

**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
<b>Gayatri Jayaraman</b> , Health Canada	
<i>FDA's Real World Evidence Program</i> .....	6
<b>John Concato</b> , Food and Drug Administration	
<i>Real World Evidence: European Medicines Agency Experience</i> .....	10
<b>Peter Arlett</b> , 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
<b>Robert W. Platt</b> and <b>Kristian B. Fillion</b> , McGill University	
<i>Medical Product Safety Surveillance: Data Quality in the Sentinel Initiative</i> .....	18
<b>Judith C. Maro</b> , 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
<b>Joseph S. Ross</b> , Yale School of Medicine	
<i>Randomized Trials and Observational Studies: What's in a Name?</i> .....	25
<b>John Concato</b> , Food and Drug Administration	
<i>Should We Replace or Enhance? Modernizing Clinical Trial Designs Using Real-World Data</i> .....	30
<b>Elodie Baumfeld Andre</b> , Pfizer Inc.	
<i>An Industry Perspective on Real-World Evidence in Regulatory Decision Making: An Example of Transparency in Pharmacoepidemiologic Research</i> .....	34
<b>James Weaver</b> , 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.

---

**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.
-