Prevnar13 (observational study). Moving forward, as expert capabilities develop, hybrid trials could bring efficiencies to clinical development by reducing cost, time, or both (depending on the intent and phase of development). Perhaps one day, external controls could be applied to common indications, in addition to rare diseases.

Conclusion

Regulators are increasingly willing to consider RWE and are encouraging its use in certain contexts. They are open to dialogue and want to understand how sponsors use RWD to support regulatory decision-making. The most promising area for success in the near-term is RWE-RCT hybrid designs. By leveraging elements of both RCTs and RWD, these designs merge the best of both worlds and eventually will help with the shift to virtual trials. Considerations for successful regulatory acceptance of hybrid trials include early strategic planning to determine “fit-for-purpose” data and feasibility for execution.

The numerous patient benefits and the potential to drive efficiencies through the use of hybrid trials are just two of the many strong, encouraging reasons to not replace but rather enhance clinical designs with RWE.

***Conflict of interest:** Dr. Elodie Baumfeld Andre is an employee of Pfizer and any opinions expressed here are her own and do not represent the views of Pfizer, Inc.*

***Acknowledgment:** Dr. Elodie Baumfeld Andre would like to thank Dr. Shane Myrick for his insightful suggestions and support in collating this presentation.*

References

1. Food and Drug Administration. (2018). *Framework for FDA's Real-World Evidence Program*. U.S. Department of Health and Human Services. Retrieved from <https://www.fda.gov/media/120060/download>
2. Food and Drug Administration. (2019). *Rare diseases: Natural history studies for drug development, guidance for industry* [Draft guidance]. U.S. Department of Health and Human Services. Retrieved from <https://www.fda.gov/media/122425/download>
3. European Medicines Agency. (2018). *EMA regulatory science to 2025: Strategic reflection*. Retrieved from https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf
4. Heads of Medicines Agencies, & European Medicines Agency. (2019). *HMA-EMA joint big data taskforce: Summary report*. Retrieved from https://www.ema.europa.eu/en/documents/minutes/hma/ema-joint-task-force-big-data-summary-report_en.pdf
5. Health Canada. (2019, April 16). *Optimizing the use of real world evidence to inform regulatory decision-making: Health Products and Food Branch notice*. Retrieved from <https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/announcements/optimizing-real-world-evidence-regulatory-decisions.html>
6. Corrigan-Curay J. Real world evidence a path forward. presented at Duke-Margolis Center for Health Policy: a framework for regulatory use of real world evidence—September 13, 2017. <https://healthpolicy.duke.edu/sites/default/files/a>
7. [Baumfeld Andre E](#), [Reynolds R](#), [Caubel P](#), [Azoulay L](#), [Dreyer NA](#). Trial designs using real-world data: The changing landscape of the regulatory approval process. *Pharmacoepidemiol Drug Saf*. 2019 Dec 10. doi: 10.1002/pds.4932. [Epub ahead of print]
8. Patorno E et al. Cardiovascular outcomes associated with canagliflozin versus other non-gliflozin antidiabetic drugs: population-based cohort study. *BMJ*. 2018 Feb 6;360:k119.
9. Patorno E et al. Using Real-World Data to Predict Findings of an Ongoing Phase IV Cardiovascular Outcome Trial: Cardiovascular Safety of Linagliptin Versus Glimepiride. *Diabetes Care*. 2019 Dec;42(12):2204-2210.
10. Collins R, Bowman L, Landray M, Peto R. *The magic of randomization versus the myth of real-world evidence*. *NEJM* (2020);382:674-678

An Industry Perspective on Real-World Evidence in Regulatory Decision Making: An Example of Transparency in Pharmacoepidemiologic Research

James Weaver, MS, MPH
Observational Health Data Analytics
Janssen R&D
jweave17@ITS.JNJ.com

Context

This discussion provides an example of the use of real-world data (RWD) to produce real-world evidence (RWE) to inform and support a US Food and Drug Administration (FDA) regulatory decision on product labeling for canagliflozin, a Janssen marketed treatment for type II diabetes mellitus (T2DM). The quality, and hence, validity of RWE relies heavily on transparency and replicability as the study presented here will demonstrate.

Canagliflozin is a selective sodium-glucose co-transporter 2 inhibitor (SGLT2i) that lowers blood glucose by increasing urinary glucose excretion. It is indicated for adults with T2DM as adjunct to diet and exercise to improve glycemic control, to reduce the risk of major cardiovascular events among patients with cardiovascular disease (CVD), and to reduce the risk of end-stage kidney disease, doubling of serum creatinine, cardiovascular death, and hospitalization for heart failure (HHF) among patients with diabetic nephropathy[1].

Cardiovascular benefit for some SGLT2i has been reported[2-6] and the findings for canagliflozin were the basis for its expanded indication to patients with T2DM and CVD. However, clinical trial results and RWE on the risk of below-knee lower-extremity amputation (BKLE) have been mixed[2, 3, 7][8][9]. Notably, the CANagliflozin cardioVascular Assessment Study (CANVAS) program reported a concerning increased risk of BKLE amputation for canagliflozin versus placebo. Additionally, a potential risk of BKLE amputation is reported in the label of ertugliflozin. In this context, OBSERVE-4D¹ was designed as the first real-world study to generate direct, head-to-head comparative evidence for HHF and BKLE amputation among SGLT2is and other antihyperglycemic agents (AHAs) to assess the extent to which the efficacy and safety profile attributed to SGLT2i exposures in experimental settings translates to effectiveness and risk in routine clinical practice.

OBSERVE-4D

OBSERVE-4D was a retrospective, observational, comparative cohort study[10]. It was “retrospective” in that the data were previously collected non-research purposes, “observational” in that treatment assignments were not randomized but selectively assigned in the course of routine clinical care, and “comparative” in that it compared outcome occurrence in pairwise comparisons of exposed populations of interest. This study was conducted in four American administrative databases that exist for the billing and insurance claim reimbursement that have been repurposed for health sciences research. The study population was defined as T2DM patients newly exposed to AHAs, including SGLT2is, where individual, de-identified patient records were returned from these RWD sources for analysis. The study also included a T2DM subpopulation with CVD to assess outcomes in patients with known risk factors for HHF and BKLE amputation.

¹ Comparative effectiveness of canagliflozin, SGLT2 inhibitors and non-SGLT2 inhibitors on the risk of hospitalization for heart failure and amputation in patients with type 2 diabetes mellitus: A real-world meta-analysis of 4 observational databases

The objectives were to assess the risk of HHF and BKLE amputation for new users of canagliflozin compared to new users of non-SGLT2i AHAs (any dipeptidyl peptidase-4 inhibitor, glucagon-like peptide-1 receptor agonist, thiazolidinedione, sulfonylurea, insulin, or other AHAs) and other SGLT2is (dapagliflozin and empagliflozin). Advanced methods to address potential confounding by imbalanced observed and unobserved population characteristics in each pairwise comparison were applied.

The OBSERVE-4D results were concordant with the CANVAS trial in that it showed a reduced risk of HHF in canagliflozin compared to non-SGLT2i AHAs as well as no difference in the risk of HHF between canagliflozin and other SGLT2is. There was no observed increased risk of BKLE amputation in canagliflozin versus non-SGLT2i AHAs or other SGLT2i. This result was discordant with the CANVAS trial. The results were consistent across the four RWD sources.

RWE Validity

Health Canada, the FDA, and European Medicines Agency (EMA) now consider the use RWE to support drug risk and benefit regulatory decisions and have published guidance documents on establishing RWD fitness for use and on assessing the validity of RWE[11][12][13][14]. For example, Health Canada has published its interpretation in “Elements of Real-World Data/Evidence Quality throughout the Prescription Drug Product Life Cycle” that includes 15 specific elements for study protocol development and general elements for data quality in retrospective data collection. By the Health Canada RWD and RWE standards, there is good reason to believe the OBSERVE-4D results are valid, as this study explicitly and independently addressed each of the guidance elements.

Two critical elements of RWE validity are bias and outcome definition and measurement. OBSERVE-4D included methods to address sources of measured confounding using propensity scores (PS) computed with large-scale regularized regression. This data-driven, machine learning approach does not use as input a set of potential confounders provided by subjective clinical expertise. Rather, it explicitly assumes that any imbalanced characteristic could potentially confound a comparative result. The method predicts treatment assignment conditioned on all observed characteristics and through matching or stratification on the PS can correct initially observed imbalance. This is not always achievable, but in OBSERVE-4D the method was successful in establishing comparisons that did not meaningfully differ on observed characteristics. Secondly, empirical calibration using negative control outcomes assessed the threat of residual systematic error from unmeasured sources. The technique lends well to a simple analogy – when standing on a force-dependent scale, the value for mass returned is computed as the force exerted on the scale proportionally calibrated to adjust for acceleration due to gravity. Similarly, a result from an epidemiologic method can be proportionally calibrated to reflect how well the method returns a null finding when one is expected. In OBSERVE-4D, little calibration was ultimately required, as the study performed as expected and did not return non-null results for negative control outcomes unrelated to the exposures.

Regarding outcome definitions and measurement, there was concern from the EMA that BKLEA amputation was not the actual outcome of interest, but rather it is a downstream effect on a causal pathway from a precipitating event that may be caused by SGLT2is. A subsequent Janssen study using the same design as OBSERVE-4D was conducted to assess lower extremity osteomyelitis, peripheral occlusive disease, gangrene, and lower extremity ulcer under the hypothesis that these events could lead to BKLE amputation. The results of this study showed no increase risks for these conditions.

In terms of RWD quality, Health Canada has asserted that curated data are always preferred. The data sources used OBSERVE-4D had all been transformed to the Observational and Medical Outcomes Partnerships (OMOP) common data model (CDM)[15], which is a standardized representation of health care experiences using a common database structure and content vocabulary that enables consistent analysis across disparate databases. The OMOP CDM is maintained by Observational Health Data Sciences and Informatics (OHDSI²), an open-science community that has also developed a large suite of rigorously tested data quality tools for assessing RWD and data analysis tools for generating RWE. Importantly, the process of converting a database from its native state to the CDM is one of extensive data curation and quality improvement. The process establishes research-ready data against which standardized and validated software tools can be applied. The foundation of OBSERVE-4D is the years of effort spent developing the OMOP CDM, transforming databases, and developing software for RWE generation.

Further, a Health Canada element of RWD quality states that “approaches should be taken to validate all relevant elements of existing administrative datasets and should be described and provided whenever possible.” The OHDSI community has developed a harmonized data quality assessment tool that provides a structured, empirical evaluation of the trustworthiness of an RWD source. Briefly, the tool includes a large set of tests that adhere to a framework of data quality categories (conformance, completeness, plausibility) and assessment contexts (verification and validation)[16]. Quantifying RWD quality allows for empirically informed assessment on whether data quality could be influencing the validity of RWE.

Lastly, OBSERVE-4D was novel in its transparency. The study included a large set of sensitivity analyses to address the multitude of design choices faced by researchers to address different forms of selection, measurement, and confounding biases inherent to observational research. All design choices were prespecified and publicly recorded by registering the study protocol and amendment history before study execution. Including the amendment history provides visibility into decisions made in response to the perspectives of the different stakeholders involved, including regulatory agency representatives. The full source code for executing the study against any database adherent to the OMOP CDM standard is also available. The complete result set including statistical diagnostics for assessing validity were subsequently made available in an online tool for other researchers to explore, discuss, and interpret³. OBSERVE-4D demonstrates that full transparency is feasible when generating RWE, is necessary for improving RWE replicability, and allows for informed debate about the validity of RWE.

Conclusion

Full transparency is both possible and necessary in generating RWE. It enables true replicability, builds confidence in RWE, and it allows for informed debate about the validity of RWE in the context of regulatory decision making regarding the risks and benefits of drugs. OBSERVE-4D is an example of a Janssen R&D approach to applying the design, implementation, and interpretation best-practices advocated by the OHDSI community to address a clinical question of importance to patients and regulatory agencies.

² www.ohdsi.org

³ <https://data.ohdsi.org/AhasHfBkleAmputation/>

References

1. INVOKANA® (canagliflozin) tablets, for oral use [package insert]. Titusville, NJ: Janssen Pharmaceuticals; 2018.
2. Neal, B., et al., *Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes*. N Engl J Med, 2017. **377**(7): p. 644-657.
3. Udell, J.A., et al., *Cardiovascular Outcomes and Risks After Initiation of a Sodium Glucose Cotransporter 2 Inhibitor: Results From the EASEL Population-Based Cohort Study (Evidence for Cardiovascular Outcomes With Sodium Glucose Cotransporter 2 Inhibitors in the Real World)*. Circulation, 2018. **137**(14): p. 1450-1459.
4. Zinman, B., et al., *Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes*. N Engl J Med, 2015. **373**(22): p. 2117-28.
5. Patorno, E., et al., *Cardiovascular outcomes associated with canagliflozin versus other non-gliflozin antidiabetic drugs: population based cohort study*. Bmj, 2018. **360**: p. k119.
6. Fitchett, D., et al., *Effects of empagliflozin on risk for cardiovascular death and heart failure hospitalization across the spectrum of heart failure risk in the EMPA-REG OUTCOME(R) trial*. Eur Heart J, 2018. **39**(5): p. 363-370.
7. Inzucchi, S.E., et al., *Empagliflozin and Assessment of Lower-Limb Amputations in the EMPA-REG OUTCOME Trial*. Diabetes Care, 2018. **41**(1): p. e4-e5.
8. STELAGRO™ (ertugliflozin) tablet, for oral use [package insert]. Whitehouse Station, NJ: Merck Sharp & Dohme Corp; 2017
9. Yuan, Z., et al., *Risk of lower extremity amputations in people with type 2 diabetes mellitus treated with sodium-glucose co-transporter-2 inhibitors in the USA: A retrospective cohort study*. Diabetes Obes Metab, 2018. **20**(3): p. 582-589.
10. Ryan, P.B., et al., *Comparative effectiveness of canagliflozin, SGLT2 inhibitors and non-SGLT2 inhibitors on the risk of hospitalization for heart failure and amputation in patients with type 2 diabetes mellitus: A real-world meta-analysis of 4 observational databases (OBSERVE-4D)*. Diabetes Obes Metab, 2018. **20**(11): p. 2585-2597.
11. Health Canada. (2019). Elements of Real World Data/Evidence Quality throughout the Prescription Drug Product Life Cycle. [online] Available at: <https://www.canada.ca/en/services/health/publications/drugs-health-products/real-world-data-evidence-drug-lifecycle-report.html> [Accessed 18 February 2020].
12. European Medicines Agency. (2019). HMA-EMA Joint Big Data Taskforce Phase II report: 'Evolving Data-Driven Regulation'. Amsterdam: EMA.
13. European Medicines Agency. (2019). HMA-EMA Joint Big Data Taskforce – Summary report. Amsterdam: EMA.
14. Food and Drug Administration. (2019). Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics. Maryland: FDA.
15. Voss, E.A., et al., *Feasibility and utility of applications of the common data model to multiple, disparate observational health databases*. J Am Med Inform Assoc, 2015. **22**(3): p. 553-64.
16. Kahn, M.G., et al., *A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data*. EGEMS (Wash DC), 2016. **4**(1): p. 1244.

SESSION 4:

**Demonstrating How Real World Data/Evidence
Can be Used in Support of Regulatory Approvals**

**Demonstrating and Communicating
Real World Data (RWD) Reliability to Support
Fitness for Use for Regulatory Decision-Making**

Nirosha Mahendraratnam Lederer, PhD, MSPH
Duke-Margolis Center for Health Policy
nirosha.mahendraratnam@duke.edu

Evaluating whether real-world data (RWD) are fit for use is critical for generating real-world evidence (RWE) to inform regulatory decision-making related to the effectiveness of medical products. Indeed, FDA emphasized the importance of RWD fitness for use as a priority challenge area in its 2018 Framework for FDA's Real-World Evidence Program.

In September 2019, the Robert J. Margolis, MD, Center for Health Policy at Duke University, informed by its multistakeholder Real-World Evidence Collaborative, released a white paper titled "[Determining Real-World Data's Fitness for Use and the Role of Reliability](#)."² The white paper is a resource for sponsors who design studies using RWD sources, for regulators who develop policy, and for researchers who develop best practices for study methods.

Stakeholders have called for the development of a minimum set of checks to assess data fitness for use; however, these checks are often difficult to define broadly due to heterogeneity within and between RWD. This white paper focuses on the development of a minimum set of verification checks to assess RWD reliability. Reliability and relevancy are components of RWD fitness for use.¹ Reliability refers to the credibility of the data, while relevancy refers to the ability of the data to answer the research question.¹ Relevancy cannot be broadly evaluated because of its dependence on the research question. Reliability can be assessed through verification and validation checks. According to a literature review across multiple data sources, the majority of checks to assess reliability are verification checks.^{3,4}

Verification checks assess conformance, completeness and plausibility. Conformance checks determine if the structure of the data matches expectations, completeness checks verify the presence of a data value, and plausibility checks assess believability of the data.⁴ Chapters 2 and 3 of the white paper focus on the application of verification checks to electronic health record (EHR) data and person-generated health data (PGHD). Verification checks must be continuously assessed throughout analysis due to the dynamic nature of data curation.

A minimum set of verification checks is a first step toward demonstrating reliability. Determining if the RWD are fit for use may require additional reliability and relevancy checks. Additionally, the utility of RWD to address the research question is dependent on the methods and regulatory and clinical contextual factors.

Several ongoing stakeholder efforts support the assessment of data fitness for use. For example, FDA, Harvard Pilgrim Health Care Institute, and Harvard Medical School are developing uniform metadata standards for assessing the quality, completeness, and stability of EHR data across data sources.^{5,6} Another project through collaboration with FDA and several government agencies focuses on developing methodology to harmonize four common data models (Sentinel, OHDSI, PCORnet, i2b2) to support access to RWD for patient-centered outcome research.^{7,8} FDA in collaboration with University of California-San Francisco released a report in October 2019 titled “Source Data Capture from EHRs: Using Standardized Clinical Research Data” which describes an approach for structured data flow from an electronic health record to an electronic data capture system.^{9,10} Additionally, CDISC launched an initiative to standardize data as part of a learning health care system.^{11,12}

While much progress has been made toward advancing data fitness for use, more work must still be done. Next steps might include improving documentation and communication of the results of data checks, identifying data curation best practices stratified by data source, identifying checks to assess relevancy, and evaluating data accrual and analysis.

Funding for this 2019 white paper was made possible through the generosity of the Margolis Family Foundation, which provides core resources for the Center, as well as a combination of financial and in-kind contributions from Real-World Evidence Collaborative members, including AbbVie; Eli Lilly and Company; Genentech, a member of the Roche Group; GlaxoSmithKline; Johnson & Johnson; Merck & Co.; Novartis; Teva; and UCB. For more information on the Real-World Evidence Collaborative, visit <https://healthpolicy.duke.edu/real-world-evidence-collaborative>.

References

1. US Food and Drug Administration. *Framework For FDA’s Real-World Evidence Program*. US Department of Health & Human Services; December 2018.
2. Mahendraratnam N, Silcox C, Mercon K, et al. *Determining Real-World Data’s Fitness for Use and the Role of Reliability*. Duke-Margolis Center for Health Policy; 2019.
3. Callahan TJ, Bauck AE, Bertoch D, et al. A Comparison of Data Quality Assessment Checks in Six Data Sharing Networks. *EGEMS (Wash DC)*. 2017;5(1):8.
4. Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Wash DC)*. 2016;4(1):1244.
5. Funding Sources. <https://www.popmednet.org/about/funding>.
6. Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data. 2018; <https://www.sentinelinitiative.org/sentinel/methods/standardization-and-querying-data-quality-metrics-and-characteristics-electronic>.

7. Harmonization of Various Common Data Models and Open Standards for Evidence Generation. 2017; <https://aspe.hhs.gov/harmonization-various-common-data-models-and-open-standards-evidence-generation>.
8. U.S. Food and Drug Administration. Real World Data and the PCORTF Common Data Model Harmonization Project.
9. Source Data Capture from Electronic Health Records (EHRs): Using Standardized Clinical Research Data. 2019; <https://www.fda.gov/science-research/advancing-regulatory-science/source-data-capture-electronic-health-records-ehrs-using-standardized-clinical-research-data>.
10. Rocca M, Asare A, Esserman L, Dubman S, Gordon G. *Source Data Capture from EHRs: Using Standardized Clinical Research Data*. 2019.
11. Clinical Data Interchange Standards Consortium. Learning Health System – Essential Standards to Enable Learning (ESTEL) Webinar.
12. Kush R. Update and Discussion on the Learning Health System Initiative: Essential Standards to Enable Learning (ESTEL). In:2018.

What's Next? Pathways to Unlock the Value of RWD from Oncology EHRs

Nicole Mahoney, PhD
Flatiron Health
nicole.mahoney@flatiron.com

Disclosure: *The author reports employment in Flatiron Health, Inc., which is an independent subsidiary of the Roche group.*

Summary

This presentation followed a number of sessions introducing frameworks to establish how and when real-world evidence (RWE) may be used to support regulatory decisions about the effectiveness of medical products. Recent research has shown that valuable clinical information contained in electronic health records (EHRs) may be unlocked to generate RWE that can complement the information collected from randomized clinical trials and support regulatory decisions in oncology. From a policy perspective, “what’s next” for stakeholders is to take steps towards the practical application of those frameworks. Collaborative experiments and demonstration projects offer an important pathway to advance our collective understanding of how to capture the full value from EHRs and other real-world data (RWD) sources.

Flatiron Health is an oncology-focused health technology company that provides technology support and services to community practices through its proprietary EHR software, OncoEMR. Flatiron Health generates RWD from two EHR-derived primary sources: a data pipeline stemming from this nationwide EHR network, and EHR data integrations with academic research centers that enable bidirectional transmission of RWD.

Flatiron Health databases aggregate deidentified data from structured (e.g., diagnosis, demographic information) and unstructured (e.g., free text in physician notes, radiology or other laboratory reports) sources. Those data are standardized and harmonized in a curation process that yields analyzable data sets, which, in turn, can be licensed to researchers for a variety of purposes. Processing information held in unstructured format is key, since unstructured sources often contain the granular clinical insights required to answer sophisticated research questions. Data organizations take different approaches to the curation of unstructured patient data. In Flatiron Health, unstructured data is processed through technology-enabled manual chart reviews. Using specific policies and procedures, highly trained clinical “abstractors,” including oncology nurses and cancer registry professionals, capture key unstructured data points and enter them into predefined, structured data models. Throughout this process, there are quality control mechanisms to ensure data integrity and the ability to trace data provenance. This approach combines machine-based efficiency to handle large amounts of data with the clinical nuance and context provided by human manual review of unstructured records (1). The result is a nationwide longitudinal, demographically and geographically diverse de-identified database currently including data from over 280 cancer clinics (~800 sites of care) representing more than 2.4 million US cancer patients available for analysis.

Interest in pairing genomic with clinical information to generate precision medicine tools is high and expected to grow, and databases of de-identified clinical information may be combined with external databases, including those holding genomic data towards this purpose. This is another example where the handling of unstructured information

is again a critical component. Reporting of genomic testing results is still a fragmented practice in need of overall standardization, and these results are often contained in large, unstructured files that are either not fully captured in EHRs or difficult to extract due to their lack of uniformity. Flatiron Health has established a partnership with Foundation Medicine, a company that provides genomic profiling via next-generation sequencing of solid and hematologic malignancies, to help address these challenges. These two organizations have created a combined clinicogenomic database that contains information on over 50,000 patients (2).

EHRs contain timely and deep clinical information, and EHR-derived data holds promise to address benefit–risk evaluations about medical products and support a range of regulatory decisions, including label expansions. To date, there are a few examples of RWD deployment in support of label expansions related to effectiveness, but drug developers, data organizations and regulators are actively learning about the benefits and limitations to using RWD for regulatory purposes. In this regard, pre-competitive demonstration projects offer a powerful platform to accelerate collective learning.

One seminal challenge in the utilization of RWD is the variable cross-applicability of clinical trial endpoints, given the potential for incompleteness and subjectivity in documentation in routine practice, compared with the regimented data collection stipulated in clinical trials. For example, while mortality is an objective clinical outcome in oncology, it is not collected routinely in EHRs. Completeness of mortality information is important, since incomplete capture of mortality would skew measurements of overall survival times. To develop a high-quality real-world mortality variable, Flatiron Health combines structured data on patient deaths with information from unstructured documents such as condolence letters or sympathy cards and links these data to external death records (such as the U.S. Social Security Death Index) and commercial obituary information. In the case of mortality, the resulting variable can be benchmarked against an external gold standard (the US National Death Index) to generate quality metrics such as sensitivity (3). The sensitivity of this variable is essentially a measure of the completeness of the information on which we would base any subsequent overall survival estimates, and studies demonstrate that the reliability of results drops when the completeness of mortality data drops below 90%.

Not all clinical effectiveness variables can be benchmarked against external gold standards, and it becomes challenging to develop real-world clinical oncology endpoints lacking such reference, and to determine whether they provide clinically meaningful information. Flatiron Health and 10 other RWD organizations are currently collaborating in the Friends of Cancer Research pilot 2.0 demonstration project to define, and assess the reliability and clinical utility of several real-world clinical endpoints in oncology. As a first step in the project, participant organizations identified cohorts of patients with non-small-cell lung cancer (NSCLC) receiving specific treatments as part of their routine clinical care, and explored the similarities and differences in the populations across the datasets. Next, the group aligned on high-level endpoint definitions feasible to be implemented across the different data sources, including overall survival and progression-free survival, as well as others that may be more readily ascertained from RWD sources – such as time to next treatment and time to treatment discontinuation (4). Future experiments in the pilot project seek to determine if these real-world endpoints may serve to detect differences in effectiveness across treatments, and how those treatment differences mimic (or not) the published clinical trial experience.

Efforts such as this assessment of validity are an important step towards fully informed policy discussions. As mentioned above, traditional clinical trials demonstrate efficacy via clinical endpoints that tend not to be applicable to RWD. If RWE is to contribute to the information about the benefits and risks of medical products, the research community needs to understand how to interpret so-called real-world effectiveness measures. Reaching alignment on methods to develop real-world clinical endpoints, characterize their quality, and potentially assess their validity

for regulatory uses is imperative; this undertaking will require new approaches across multiple stakeholders in drug development and regulatory organizations. Projects such as the one spearheaded by The Friends of Cancer Research will help advance thinking on how to do so. Ideally, such collaborative experiments will contribute to our collective learning and inform “what’s next,” namely, much needed guidance to help drug developers transition from high-level frameworks to actual submissions of RWD for regulatory purposes.

References

1. Berger ML, Curtis MD, Smith G, Harnett J, Abernethy AP. Opportunities and challenges in leveraging electronic health record data in oncology. *Futur Oncol.* 2016;12(10):1261-1274. <https://doi.org/10.2217/fon-2015-0043>
2. Singal G, Miller PG, Agarwala V, Li G, Kaushik G, Backenroth D, Gossai A, Frampton GM, Torres AZ, Lehnert EM, Bourque D, O'Connell C, Bowser B, Caron T, Baydur E, Seidl-Rathkopf KN, Ivanov I, Alpha-Cobb G, Guria A, He J, Frank S, Nunnally AC, Bailey M, Jaskiw A, Feuchtbaum D, Nussbaum NC, Abernethy AP, Miller VA. Association of patient characteristics and tumor genomics with clinical outcomes among patients with non-small cell lung cancer using a clinicogenomic database. *JAMA.* 2019;321(14):1391-1399.
3. Curtis MD, Griffith SD, Tucker MG, Taylor MD, Capra WB, Carrigan G, Holzman B, Torres AZ, You P, Arneri B, Abernethy AP. Development and validation of a high-quality composite real-world mortality endpoint. *Health Serv Res.* 2018;53(6):4460-4476.
4. Stewart M, Norden AD, Dreyer N, Henk HJ, Abernethy AP, Chrischilles E, Kushi L, Mansfield AS, Khozin S, Sharon E, Arunajadai S, Carnahan R, Christian JB, Miksad RA, Sakoda LC, Torres AZ, Valice E, Allen J. An exploratory analysis of real-world end points for assessing outcomes among immunotherapy-treated patients with advanced non-small-cell lung cancer. *JCO Clin Cancer Informatics.* 2019. <https://doi.org/10.1200/CCI.18.00155>

**Real-world data curation to transform medical investigation:
Technology to reimagine the economics of evidence
gathering and support regulatory decision-making**

Aaron Leibtag
Pentavere Research Group Inc.
aleibtag@pentavere.com

Summary

In this article, Aaron Leibtag – CEO and co-founder of Pentavere Research Group Inc., a highly-regarded healthcare technology company – describes the challenges and limitations of existing, manual data abstraction for medical research, including the economic constraints of this labor-intensive approach and its vulnerability to imperfect or erroneous conclusions. He outlines his organization’s application of real-world data curation to transform the practice of medical research. He argues that such data curation engines can reduce the economic barriers to developing comprehensive research, facilitate regulatory approvals of medical innovation, and provide patients with more personalized, meaningful care information.

The rapid emergence of ‘big data’ is driving the need for rapid, accurate, cost-effective collection, distillation, and analysis of previously inaccessible data, unleashing unimaginable discovery across every sector, including the field of scientific investigation. This data availability, when combined with Artificial Intelligence (AI) and machine learning, has fueled a bold vision for healthcare. There is much optimism that the tools are within reach to deliver precision medicine and highly personalized care, solving many challenges faced by the industry, payers, regulators, clinicians, and patients.

Unfortunately, this vision is yet unrealized since the underlying information required to conduct this caliber of data analytics still depends on manual extraction of clinical variables from patient charts and research databases. Although it is fair to say that researchers today are drowning in data – dispersed across thousands of files, reports, patient records and decades of completed studies – the data cannot be accessed and abstracted in an economically viable manner.

By current research practices, the standard model of data abstraction sees costs rise in direct correlation to the number of patients and variables examined, since researchers must manually gather these data points from the population cohort.

This reality is vividly clear when you consider the electronic health record of a single lung cancer patient. A compilation of the patient’s clinical notes and reports, in order to obtain the necessary biomarkers, lab and imaging reports, metastases locations, treatment records and outcomes, is typically hundreds of pages. The resulting cost of manual data collection is significant making it difficult for researchers to secure funding to conduct such time-consuming and costly patient chart reviews for patients who are the most vulnerable with the highest mortality.

Even when such data gathering is approved for investigation, researchers face numerous challenges. Among them, manual data abstraction is vulnerable to human error in interpretation of diverse data inputs and formats. For example, manual data abstractors who are tasked with collecting patient diagnostic scores such as ECOG might find that a

single patient's record lists ECOG scores at different stages of treatment. Thus, the collected data could have great variability in ECOG values, due to the varying interpretations of the manual abstractors, resulting in conflicting findings. Even if the ECOG value of interest to be extracted was defined as that at stage IV, whether the score before or after the date that stage IV was determined could complicate the extraction process.

Similarly, inconsistent information may exist in patient records, such as conflicting information about a patient's history of smoking, per the lung cancer population cohort. Manual data abstractors may never realize that these inconsistencies exist; if they do identify these issues, they must then make judgements in terms of which data points to input, often without that variability being documented or validated.

The alternative: applying a data curation engine

As an alternative to the cost, time and inaccuracies that result from manual chart reviews, imagine the potential transformation of current research processes if a health technology provider can introduce a data curation engine that extracts real-world evidence from medical records in a way that is research grade and, ultimately, regulatory-grade once the regulators decide on the scope and applicability of such evidence.

Pentavere Research Group uses both traditional and cutting-edge NLP and AI methodologies to label, extract, and curate variables from structured and unstructured clinical texts, producing high-quality row-column datasets [1]. We are working with health agencies to design the frameworks and methodologies to gather, synthesize and amalgamate their data inputs. This collaborative process will enable regulators to process vast quantities of disparate data and produce data that will be considered regulatory grade.

The process begins by establishing comprehensive protocols for data handling including data and statistical management plans and quality assurance processes. Since it is critical that client research projects attain research-grade data, which can earn regulatory-grade acknowledgement, an important part of the approach is designing natural language processing algorithms for data collection. The 'gold standard rules' that define exactly what features will be abstracted are generated in collaboration with the client to ensure that the maximum consistency is attained in data collection [1].

Then, Pentavere conducts a data discovery process with the clinical inputs and the client to refine these data collection rules and algorithms that will be applied by the data engine. These precise, customized criteria govern what data variables are extracted, and these rules are specifically documented for transparency and consideration by the research team policymakers, and regulatory bodies.

This curation process enables meticulous review and auditing of the abstracted data, to ensure accuracy of findings and allow researchers to have confidence in the final data sets. The transparency of the process enables any necessary regulatory reviews including adjudication and validation of the data in a manner that's impossible with current manual data extraction and curation.

As an example, Pentavere performed this activity for the University Health Network, successfully aggregating more than 1,400 records for lung cancer patients [2], including 600 individuals who received treatment at other institutions. In addition to generating unique insights on treatments [3] and outcomes [4] in this vulnerable patient population the process identified potential areas for improvement in the clinical documentation as well as opportunities for improved data generation.

Although the practice of data analytics typically elicits much discussion regarding the protection of customer data and confidentiality – a paramount concern in healthcare – the model applied by Pentavere effectively eliminates these challenges. As a technology company, at no point does Pentavere remove the data from client systems. Pentavere only provides the technology and expertise to scale the curation of a client’s data on their own premises and servers. They deploy their tools behind the client agency’s firewall, and they serve as the agent for that agency to curate and extract their data onsite.

The natural next step is to take real-world data curation beyond the walls of a single healthcare provider or institution and generate mass scale data aggregation across multiple providers, enhancing the richness and detail of population data without compromising data privacy. This ability to compile greater volumes of data is invaluable when one considers the previous example of lung cancer patients. Despite being the number one cause of cancer-related deaths, researchers often lack access to adequate data to conduct the highest quality of investigation for this population cohort.

To help resolve such challenges, Pentavere is partnering with multiple health care stakeholders to enable data aggregation on a larger scope and scale. One collaboration is with the Canadian Personalized Health Innovation Network (CPHIN), a pan-Canadian health network that is focused on generating real-world evidence from across the country for use by researchers, clinicians and the government.

By introducing Pentavere’s data enrichment engine to CPHIN sites across the country, they can aggregate data on a scale that has not been possible in the past with the complexity and significant expense of creating large data sets. Now, the Canadian healthcare system can come together to use the power of data to accelerate system transformation and improve health outcomes for all Canadians.

Transforming clinician-patient interactions

While Pentavere’s approach to data curation enables more cost-effective research techniques and supports the creation of readily verified, auditable data for regulatory decision-making and approvals, perhaps the most powerful benefit of this technology lies in its ability to transform the way clinicians can serve their patients.

Clinicians today are constantly faced with the scenario in which a patient asks, “What are the likely outcomes of this treatment for patients like me?” While the physician can offer their best opinion, based on their years of experience, recall of similar cases, and reference to clinical trial data, they are unable to answer the patient’s question with specific data that share the same variables. With Pentavere’s approach, clinicians could access specific, tailored data that can enhance both patient engagement and the effectiveness of their treatments.

In summary, there is great potential in the application of real-world data curation. It can dramatically lower the costs of healthcare data collection and enhance the scope, scale and specificity of data available for study, which will enable more precise, accurate and high-quality research.

Ultimately, this technique can place more relevant, meaningful information in the hands of dedicated clinicians as they remain steadfastly focused on providing the best level of care to their patients.

References

1. Petch, J.B., J.; Murray, J.; Mamdani, M., *Extracting clinical features from dictated ambulatory consult notes using natural language processing: a pilot validation study*. JMIR, 2018.
 2. Law, J.H., et al., *Generating real-world evidence: Using automated data extraction to replace manual chart review*. Journal of Clinical Oncology, 2019. **37**(15_suppl): p. e18096-e18096.
 3. Pettengell, C., et al., *Real world evidence of the impact of immunotherapy in patients with advanced lung cancer*, in *World Conference on Lung Cancer*. 2019: Barcelona.
 4. Law, J.H., et al., *Real world outcomes of advanced NSCLC patients with liver metastases*, in *World Conference on Lung Cancer*. 2019: Barcelona.
-