Federated Analysis:
State of the Science

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Welcome

Federated Analysis:
State of the Science

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Housekeeping

Acknowledgements

Organized with support of a steering committee

Kim McGrail | HDRN Canada

Charles Burchill | HDRN Canada

Daryl Fung | HDRN Canada

Amy Freier | HDRN Canada

Johanne Provençal | CRDCN

Grant Gibson | CRDCN

Robert Platt | CNODES

Therese Stukel | ICES

Michael Paterson | ICES

Stephanie Dixon | ICES

Claudia Sanmartin | StatCan

Marc Fiume | DNAstack

Bryce Pickard | Integrate AI

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Why "federated analysis"?

Capabilities and possibilities are changing rapidly

One possible approach that can help within Canada

Enable international collaborations

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Why a "collective learning series"?

Share knowledge from experts on the state of the science of federated analytics.

Translate lessons learned into tools and resources that can be available to the broader research community.

Develop and/or advocate for policies and tools for researchers conducting multi-regional research.

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Overview

1. Definitions / common language
2. Canadian context
3. Current practice for non-pooled data
4. Enablers for federated analysis
5. What is next in the the federated analysis learning series?

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Pooled Analysis

Analysis of individual level data that are combined from multiple locations and/or sources.

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Distributed Data

Data stored across multiple organizations, institutions,
or data centres.

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Federated Data

Distributed data that are able to be analyzed together while remaining separate.

Horizontal Federation

Partitions that include the same features/measures but for different people (e.g. different provinces, different countries).
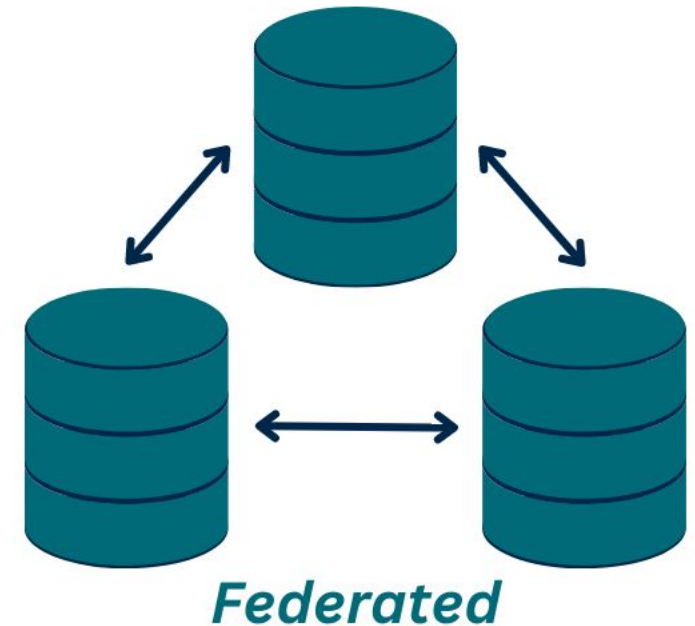
Vertical Federation

Partitions that include the same people but different features (e.g. different data sets, one on health care use and the other a survey, or genomic information)

Federated Analysis

Analysis of data across multiple datasets in a fast and secure manner, where the data are not co-located (i.e., distributed data).

Federated

A spectrum of options

Distributed   **Federated**   **Pooled**

Réseau de recherche sur les données de santé du Canada

Health Data Research Network Canada

Canadian Context

Why federation?

- Constitutional federation
- Local responsibility for delivery of services
- UNDRIP and commitments to Calls to Action and Reconciliation
- Expands opportunities

Réseau de recherche sur les données de santé du Canada

Health Data Research Network Canada

Not all data need to be federated

- Canadian Institute for Health Information
- Statistics Canada
- Canadian Research Data Centre Network
- Canadian Partnership for Tomorrow's Health
- Canadian Longitudinal Study on Aging

Réseau de recherche sur les données de santé du Canada

Health Data Research Network Canada

Current Practice

All analyses need consistency across sites

Common Protocol/Analytic Plan: Common plan across sites; allows amendments for data variation between sites.

Harmonized Data: Data from different sources that have been standardized so that comparisons can be made using a common protocol

Common Data Model: Allows for systematic analysis of disparate observational databases across jurisdictions by transforming the data contained within each jurisdiction's source databases into an agreed set of standardized data elements and data tables.

Réseau de recherche sur les données de santé du Canada

Health Data Research Network Canada

Examples

Réseau de recherche sur les données de santé du Canada

Health Data Research Network Canada

CNODES - Incretin drugs for type II diabetes

JAMA Internal Medicine | Original Investigation

Association Between Incretin-Based Drugs and the Risk of Acute Pancreatitis

Laurent Azoulay, PhD; Kristian B. Filion, PhD; Robert W. Platt, PhD; Matthew Dahl, BSc; Colin R. Dormuth, ScD; Kristin K. Clemens, MD, MSc; Madeleine Durand, MD, MSc; Nianping Hu, MD, PhD; David N. Juurlink, MD, PhD; J. Michael Paterson, MSc; Laura E. Targownik, MD, MSHS; Tanvir C. Turin, MD, PhD; Pierre Ernst, MD, MSc; and the Canadian Network for Observational Drug Effect Studies (CNODES) Investigators

JAMA Intern Med. 2016;176(10):1464-1473. doi:10.1001/jamainternmed.2016.1522
Published online August 1, 2016.

The NEW ENGLAND JOURNAL *of* MEDICINE

ORIGINAL ARTICLE

A Multicenter Observational Study of Incretin-based Drugs and Heart Failure

Kristian B. Filion, Ph.D., Laurent Azoulay, Ph.D., Robert W. Platt, Ph.D., Matthew Dahl, B.Sc., Colin R. Dormuth, Sc.D., Kristin K. Clemens, M.D., Nianping Hu, M.D., Ph.D., J. Michael Paterson, M.Sc., Laura Targownik, M.D., M.S.H.S., Tanvir C. Turin, M.D., Ph.D., Jacob A. Udell, M.D., M.P.H., and Pierre Ernst, M.D., for the CNODES Investigators*

4;12   NEJM.ORG   MARCH 24, 2016

RESEARCH

OPEN ACCESS

CrossMark
click for updates

Incretin based drugs and the risk of pancreatic cancer: international multicentre cohort study
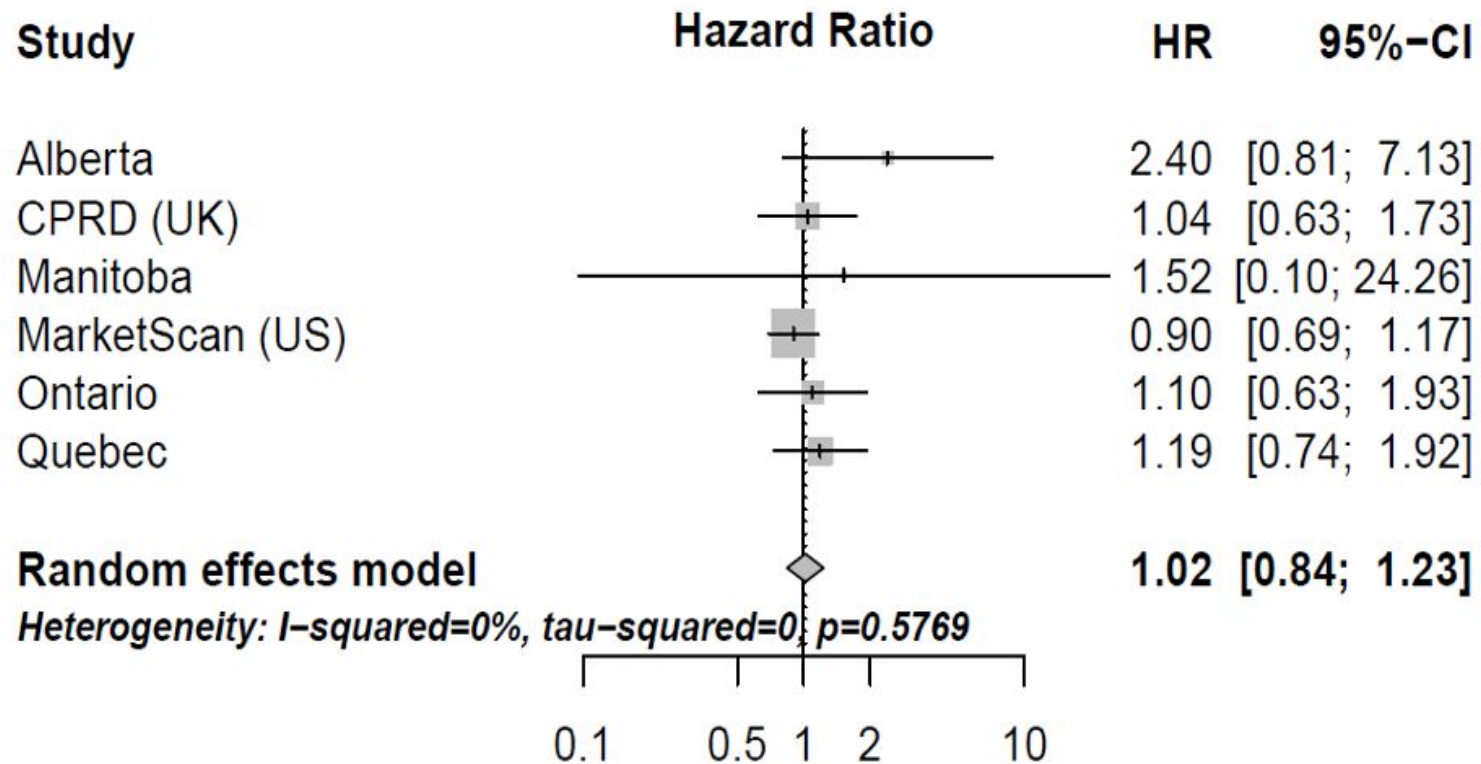
Laurent Azoulay,[1,2] Kristian B Filion,[1,3] Robert W Platt,[4] Matthew Dahl,[5] Colin R Dormuth,[6] Kristin K Clemens,[7] Madeleine Durand,[8] David N Juurlink,[9] Laura E Targownik,[5,10] Tanvir C Turin,[11] J Michael Paterson,[9,12] Pierre Ernst[1,3] for the Canadian Network for Observational Drug Effect Studies (CNODES) Investigators

Cite this as: *BMJ* 2016;352:i581
http://dx.doi.org/10.1136/bmj.i581

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Pancreatic cancer (Incretins vs sulfonylureas)

Study	Hazard Ratio	HR	95%-CI
Alberta		2.40	[0.81; 7.13]
CPRD (UK)		1.04	[0.63; 1.73]
Manitoba		1.52	[0.10; 24.26]
MarketScan (US)		0.90	[0.69; 1.17]
Ontario		1.10	[0.63; 1.93]
Quebec		1.19	[0.74; 1.92]
Random effects model		1.02	[0.84; 1.23]

Heterogeneity: I-squared=0%, tau-squared=0, p=0.5769

0.1 0.5 1 2 10

Exposure: Ever use with 1 year lag

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

OHDSI - LEGEND

Large-scale **E**vidence **G**eneration and **E**valuation in a **N**etwork of **D**atabases

1. Select multiple target and comparator cohorts

2. Carefully design a study, including sensitivity analyses

3. Run (simultaneously) on multiple databases

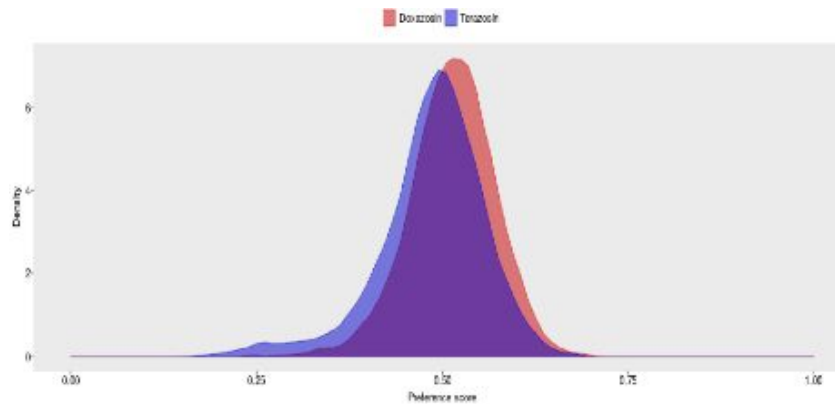4. Diagnostics and sensitivity analyses

5. Publish ALL results

https://www.ohdsi.org/legend-oct2021-update/

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

LEGEND - Example

- Each small box gives a treated/control propensity score distribution for a pair of drugs and an outcome

Réseau de recherche sur les données de santé du Canada

Health Data Research Network Canada

Enablers

What is needed to support federated analysis?

Acknowledge (current) differences:

- Privacy laws
- Policies and practices
- Data collection

Address challenges:

- Cultural / Organizational
- Stewardship / approach to risk
- Size / Complexity of data
- Data curation / preparation and documentation

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

What is needed to support federated analysis?

- Common governance framework for "nodes"
- Connected data infrastructure - ideally integrated with existing structures
- Agreement on data standards for harmonization
- Adequate local AND centralized computing, software and analytic capabilities
- Training and learning resources for users
- Ongoing commitment to IDEA, public engagement, Indigenous data sovereignty, and other fundamental features of good health data analysis
- Sufficient financial support

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

What is next in the Federated Analysis collective learning series?

Acknowledgements

Organized with support of a steering committee

Kim McGrail | HDRN Canada

Charles Burchill | HDRN Canada

Daryl Fung | HDRN Canada

Amy Freier | HDRN Canada

Johanne Provençal | CRDCN

Grant Gibson | CRDCN

Robert Platt | CNODES

Therese Stukel | ICES

Michael Paterson | ICES

Stephanie Dixon | ICES

Claudia Sanmartin | StatCan

Marc Fiume | DNAstack

Bryce Pickard | Integrate AI

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Topics & Timeline

FEBRUARY & MARCH 2024

Topic 2: Current Approaches for Distributed Analysis

Session A: Lecture / *Session B*: Panel Discussion

Purpose: To explore current approaches to federated analysis, and their respective benefits and limitations, through practical use cases and discussions.

APRIL 2024

Topic 3: Statistical Review with Distributed Data

Purpose: To demonstrate the possible statistical analyses with distributed data through practical examples and discuss key aspects around the value and challenges.

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Topics & Timeline

MAY 2024

Topic 4: Panel Discussion: Artificial Intelligence in a Federated Landscape

Purpose: To provide an overview of possible AI analytics in a federated landscape and other artificial intelligence considerations within the Canadian health data context.

JUNE 2024

Topic 5: Trusted/Secure Environments & Their Role in Supporting Federated Analysis -

Purpose: To describe the trusted/secure environments for federated analysis that exist in Canada and internationally, how they evolved, and their risks, best practices and arising opportunities.

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Topics & Timeline

Topic 6: Possibilities, Uncertainty, and the Future of Federation

Purpose: To explore the possibilities of what can be achieved using federated analysis, while discussing what is needed to establish a better federated analysis environment across Canada.

Summing up: Next steps for federated analysis

Possible hybrid event to consolidate learning and set goals and plans to support federated data analysis

Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Thank you!

For events & updates, follow HDRN Canada:

hdrnrrds   @hdrn_rrds   @hdrncanada   hdrn.ca