Adventures in Replication: Initial Experiences related to the Joint ISPE/ISPOR Guidance

Judith C. Maro[1], Shirley V. Wang[2], Ting-Ying (Jane) Huang[1], Robert W. Platt[3], Xavier Kurz[4]

1. Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA
2. Harvard Medical School and Brigham and Women's Hospital, Boston, MA
3. McGill University, Montreal, QC, Canada
4. European Medicines Agency, London, United Kingdom

Agenda

- Overview of ISPE/ISPOR Guidance
- Applied Example: REPEAT Initiative
- Applied Example: Sentinel System
- Applied Example: Canadian Network for Observational Drug Effect Studies
- Regulatory Perspective: Why Replication is So Important

Adventures in Replication:

Overview of ISPE/ISPOR Joint Task Force papers on ‘real world’ evidence to guide decision making

Shirley V Wang PhD, ScM

Division of Pharmacoepidemiology, Department of Medicine,

Brigham and Women’s Hospital, Harvard Medical School

Disclosures

- This work was not funded
- At the time that this work was conducted, Dr. Wang was principal investigator on other grants from:
  - Agency for Healthcare and Research Quality
  - National Institute of Aging
  - Laura and John Arnold Foundation
  - FDA Sentinel Initiative
  - Investigator initiated grants to Brigham and Women's Hospital from Novartis, J & J, Boehringer Ingelheim for unrelated work
- She is a consultant to Aetion Inc, for unrelated work

Timely Topic

Recent legislation with sections focused on evaluating when and how to make greater use of **'real world' evidence** from **'real world' data** (*administrative, clinical healthcare databases*) to support regulatory decisions:

- 21st Century Cures Act (FDA)
- PDUFA VI (FDA)
- Adaptive Pathways (EMA)

Joint Task Force ISPE/ISPOR

Improving the Confidence of Decision-Makers in Utilizing Real World Evidence By Increasing Transparency and Reproducibility

Received: 21 July 2017 | Revised: 26 July 2017 | Accepted: 28 July 2017
DOI: 10.1002/pds.4297

WILEY

ORIGINAL REPORT

Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making

Marc L. Berger[1] | Harold Sox[2] | Richard J. Willke[3] | Diana L. Brixner[4] | Hans-Georg Eichler[5] | Wim Goettsch[6] | David Madigan[7] | Amr Makady[6] | Sebastian Schneeweiss[8] | Rosanna Tarricone[9] | Shirley V. Wang[8] | John Watkins[10] | C. Daniel Mullins[11]

Received: 21 July 2017 | Revised: 25 July 2017 | Accepted: 25 July 2017
DOI: 10.1002/pds.4295

WILEY

ORIGINAL REPORT

Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0

Shirley V. Wang[1,2] | Sebastian Schneeweiss[1,2] | Marc L. Berger[3] | Jeffrey Brown[4] | Frank de Vries[5] | Ian Douglas[6] | Joshua J. Gagne[1,2] | Rosa Gini[7] | Olaf Klungel[8] | C. Daniel Mullins[9] | Michael D. Nguyen[10] | Jeremy A. Rassen[11] | Liam Smeeth[6] | Miriam Sturkenboom[12] |
on behalf of the joint ISPE-ISPOR Special Task Force on Real World Evidence in Health Care Decision Making

Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0

- Importance of achieving reproducible research is well recognized in many reporting guidelines
  - STROBE, RECORD, PCORI Methodology Report, EnCePP, ISPE Guidelines for Good Pharmacoepidemiology Practice (GPP)

- These increase transparency, but even strict adherence would not necessarily provide all the information for full reproducibility

- **Goal: catalogue specific decisions made during study implementation**

Reproducibility

What is reproducibility in database studies?

Reproducibility

	Data Source	Methods
Analytic reproduction *Re-running the same code on same data*	**Same**	**Same**
Direct replication *Independent implementation of a specific study*	**Same**	**Same**
Conceptual replication (robustness) *Implementing a study of the same exposure (and comparator), outcome and estimand of interest*	**Different**	**Same**
	Same	**Different**
	Different	**Different**

Reproducibility

What is reproducibility in database studies?

	Data Source	Methods
Analytic reproduction *Re-running the same code on same data*	**Same**	**Same**

Reproducibility

Hazard ratio = 2.0
Hazard ratio = 2.0
Hazard ratio = 2.0

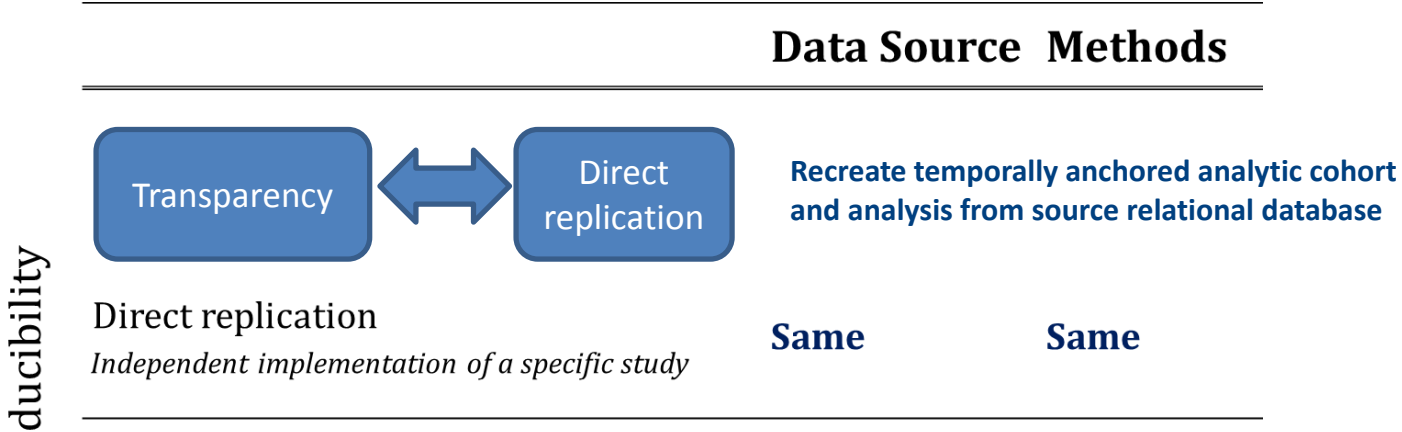
Important but not transparent by itself

Thousands of lines of code to create a temporally anchored analytic cohort from raw longitudinal data streams

What scientific decisions is the code implementing? Agree with the validity and/or relevance for the question of interest?

Reproducibility

What is reproducibility in database studies?



Transparency in design and implementation decisions are a necessary pre-condition for direct replication

∴ **Ability to directly replicate a study is a proxy for transparency of study methodology**

Need transparency to assess validity and relevance of evidence

Reproducibility

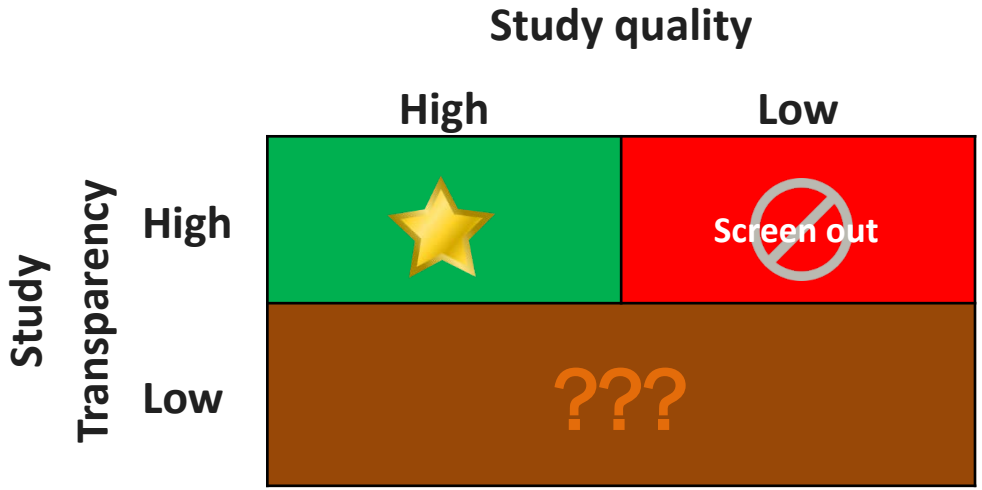
What is reproducibility in database studies?

Reproducibility	Data Source	Methods
	Most common, most interesting? **Why do results differ or converge?**	**Need transparency to understand** • **Subtle design/implementation differences** • **Differences in data** • **Differences in population**
Conceptual replication (robustness) *Implementing a study of the same exposure (and comparator), outcome and estimand of interest*	**Different**	**Same**
	Same	**Different**
	Different	**Different**

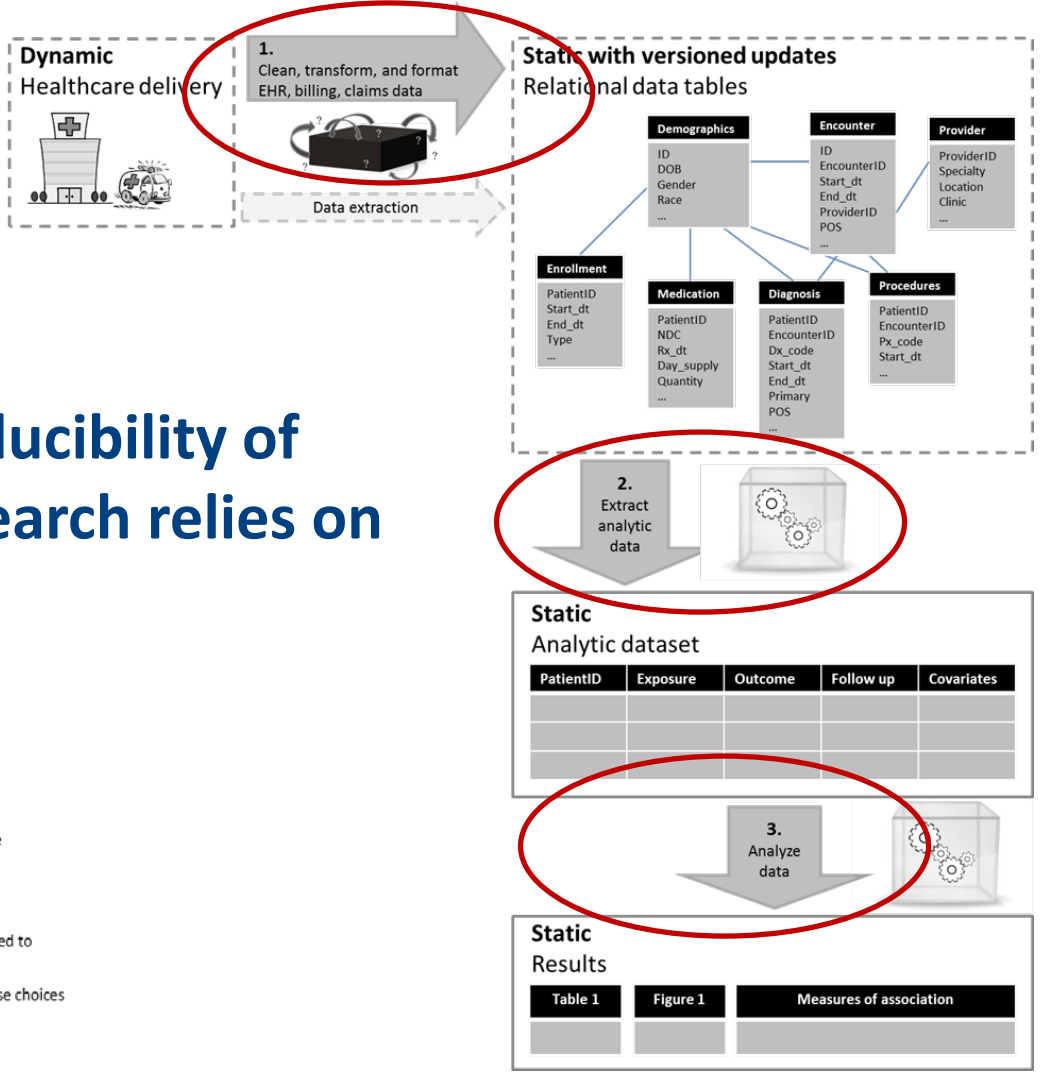
Important point to keep in mind

Transparency facilitates **assessment** of validity, relevance, replicability

Transparent and replicable ≠ Scientifically valid and robust



Transparency and reproducibility of healthcare database research relies on data provenance:



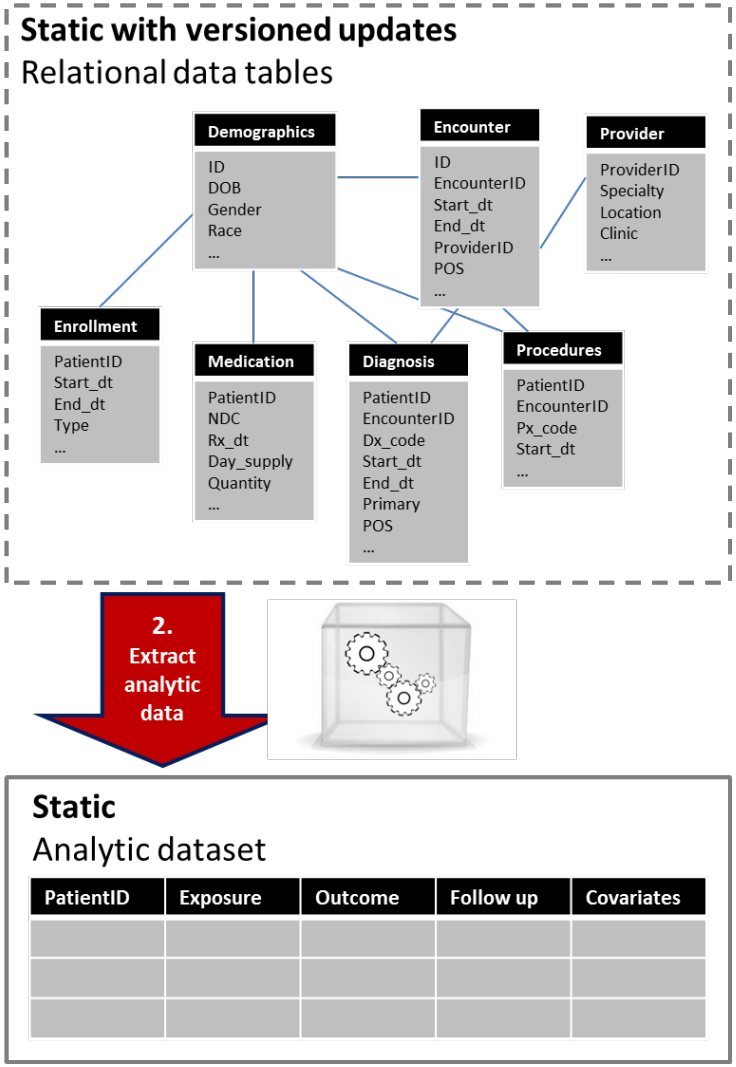
Dashed line boxes represent a snapshot of a dynamically updated data model at a moment in time

Solid line boxes represent a static data model

Gray arrows represent transition from one data model to another

Black boxes represent operational choices made to extract and reshape data before data is provided to researcher

Clear boxes represent operational choices performed by researchers after data was provided. These choices should be made transparent by the researchers with reporting of study findings

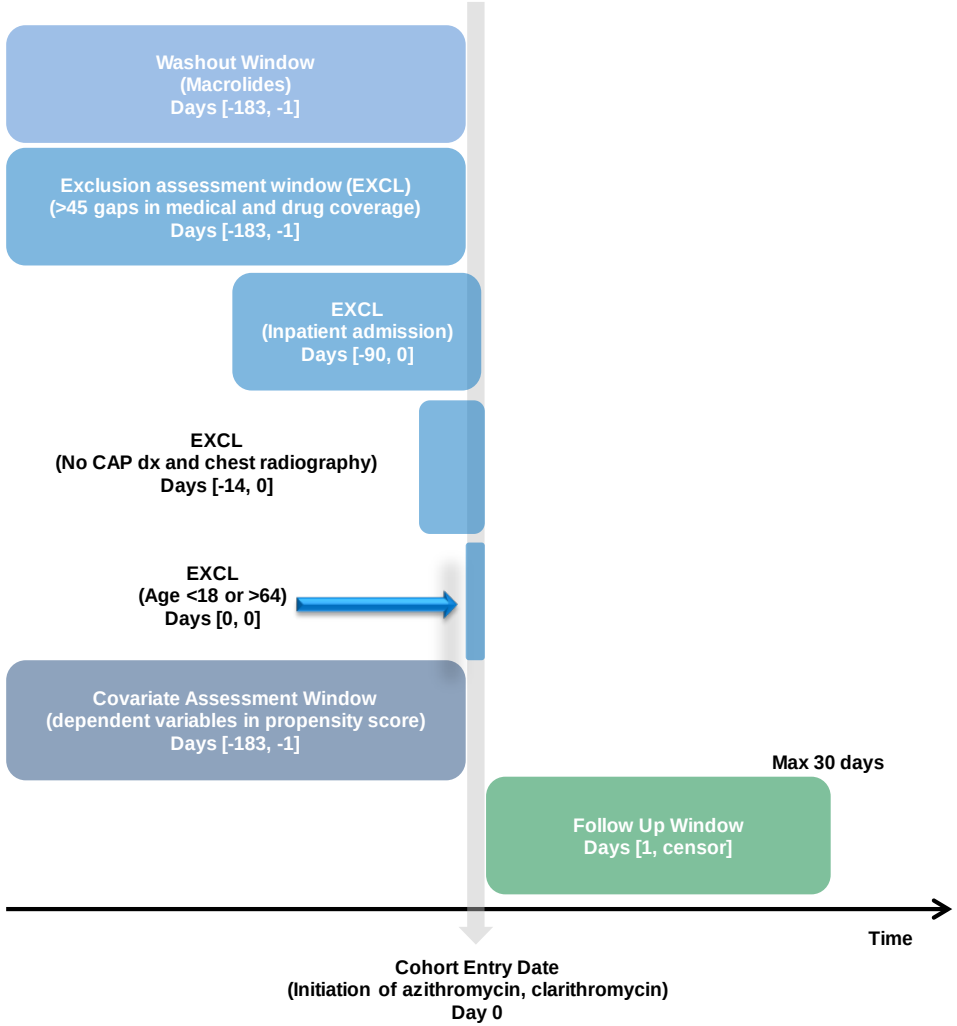


Step 2
Parameters for creation of a study population

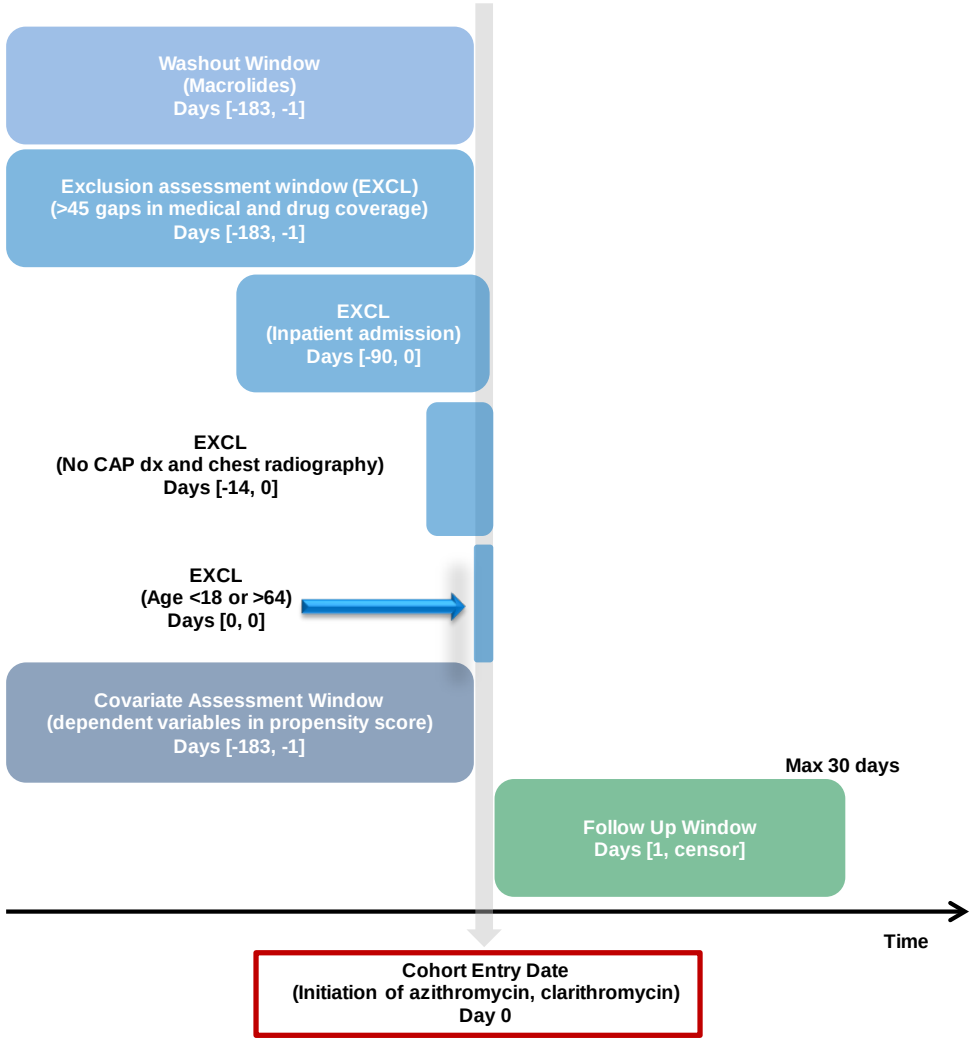
Comprehensive catalogue with 9 sections:

A. Data source
B. Design diagram
C. Inclusion/exclusion criteria (attrition table)
D. Exposure definition
E. Follow up definition
F. Outcome definition
G. Covariates
H. Control sampling
I. Software

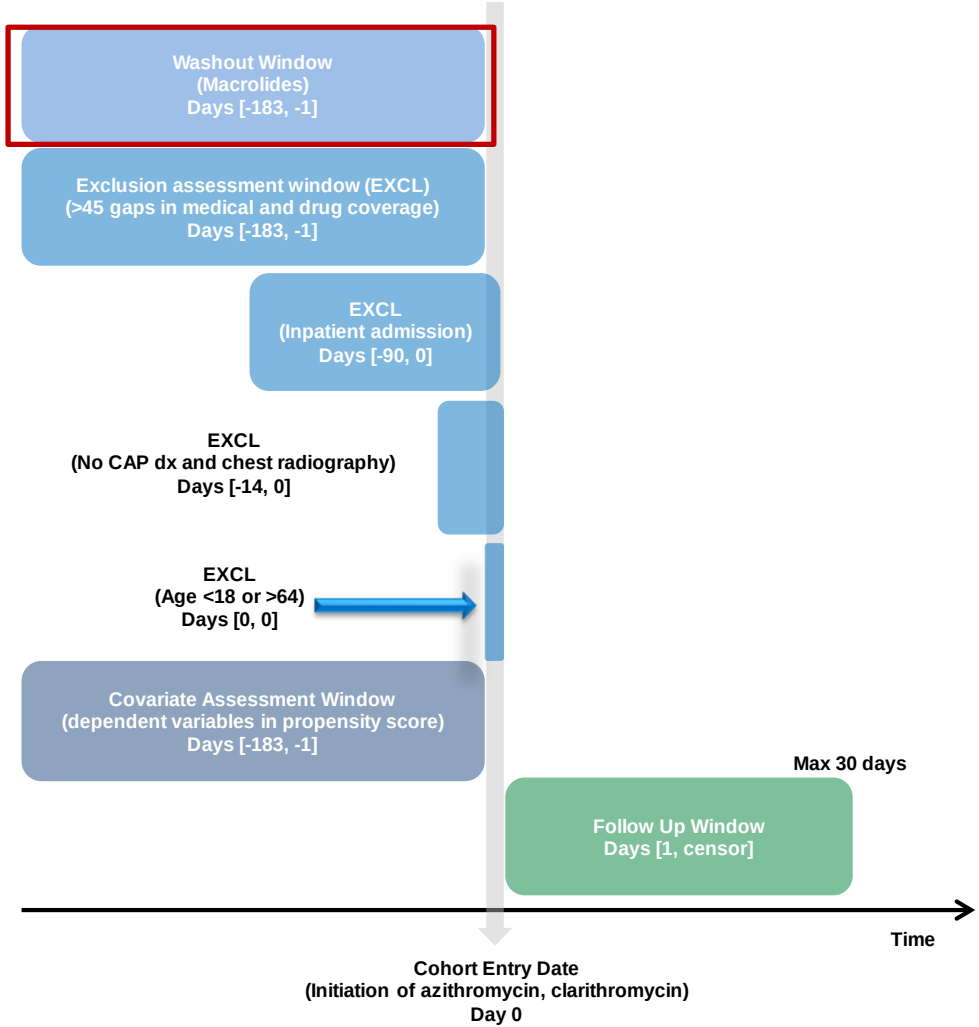
Design diagram



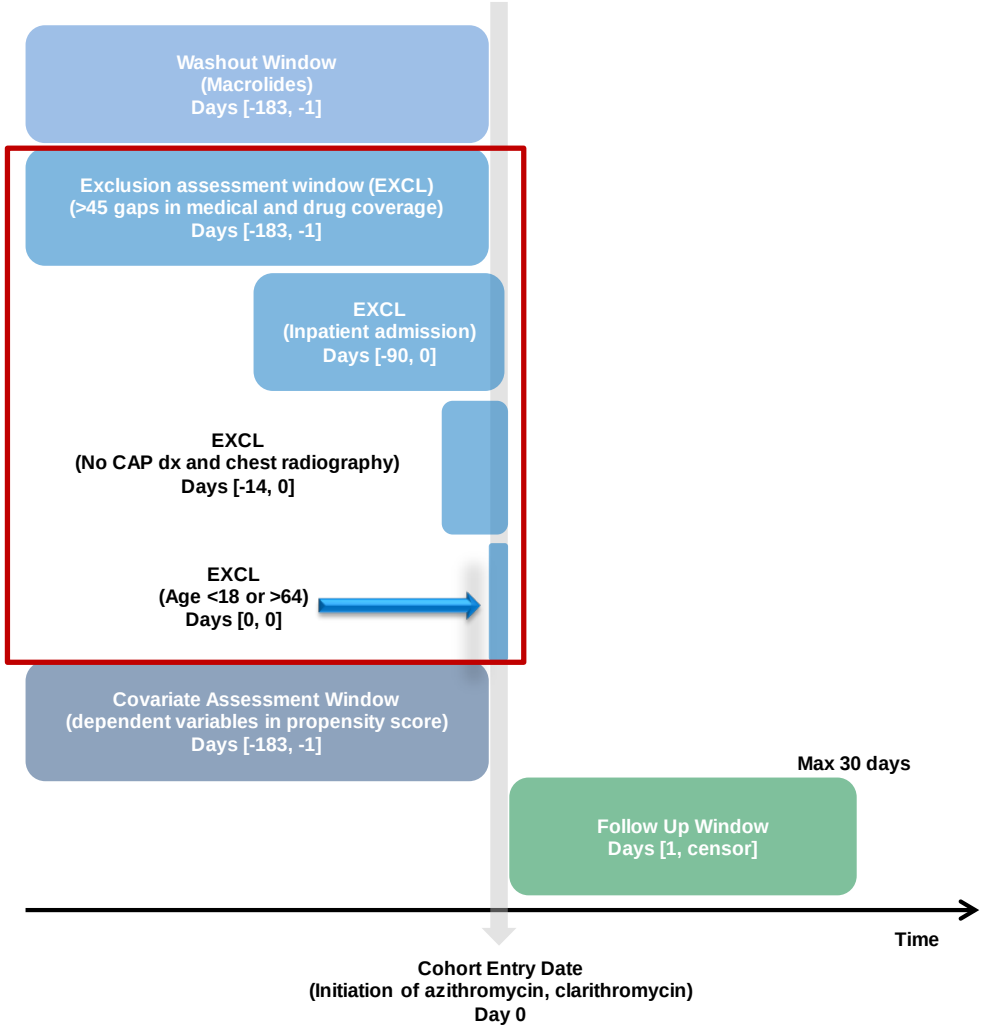
Design diagram



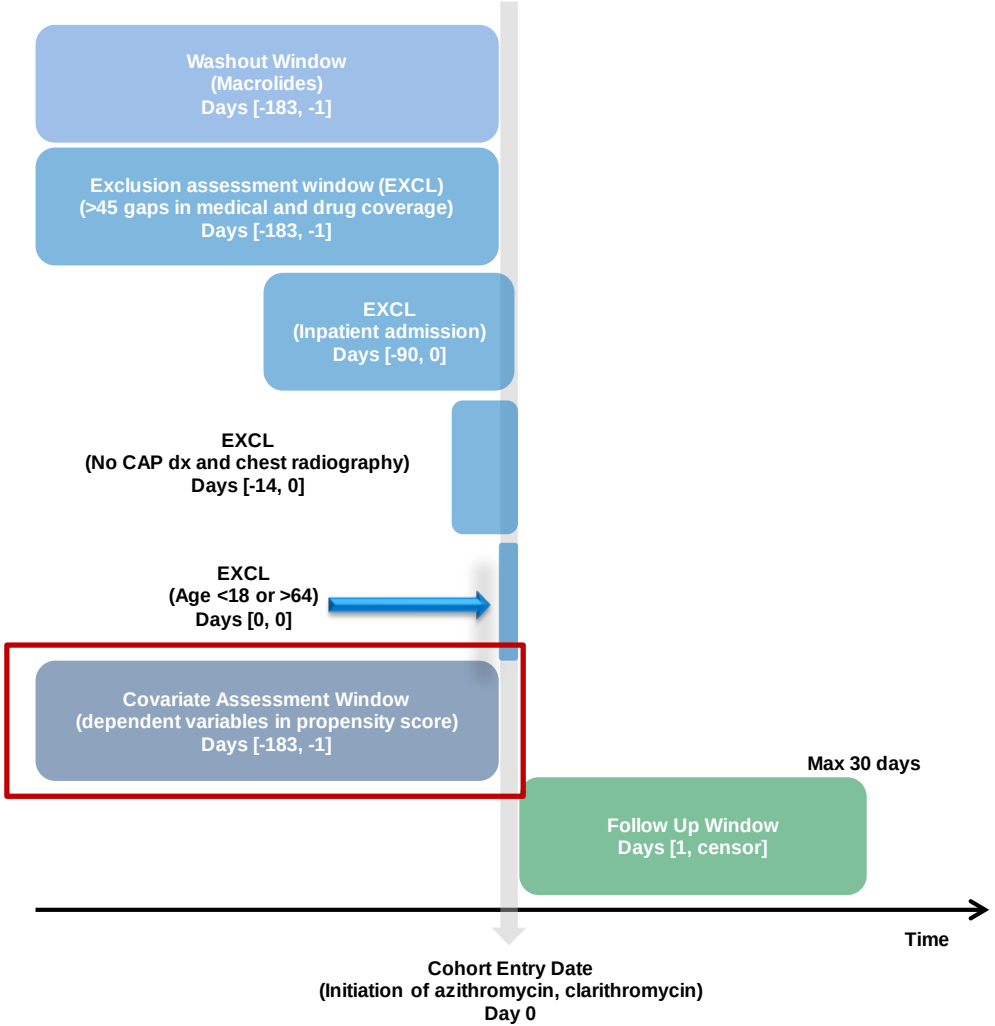
Design diagram



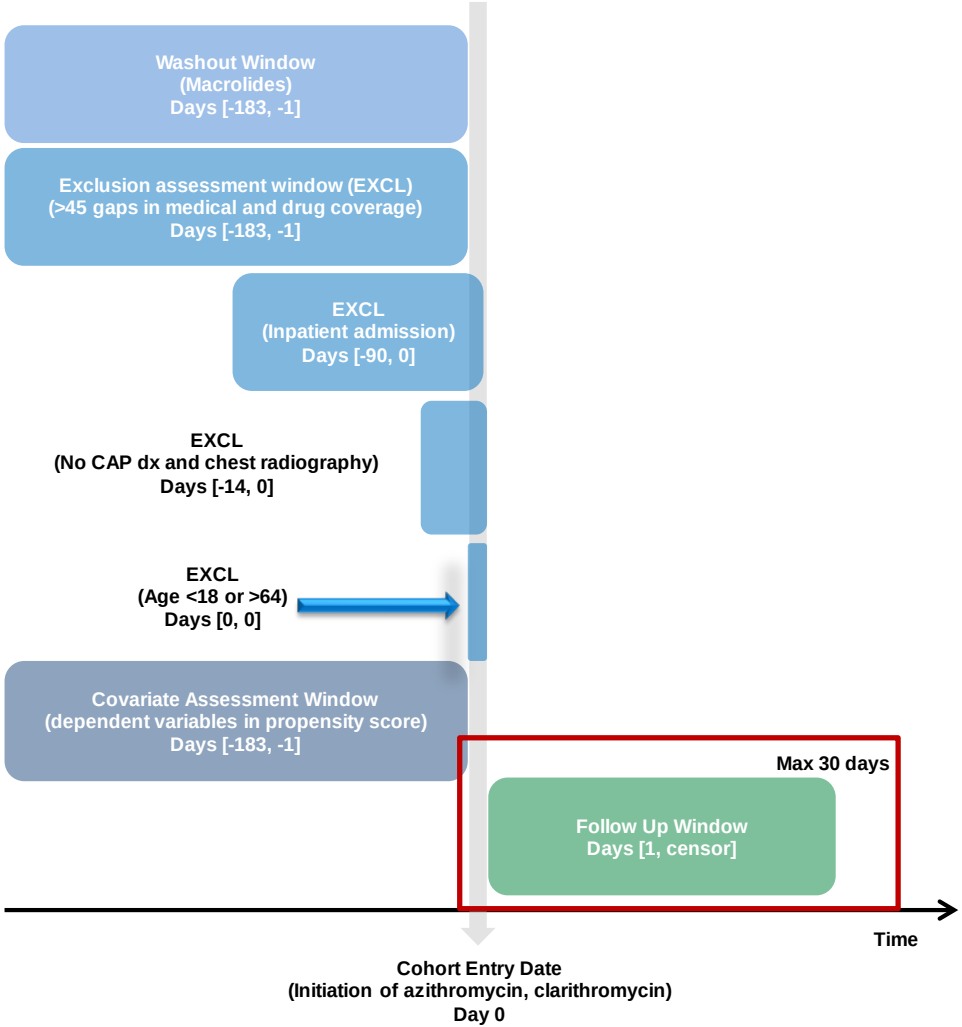
Design diagram



Design diagram

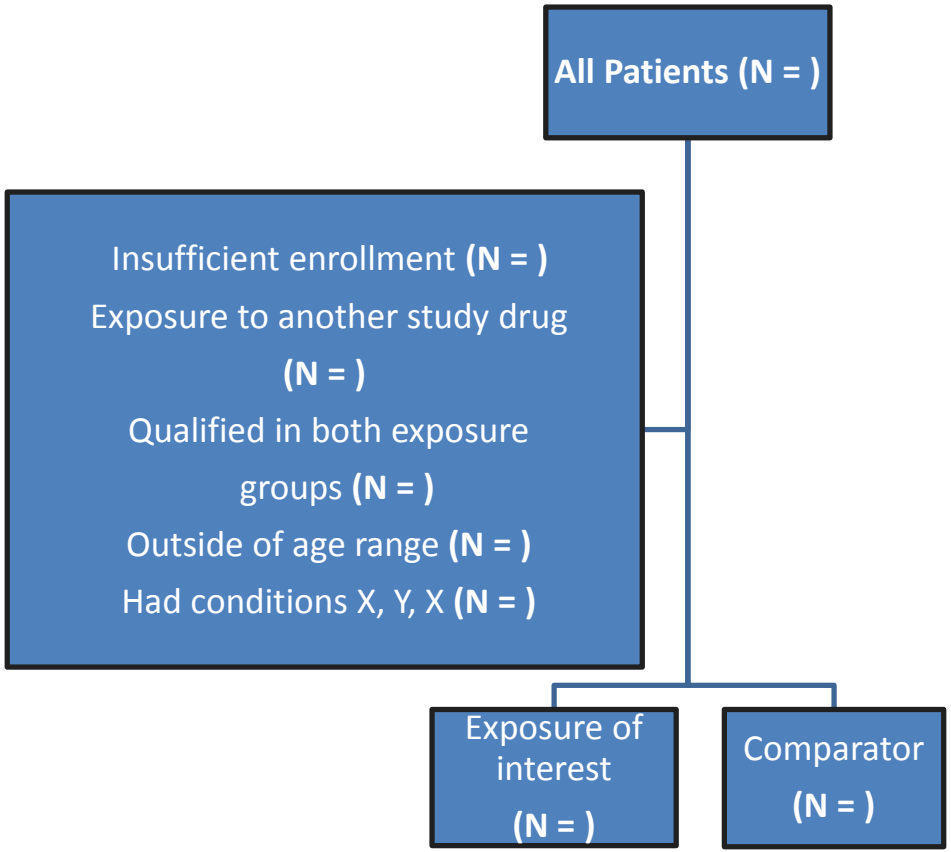


Design diagram



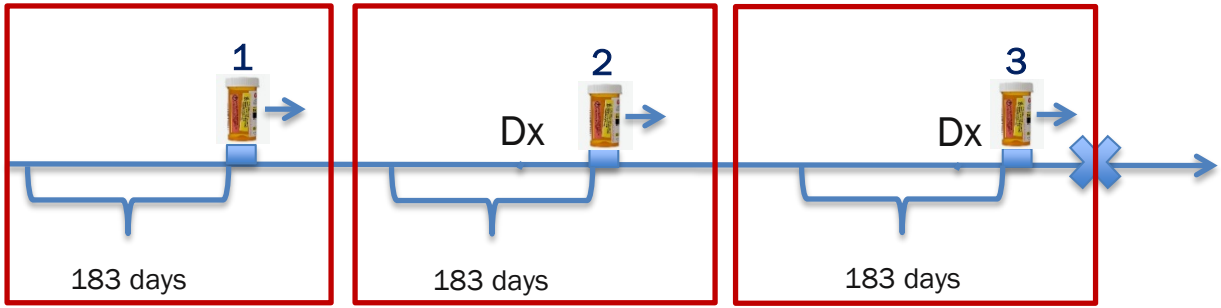
Attrition table

showing how the study population was derived



Example specificity in reporting on exclusions

Patients entered the study at initiation of metformin after a 183 day washout without dispensation of any anti-diabetic agents. Patients were required to have diabetes, defined by ICD9 codes 250. * recorded in any care setting and any diagnosis position within 183 days prior to but not including study entry date.

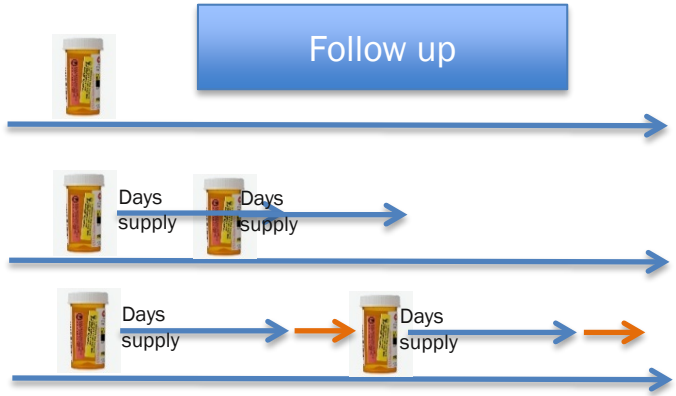


What is the study entry date?

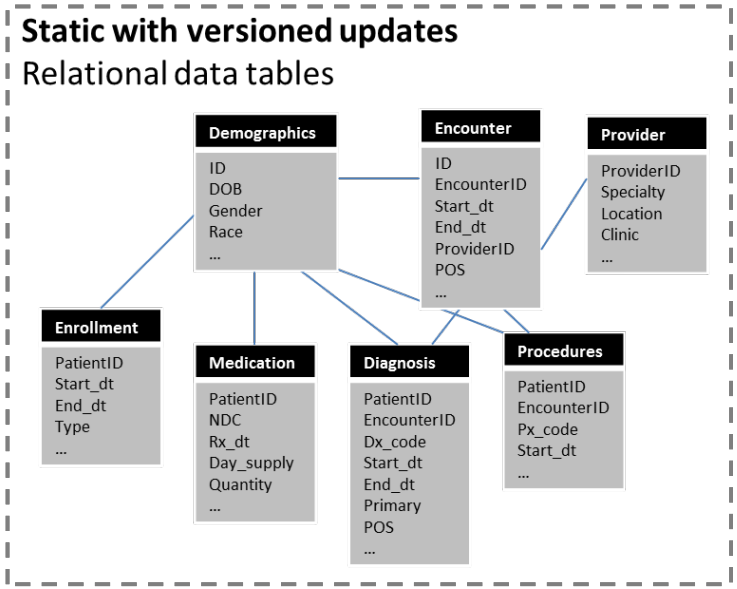
- Identify first new initiation date (1)
  - *Patient does not contribute*
- Consider all new initiation dates (1,2,3), use first that meets inclusion/exclusion
  - *Patient contributes (2)*
- Consider all new initiation dates (1,2,3), use all that meet inclusion/exclusion
  - *Patient contributes (2, 3)*

Example specificity defining exposure

- Codes
  - Frequency and temporality
  - Diagnosis position
  - Care setting
- Type of exposure (e.g. incident, cumulative, time-varying)
- Induction period
- Exposure risk window
- Stockpiling
- Bridging exposure episodes
- Exposure extension
- Switching/add on



Wang 2017, ISPE/ISPOR Joint Task Force



Static

Analytic dataset

PatientID	Exposure	Outcome	Follow up	Covariates

Step 2
Parameters for creation of study population

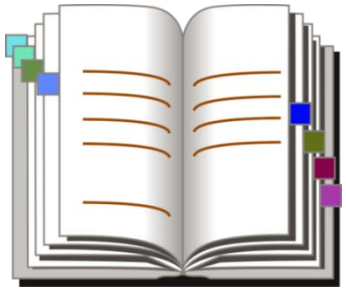
Comprehensive catalogue with 9 sections:

A. Data source
B. Design diagram
C. Inclusion/exclusion criteria (attrition table)
D. Exposure definition
E. Follow up definition
F. Outcome definition
G. Covariates
H. Control sampling
I. Software

Summary

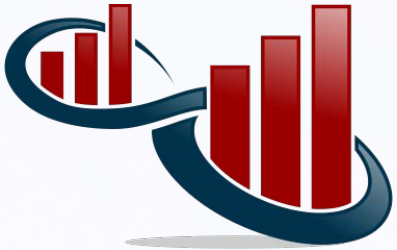
Comprehensive catalogue of **specific operational parameters** that represent scientific decisions made when defining a study population from longitudinal data captured in claims and EHRs

- Reporting these will facilitate replicability and validity assessment
- Expect catalogue will grow and change over time



Consensus - limited set of parameters are absolutely necessary to recreate a study population

Which? *Debatable.*



REPEAT

Reproducible Evidence: Practices to Enhance and Achieve Transparency

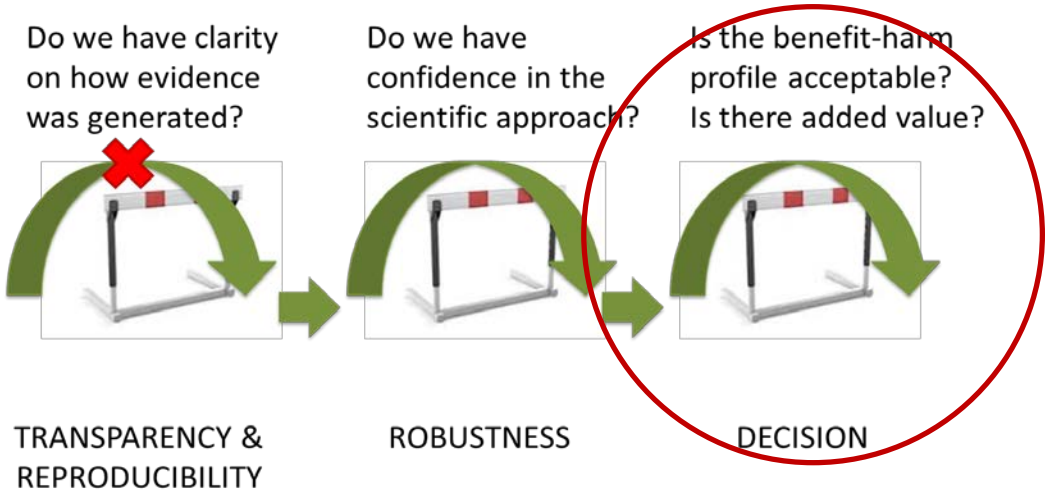
www.repeatinitiative.org

Disclosures

- This work was funded by the Laura and John Arnold Foundation
- At the time that this work was conducted, Dr. Wang was principal investigator on other grants from:
  - Agency for Healthcare and Research Quality
  - National Institute of Aging
  - Laura and John Arnold Foundation
  - FDA Sentinel Initiative
  - Investigator initiated grants to Brigham and Women's Hospital from Novartis, J & J, Boehringer Ingelheim for unrelated work
- She is a consultant to Aetion Inc, for unrelated work

Objective

To increase the confidence of decision makers in using evidence from healthcare databases by producing empirically based recommendations on how to transparently report on study implementation, achieve reproducible and robust findings



Aim 1. To quantify the current state of healthcare database study reproducibility via direct replication

1. Systematic search using Google Scholar

Top h-5 clinical, epidemiology journals

- Published after Jan 1, 2011
- “cohort” + “claims” + database name

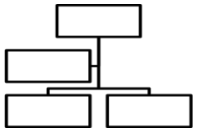
Aim 1. To quantify the current state of healthcare database study reproducibility via direct replication

1. Systematic search using Google Scholar

Top h-5 clinical, epidemiology journals

- Published after Jan 1, 2011
- “cohort” + “claims” + database name

2. Apply exclusion criteria



CONSORT style diagram

- Include descriptive, comparative safety/effectiveness cohort studies
- Exclude if data source mismatch, PDF unavailable, methods study, etc.

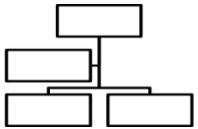
Aim 1. To quantify the current state of healthcare database study reproducibility via direct replication

1. Systematic search using Google Scholar

Top h-5 clinical, epidemiology journals

- Published after Jan 1, 2011
- "cohort" + "claims" + database name

2. Apply exclusion criteria



CONSORT style diagram

- Include descriptive, comparative safety/effectiveness cohort studies
- Exclude if data source mismatch, PDF unavailable, methods study, etc.

Random sample **250 studies**

3. Evaluate transparency considering all publicly available information

Standardized extraction form

- Based on ISPE/ISPOR catalogue
- Measure/describe how often specific parameter decisions were unclear

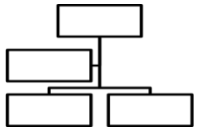
Aim 1. To quantify the current state of healthcare database study reproducibility via direct replication

1. Systematic search using Google Scholar

Top h-5 clinical, epidemiology journals

- Published after Jan 1, 2011
- "cohort" + "claims" + database name

2. Apply exclusion criteria



CONSORT style diagram

- Include descriptive, comparative safety/effectiveness cohort studies
- Exclude if data source mismatch, PDF unavailable, methods study, etc.

Random sample **250 studies**

3. Evaluate transparency considering all publicly available information

Standardized extraction form

- Based on ISPE/ISPOR catalogue
- Measure/describe how often specific parameter decisions were unclear

4. Replicate 150 studies
80% comparative
(blind to original results)

Metrics to quantify replicability

- Abs. Diff, Std. Diff, "calibration", etc.

Aim 1. To quantify the current state of healthcare database study reproducibility via direct replication

1. Systematic search using Google Scholar

2. Apply exclusion criteria

Random sample **250 studies**

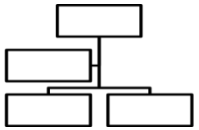
3. Evaluate transparency considering all publicly available information

4. Replicate 150 studies
80% comparative
(blind to original results)

5. Contact original authors to discuss assumptions, understand differences

Top h-5 clinical, epidemiology journals

- Published after Jan 1, 2011
- "cohort" + "claims" + database name



CONSORT style diagram

- Include descriptive, comparative safety/effectiveness cohort studies
- Exclude if data source mismatch, PDF unavailable, methods study, etc.

Standardized extraction form

- Based on ISPE/ISPOR catalogue
- Measure/describe how often specific parameter decisions were unclear

Metrics to quantify replicability

- Abs. Diff, Std. Diff, "calibration", etc.

Aim 2. To evaluate the robustness of evidence currently found in healthcare database studies

Involve original investigators

1. Identify random sample of 50 comparative studies

- **Closely replicated**
- **Noted design/analysis issue**
- **Implementation parameters ≠ intended question?**

Aim 2. To evaluate the robustness of evidence currently found in healthcare database studies

Involve original investigators

1. Identify random sample of 50 comparative studies

- **Closely replicated**
- **Noted design/analysis issue**
- **Implementation parameters ≠ intended question?**

2. Conduct numerous sensitivity analyses

- **Plausible alternative parameters**
- **Address design/analysis issues**
- **Assay sensitivity – e.g. negative control outcomes**

Aim 2. To evaluate the robustness of evidence currently found in healthcare database studies

Involve original investigators

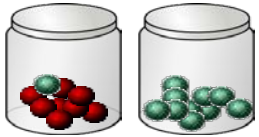
1. Identify random sample of 50 comparative studies

- Closely replicated
- Noted design/analysis issue
- Implementation parameters ≠ intended question?

2. Conduct numerous sensitivity analyses

- Plausible alternative parameters
- Address design/analysis issues
- Assay sensitivity – e.g. negative control outcomes

3. Conduct external adjustment under varying assumptions



- Quantitative bias adjustment (misclassification)
- Residual confounding

Aim 2. To evaluate the robustness of evidence currently found in healthcare database studies

Involve original investigators

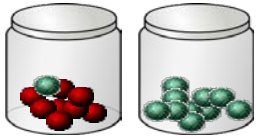
1. Identify random sample of 50 comparative studies

- Closely replicated
- Noted design/analysis issue
- Implementation parameters ≠ intended question?

2. Conduct numerous sensitivity analyses

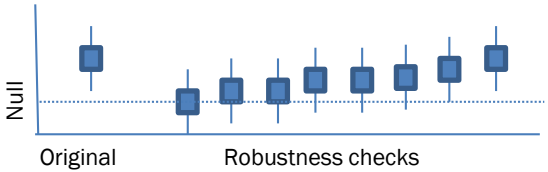
- Plausible alternative parameters
- Address design/analysis issues
- Assay sensitivity – e.g. negative control outcomes

3. Conduct external adjustment under varying assumptions



- Quantitative bias adjustment (misclassification)
- Residual confounding

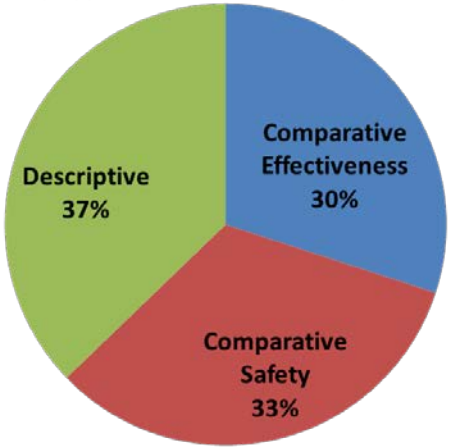
4. Evaluate robustness of evidence



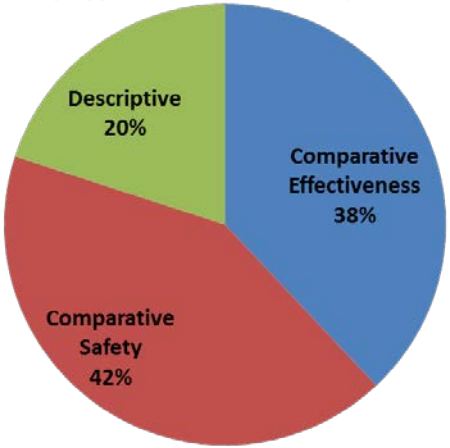
- Vibration ratio

Random Sample of Peer-Reviewed, Published Database Studies

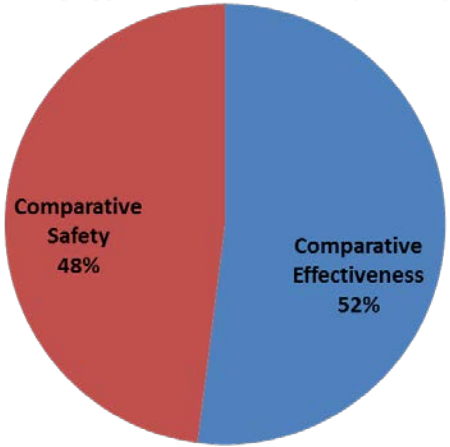
Study types for transparency (N = 250)



Study types for replication (N = 150)



Study types for robustness (N = 50)



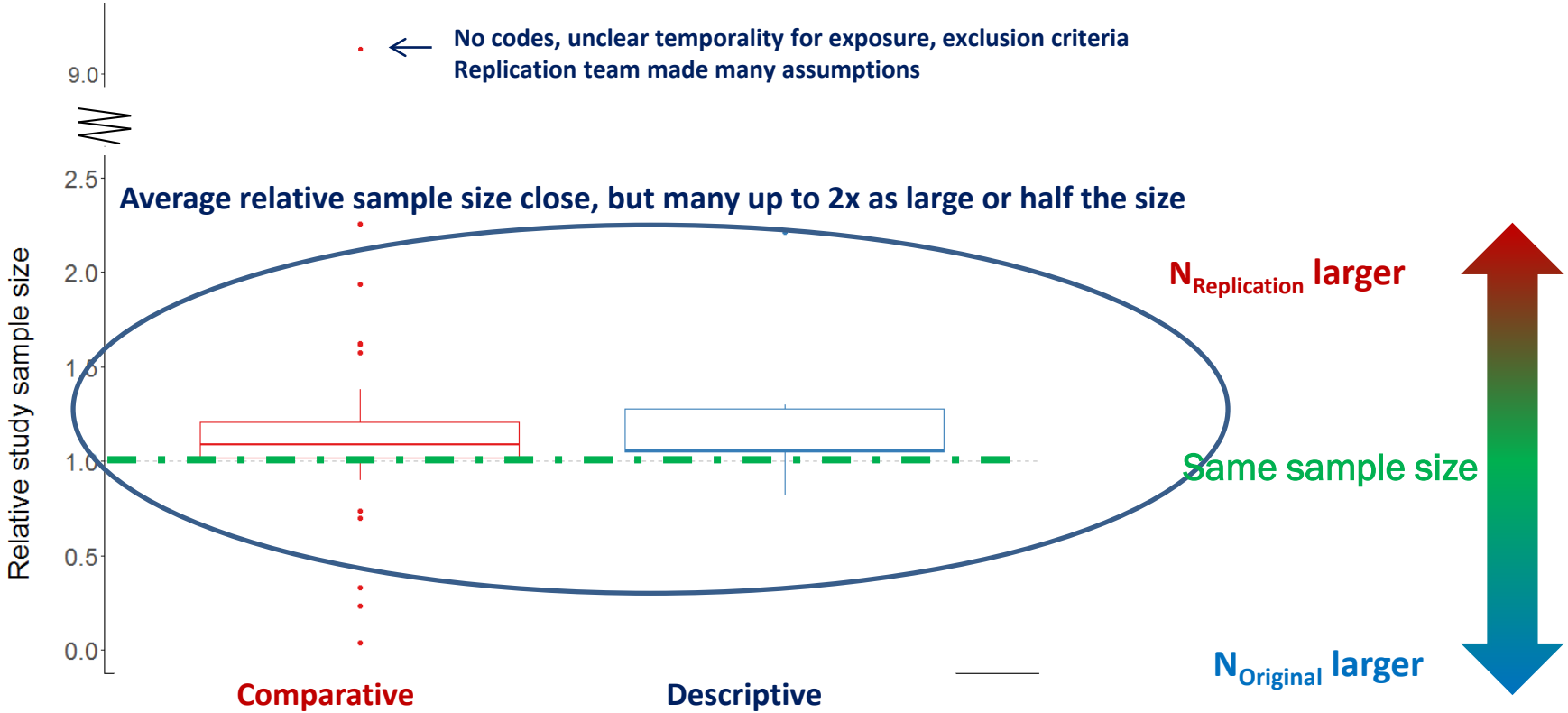
Current progress

Transparency Evaluation	Replication	Robustness	Author Contacts
128 of 250	56 of 150	1 of 50	1 of 150 (only 10 attempted contacts)

INTERIM RESULTS

Relative sample size of replication versus original

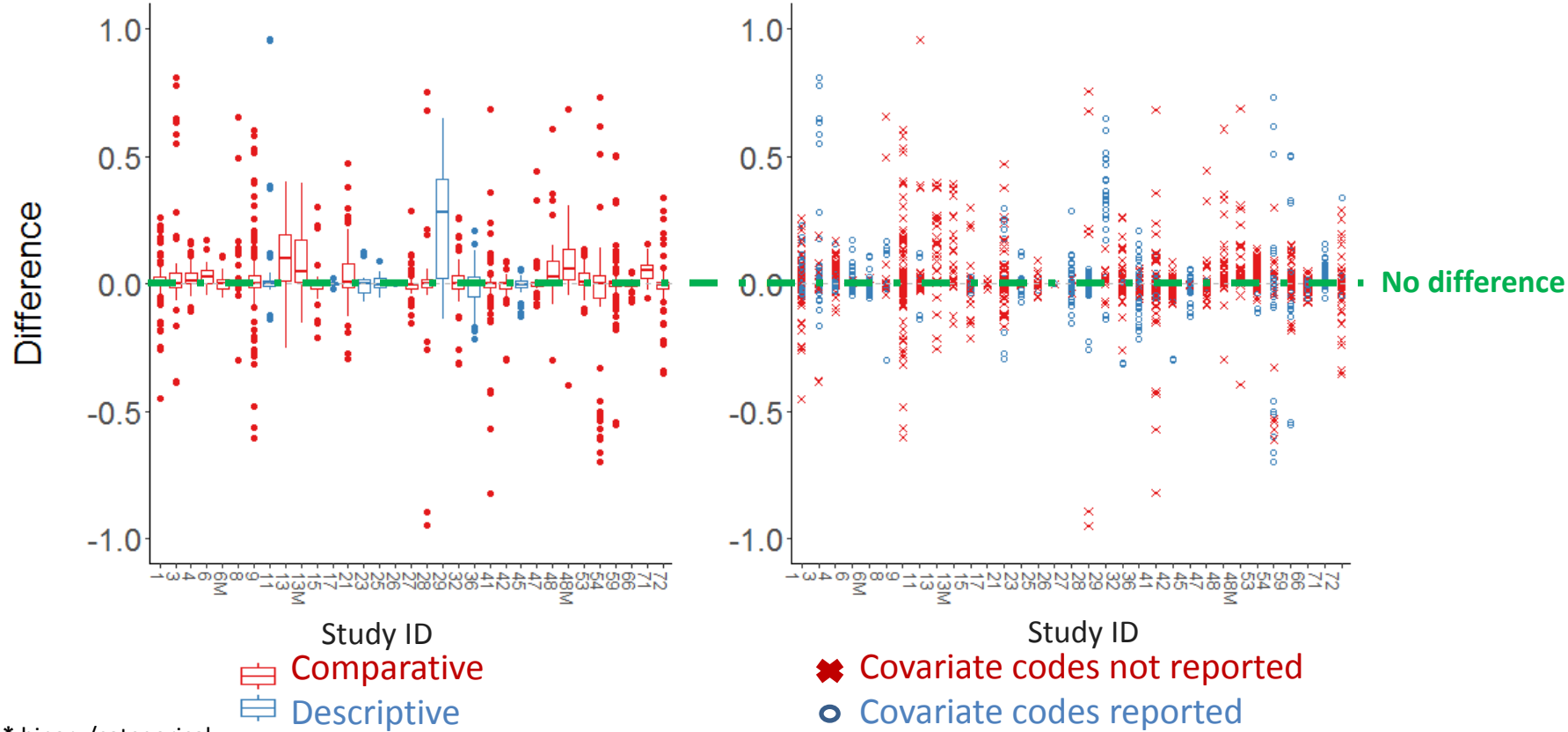
($N_{replication}/N_{original}$)



INTERIM RESULTS

Difference in baseline characteristics* of cohort

(% original – % replication)

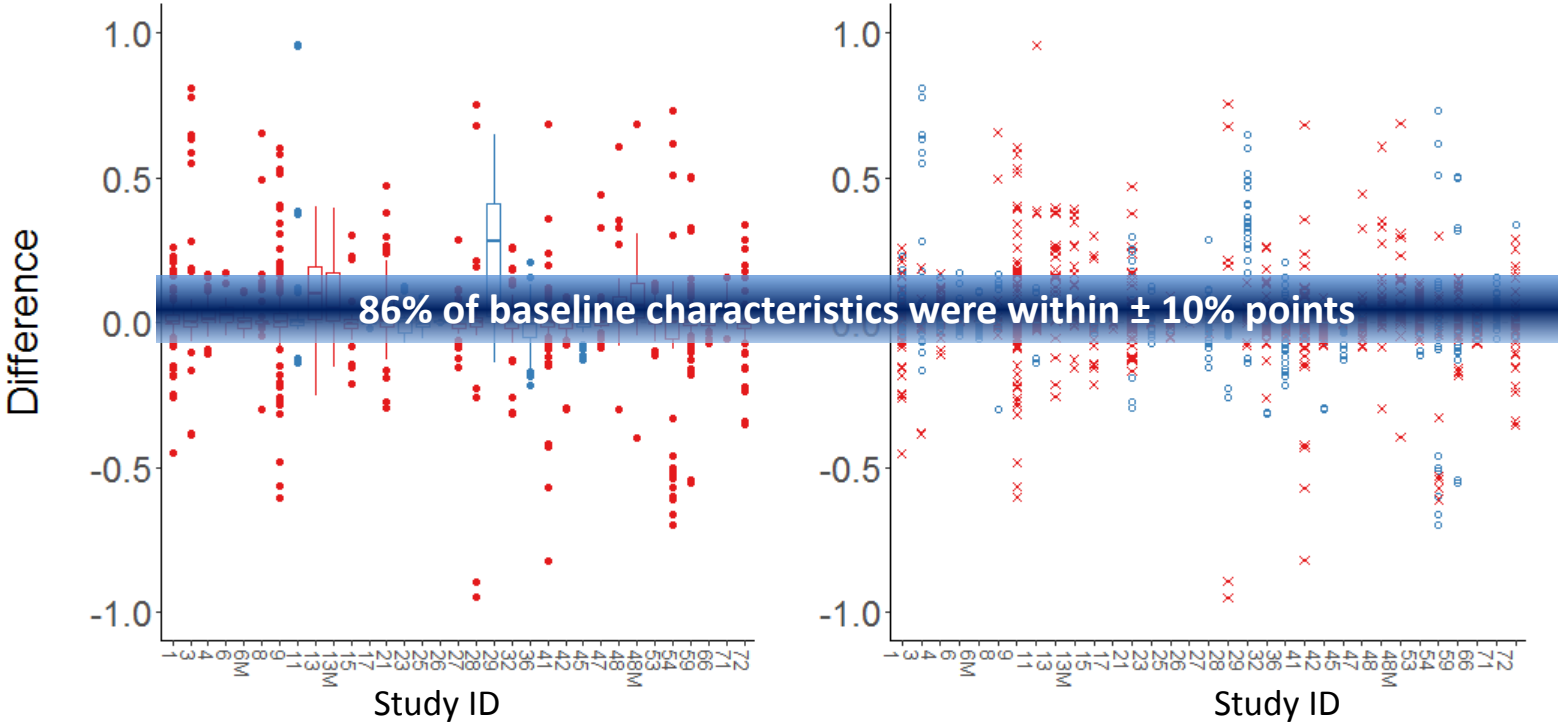


* binary/categorical

INTERIM RESULTS

Difference in baseline characteristics* of cohort

(% original – % replication)



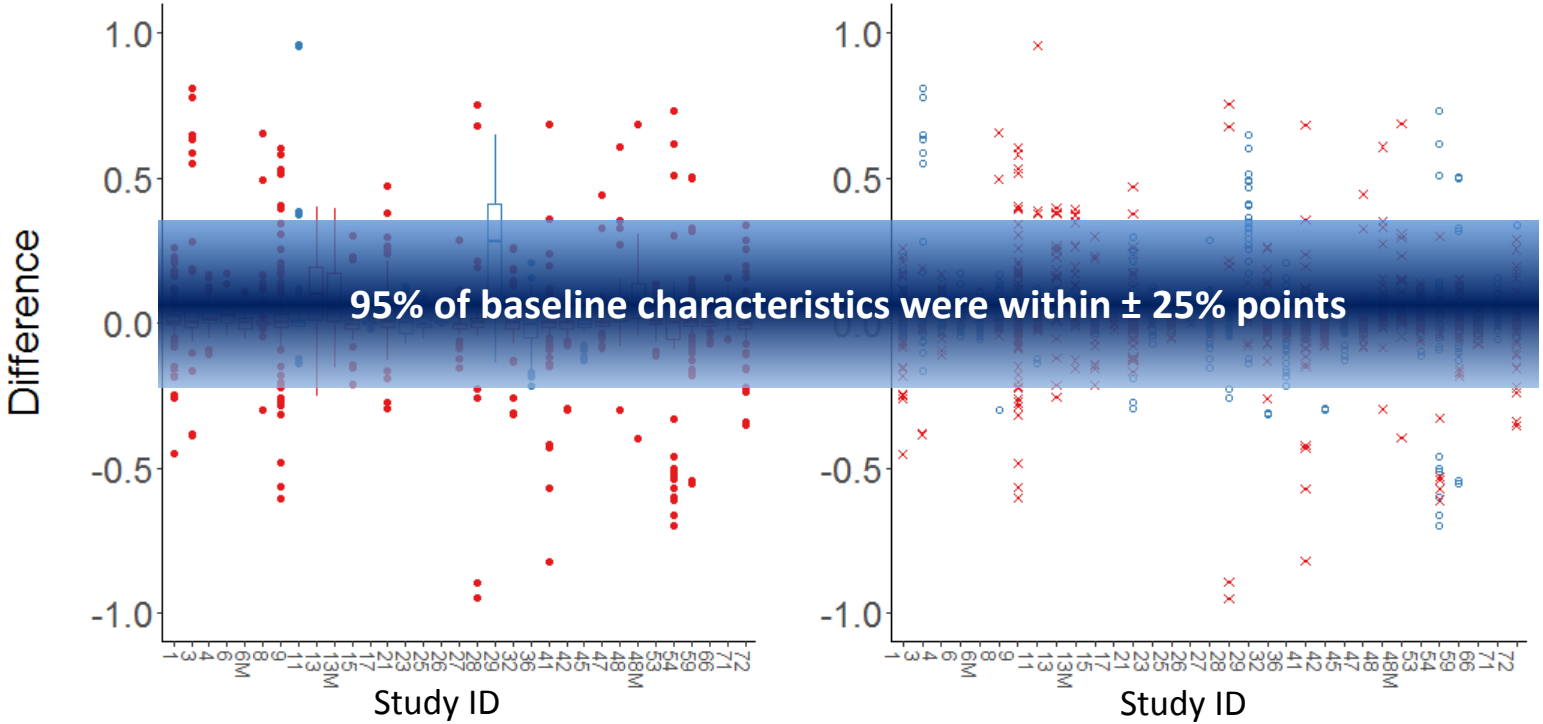
86% of baseline characteristics were within ± 10% points

* binary/categorical

INTERIM RESULTS

Difference in baseline characteristics* of cohort

(% original – % replication)

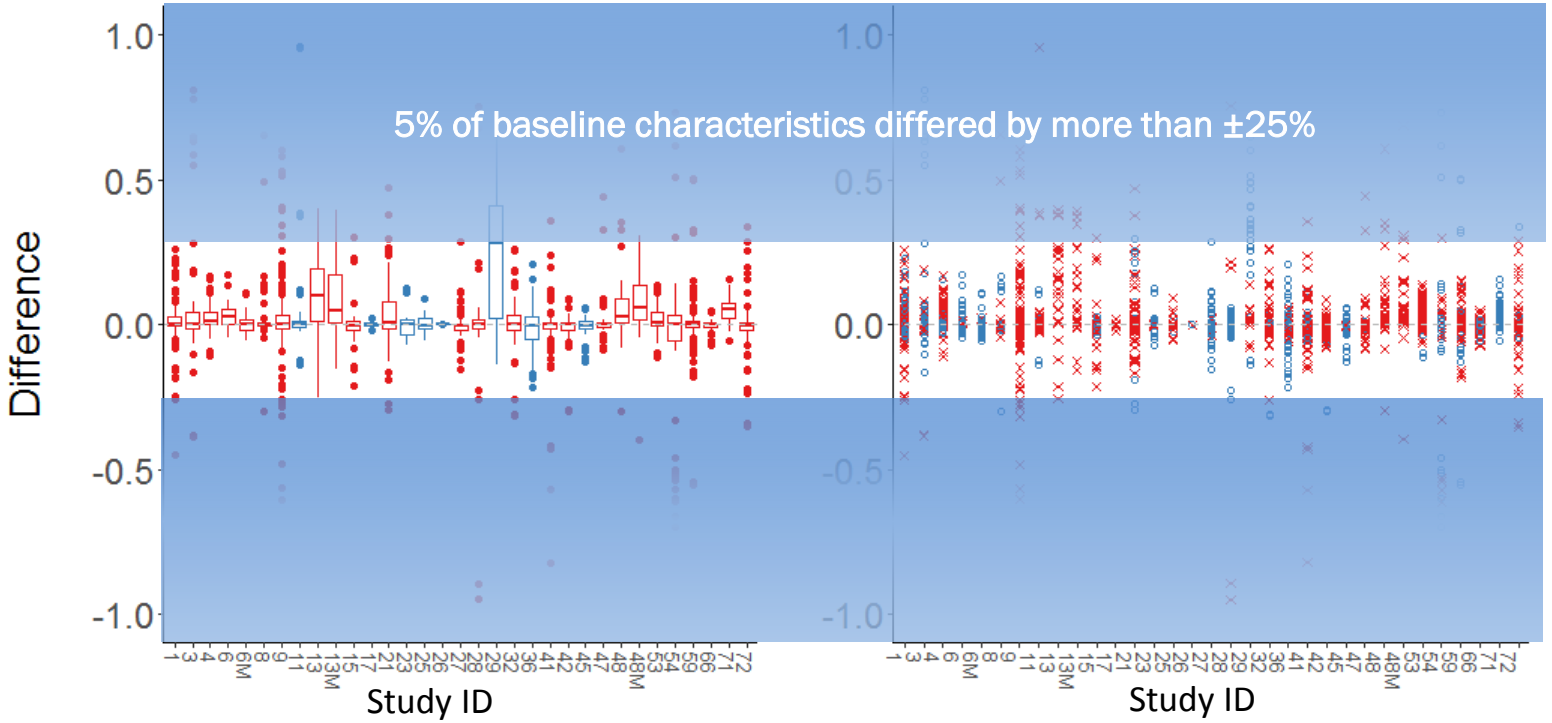


* binary/categorical

INTERIM RESULTS

Difference in baseline characteristics* of cohort

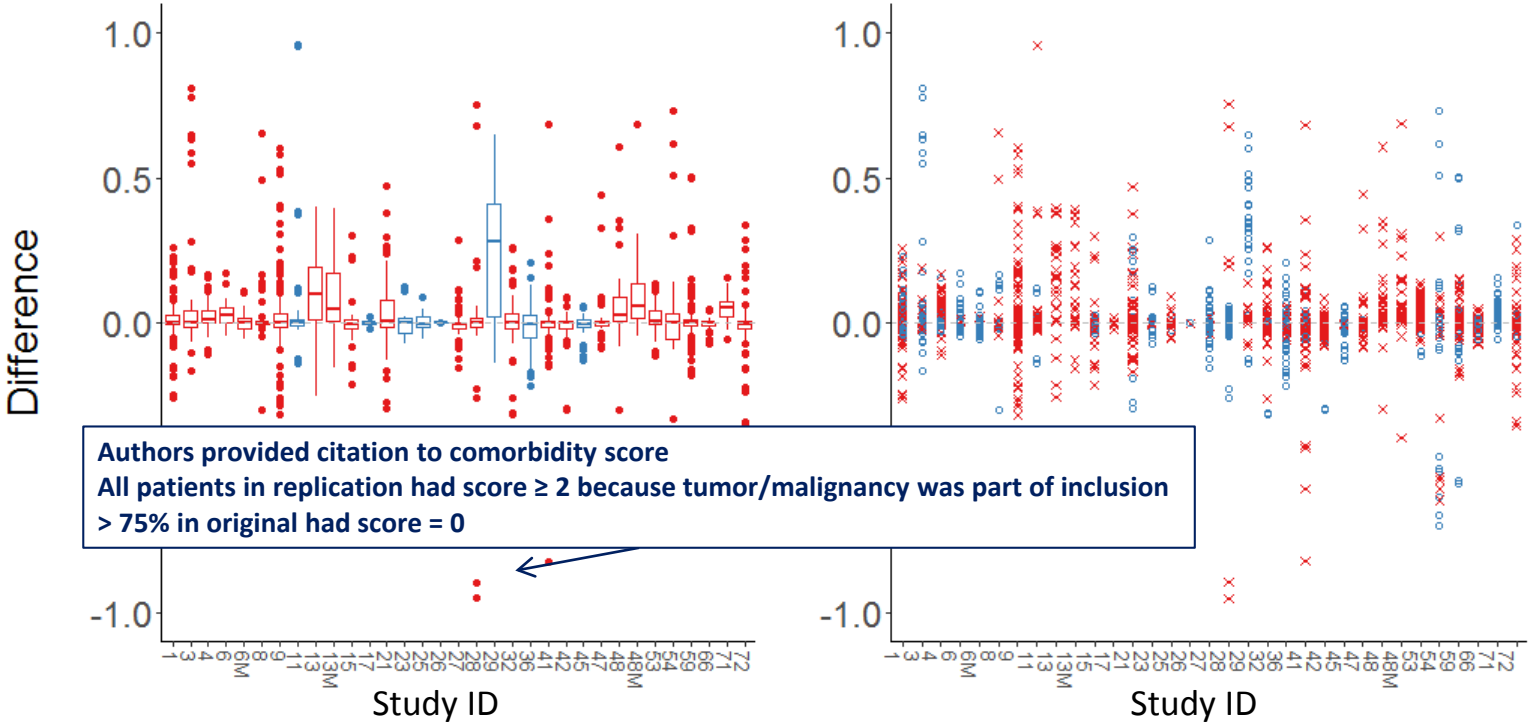
(% original – % replication)



* binary/categorical

INTERIM RESULTS

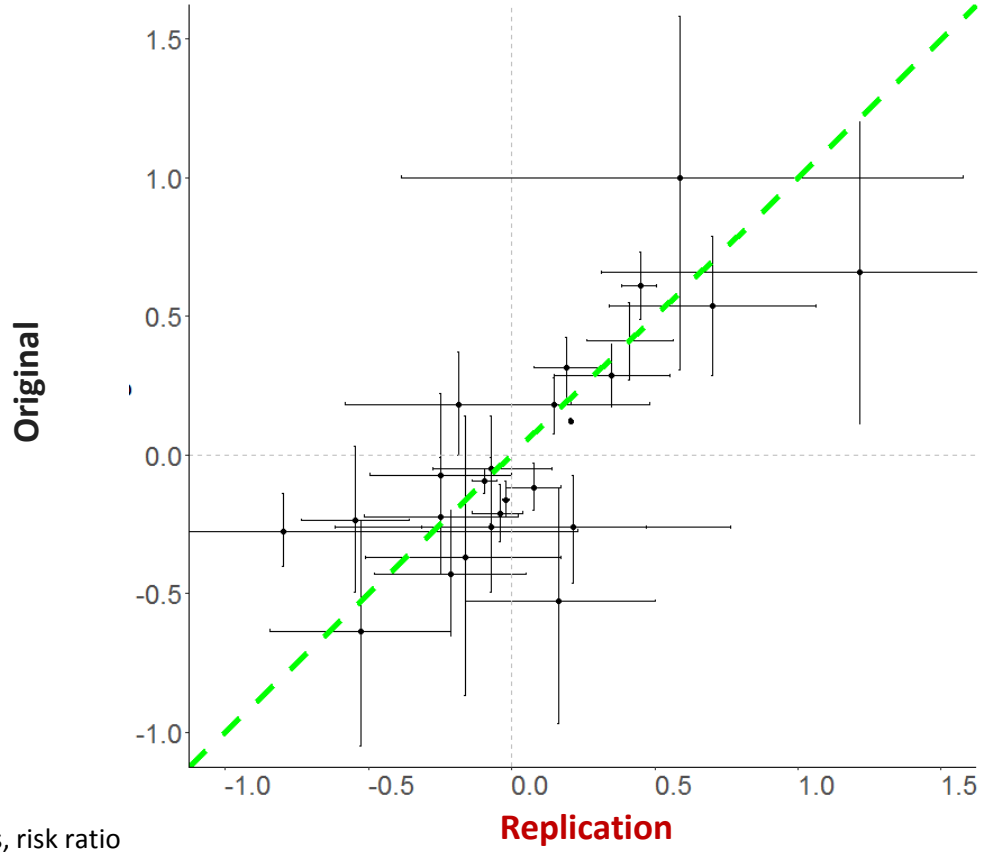
Why did the replication differ so much from the original for some baseline characteristics?



* binary/categorical

INTERIM RESULTS

Calibration of effect estimates* for original versus replication

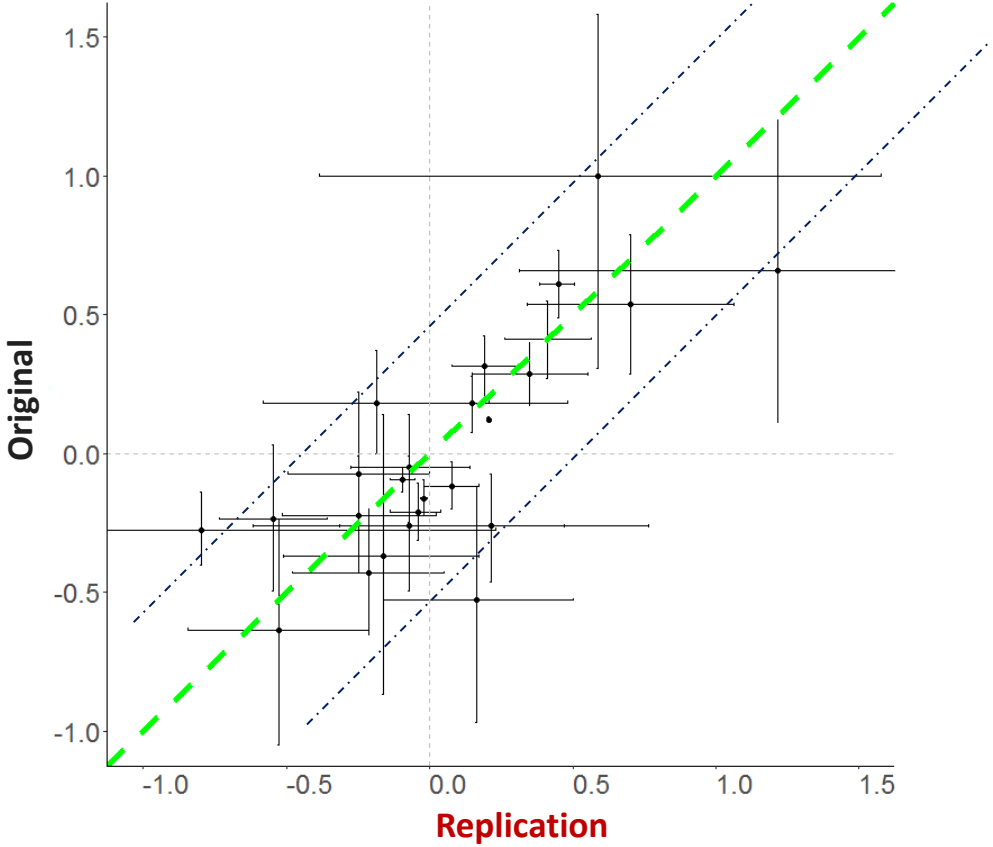


Replication = Original

* Hazard, odds, risk ratio

INTERIM RESULTS

Calibration of effect estimates* for original versus replication

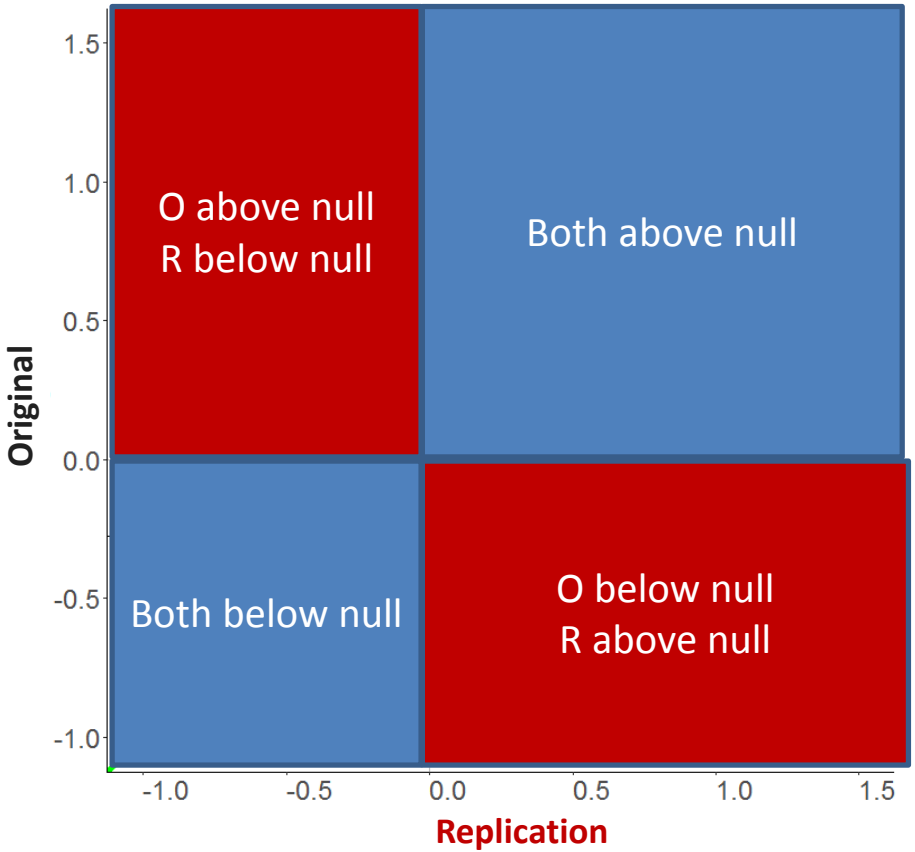


Estimates follow diagonal

* Hazard, odds, risk ratio

INTERIM RESULTS

Effect estimate agreement between original and replication



Same side of null?

84% of effect estimates were on the same side of null
16% were not

52% of effect estimates *and* confidence intervals were on same side of null

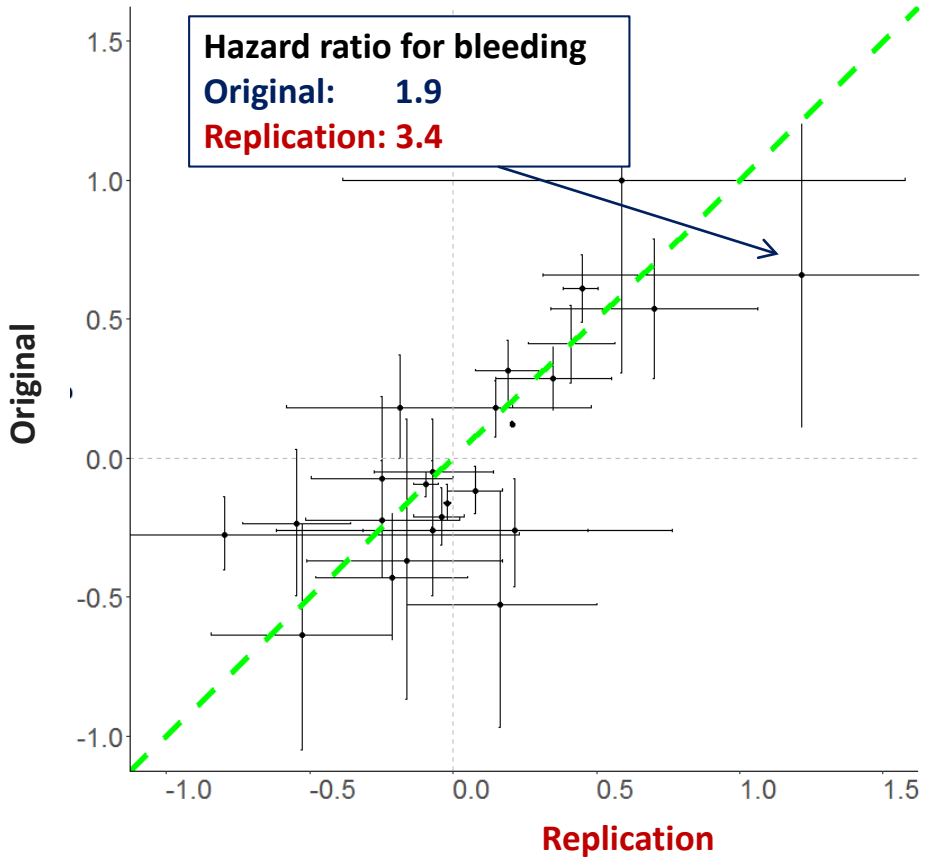
Difference in effect estimate
log(original) – log(replication)

- **Mean: 0.0**
- **29% within ± 0.1**
- **Range: -0.6, 0.4**

* Hazard, odds, risk ratio

INTERIM RESULTS

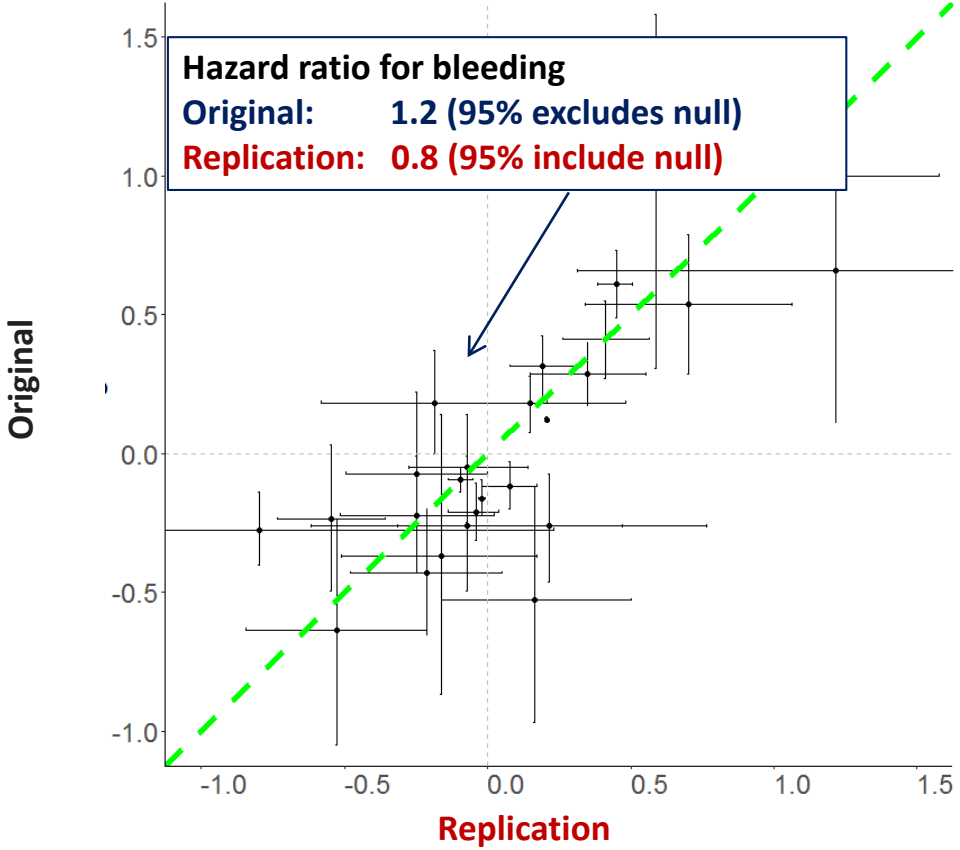
Why is the replication estimate substantially larger?



* Hazard, odds, risk ratio

INTERIM RESULTS

Why are the effect estimates on opposite sides of null?



* Hazard, odds, risk ratio

Work in progress…

- **Empirical evaluation**
  - Describe frequency of reporting, impact of transparency of specific study parameters
  - Prioritize reporting on parameters with demonstrable influence on replicability or robustness
  - **Hard to replicate analysis results if unable to replicate base cohort**
    - Majority of internal debate over vague prose on temporality (slower timeline for replication)
    - Exclusion criteria not detailed, selection of study entry date before or after applying exclusions
    - How much do assumptions on these paramters matter? Context dependent, robustness next…
- **Shared terminology and structured reporting templates**
  - Simplify reporting - terminology used for the same concepts varies
  - Visualization of study design implementation
- **Reporting on research using unstructured data (NLP, machine learning)**

REPEAT Core Team (alphabetical)

7 groups working in parallel on different studies (1+ faculty, 2+ research staff)

- Adrian Ortiz Santiago BS
- Ajinkya Pawar PhD MS
- Elisabetta Patorno MD DrPH
- Elizabeth M. Garry PhD MPH
- Emma Payne BS
- Jessica Franklin PhD
- Joshua Gagne PharmD ScD
- Krista Huybrechts PhD MS
- Kristina Stefanini BA
- Lily Bessette BS
- Mimi Zakarian BS
- Monica L. Gierrada MPH
- Mufaddal Mahresi MD MPH
- Nileesa Gautam BS
- Sebastian Schneeweiss MD ScD
- Shirley V Wang PhD ScM
- Sushama Kattinakere MBBS MSPH
- Yinzhu Jin MS MPH

REPEAT

Reproducible Evidence: Practices to Enhance and Achieve Transparency

www.repeatinitiative.org

Scientific Advisory Board (alphabetical)

Regulators, HTA, delivery systems, patients, payers, industry, journals, research societies…

- Jeffrey Brown PhD
- Alison Bourke MSc FRPharm.S
- Amr Makady PharmD PhD
- Andrew Bate PhD
- Brian Bradbury DSc
- Brian Nosek PhD
- Christine Laine MD MPH FACP
- David Martin MD MPH
- Deborah Zarin MD
- Dick Willke PhD
- Dorothee Bartels MSc PhD
- Elizabeth Loder MD MPH
- Frank de Vries PharmD PhD
- Hans-Georg Eichler MD, MSc
- Henrik Toft Sørensen MD PhD
- Javier Jimenez MD MPH
- Jesper Hallas MD PhD
- Joanne Waldstreicher MD
- John Ioannidis MD DSc
- John Seeger PharmD DrPh
- K. Arnold Chan MD ScD
- Karen Burnett MBA MS
- Kris Kahler PhD
- Laura Happe PharmD MPH
- Liam Smeeth PhD
- Lisa Freeman
- Michael Nguyen MD
- Nam-Kyong Choi B. Pharm PhD
- Páll Jónsson PhD Mres
- Peter Arlett BSc MBBS MRCP FFPM
- Peter Tugwell MSc MD FRCPC
- Richard Platt MD MSc
- Sarah Priddy Alwardt PhD
- Sean Hennessy PharmD, PhD
- Troyen Brennan MD
- Will Shrank MD
- Wolfgang Winkelmayer MD MPH ScD FASN
- Yoshiaki Uyama PhD

Adventures in Replication

The Sentinel System Experience

Ting-Ying Jane Huang, PhD

Research Scientist

Department of Population Medicine

Harvard Medical School and Harvard Pilgrim Health Care Institute

8/26/2018

Disclosure

- Studies used as examples in this presentation were supported by the U.S. Food and Drug Administration (FDA) through the Department of Health and Human Services Contract No. HHSF223201400030I

- This presentation reflects the views of the authors and not necessarily those of the U.S. FDA

- The author has no conflicts of interest to disclose

Sentinel System

Sentinel

sentinelinitiative.org

- A component of the U.S. FDA Sentinel Initiative
- Active safety surveillance system to monitor regulated products
  - **Pre-existing electronic healthcare data from multiple sources**
  - **Routine querying tools (pre-tested, parameterizable *modular programs*)**
- Sentinel Distributed Database
  - 66.9 million members with medical and drug coverage currently accruing new data
  - 14.4 billion pharmacy dispensings
  - 13.3 billion medical encounters

Sentinel Common Data Model v6.0

Administrative

Enrollment	Demographic	Dispensing	Encounter	Diagnosis	Procedure
Person ID	Person ID	Person ID	Person ID	Person ID	Person ID
Enrollment start & end dates	Birth date	Dispensing date	Service date(s)	Service dates	Service date(s)
Drug coverage	Sex	National drug code (NDC)	Encounter ID	Encounter ID	Encounter ID
Medical coverage	Zip code	Days supply	Encounter type and provider	Encounter type and provider	Encounter type & provider
Medical record availability	Etc.	Amount dispended	Facility	Diagnosis code & type	Procedure code & type
			Etc.	Principal discharge diagnosis	Etc.

Clinical

Lab Result	Vital Signs
Person ID	Person ID
Result and specimen collection dates	Measurement date & time
Test type, immediacy & location	Height & weight
Logical Observation Identifiers Names and Codes (LOINC®)	Diastolic & systolic BP
Test result & unit	Tobacco use & type
Etc.	Etc.

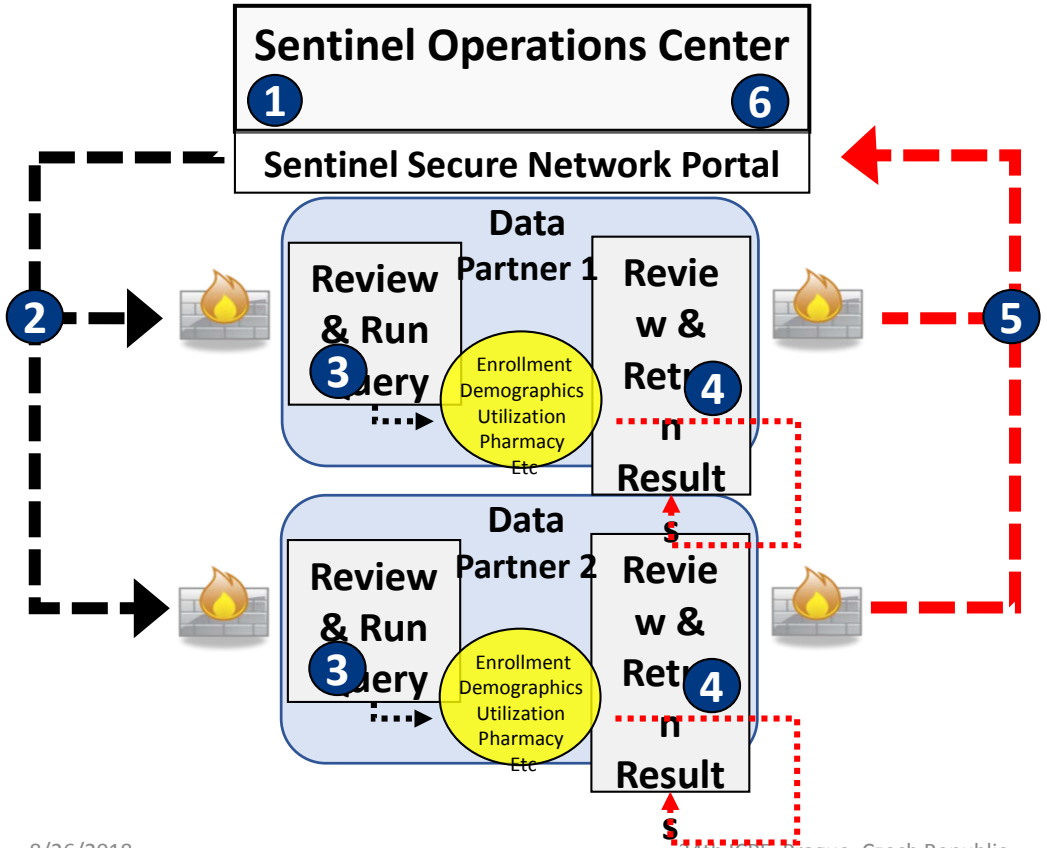
Registry

Death	Cause of Death	State Vaccine
Person ID	Person ID	Person ID
Death date	Cause of death	Vaccination date
Source	Source	Admission type
Confidence	Confidence	Vaccine code & type
Etc.	Etc.	Provider
		Etc.

Inpatient

Inpatient Pharmacy	Inpatient Transfusion
Person ID	Person ID
Administration date & time	Administration start & end date & time
Encounter ID	Encounter ID
National Drug Code (NDC)	Transfusion administration ID
Route	Transfusion product code
Dose	Blood type
Etc.	Etc.

Sentinel Distributed Database



1- User creates and submits query

2- Data Partners retrieve query

3- Data Partners review and run query against their local data

4- Data Partners review results

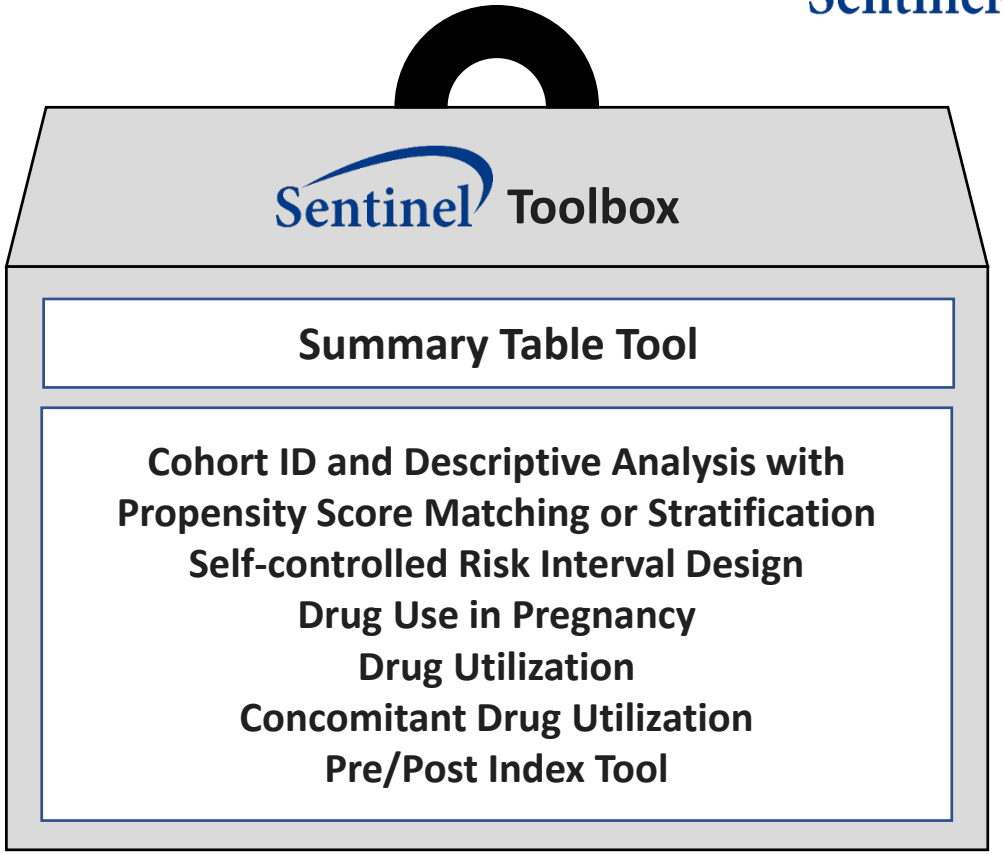
5- Data Partners return results via secure network

6- Results are aggregated and returned

Sentinel Toolbox

Modular Programs

- Pre-tested, parameterizable SAS macros
- Summary Table
- Cohort Identification and Descriptive Analysis
- Propensity Score Analysis



Query Parameterization

Design:

Identify patients ___ with a ___ dispensing of a ___**.** To be eligible, patients must have met the following criteria in the ___ days before the index dispensing: (1) continuous enrollment in ___ benefits, (2) no prescription for ___ or ___, and (3) no diagnosis of ___ in ___ care setting.

The primary outcome of interest is ___ identified with ___ in ___ position during an ___ encounter**.**

Standardized Reporting

Specifications for Request ID cder_mpl2p_wp004

Query Period: January 1, 2001 - September 3
Coverage Requirement: Medical and Drug Coverage
Enrollment Requirement: 183 days
Enrollment Gap: 45 Days
Age Group(s): 18-64 years

	Primary Analysis: Exp
Drug/Exposure	
Incident Exposure/Comparator	All typical antipsychotics
Incident w/ Respect to:	All atypical and typical
Washout	183 days
Cohort Definition	Cohort includes only the firs
Episode Gap	30 days
Episode Extension Period	None
Minimum Episode Duration	1 day
Maximum Episode Duration	None
Minimum Days Supplied	1 day
Episode Truncation at Death	Yes
Episode Truncation for Exposure	All atypical antipsychotics

	Primary Analysis: Exposure/Comparator Pair 1		Sensitivity Analysis 1: Exposure/Comparator Pair 2	
Pre-Existing Condition	Dementia	Dementia	Dementia	Dementia
Include/Exclude	Exclude	Exclude	Exclude	Exclude
Care Settings/PDX	Any	Any	Any	Any
Lookback Period	-183, -1	-183, -1	-183, -1	-183, -1
Event/Outcome				
Event/Outcome	emorrhagic and ischemic	Hemorrhagic and ischemic	Hemorrhagic and ischemic	Hemorrhagic and ischemic
Care Setting/PDX	IPP	IPP	IPP	IPP
Washout	0	0	0	0
Blackout Period	None	None	None	None
Propensity Score Matching				
Covariates	See Covariates tab	See Covariates tab	See Covariates tab	See Covariates tab
Covariate Evaluation Window	-183, -1	-183, -1	-183, -1	-183, -1
Matching Ratio	1:1	1:1	1:1	1:1
Matching Caliper	0.050	0.050	0.050	0.050
Analysis Type	Unconditional	Unconditional	Unconditional	Unconditional
Propensity Score Percentile				
Percentiles	5	5	5	5
Additional Covariates to Adjust for	None	None	None	None
Subgroup Analysis	None	None	None	None
Kaplan Meier Plot	Yes	Yes	No	No

Standardized Reporting

Glossary of Terms for Analyses Using Cohort Identification and Descriptive Analysis (CIDA) Tool*

Amount Supplied - number of units (pills, tablets, vials) dispensed. Net amount per NDC per dispensing. This is equivalent to the "RxAmt" value in the Sentinel Common Data Model.

Blackout Period - number of days at the beginning of a treatment episode that events are to be ignored. If an event occurs during the blackout period, the episode is excluded.

Care Setting - type of medical encounter or facility where the exposure, event, or condition code was recorded. Possible care settings include: Inpatient Hospital Stay (IP), Non-Acute Institutional Stay (IS), Emergency Department (ED), Ambulatory Visit (AV), and Other Ambulatory Visit (OA). For laboratory results, possible care settings include: Emergency department (E), Home (H), Inpatient (I), Outpatient (O), or Unknown or Missing (U). Along with the Principal Diagnosis Indicator, forms the Care Setting/PDX parameter.

Ambulatory Visit (AV) - includes visits at outpatient clinics, same-day surgeries, urgent care visits, and other same-day ambulatory hospital encounters, but excludes emergency department encounters.

Emergency Department (ED) - includes ED encounters that become inpatient stays (in which case inpatient stays would be c

Glossary of Terms for Analyses Using Propensity Score Matching (PSM) Tool*

Covariate Evaluation Window - specified number of days relative to index date to evaluate the occurrence of covariates of interest. Note: members are required to have continuous enrollment during the covariate evaluation window, regardless of the value included in the "Continuous enrollment before exposure" field.

Mahalanobis Distance - provides a measure of balance across all variables while accounting for their correlation.

Matching Caliper - maximum allowed difference in propensity scores between treatment and control patients. Requester may select any caliper (e.g., 0.01, 0.025, and 0.05).

Matching Ratio - patients in exposed and comparator groups are nearest neighbor matched by a 1:1 or 1:n (up to 10) matching ratio.

Matched Conditional and Unconditional Analysis - in a conditional matched analysis, a Cox model, stratified by Data Partner site and matched set, is run on the matched population. This can be done for both the both 1:1 and 1:n matched cohorts. In an unconditional analysis, a Cox model, stratified by Data Partner site only, is run on the matched population. This can be done for the 1:1 matched cohort only.

Propensity Score Stratification - option to stratify propensity scores based on requester-defined percentiles in the unmatched population. In a stratified analysis, a Cox model, stratified by Data Partner site, is run on the stratified population. Note that all patients identified in exposure and comparator cohorts are used in the analysis.

PSM Tool - performs effect estimation by comparing exposure propensity-score matched parallel new user cohorts. Propensity score estimation and matching are conducted within each Sentinel Data Partner site via distributed programming code; data are returned to the SOC, aggregated, and used to calculate effect estimates.

Standardized Reporting

Table 1a. Cohort of New Initiators of Typical Antipsychotics and Atypical Antipsychotics (Unmatched, Aggregated), Ratio = F, Caliper = 0.05

Characteristic	Medical Product: Typical Antipsychotics N/Mean	Typical Antipsychotics %/Std Dev[1]	Atypical Antipsychotics N/Mean	Atypical Antipsychotics %/Std Dev[1]	Covariate Balance: Absolute Difference	Covariate Balance: Standardized Difference
Patients (N)	45,576	100.0%	806,611	100.0%	-	-
Patient Characteristics						
Mean age	44.0	12.6	39.9	12.8	4.121	0.324
Age: 18-64	45,576	100.0%	806,611	100.0%	0	-
Gender (Ambiguous)	-	0.0%	3	0.0%	0	-
Gender (Female)	21,206	46.5%	489,469	60.7%	-14.153	-0.287
Gender (Male)	24,368	53.5%	317,090	39.3%	14.155	0.287
Gender (Unknown)	2	0.0%	49	0.0%	-0.002	-0.002
Recorded						
Prior						
Atria						
Acut						
Coag						
Diab						
Hear						
Hype						

Table 2. Effect Estimates for Typical Antipsychotics and Atypical Antipsychotics and Stroke by Analysis Type

Medical Product	Number of New Users	Person Years at Risk	Average Person Days at Risk	Average Person Years at Risk	Number of Events	Incidence Rate per 1000 Person Years	Risk per 1000 New Users	Incidence Rate Difference per 1000 Person Years	Difference in Risk per 1000 New Users	Hazard Ratio (95% CI)	Wald P-Value
Unmatched Analysis (Site-adjusted only)											
Typical Antipsychotics	45,576	10,125.82	81.15	0.22	25	2.47	0.55	1.30	0.06	1.75 (1.17, 2.63)	0.007
Atypical Antipsychotics	806,611	338,987.22	153.50	0.42	396	1.17	0.49				
1:1 Matched Unconditional Predefined Analysis; Caliper=0.05											
Typical Antipsychotics	45,495	10,113.92	81.20	0.22	25	2.47	0.55	-0.10	-0.62	0.87 (0.54, 1.41)	0.566
Atypical Antipsychotics	45,495	20,636.19	165.67	0.45	53	2.57	1.16				

Sentinel Replications

- Tool verification: function performance and result comparability
- Sentinel protocol-based assessments
  - ACEI and angioedema: Toh 2012 ← Gagne 2016
  - **Dabigatran/warfarin and bleeding: Go 2017 ← Dabigatran variability**
- Known positives
  - Glyburide, Glipizide, and Serious Hypoglycemia: Zhou 2017
  - Clindamycin/Penicillin and Clostridium Difficile: Carnahan, 2018
- Canadian Network for Observational Drug Effect Studies (CNODES) protocol-based assessments
  - **Incretins and acute pancreatitis: Azoulay 2016 ← Incretin replication**
  - **Incretins and heart failure: Filion 2016 ← Incretin replication**

Example 1: Dabigatran Variability

Example 1: Dabigatran Variability

Reproducibility

	Data Source	Methods
Analytic reproduction *Re-running the same code on same data*	**Same**	**Same**
Direct replication *Independent implementation of a specific study*	**Same**	**Same**
Conceptual replication (robustness) *Implementing a study of the same exposure (and comparator), outcome and estimand of interest*	**Different**	**Same**
	Same	**Different**
	Different	**Different**

Example 1: Dabigatran Variability

Annals of Internal Medicine®

LATEST ISSUES CHANNELS CME/MOC IN THE CLINIC JOURNAL CLUB WEB EXCLUSIVES AUTHOR INFO

ADVANCED SEARC

ORIGINAL RESEARCH | 14 NOVEMBER 2017

Outcomes of Dabigatran and Warfarin for Atrial Fibrillation in Contemporary Practice: A Retrospective Cohort Study

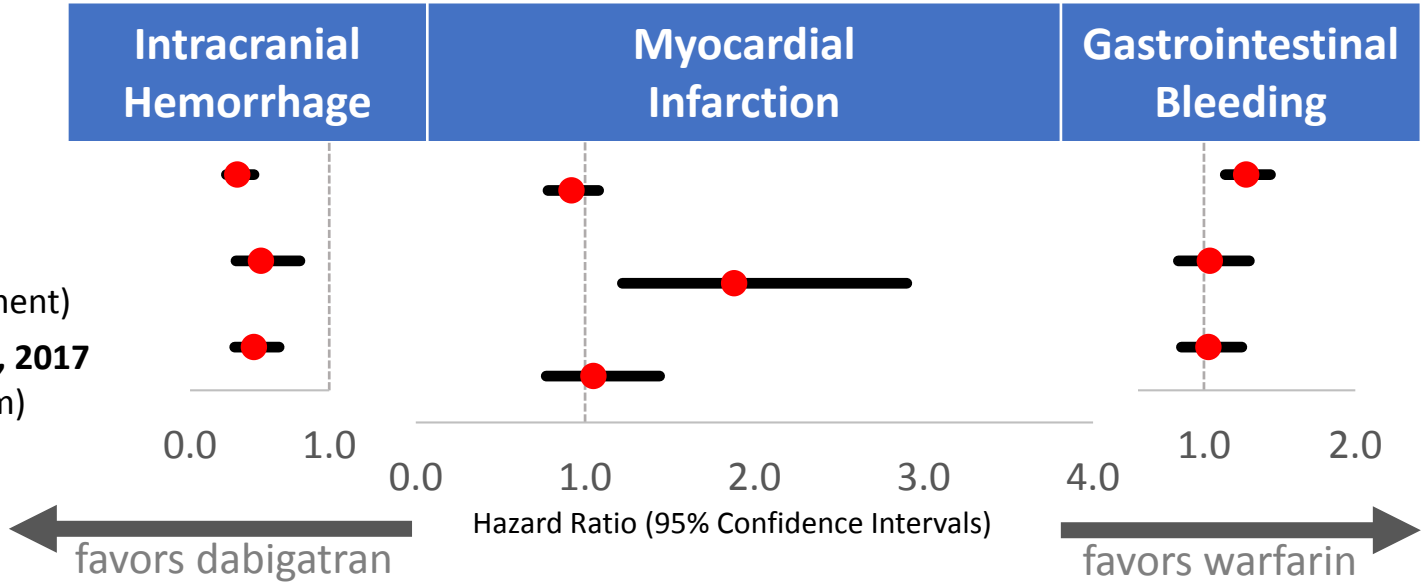
Alan S. Go, MD; Daniel E. Singer, MD; Sengwee Toh, ScD; T. Craig Cheetham, PharmD, MS; Marsha E. Reichman, PhD; David J. Graham, MD, MPH; Mary Ross Southworth, PharmD; Rongmei Zhang, PhD; Rima Izem, PhD; Margie R. Goulding, PhD; Monika Houstoun, PharmD; Katrina Mott, MS; Sue Hee Sung, MPH; Joshua J. Gagne, PharmD, ScD

Article, Author, and Disclosure Information

Example 1: Dabigatran Variability

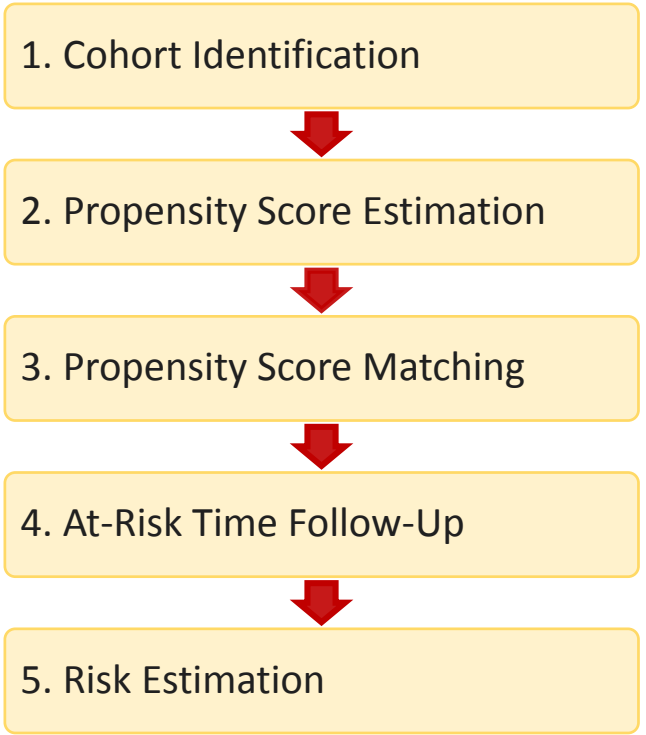
Go, 2017
(Sentinel, protocol assessment)

Sentinel Replication of Go, 2017
(Sentinel, modular program)



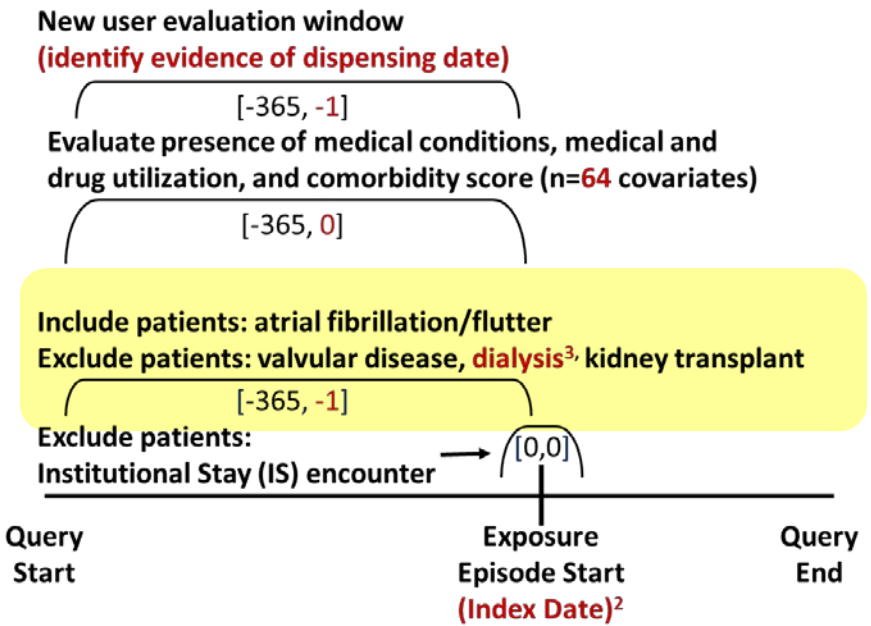
Example 1: Dabigatran Variability

- Strategies to identify discrepancy
- Programming code audit
  - Study parameter differences, including subsequent interpretation and operational implementation, in computer programs
- Anchor cohort creation and output review
  - Isolate divergence in each of five main processing steps
  - 2010-2014 Truven Health MarketScan® Research Databases (formatted to Sentinel Common Data Model)
  - Patients matched in both anchor cohorts

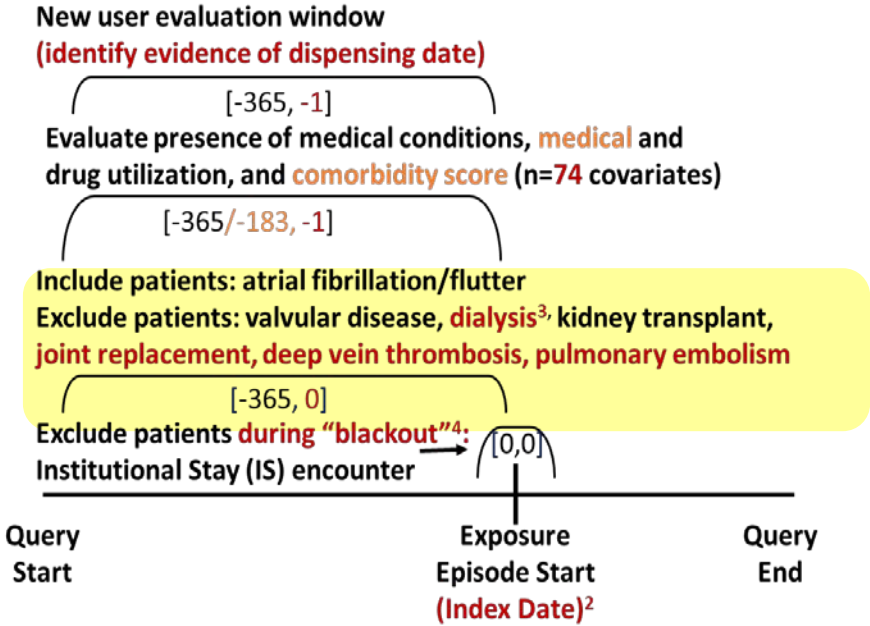


Investigation Results

Go 2017



Modular Program Replication

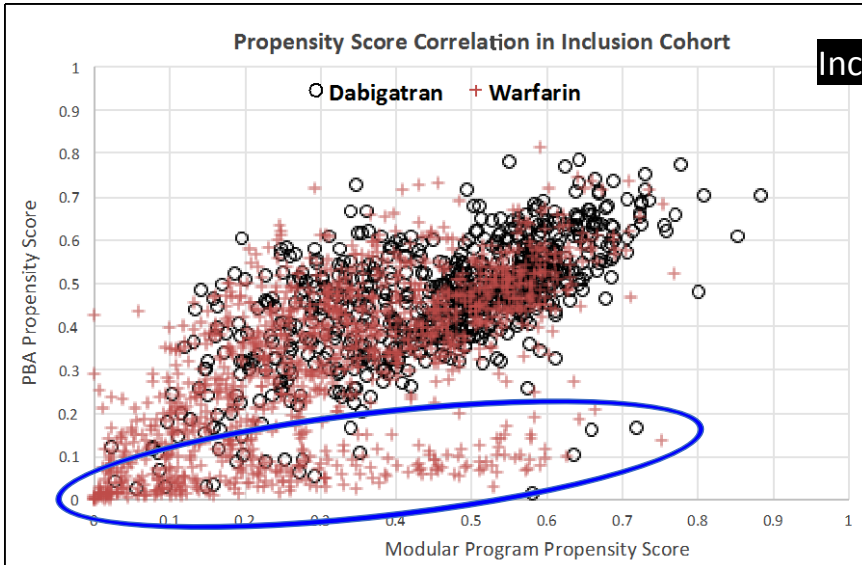


1. Maximum enrollment gap to bridge consecutive enrollment spans: 30 vs 45 days
2. Dispensings with zero days supply or supply amount were vs were not considered to define the index date
3. Dialysis patients defined using CPTs in any vs outpatient-only care setting
4. If patient disenrolls or has outcome on day 0, they are removed from the cohort

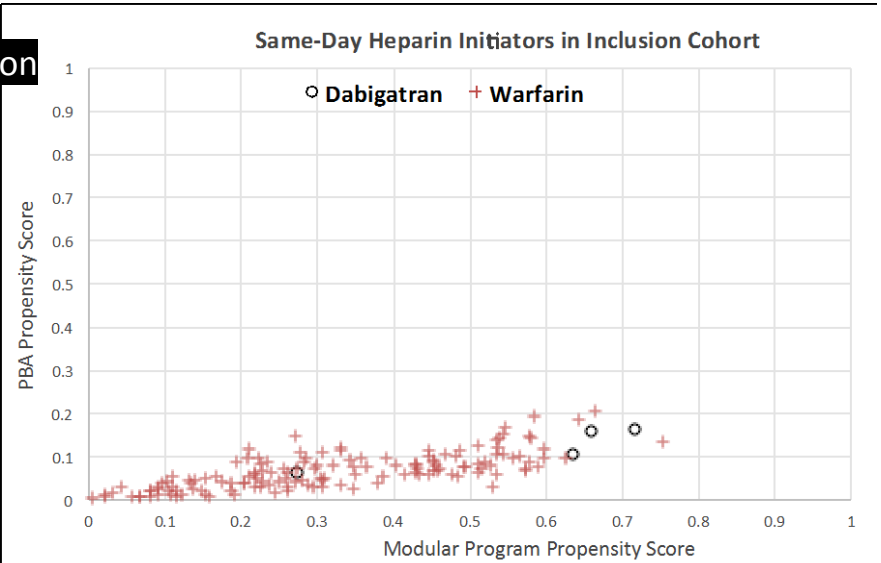
Anchor Cohorts

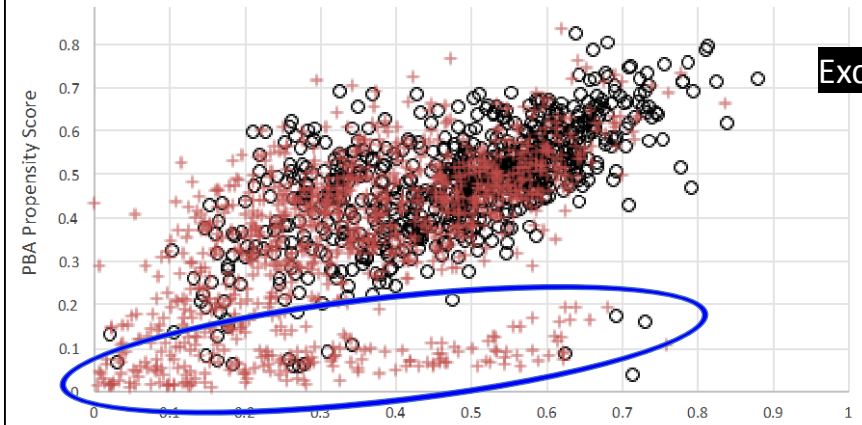
Parameter	Inclusion Cohort	Exclusion Cohort
Enrollment gap bridging	Set gap to 30 days	
New user definition	Use evidence of days supply during washout period	
Valid Dispensing Definition	Exclude dispensings with zero pills or zero days supply from consideration	
Cohort Inclusion Definitions	Inclusion/exclusion lookback period: (-365, -1) Exclude patients censored on Day 0 from the cohort Define dialysis using codes in any care setting	
	Include patients with history of joint replacement, pulmonary embolism and deep vein thrombosis	**Exclude** patients with history of joint replacement, pulmonary embolism, and deep vein thrombosis

Propensity Score Estimation

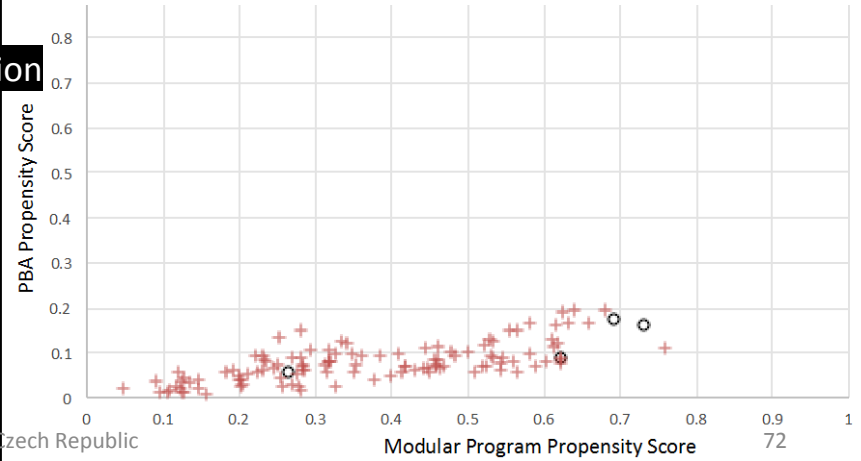


Inclusion

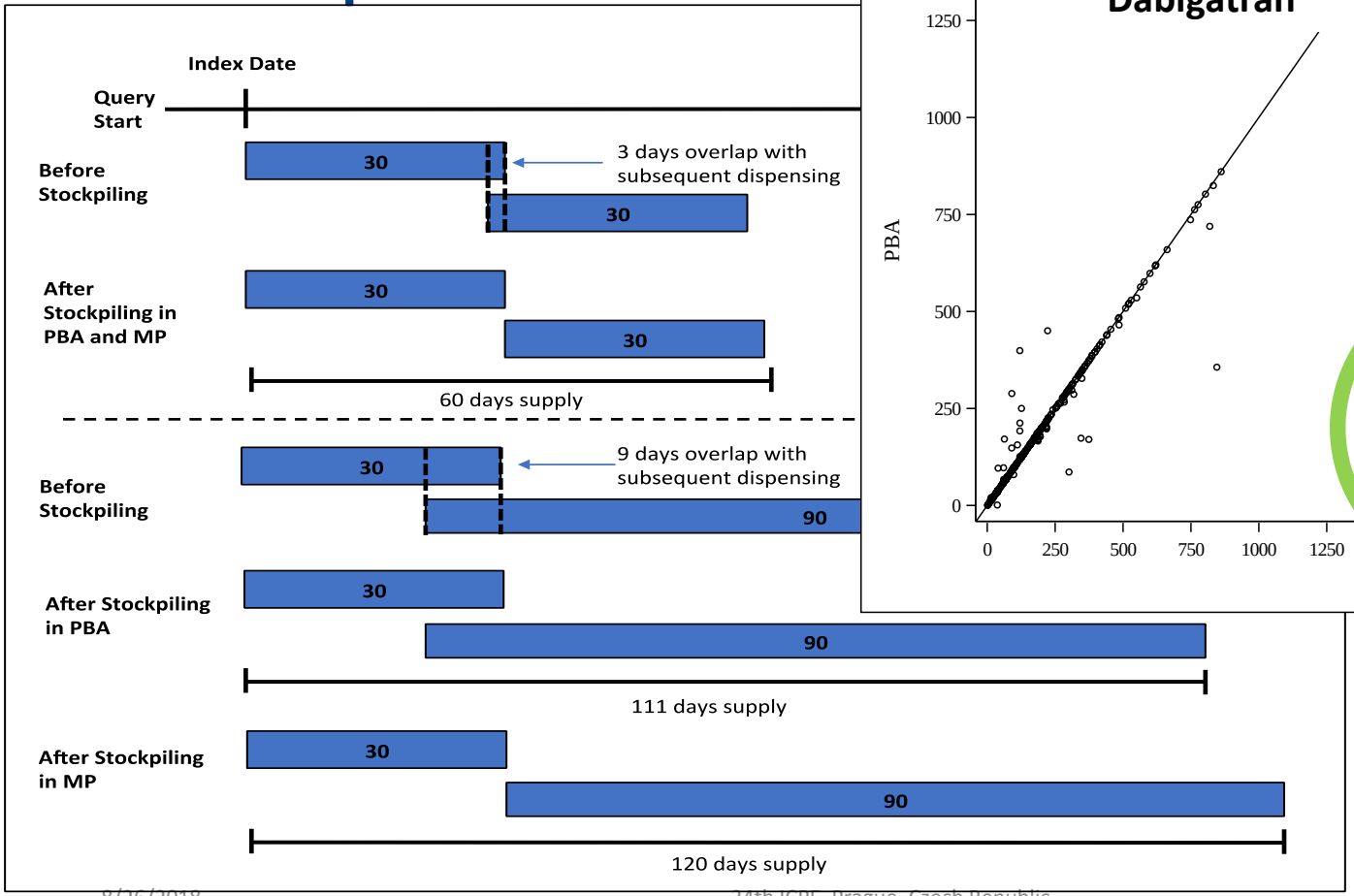


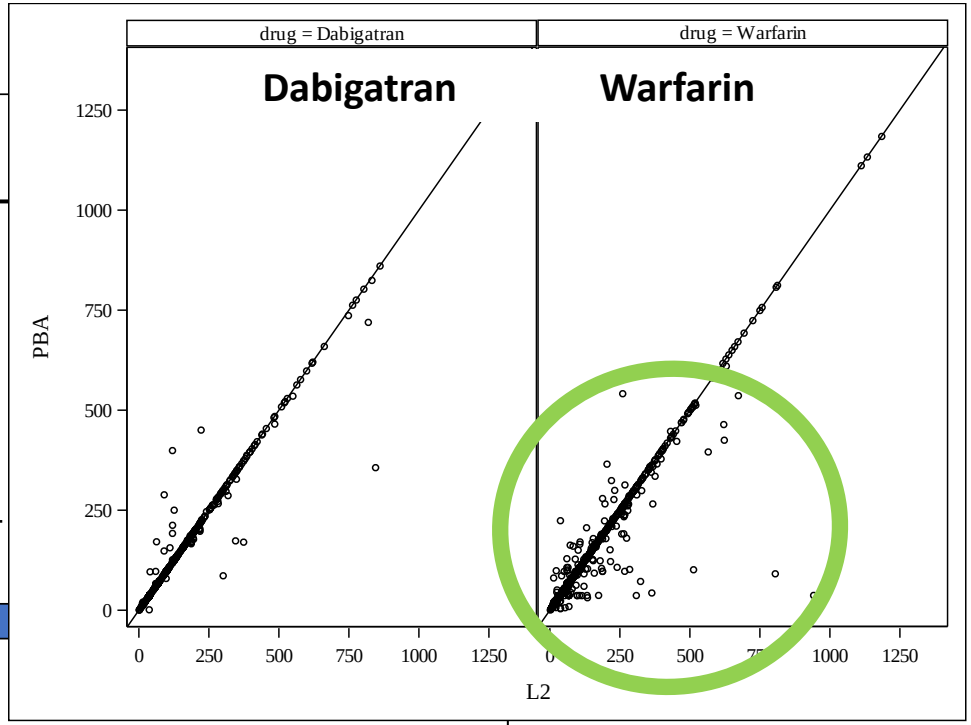


Exclusion



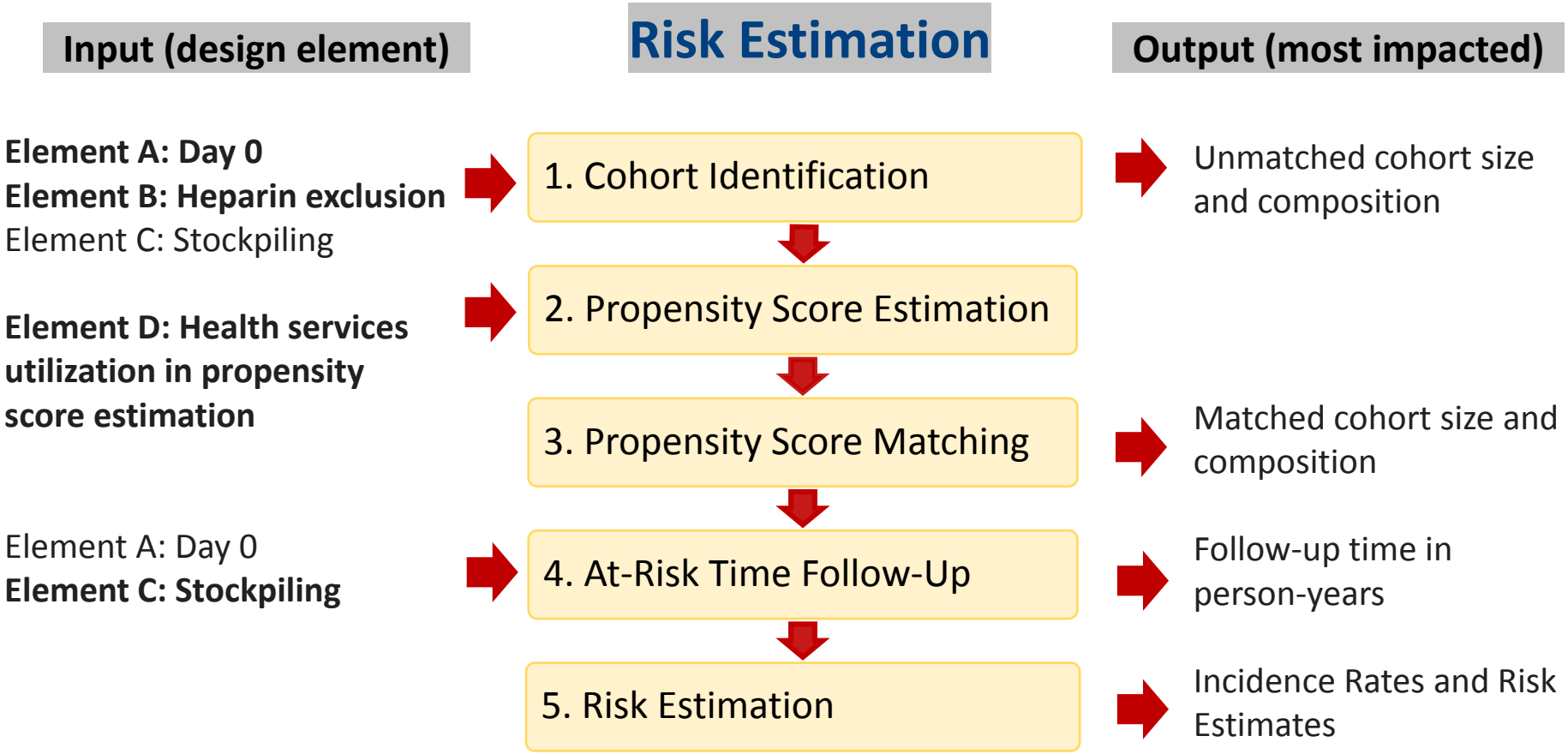
Follow-Up


Index Date
Query Start
Before Stockpiling
30
30
3 days overlap with subsequent dispensing
After Stockpiling in PBA and MP
30
30
60 days supply
Before Stockpiling
30
90
9 days overlap with subsequent dispensing
After Stockpiling in PBA
30
90
111 days supply
After Stockpiling in MP
30
90
120 days supply


drug = Dabigatran
drug = Warfarin
Dabigatran
Warfarin
PBA
L2
0
250
500
750
1000
1250

Example 1: Lessons Learned

- Comprehensive programming review was necessary to identify sources of discrepancy
- Difference in findings may be caused by
  - Intentional scientific decisions: additional exclusion criteria
  - Unintentional, alternate interpretation of operational definitions: Day 0 management
  - Limitation of Sentinel modular programs: stockpiling rules
- Changes in upstream study design decisions (e.g., cohort identification and propensity score estimation) are expected to have downstream influence that could lead to discrepant findings

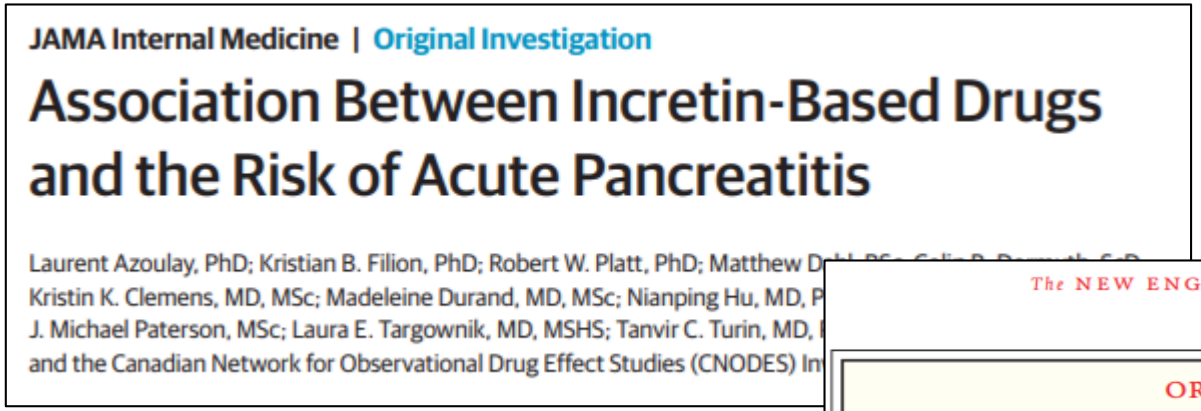

Input (design element)
Risk Estimation
Output (most impacted)
Element A: Day 0
Element B: Heparin exclusion
Element C: Stockpiling
1. Cohort Identification
Unmatched cohort size and composition
Element D: Health services utilization in propensity score estimation
2. Propensity Score Estimation
3. Propensity Score Matching
Matched cohort size and composition
Element A: Day 0
Element C: Stockpiling
4. At-Risk Time Follow-Up
Follow-up time in person-years
5. Risk Estimation
Incidence Rates and Risk Estimates

Example 2: Incretin Replication

Example 2: Incretin Replication

		Data Source	Methods
Reproducibility	Analytic reproduction *Re-running the same code on same data*	**Same**	**Same**
	Direct replication *Independent implementation of a specific study*	**Same**	**Same**
	Conceptual replication (robustness) *Implementing a study of the same exposure (and comparator), outcome and estimand of interest*	**Different**	**Same**
		Same	**Different**
		Different	**Different**

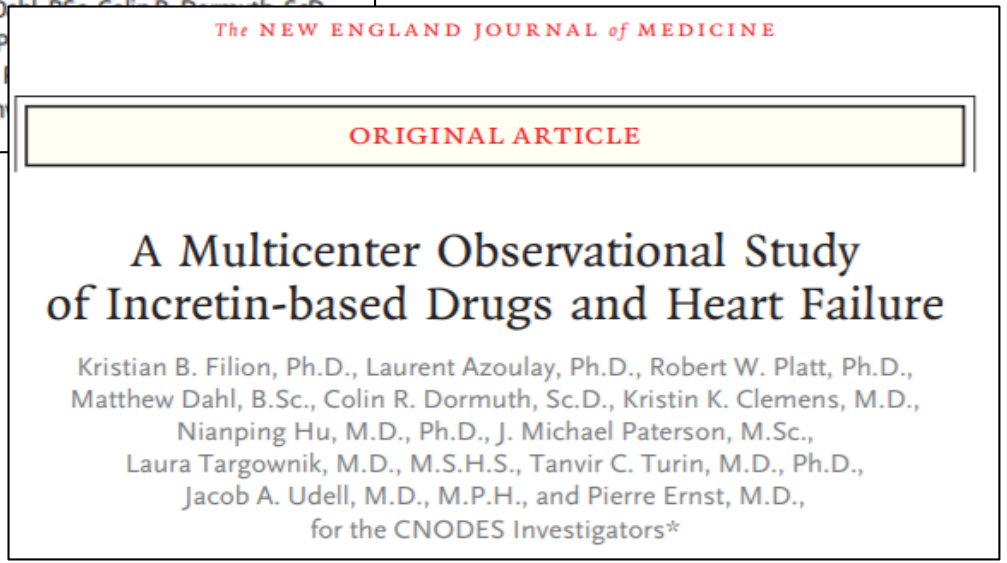
Example 2: Incretin Replication



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 D
Kristin K. Clemens, MD, MSc; Madeleine Durand, MD, MSc; Nianping Hu, MD, P
J. Michael Paterson, MSc; Laura E. Targownik, MD, MSHS; Tanvir C. Turin, MD,
and the Canadian Network for Observational Drug Effect Studies (CNODES) In



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*

Example 2: Incretin Replication

	CNODES	Sentinel
Data source	Administrative claims data from 5 Canadian provinces, UK CPRD, US MarketScan databases	US MarketScan Databases (converted to Sentinel Common Data Model)
Study design	**Nested case-control study**	**Retrospective cohort study**
Data year	2007-2014	2010-2016
Study treatment	Incretin-based drugs	
Comparator treatment	≥2 oral hypoglycemic agents (OHAs)	
Outcome	Hospitalized acute pancreatitis, heart failure	
Exposure assessment	**Concurrent treatment overlapping with outcome occurrence**	**New treatment preceded by a 365-day washout period**
Confounding adjustment	1:n (n≤20) matching and outcome model regression	Propensity score stratification
Outcome model	Conditional logistic regression	Cox proportional hazards model

Example 2: Incretin Replication

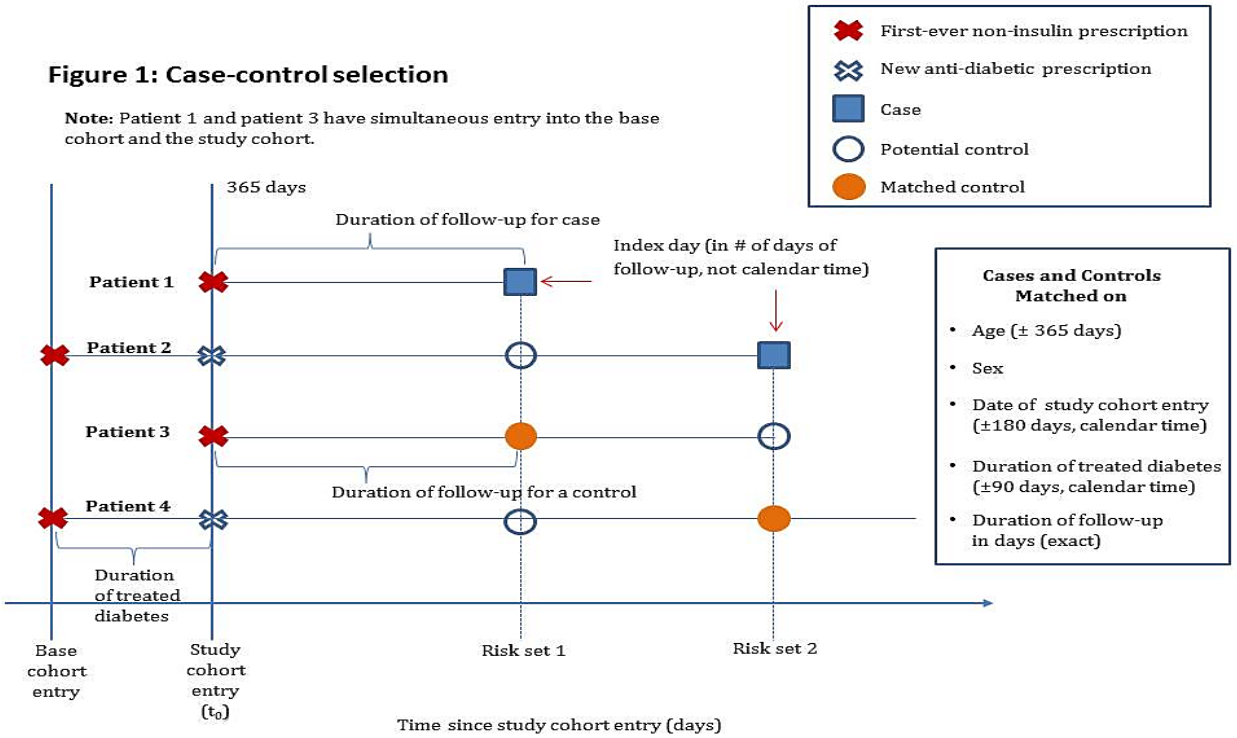
CNODES cohorts

Exposure hierarchy on index (outcome) day

i. **Incretins**
ii. Insulins
iii. **2+ OHAs**
iv. Single OHA
v. Not exposed

Figure 1: Case-control selection

Note: Patient 1 and patient 3 have simultaneous entry into the base cohort and the study cohort.



Example 2: Incretin Replication

Sentinel

Sentinel cohorts

New exposure based on pre-index washout qualification only

- Does not guarantee ongoing concurrent exposure for 2+ OHAs

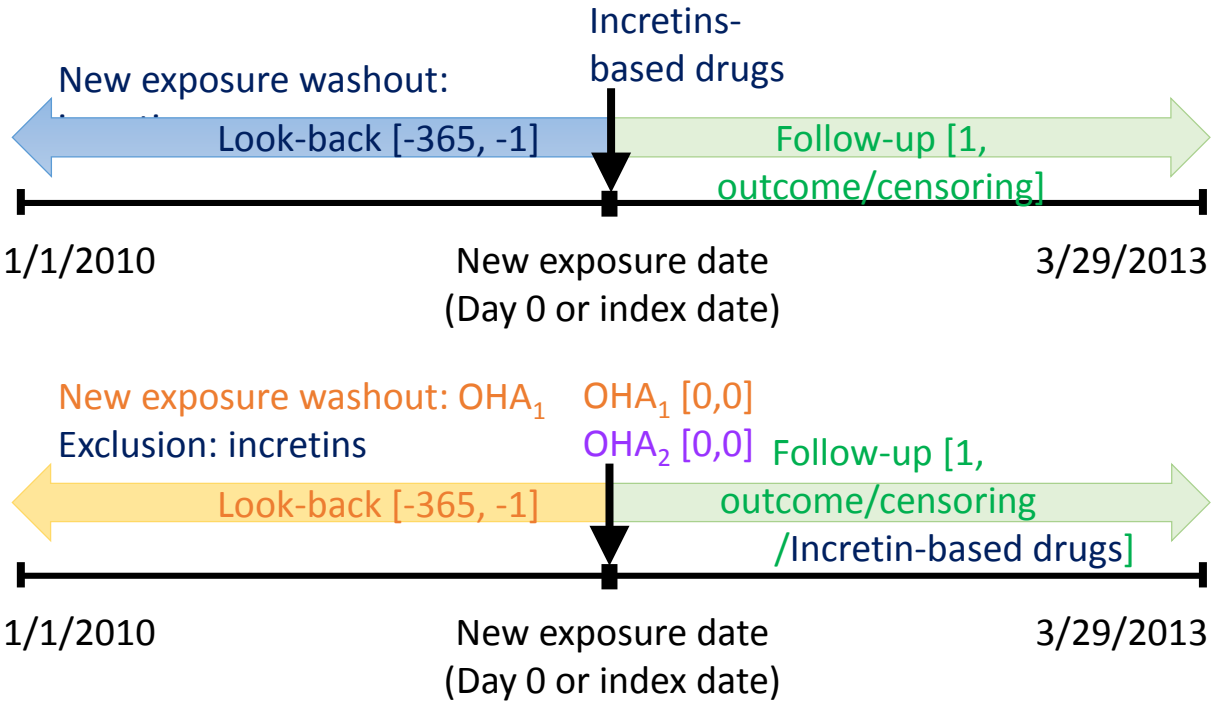
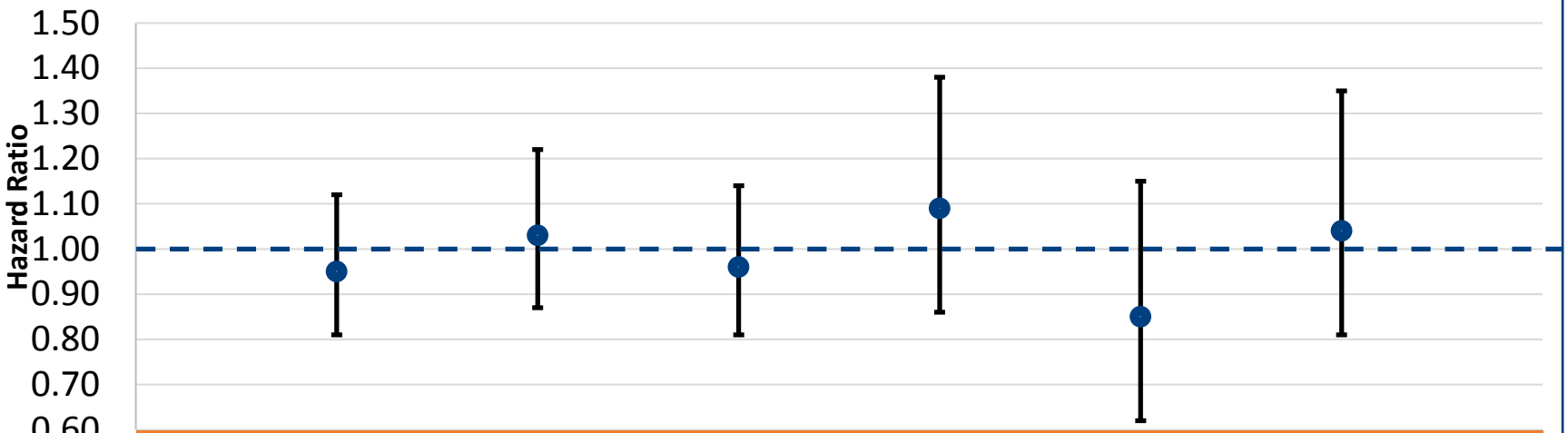
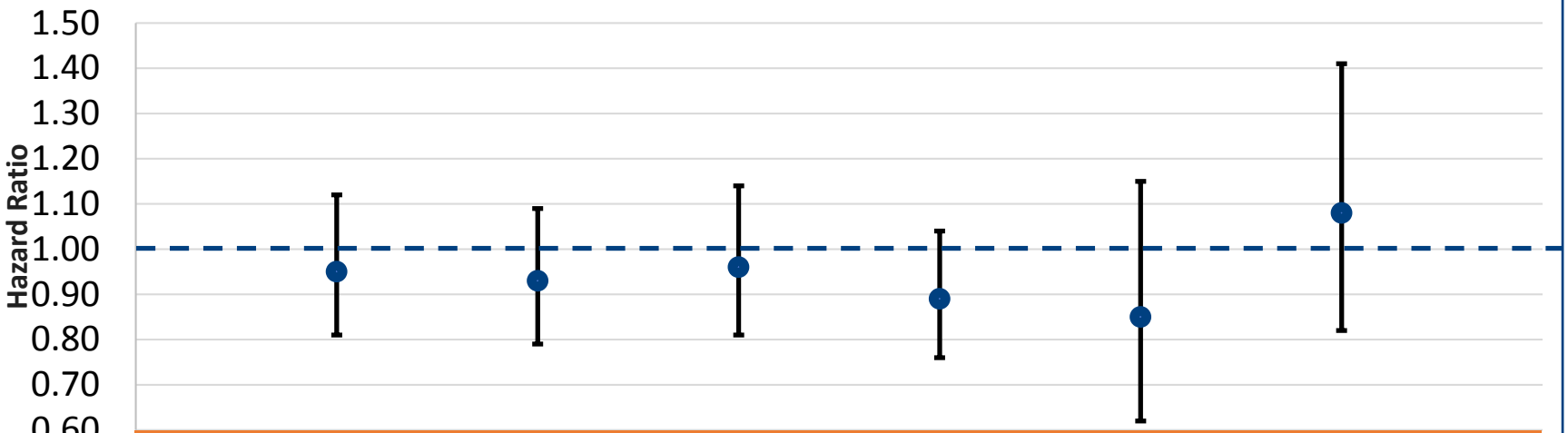


Figure 1. Association Between the Use of Incretin-Based Drugs Compared to Use of 2+ Oral Hypoglycemic Agents and the Risk of Acute Pancreatitis



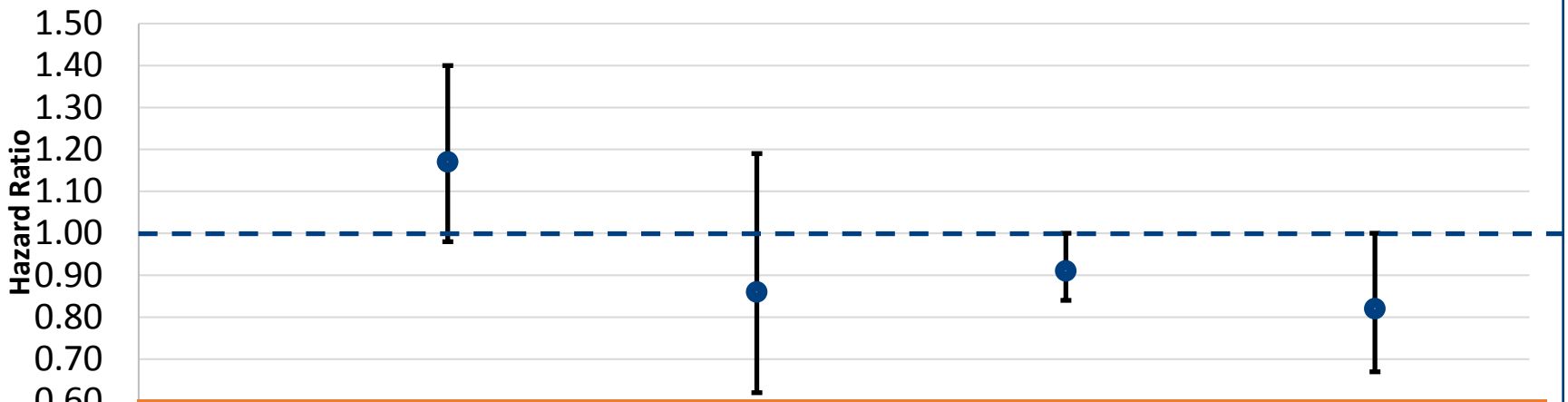
Sentinel Main	CNODES Main	Sentinel DPP-4 Inhibitors	CNODES DPP-4 Inhibitors	Sentinel GLP-1 Analogs	CNODES GLP-1 Analogs
0.95 (0.81, 1.12)	1.03 (0.87, 1.22)	0.96 (0.81, 1.14)	1.09 (0.86, 1.38)	0.85 (0.62, 1.15)	1.04 (0.81, 1.35)

Figure 2. Association Between the Use of Incretin-Based Drugs Compared to Use of 2+ Oral Hypoglycemic Agents and the Risk of Acute Pancreatitis (MarketScan only)



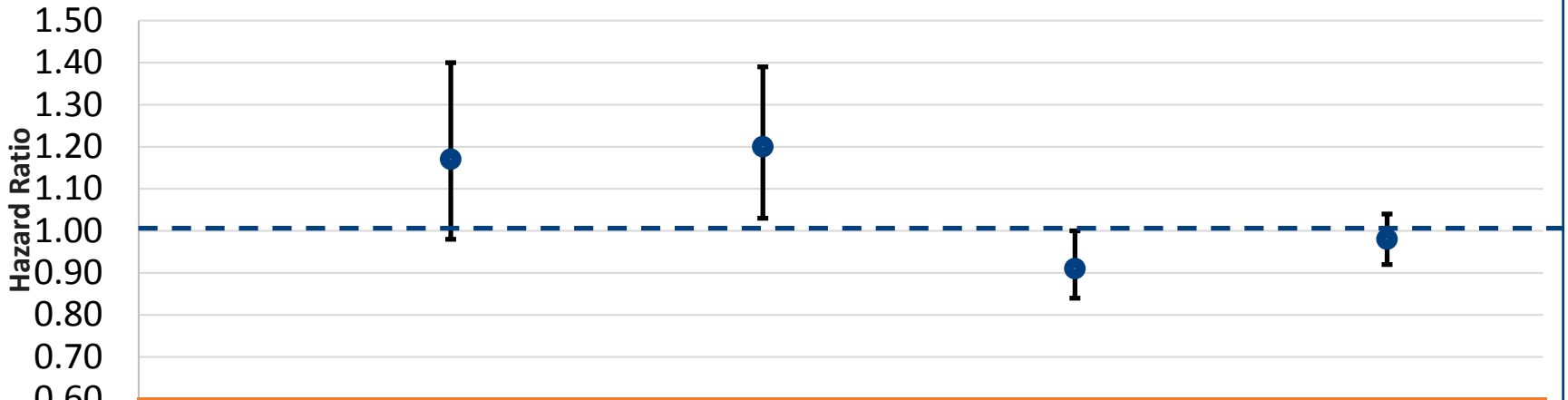
Sentinel Main	CNODES Main	Sentinel DPP-4 Inhibitors	CNODES DPP-4 Inhibitors	Sentinel GLP-1 Analogs	CNODES GLP-1 Analogs
0.95 (0.81, 1.12)	0.93 (0.79, 1.09)	0.96 (0.81, 1.14)	0.89 (0.76, 1.04)	0.85 (0.62, 1.15)	1.08 (0.82, 1.41)

Figure 3. Association Between the Use of Incretin-Based Drugs Compared to Use of 2+ Other Hypoglycemic Agents and the Risk of Heart Failure



Sentinel History CHF	CNODES History CHF	Sentinel No History CHF	CNODES No History CHF
1.17 (0.98, 1.40)	0.86 (0.62, 1.19)	0.91 (0.84, 1.00)	0.82 (0.67, 1.00)

Figure 4. Association Between the Use of Incretin-Based Drugs Compared to Use of 2+ Oral Hypoglycemic Agents and the Risk of Heart Failure (MarketScan only)



Sentinel History CHF	CNODES History CHF	Sentinel No History CHF	CNODES No History CHF
1.17 (0.98, 1.40)	1.20 (1.03, 1.39)	0.91 (0.84, 1.00)	0.98 (0.92, 1.04)

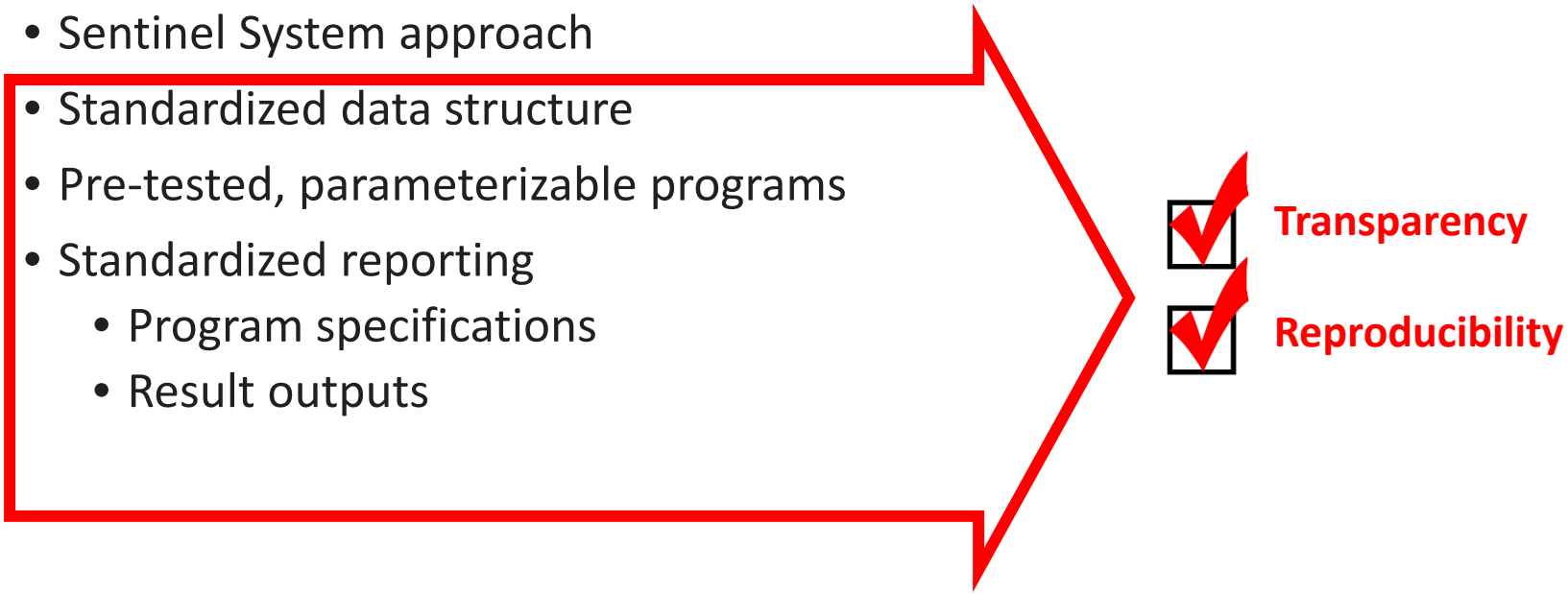
Example 2: Lessons Learned

- Comprehensive programming review was not necessary
  - Thanks to CNODES readily-available protocols
  - Yet, author contact is needed to access study design and operational details
- Tool limitations in customizability
  - Fixed study design options
  - Limited capability to identify and characterize drug utilization following complex therapeutic regimen
- Comparable findings under altered but robust design
- Fast analysis turn-around time
- Ready to replicate in Sentinel Distributed Database

Summary and Discussion

Summary and Discussion

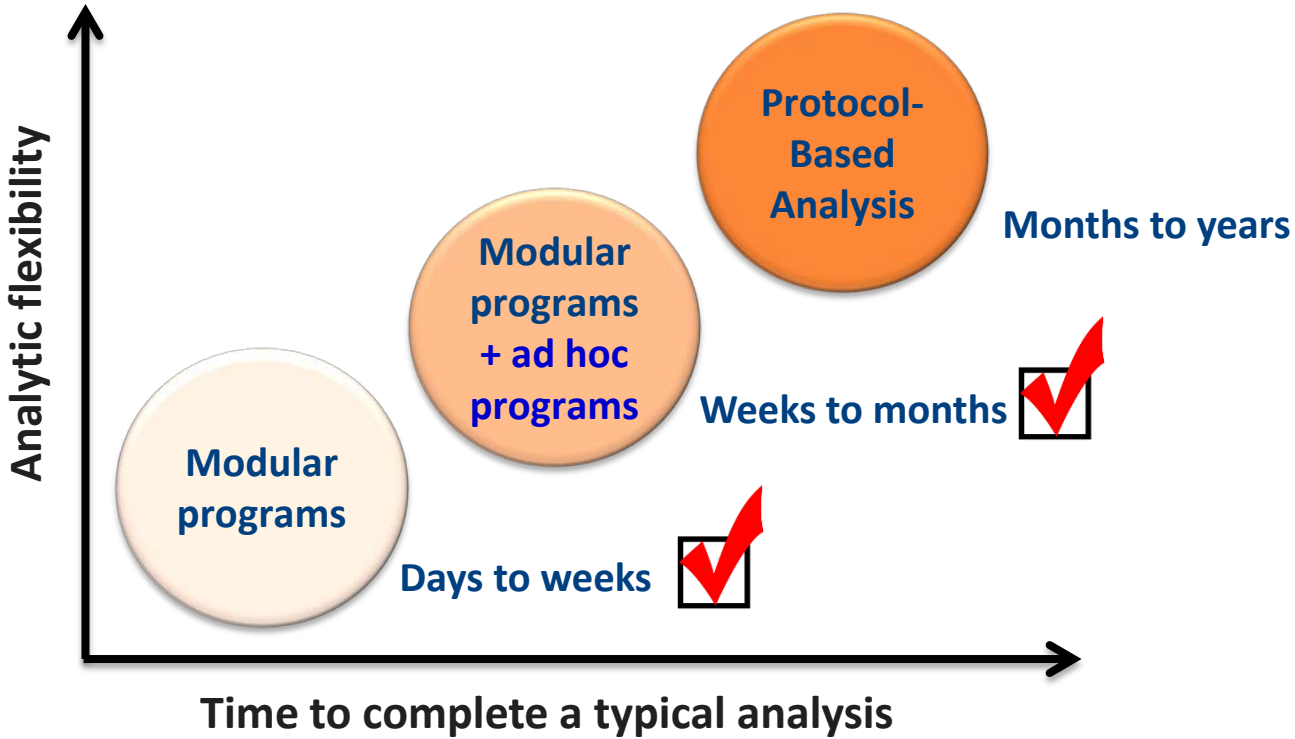
- Common data model plus modular programs
- Sentinel System approach
- Standardized data structure
- Pre-tested, parameterizable programs
- Standardized reporting
  - Program specifications
  - Result outputs



Transparency

Reproducibility

Safety Assessment in Sentinel



Summary and Discussion

- However, standardization also implies…
- Moderate flexibility: methods, measures, and reporting
  - Ad hoc programs offer some degree of customization
- Upfront investment to build and maintain the system
  - Deploy a common data model
  - Routine data refresh and quality assurance
  - Continuous enhancement of the modular programs

References

1. Sentinel System and Routine Querying Tools: https://www.sentinelinitiative.org
2. Toh, S., Reichman, M.E., Houstoun, M., Southworth, M.R., Ding, X., Hernandez, A.F., Levenson, M., Li, L., McCloskey, C., Shoaibi, A. and Wu, E., 2012. Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. Archives of internal medicine, 172(20), pp.1582-1589.Go, A.S., Singer, D.E., Toh, S., Cheetham, T.C., Reichman, M.E., Graham, D.J., Southworth, M.R., Zhang, R., Izem, R., Goulding, M.R. and Houstoun, M., 2017. Outcomes of dabigatran and warfarin for atrial fibrillation in contemporary practice: a retrospective cohort study. Annals of internal medicine, 167(12), pp.845-854.
3. Zhou, M., Wang, S.V., Leonard, C.E., Gagne, J.J., Fuller, C., Hampp, C., Archdeacon, P., Toh, S., Iyer, A., Woodworth, T.S. and Cavagnaro, E., 2017. Sentinel Modular Program for Propensity Score–Matched Cohort Analyses. Epidemiology, 28(6), pp.838-846.
4. Carnahan, R.M., Kuntz, J.L., Wang, S.V., Fuller, C., Gagne, J.J., Leonard, C.E., Hennessy, S., Meyer, T., Archdeacon, P., Chen, C.Y. and Panozzo, C.A., 2018. Evaluation of the US Food and Drug Administration sentinel analysis tools in confirming previously observed drug-outcome associations: The case of clindamycin and Clostridium difficile infection. Pharmacoepidemiology and drug safety.
5. Azoulay L, Filion KB, Platt RW, Dahl M, Dormuth CR, Clemens KK, Durand M, Hu N, Juurlink DN, Paterson JM, Targownik LE, Turin TC, Ernst P; and the Canadian Network for Observational Drug Effect Studies (CNODES) Investigators., Suissa S, Dormuth CR, Hemmelgarn BR, Teare GF, Caetano P, Chateau D, Henry DA, Paterson JM, LeLorier J, Levy AR, Ernst P, Platt RW, Sketris IS. Association between Incretin-Based Drugs and the Risk of Acute Pancreatitis. JAMA Intern Med. 2016 Oct 1;176(10):1464-1473.
6. Filion KB, Azoulay L, Platt RW, Dahl M, Dormuth CR, Clemens KK, Hu N, Paterson JM, Targownik L, Turin TC, Udell JA, Ernst P; CNODES Investigators.. A Multicenter Observational Study of Incretin-based Drugs and Heart Failure. N Engl J Med. 2016 Mar 24;374(12):1145-54.

CANADIAN NETWORK FOR OBSERVATIONAL
DRUG EFFECT STUDIES (CNODES)

Replication in CNODES

Robert Platt

August 26, 2018
34th ICPE, Prague

Disclosures

- **Albert Boehringer I Chair in Pharmacoepidemiology**
- **Canadian Institutes of Health Research (CIHR) grants**
- **No conflicts relevant to this presentation to disclose**

CNODES funding and investigators

Canadian Network for Observational Drug Effect Studies (CNODES), a collaborating center of the Drug Safety and Effectiveness Network (DSEN), is funded by the Canadian Institutes of Health Research (CIHR, Grant #DSE – 146021).

CNODES INVESTIGATORS	
Executive:	Samy Suissa (NPI*), Robert Platt
British Columbia:	Colin Dormuth
Alberta:	Brenda Hemmelgarn
Saskatchewan:	Gary Teare
Manitoba:	Patricia Caetano, Dan Chateau
Ontario:	David Henry, Michael Paterson
Québec:	Jacques LeLorier
Atlantic (NB, NL, NS, PEI):	Adrian Levy, Ingrid Sketris
UK CPRD:	Pierre Ernst, Kristian Filion

*Nominated Principal Investigator

CNODES at a glance

The Canadian Network for Observational Drug Effect Studies (CNODES) uses **population-based administrative healthcare data** to provide **timely responses** to queries for Canadian public stakeholders regarding drug safety and effectiveness

Reproducibility

Reproducibility

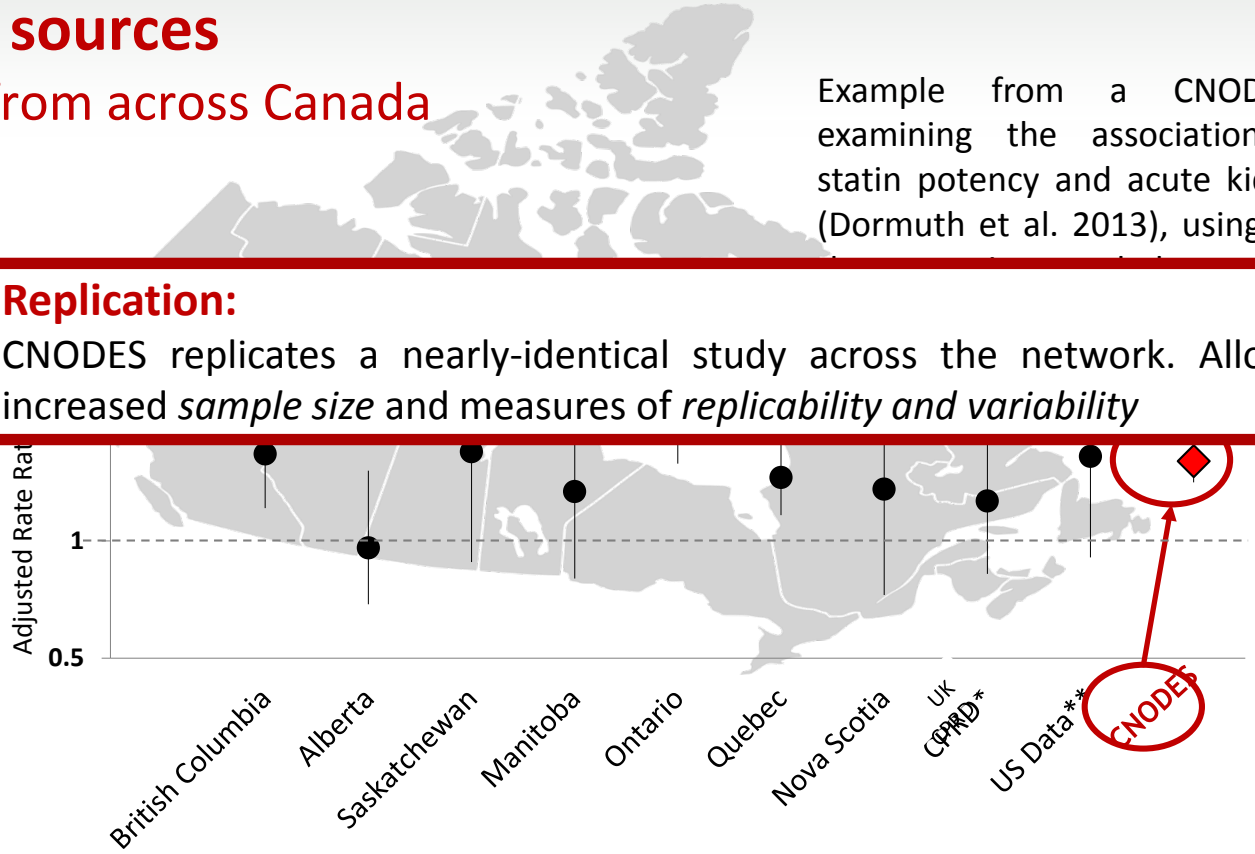
	Data Source	Methods
Analytic reproduction *Re-running the same code on same data*	**Same**	**Same**
Direct replication *Independent implementation of a specific study*	**Same**	**Same**
Conceptual replication (robustness) *Implementing a study of the same exposure (and comparator), outcome and estimand of interest*	**Different**	**Same**
	Same	**Different**
	Different	**Different**

Data sources

Data from across Canada

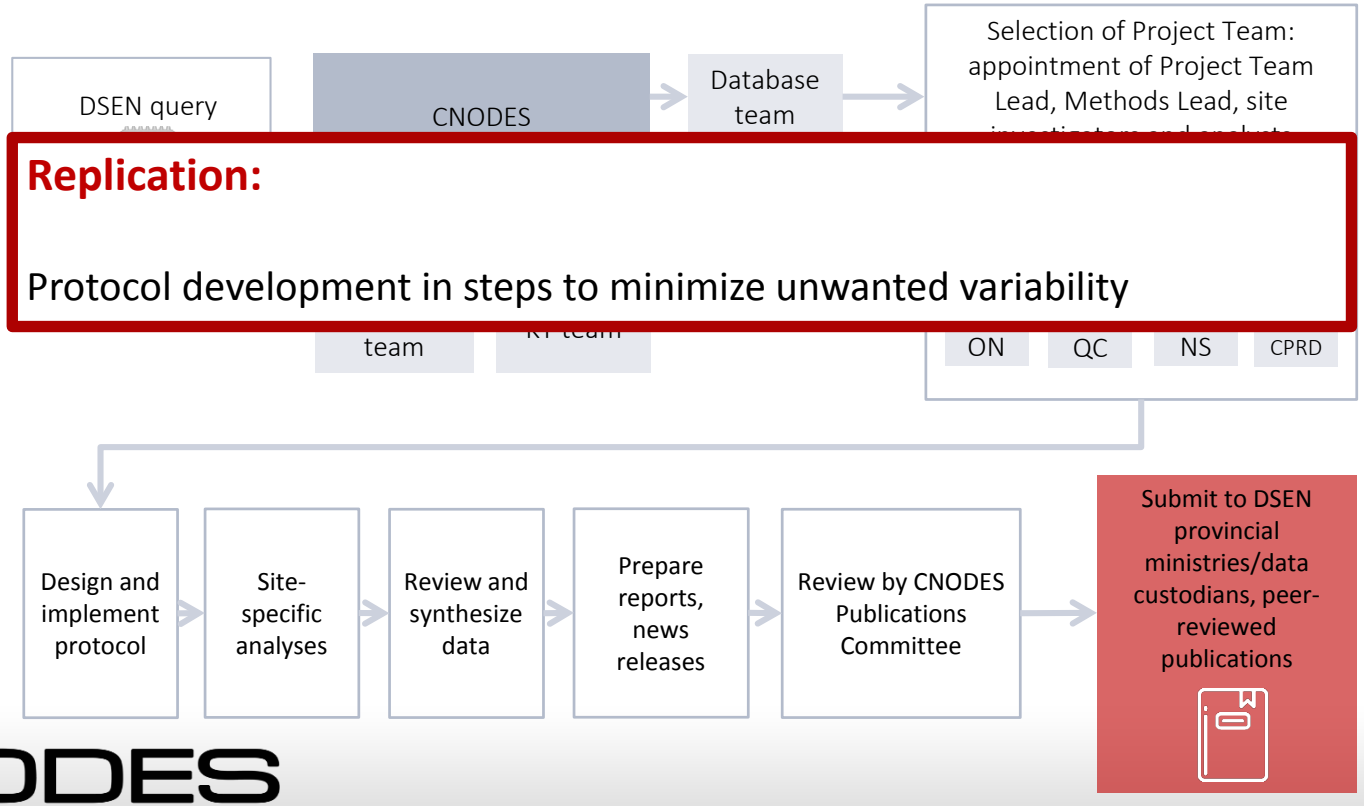
Example from a CNODES study examining the association between statin potency and acute kidney injury (Dormuth et al. 2013), using data from two

Replication:
CNODES replicates a nearly-identical study across the network. Allows with increased *sample size* and measures of *replicability and variability*



The CNODES process

From query submission to project completion and knowledge translation



Replication:

Protocol development in steps to minimize unwanted variability

Why does CNODES need Replication?

- Sample size
  - Canada: ~30 million covered lives
  - But provinces keep their own data
  - Replication/meta-analysis allows increased power
- Robustness
  - Natural (sampling) variability
  - Variability in
    - Population served
    - Formulary
  - Province-specific estimates

Reproducible Research

- To what extent do differences between studies reflect:
  - **Population**
    - **Characteristics**
    - **Prescription/formulary patterns**
  - Question
  - Data source/structure
  - Analysis plan
  - Analyst/code
  - Interpretation?

Ensuring Replicability in CNODES

1. Protocol approved by every site plus methods and content experts
   - Feasibility, scientific content, power
2. **Statistical analysis plan**
   - Very detailed protocol for distribution across sites
3. Iterative analytic process
   - Ensure reproducibility but avoid contamination
4. Meta-analysis and outlier checking

The CNODES Statistical Analysis Plan

- Step-by-step guide that sets out how each CNODES site will design their study and analyze their data.
- Written after the scientific protocol has been developed, in consultation with site liaisons and analysts
- Created in phases:
  - Cohort construction; definitions of exposures, outcomes, and measures of confounding; descriptive statistics
  - Primary analyses
  - Sensitivity analyses in addition to the primary statistical analyses
- Guiding principles:
  - Given this protocol, any two analysts should produce the same results from the same dataset.
  - Minimize unwanted heterogeneity

Reproducible Research

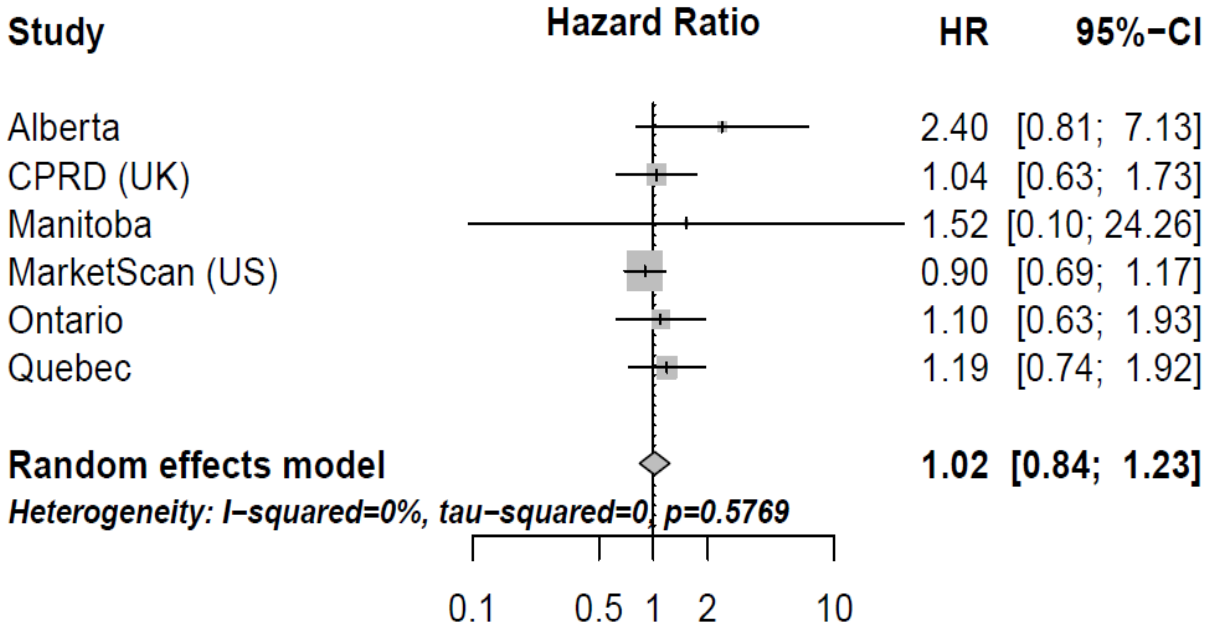
- To what extent do differences between studies reflect:
  - Population
    - Characteristics
    - Prescription/formulary patterns
  - Question
  - Data source/structure
  - Analysis plan
  - **Analyst/code?**
  - Interpretation?

Variability Across Analysts

- Some code-sharing
- Sharing of generic (fake) data
  - Conduct basic analyses to ensure replication
  - Analytic reproduction
- Frequent contact between analysts
  - Discussion groups
  - Structured review
  - Blinded checks

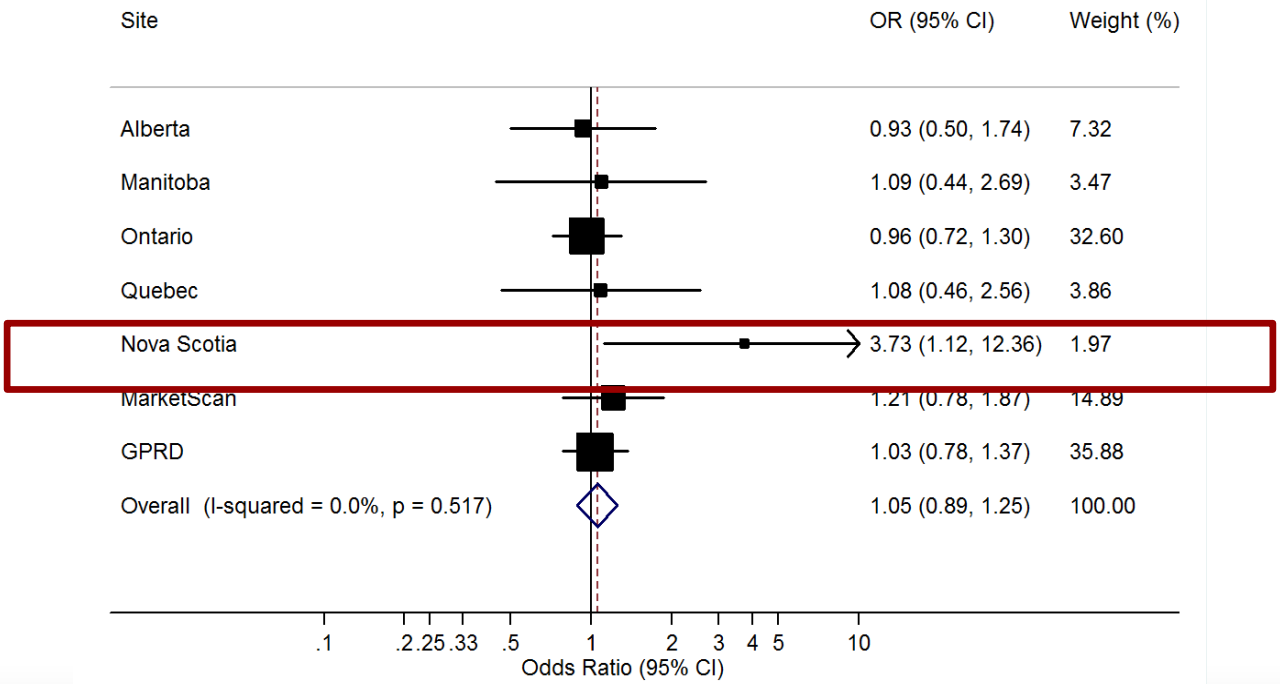
Examples

Pancreatic cancer
(Incretin-based drugs vs sulfonylureas)



Exposure: Ever use with 1 year lag

PPIs and HCAP



CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

Filion KB. Gut 2014.

Reproducibility

Reproducibility

	Data Source	Methods
Analytic reproduction *Re-running the same code on same data*	**Same**	**Same**
Direct replication *Independent implementation of a specific study*	**Same**	**Same**
Conceptual replication (robustness) *Implementing a study of the same exposure (and comparator), outcome and estimand of interest*	**Different**	**Same**
	Same	**Different**
	Different	**Different**

International Replication Collaborations

- CNODES-EMA collaboration on DOACs
  - Concurrent replication of a common protocol
  - Multiple sites in Europe plus Canadian sites
- CNODES-Sentinel replication – Incretins
  - Sentinel replicated a CNODES study
  - Modified protocol to use semi-automated systems
- Both allow testing robustness of findings across much wider populations.

Concluding Thoughts

- Replication an important component of CNODES' processes
- Key to
  - Quality control/detecting data and/or analysis errors
  - Understanding of population/formulary differences
  - Understanding inherent variability

Thank you

Visit us at www.cnodes.ca

CNODES

CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

CIHR IRSC

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Adventures in Replication:

a regulatory perspective

ICPE Prague, 26 August, 2018

Presented by Xavier Kurz, European Medicines Agency

An agency of the European Union

Disclosure

Employee of the European Medicines Agency

The views expressed in this presentation are my personal views and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties.

No other relationships to disclose.

Why is replication relevant for regulators ?

- Regulators need to rely on best evidence to take and communicate informed decisions; different studies are often required and complementary to meet regulatory needs for information.
- Generalisability, particularly for EU: inferences from study results to general population (judgmental decision) stronger if based on more than one study
- Post-authorisation studies often based on secondary data collection with inherent limitations; confirmation of results may be needed
- Need to identify and control for determinants of risk which are not all collected in all data sources, e.g. confounders, effect modifiers.

- Regulatory decisions on safety are therefore rarely based on single study

  Assessment of CHC and VTE: ~ 40 observational studies from 13 countries, incl. meta-analyses

In this presentation:

Lessons learned from three examples of safety issues with replication:

- Hydrochlorothiazide and risk of skin and lip cancer
- Fluoroquinolones and risk of tendon disorders
- Direct oral anticoagulants and risk of bleeding

Example 1: Hydrochlorothiazide and risk of skin and lip cancer

Pottegard et al. Hydrochlorothiazide use is strongly associated with risk of lip cancer. J Intern Med 2017; 282: 322–331.

- HCTZ is photosensitizing and has previously been linked to lip cancer in screening study in the US
- Association between HCTZ and risk of lip cancer studied in nested case-control analysis using the Danish Registries between 2004-2012- Conditional logistic regression.

Table 2 *Association between exposure to hydrochlorothiazide and risk of squamous cell carcinoma of the lip, according to cumulative amount of hydrochlorothiazide use*

Subgroup	Cases	Controls	Crude OR[a]	Adjusted OR[b]
Nonuse	494	55 666	1.0 (ref.)	1.0 (ref.)
Ever-use	139	7401	2.2 (1.8–2.6)	2.1 (1.7–2.6)
High use (≥25 000 mg)	94	2771	4.0 (3.2–5.0)	3.9 (3.0–4.9)
Cumulative amount				
1–4999 mg	16	1745	1.0 (0.6–1.7)	1.0 (0.6–1.7)
5000–9999 mg	12	1083	1.2 (0.7–2.2)	1.2 (0.7–2.2)
10 000–24 999 mg	17	1802	1.1 (0.7–1.7)	1.1 (0.7–1.7)
25 000–49 999 mg	20	1253	1.9 (1.2–2.9)	1.8 (1.2–2.9)
≥50 000 mg	74	1518	5.8 (4.5–7.5)	5.5 (4.2–7.2)

Pedersen et al. Hydrochlorothiazide use and risk of nonmelanoma skin cancer: A nationwide case-control study from Denmark. J Am Acad Dermatol 2018;78:673-81.

- HCTZ is photosensitizing and has previously been linked to lip cancer
- Association between HCTZ and risk of basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) studied in the Danish Cancer Registry during 2004-2012- Conditional logistic regression.

Subgroup	Case patients	Controls	Adjusted OR (95% CI)*	Adjusted OR (95% CI)†
Basal cell carcinoma				
Nonuse	63,653	1,281,894	1.0 (ref)	1.0 (ref)
Ever use	7900	148,989	1.07 (1.04-1.10)	1.08 (1.05-1.10)
High use (≥50,000 mg)	1897	30,075	1.28 (1.22-1.34)	1.29 (1.23-1.35)
Squamous cell carcinoma				
Nonuse	6817	149,944	1.0 (ref)	1.0 (ref)
Ever use	1812	22,518	1.80 (1.70-1.90)	1.75 (1.66-1.85)
High use	862	4802	4.05 (3.75-4.39)	3.98 (3.68-4.31)

Safety signal assessed by the EU Pharmacovigilance Risk Assessment Committee (PRAC)

- Potentially significant public health consequences.
- Only photosensitisation listed as a rare adverse reaction in the SPC.
- Limited quality and size of previous studies with missing data on confounders and no data on absolute risks. Importance of lag time periods.
- Two studies in same population and same databases (Danish registries). Uncertainty in the applicability or generalisability of the findings given variation between EU countries in:
  - incidence of NMSC and risk profiles
  - marketing status, approved indications and treatment strategies for antihypertensives
  - HCTZ exposure.

Replication study done by EMA in THIN database (UK) using same design *(D. Morales et al., submitted for publication).*

Study design differences

Definition	Danish studies	THIN studies
Population		
▪ No previous history of skin or other cancer*	Yes	Yes
▪ No record of organ transplantation	Yes	Yes
▪ No record of HIV diagnosis	Yes	Yes
▪ No previous azathioprine/cyclosporine/mycophenolate use	Yes	Yes
▪ Only patients included with at least 10 years follow-up	Yes	As secondary analysis
Time period of study	2004-2012	1999-2016
Outcomes		
▪ Method of identifying diagnoses	Histology	Read codes
Control selection		
▪ Up to 100 controls and 20 controls sampled respectively	Yes	Yes
▪ Selection of matched controls using risk set sampling	Yes	Yes
Exposures		
▪ HCTZ excluded within 2 years of index date	Yes	Yes
▪ High dose HCTZ definition in lip cancer studies	25,000 mg	25,000 mg
▪ High dose HCTZ definition in other cancer studies	50,000 mg	50,000 mg
Covariates included as potential confounders		
▪ Retinoids/tetracyclines/macrolides/quinolones/amiodarone	Yes	Yes
▪ Aspirin, NSAIDs, statins	Yes	Yes
▪ Diabetes, COPD and alcohol abuse	Yes	Yes
▪ Charlson comorbidity index score	Yes	Yes
▪ Highest achieved education	Yes	No
▪ Smoking and body mass index	No	Yes

5 year exposure lag period

Lip cancer	Adjusted OR[a]	Adjusted OR[b]	Adjusted with smoking & BMI[e]
All patients irrespective of follow-up duration			
Non-use	1.0 (ref)	1.0 (ref)	1.0 (ref)
Ever use	**2.48 (1.16-5.29)**	**2.15 (1.00-4.63)**	**2.31 (1.07-4.97)**
Cumulative amount (mg)			
▪ 1-24,999	**2.96 (1.38-6.32)**	**2.59 (1.20-5.60)**	**2.85 (1.32-6.15)**
▪ >=25,000	-	-	-

Table 2 *Associ… lip, according to cumulative an…*

Subgroup				Adjusted OR[b]
Nonuse				1.0 (ref.)
Ever-use				2.1 (1.7–2.6)
High use (≥25…				3.9 (3.0–4.9)
Cumulative a…				
1–4999 mg				1.0 (0.6–1.7)
5000–9999 mg	12	1083	1.2 (0.7–2.2)	1.2 (0.7–2.2)
10 000–24 999 mg	17	1802	1.1 (0.7–1.7)	1.1 (0.7–1.7)
25 000–49 999 mg	20	1253	1.9 (1.2–2.9)	1.8 (1.2–2.9)
≥50 000 mg	74	1518	5.8 (4.5–7.5)	5.5 (4.2–7.2)

Basal cell carcinoma	Adjusted OR[a]	Adjusted OR[b]	Adjusted with smoking & BMI[e]
All patients irrespective of follow-up duration			
Non-use	1.0 (ref)	1.0 (ref)	1.0 (ref)
Ever use	**1.24 (1.16-1.32)**	**1.08 (1.01-1.15)**	**1.10 (1.03-1.17)**
Cumulative amount (mg)			
▪ 1-24,999	**1.13 (1.04-1.24)**	1.06 (0.98-1.14)	1.08 (0.995-1.16)
▪ 25,000-49,999	**1.21 (1.03-1.43)**	1.08 (0.92-1.26)	1.10 (0.94-1.29)
▪ >=50,000	**1.42 (1.10-1.84)**	**1.30 (1.03-1.65)**	**1.34 (1.06-1.69)**

Squamous cell carcinoma	Adjusted OR[a]	Adjusted OR[b]	Adjusted with smoking & BMI[e]
All patients irrespective of follow-up duration			
Non-use	1.0 (ref)	1.0 (ref)	1.0 (ref)
Ever use	**1.50 (1.26-1.78)**	**1.22 (1.02-1.45)**	**1.25 (1.05-1.48)**
Cumulative amount (mg)			
▪ 1-24,999	**1.28 (1.03-1.59)**	1.03 (0.83-1.28)	1.05 (0.84-1.30)
▪ 25,000-49,999	**1.69 (1.15-2.47)**	1.38 (0.95-2.03)	1.44 (0.98-2.11)
▪ >=50,000	**3.40 (2.16-5.35)**	**2.93 (1.85-4.62)**	**3.05 (1.93-4.81)**

Subgroup djusted OR (95% CI)†

Basal cell carcin
Nonuse 1.0 (ref)
Ever use 1.08 (1.05-1.10)
High use (≥5 1.29 (1.23-1.35)

Squamous cell
Nonuse 1.0 (ref)
Ever use 1.75 (1.66-1.85)
High use 3.98 (3.68-4.31)

Replication of results in another population and database provides strong support for regulatory decision-making

Example 2: Risk of tendon rupture with fluoroquinolones

- Following a FDA review, the PRAC started a review of the persistence of fluoroquinolones adverse reactions to help determine the need for any restriction of indication.
- Risk of tendon rupture associated with fluoroquinolone is recognised but there is inconsistency in its reported size, with residual confounding in existing studies and limited information on absolute risk.
- Data were needed to better characterise the risk and how it is influenced by timing of exposure

- A study was performed by EMA in the THIN database
- Co-amoxiclav chosen as negative control to circumvent problems related to confounding by indication and severity.
- Manuscript submitted for publication (*D. Morales et al.*)

Table 2. Incidence rate ratios for the association between tendon rupture and current systemic fluoroquinolone and co-amoxiclav exposure.

Tendon rupture	Exposed cases/total	Exposed controls/total	Crude IRR	Adjusted IRR	Adjusted p-value
Any tendon rupture					
▪ Fluoroquinolones	111/4836	236/18356	1.79 (1.41-2.27)	1.61 (1.25-2.09)	<0.001
▪ Co-amoxiclav	98/4836	314/18356	1.15 (0.90-1.45)	1.02 (0.79-1.31)	0.900
Achilles tendon rupture					
▪ Fluoroquinolones	67/1577	82/6007	3.50 (2.45-5.02)	3.14 (2.11-4.65)	<0.001
▪ Co-amoxiclav	38/1577	114/6007	1.19 (0.81-1.77)	1.00 (0.64-1.57)	0.989
Biceps tendon rupture					
▪ Fluoroquinolones	20/1316	62/4946	1.19 (0.71-2.00)	1.07 (0.61-1.89)	0.804
▪ Co-amoxiclav	23/1316	74/4946	1.16 (0.72-1.88)	1.01 (0.61-1.66)	0.978
Other tendon rupture					
▪ Fluoroquinolones	24/1943	92/7403	0.94 (0.59-1.50)	0.82 (0.50-1.35)	0.439
▪ Co-amoxiclav	37/1943	126/7403	1.09 (0.75-1.60)	1.01 (0.68-1.50)	0.946

Replication of previous results in another population and database with negative control and taking into account additional confounders and interactions (not shown) provides support to regulatory decision-making

Example 3: Direct oral anticoagulants (DOACs) and the risk of bleeding

- From RCTs, the risk of haemorrhagic stroke and intracranial bleeding is lower for DOACs compared to vitamine K antagonists (VKA), but the risk of gastro-intestinal bleeding is increased, with differences between substances.
- Patients in RCTs are very different from real life populations. Despite several observational studies, evidence remains inconclusive for specific patient populations, e.g. those with older age, impaired renal function and other comorbidities.
- EMA-funded study to characterise the risk of major bleeding in DOAC users in a real-world setting to help establish the effectiveness of risk minimization measures.
- Requirement for common protocol study in several databases and settings in Europe.
- Preliminary results available.

Preliminary results: Risk of bleeding of DOACs vs. VKA in four large European databases

Database	Gastrointestinal bleeding			Intracranial Haemorrhage			Stroke		
	# events	Incidence rate per 1000 person years	Adjusted Hazard Ratio (95% CI)*	# events	Incidence rate per 1000 person years	Adjusted Hazard Ratio (95% CI)*	# events	Incidence rate per 1000 person years	Adjusted Hazard Ratio (95% CI)*
A	168	24.2	1.40 (1.17-1.67)	15	2.1	1.65 (0.90-3.03)	205	30.6	1.76 (1.50-2.08)
B	232	18.9	1.36 (1.17-1.58)	11	0.9	0.57 (0.30-1.08)	190	14.6	1.18 (1.00-1.39)
C	890	24.9	1.26 (1.15-1.39)	17	0.5	0.63 (0.32-1.23)	714	20.0	0.88 (0.81-0.95)
D	368	14.4	0.87 (0.76-1.00)	55	2.1	0.46 (0.33-0.63)	428	16.7	1.00 (0.88-1.13)

With more discrepancies at substance level, replication of same study design in several databases raises questions about inferences that can be made from the results.

PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2016; **25**(Suppl. 1): 156–165
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3968

SUMMARY

Multi-centre, multi-database studies with common protocols: lessons learnt from the IMI PROTECT project

Olaf H. Klungel[1,8]*, Xavier Kurz[2], Mark C. H. de Groot[1,3], Raymond G. Schlienger[4], Stephanie Tcherny-Lessenot[5], Lamiae Grimaldi[6], Luisa Ibáñez[7], Rolf H. H. Groenwold[1,8] and Robert F. Reynolds[9†]

- Use of positive controls: 6 drug-adverse event pairs with known association
- Using common protocols: implementation of different study designs for each drug-adverse event pair in different databases
- Analysis of sources of variability in results

Sources of variability: antibiotics and liver toxicity

Study design

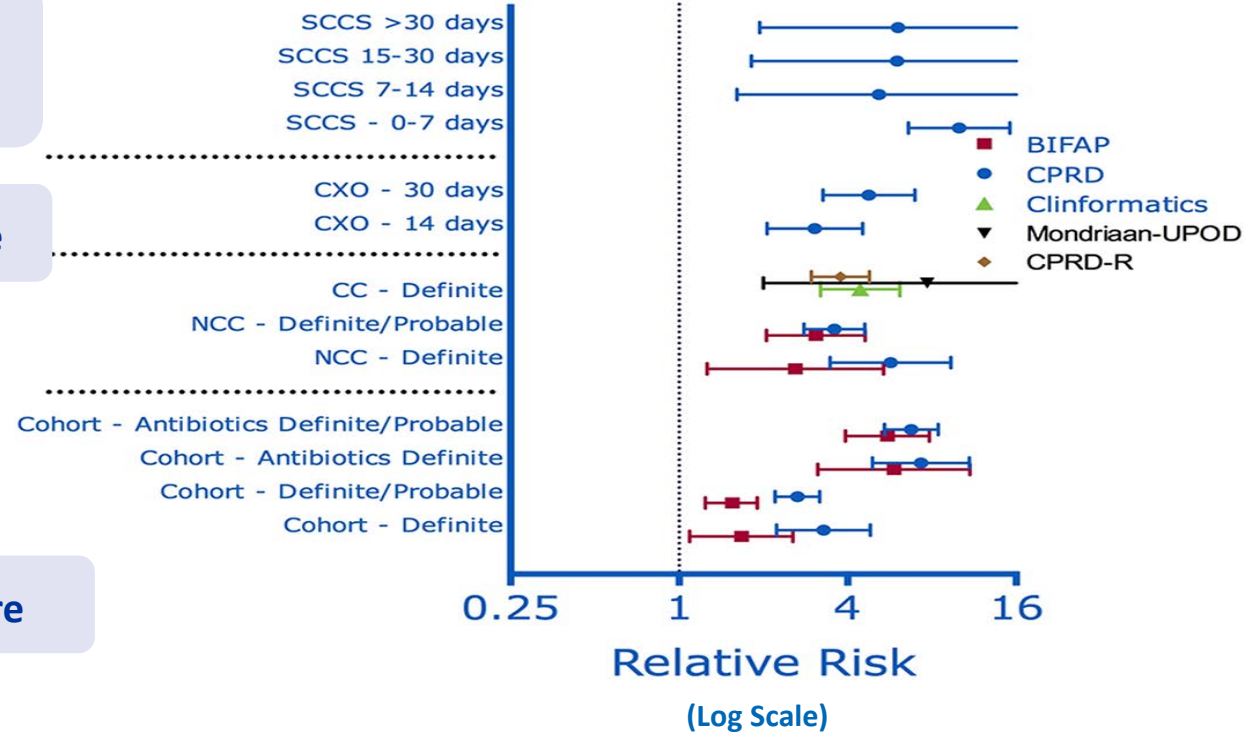
Outcome

Exposure

Study population

Confounding adjustment

Database



SCCS: self-controlled case series, CXO: case-crossover, CC: Case-control, NCC: nested case-control

Lessons learned from a regulatory perspective

- Given limitations of observational designs, multidatabase studies are useful for regulatory decision-making, if variability between results can be investigated and understood.
- Replication study most useful if it provides added value to current evidence, e.g. control for unmeasured confounding, measures of interactions, stratification by categories of effect modifiers, detailed analysis of dose effect and time factors
- Replication study requires in-depth knowledge of the data source(s)
- Careful choice of data source(s) for replication study to be made based on relevance and usefulness (not availability)

Potential barriers to replication studies

- Inadequate information available to replicate study design - differences still exist (~50% of studies registered in EU PAS Register have protocol posted)
- Time factor: replication study(-ies) may add burden on regulators and delay decision-making.
    - *How can we accelerate replication studies? Could CDM help?*
- Access to relevant data sources
- Replication studies less appealing to academics; harder to get funding and to get published.

Thank you

Acknowledgments:

Daniel Morales, Alison Cave

Further information

Xavier.kurz@ema.europa.eu

European Medicines Agency
30 Churchill Place • Canary Wharf • London E14 5EU • United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

Follow us on **@EMA_News**

Questions and Discussion