Long Live the "Medical Data Janitors": International Data Quality Assurance Practices in Distributed Data Networks

Judith C. Maro[1], Christian G. Reich[2], Keith Marsolo[3], Yoshiaki Uyama[4], Kristian B. Filion[5], Miriam C.J.M. Sturkenboom[6]

1. Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA
2. IQVIA, Cambridge, MA
3. Cincinnati Children's Hospital Medical Center, Cincinnati, OH
4. Pharmaceuticals and Medical Devices Agency (PMDA) Tokyo, Japan
5. McGill University, Montreal, QC, Canada
6. University Medical Center Utrecht, Utrecht, Netherlands

Disclosures

- The authors have the following conflicts of interest to disclose:
  - None.

Inspired by Dr. Califf's Comments…

Keynote Address

Robert Califf, Vice Chancellor for Health Data Science, Duke Health

Benefits and Risks of Medical Products: A Systematic Approach to Continuous Evidence Generation

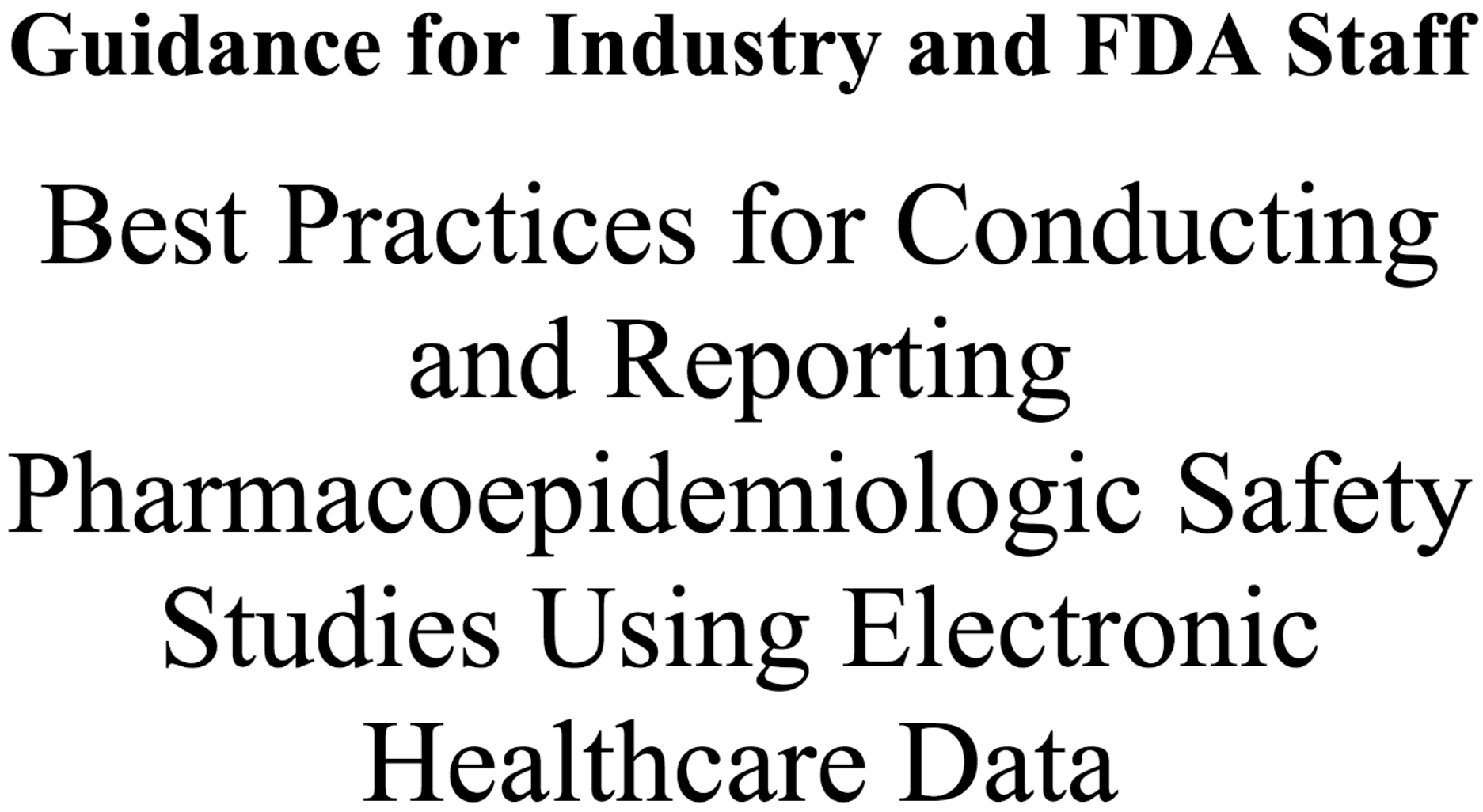
"…We've got to glorify the cleaning-up of data… Analytical techniques are increasingly automated, but understanding the context of the information and how to store it in a way that it's used for the right purpose is an art. I still use the word **data janitor**… and I think the most profound society should be the **Medical Data Janitorship** society because these are the people who are really going to make the difference…" (1:18:12)

Long Live the “Medical Data Janitors“: International Data Quality Assurance Practices in Distributed Data Networks

Judith C. Maro[1]

1. Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA

Quality Assurance Guidance Before Sentinel…



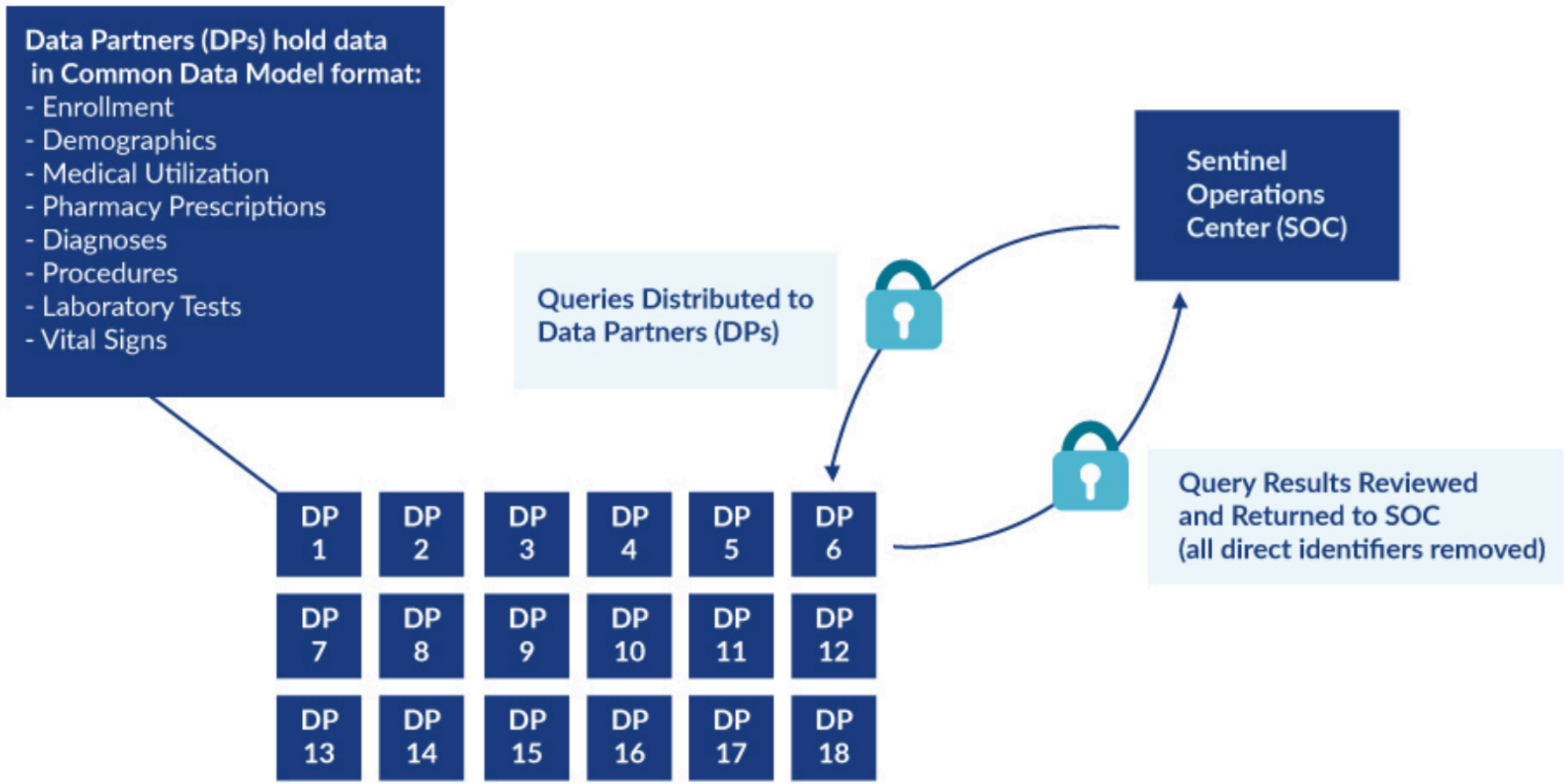
https://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf

Quality Assurance Envisioned as Project-Specific

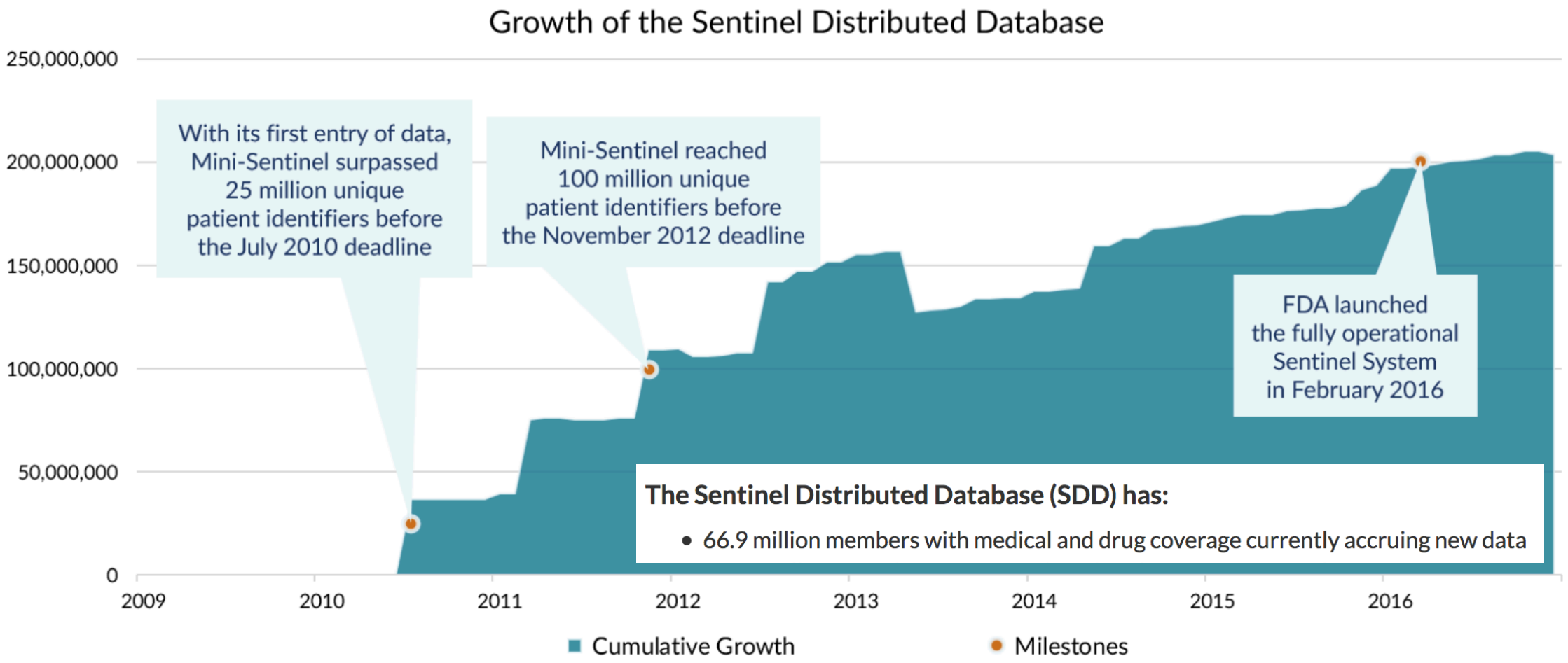
- The general procedures used by the data holders to ensure completeness, consistency, and accuracy of data collection and management.
- The frequency and type of any data error corrections or changes in data adjudication policies implemented by the data holders during the relevant period of data collection;
- A description of any peer-reviewed publications examining data quality and/or validity, including the relationships of the investigators with the data source(s);
- Any updates and changes in coding practices (e.g., ICD codes) across the study period that are relevant to the outcomes of interest
- Any changes in key data elements during the study time frame and their potential effect on the study
- A report on the extent of missing data over time (i.e., the percentage of data not available for a particular variable of interest) and a discussion on the procedures (e.g., exclusion, imputation) employed to handle this issue. Investigators should also address the implications of the extent of missing data on study findings and the missing data methods used

https://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf

Sentinel Distributed Database



Sentinel Distributed Database Characteristics



The area above depicts the cumulative number of unique patient identifiers in the Sentinel Distributed Database from 2010 to present. If patients move health plans, they may have more than one patient identifier.

Sentinel Common Data Model Guiding Principles

- Includes claims, electronic health record (EHR), and registry data and flexible enough to accommodate new data domains (e.g., free text).
- Data are stored at most **granular/raw level possible** with minimal mapping.
  - Distinct data types should be kept separate (e.g., prescriptions, dispensings)
  - Construction of medical concepts (e.g., outcome algorithms) from these elemental data is a **project-specific** design choice.
  - Sentinel stores these algorithms in a library for future use.
- Appropriate use and interpretation of local data requires the Data Partners' local knowledge and data expertise.
  - Not all tables are populated by all Data Partners→site-specificity is allowed.
- Designed to meet FDA needs for analytic flexibility, transparency, and control.

Sentinel Common Data Model v 6.0

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.

Adaptation of Guidance to a System Basis

Submit Comment

Sentinel Data Quality Assurance Practices

Project Title	**Sentinel Data Quality Assurance Practices**
Date Posted	*Thursday, March 23, 2017*
Status	Complete
Deliverables	Sentinel Data Quality Assurance Practices
Description	The Food and Drug Administration (FDA) set forth its current recommendations for data quality assurance (QA) in the following document: "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data" (Guidance), section IV.E "Best Practices – Data Sources: Quality Assurance (QA) and Quality Control (QC)," in May 2013. This Guidance describes best practices that particularly apply to observational studies designed to assess the risk associated with a drug exposure using electronic healthcare data.

https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model/sentinel-data-quality-assurance-practices

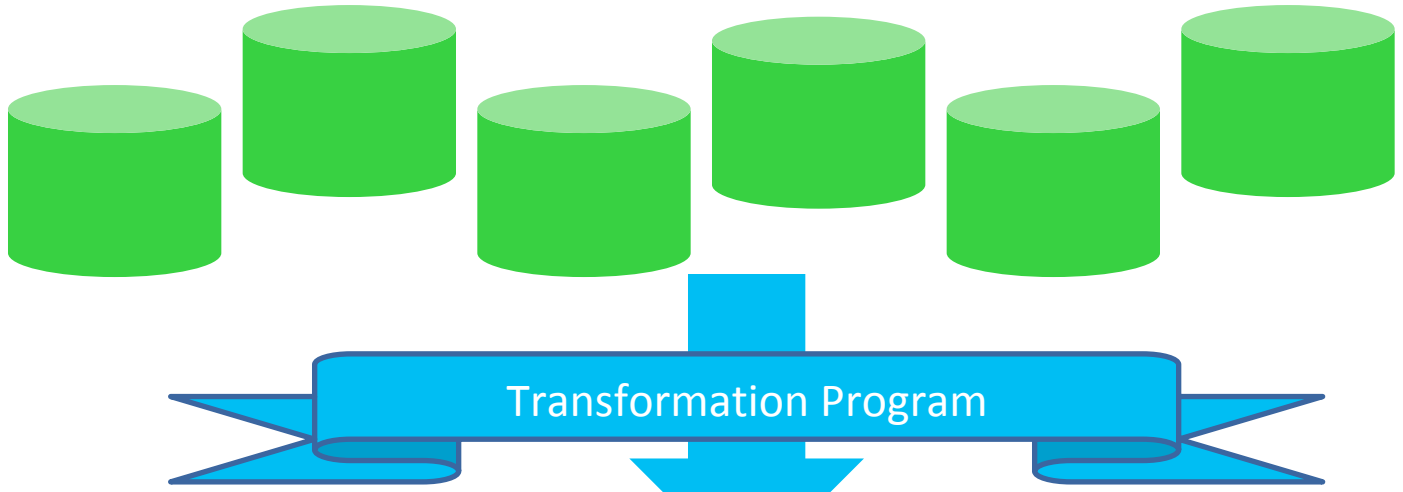
Project-Specific v. System Data Characterization

Project-Specific	System
"As needed / as you go"	"Always Ready"
Burden on Study Team	Burden on Quality Assurance Team
Ad hoc	Repeatable, Systematic
Cost is included in the cost of a study	Cost is front-loaded for studies that use system
Variable amount of data cleaning	1400+ checks to pass each dataset

Takehome: "Making data fit for purpose" at scale entails cost and time trade-offs.

Every Data Partner Transforms their Source Data into the Sentinel Common Data Model

Unique Data Partner's Source Database Structure



Data Partner's Database Transformed into SCDM Format (Refresh)

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

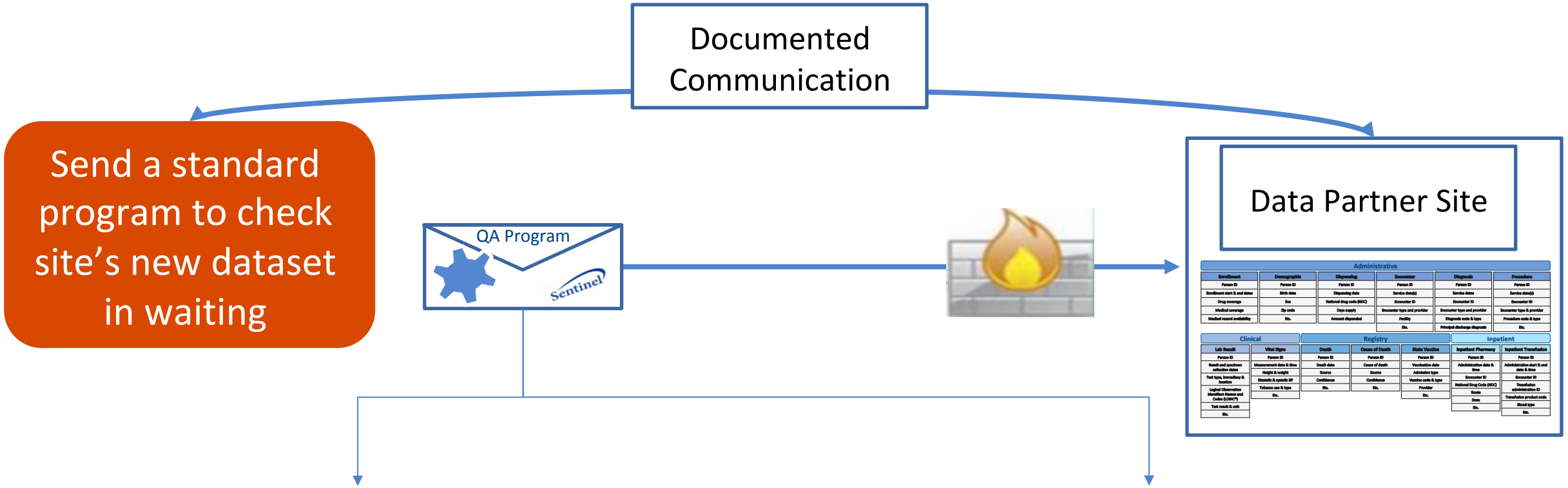
Inpatient Pharmacy	Inpatient Transfusion
Person ID	Person ID
Administration date & time	Administration start & end date & time
Encounter ID	Encounter ID
National Drug Code (NDC)	Transfusion administration ID
Route	Transfusion product code
Dose	Blood type
Etc.	Etc.

Transparent and Reproducible Quality Assurance

Data Quality Review and Characterization Programs v4.1.0

Project Title	Data Quality Review and Characterization Programs v4.1.0
Description	The Sentinel Data Quality Review and Characterization Programs are used by the Sentinel Operations Center (SOC) for data quality review and characterization of the Sentinel Distributed Database (SDD). To create the SDD, each Data Partner transformed local source data into the Sentinel Common Data Model (SCDM) format. The SOC created a set of data quality review and characterization programs to ensure that the SDD meets reasonable standards for data transformation consistency and quality and that the SDD data meets expectations needed for a distributed health data network.
Link	Sentinel Data Quality Review and Characterization Programs v4.1.0 – Overview Sentinel Data Quality Review and Characterization Programs v4.1.0 – Appendix A Sentinel Data Quality Review and Characterization Programs v4.1.0 – Appendix B Sentinel Data Quality Review and Characterization Programs v4.1.0 – SAS Programs View more details here.

Data Quality Review and Characterization Process



Compliance Checks

Level 1: Completeness, validity, accuracy

Level 2: Cross-variable and cross-table integrity

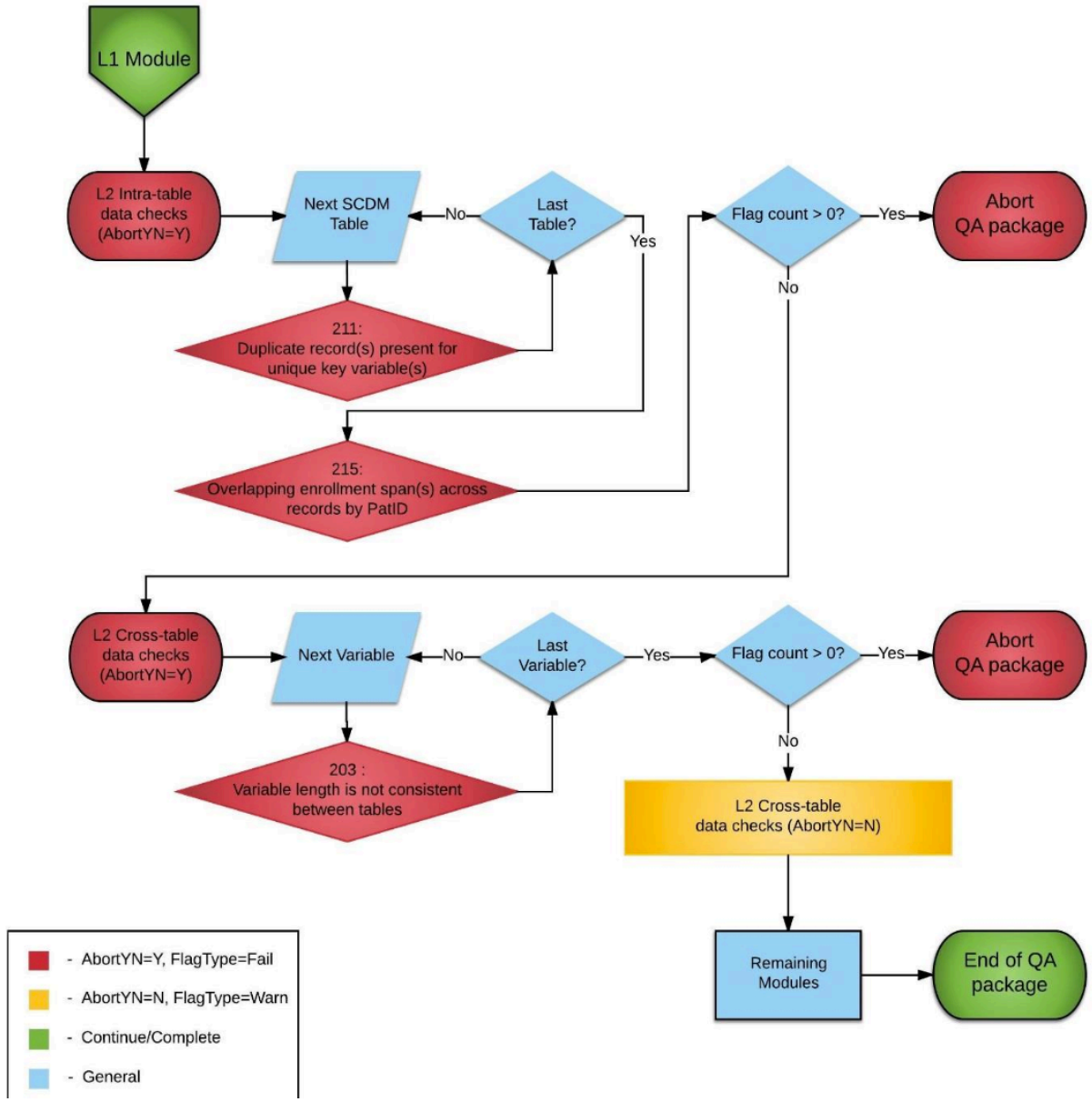
Judgment Call Checks

Level 3: Trends: consistency

Level 4: Logical: plausibility, convergence

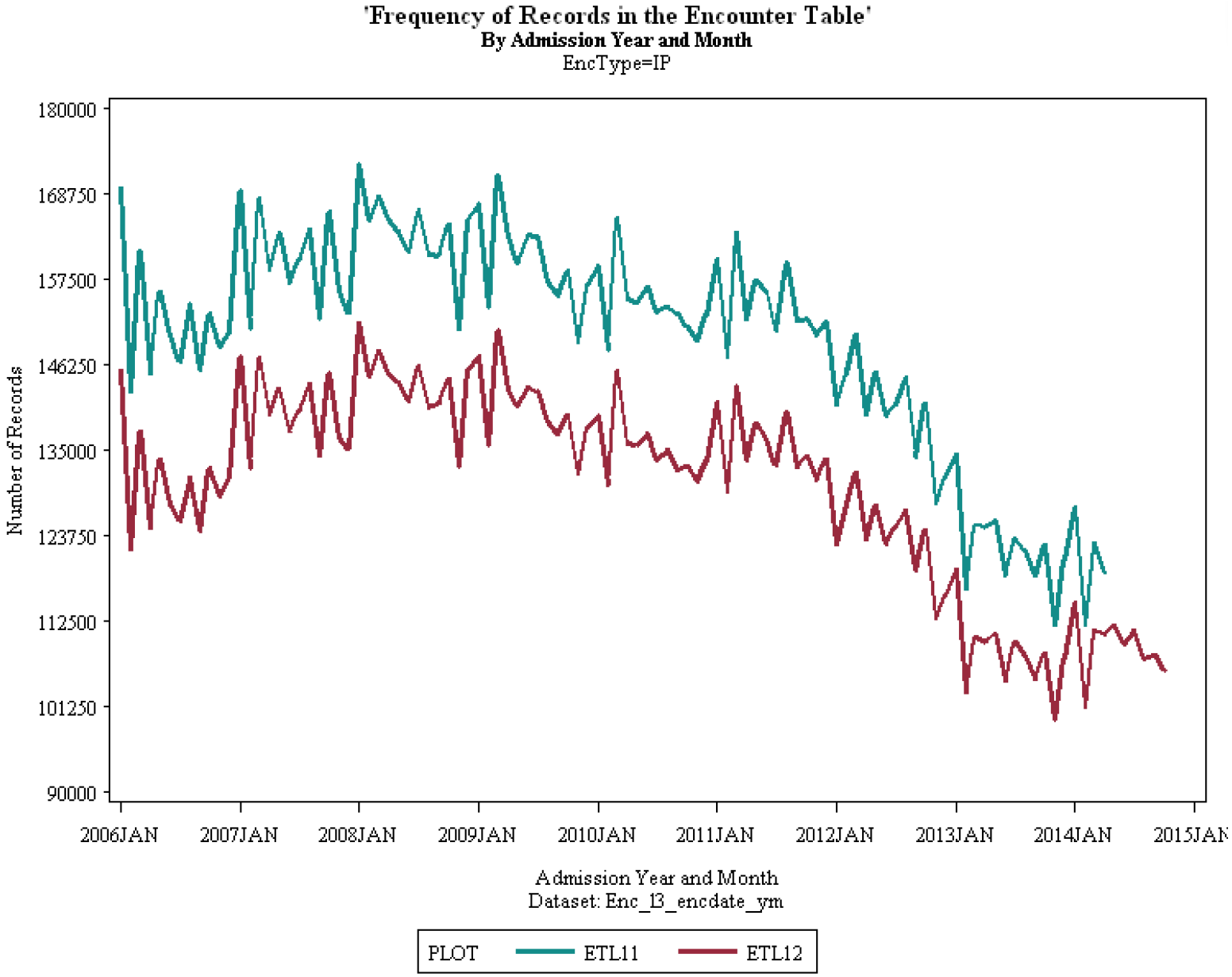
Quality Review and Characterization Program Logic

- Compliance checks for all tables are mandatory.
- Quality Review and Characterization Program will abort after it runs through all compliance checks, producing an automatically created report on failures.



Judgment Call Checks : What Do We See?

- Data Partner identified procedures done in an **outpatient** setting that were previously classified as inpatient or emergency department. These were re-assigned.
- Inpatient encounters decreased 19%.



Some Data Elements Require Additional Project-Specific Data Characterization

Platelet count original result units[‡]

Blank	FL	TH/UL	X10(3)
%	K/CMM	THOU/CMM	1000/UL
/100 W	k/cmm	thou/cmm	X10(3)/MCL
/CMM	K/CU MM	thou/mm3	X10(3)/UL
CMM	K/CUMM	THOU/UL	X10(6)/MCL
10 3L	K/MCL	THOUS/CU.MM	X10*9/L
10X3UL	K/mcL	THOUS/MCL	X10E3/UL
10^3/UL	K/UL	THOU/mcL	X1000
10*3/uL	k/uL	THOUS/UL	X10X3
10?3/uL	KU/L	Thou/uL	X10^3/UL
10E3/uL	K/MM3	THOUSA	x10
10e3/uL	K/mm3	THOUSAND	X10?3/ul
10e9/L	LB	THOUSAND/UL	X10E3/UL
E9/L	PLATELET CO	U	X10E3
BIL/L	T/CMM	X 10-3/UL	K/A?L
bil/L	TH/MM3	X 10(3)/UL	K/B5L
CU MM	th/mm3	X10 3	

- Supplementary Project-Specific Data Characterization is needed for less structured data elements (largely EHR-based elements).

Takeaways

- Sentinel's approach to Data Quality Review and Characterization shifts some of the burden away from study teams but project-specific data quality assurance may still be required.
  - New approach adheres to FDA requirements while making best use of finite resources.
  - More structured data elements are the most amenable to **system level** data characterization.
- TEAM approach (coordinating center + local experts) is needed.
- Per Dr. Califf, "Understanding the context of the information and how to store it in a way that it's used for the right purpose is an art," but transparent, repeatable programs and best practices make it more of a science.

Acknowledgements

- Data Management and Quality Assurance Team at the Sentinel Operations Center
- Sentinel Data Partners and their Data Management teams

IMS Health & Quintiles are now

IQVIA™

FDA CBER Biologics Effectiveness and Safety (BEST) Initiative

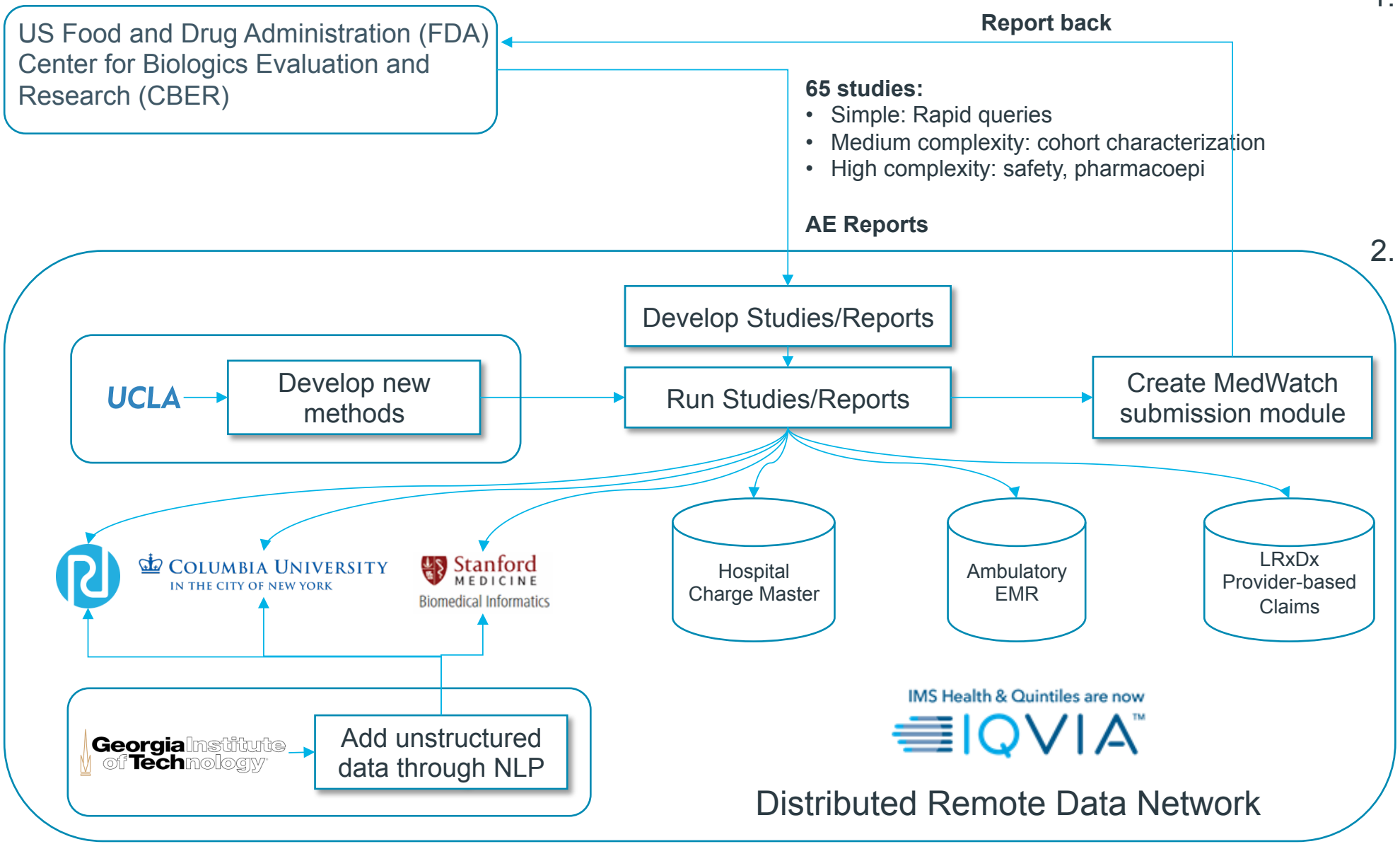
Christian Reich, MD, PhD
VP Real World Analytic Solutions

August 24, 2018

FDA Center for Biologics Evaluation and Research (CBER) Biologics Safety and Effectiveness (BEST) Initiative

- IQVIA (IMS Health & Quintiles)
- Observation Health Data Sciences and Informatics (OHDSI) Collaborative
- Columbia University
- Regenstrief Institute
- Stanford University
- Georgia Institute of Technology
- University of California Los Angeles (UCLA)

FDA/CBER BEST Initiative



1 year contract Sep 2017 – Oct 2018, two contracts:

1. Blood and Blood Product Safety Surveillance
   - OMOP CDM
   - EHR with blood products, components and vaccines
   - Tools and experts
2. New Innovative Methods for AE Reporting
   - EHR with blood products, components
   - Datamining and automated reporting of AEs from EHR

Systematic Approach to Quality

Data

- **Do data correctly represent clinical events?**
- **Metrics:**
  - Sensitivity
  - Specificity
  - Positive predictive value
  - Timeliness and temporality

Tools for manual review
Tools for automatic metrics

Processing

- **Does the process of making data available to analytics introduce errors?**
- **Metrics:**
  - Automated test results
  - Preservation of record counts
  - Mapping rates

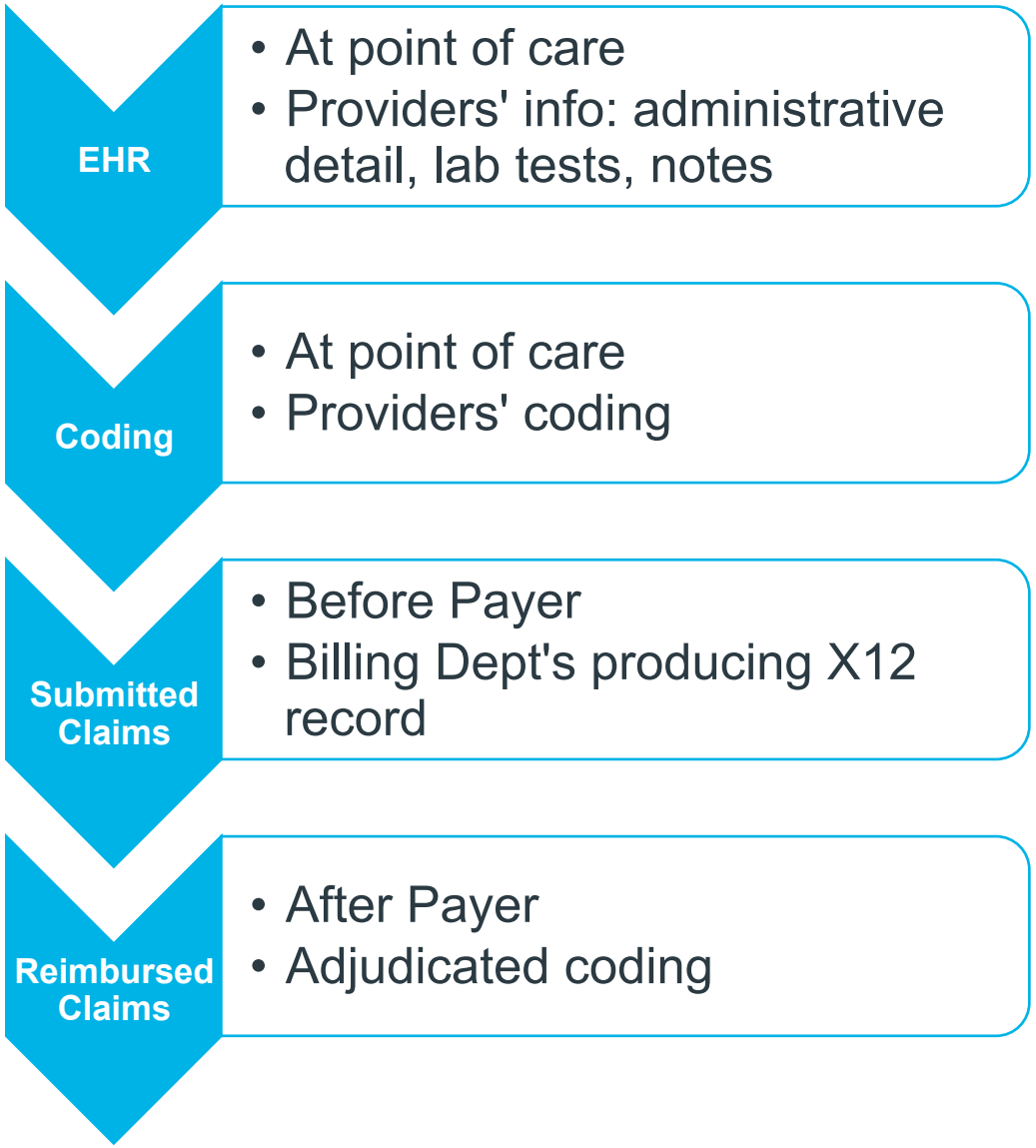
Tools for manual review
Tools for automatic metrics

Software and Methods

- **Do software tools and statistical methods reliably conform with specifications?**
- **Metrics:**
  - Software Development Life Cycle artifacts
  - Performance charac-teristics using positive and negative controls

Tools for manual review
Tools for automatic metrics

Generation of Codes and Reasons for Deviation



Quality Measures

- Sensitivity (0-100%)
- Specificity (0-100%)
- Timeliness (± hours-weeks)

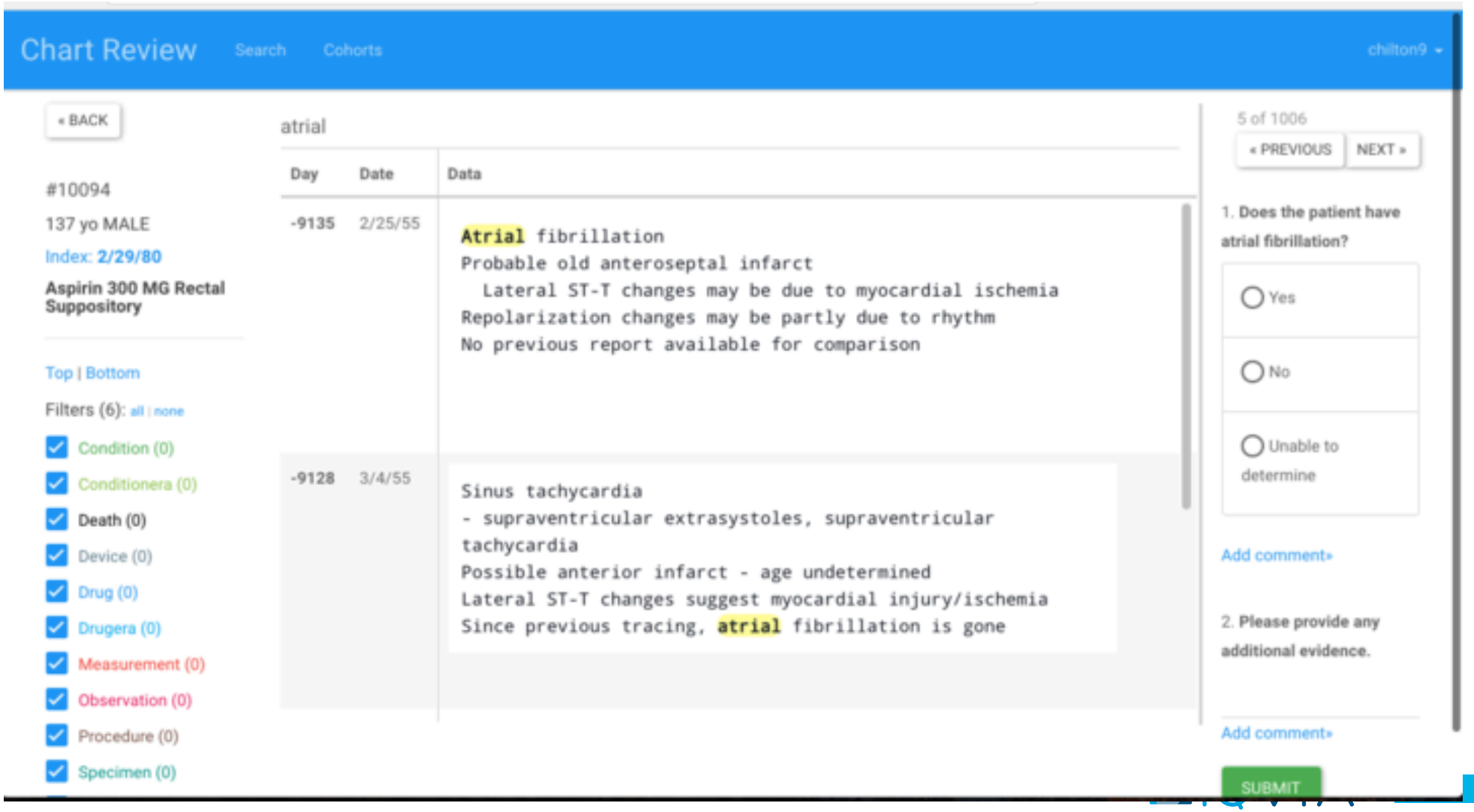
} over time

- Reasons for deviation
  - Relevance of condition
  - Amount of healthcare activity
  - Rules for reimbursement
  - Information is hierarchical
  - Bias
  - Fraud

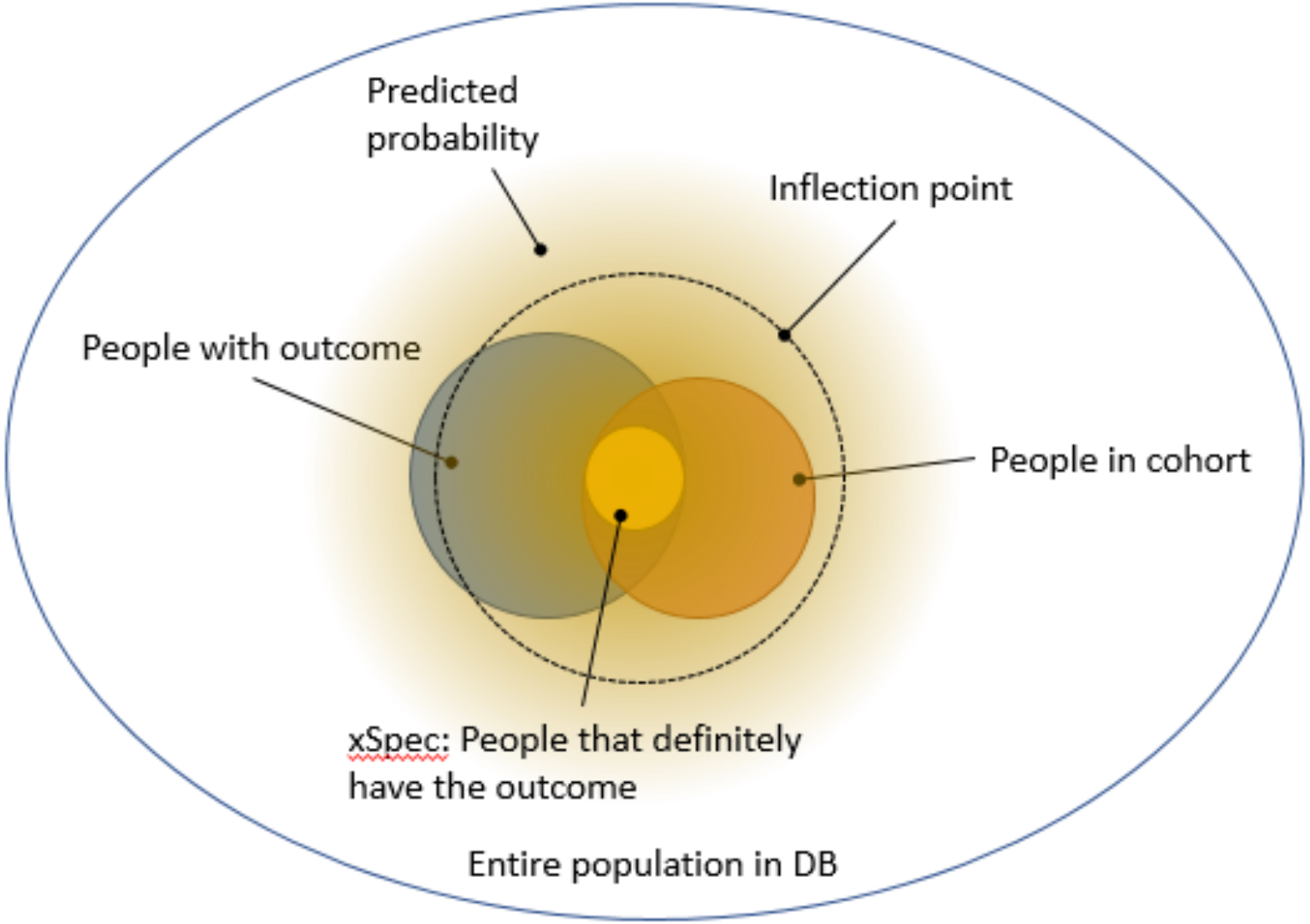
Data Quality Review: Remote Electronic Chart Validation

Standardized Evaluation Process

- Precision
- Recall
- PPV
- F-measure
- AUC



Research: Probabilistic Estimation of Quality Metrics



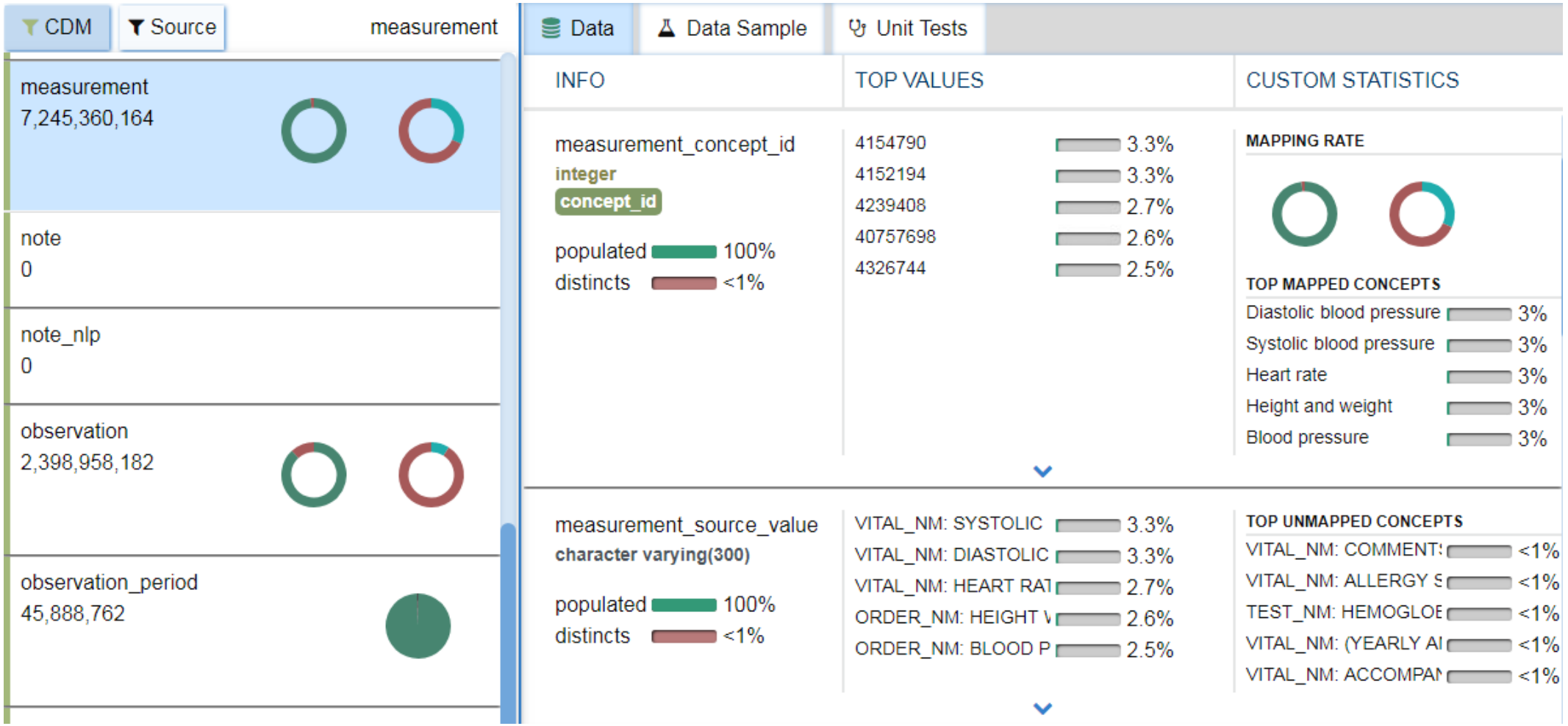
Approach:

- Create seed population wtih very high specificity (chart review or very stringent criteria)
- Build probabilistic model
- Find inflection point where cohort cuts over to background.
- Use this for sensitivity/ specificity/PPV estimation of codes and cohorts.

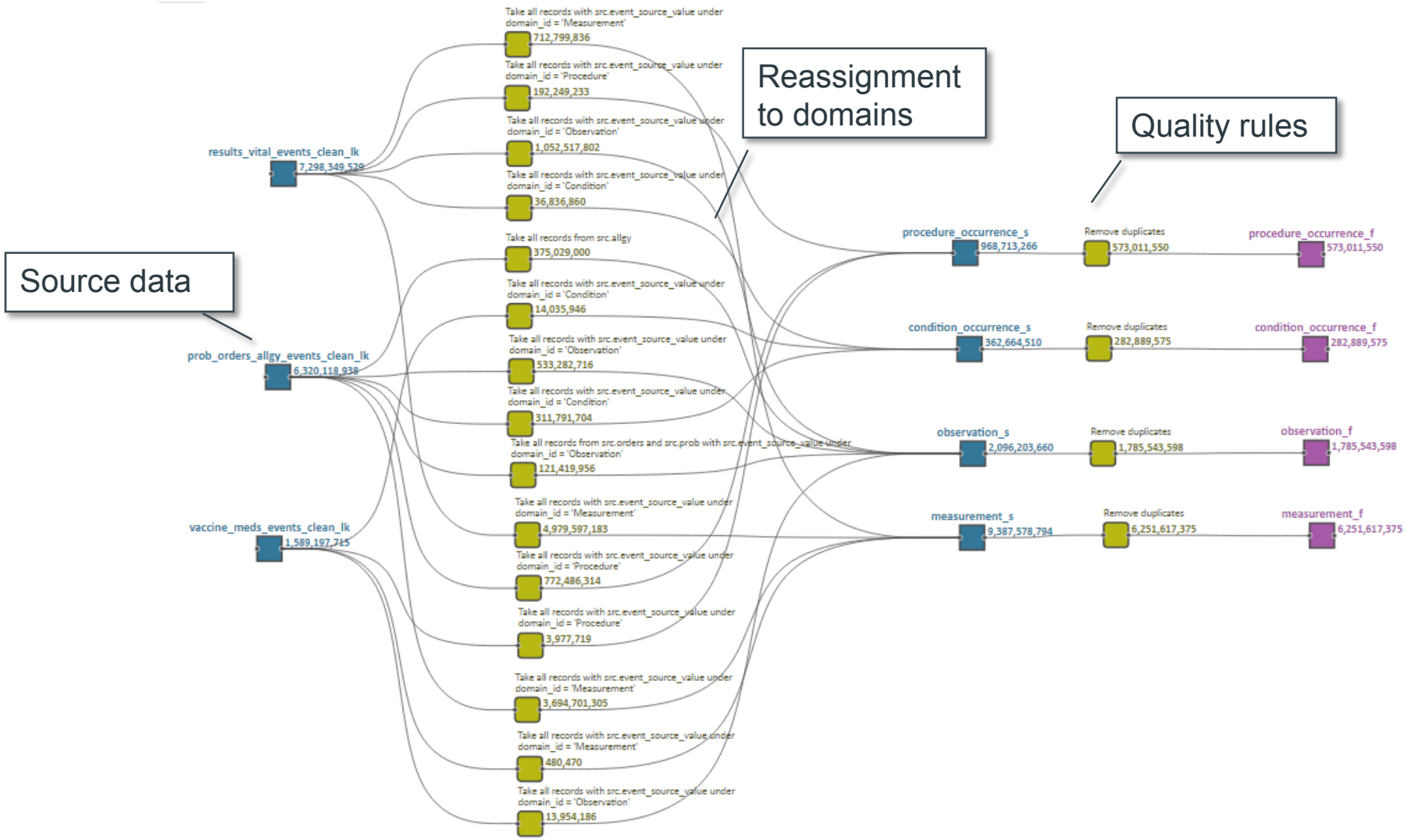
Automated Processing Quality Metrics

Test Type	Description	Tools
OMOP CDM schema compliance	Check schema is compliant with OHDSI DDL as required for a specific database type	STATIUS, ACHILLES
Adherence to business rules	Transformed data conformance to a set of standard business rules	STATIUS
Edit checks	Transformed data fits requires database quality constraints	STATIUS, ACHILLES
Data completeness	Test referential integrity and record completeness as a whole	STATIUS, ACHILLES
Mapping coverage	Test for % mappings coverage	Rabbit-in-A-Hat, USAGI, STATIUS, ACHILLES
Load coverage	Test ETL for % load coverage	STATIUS

Manual Processing Quality Review with Tools – Dashboard

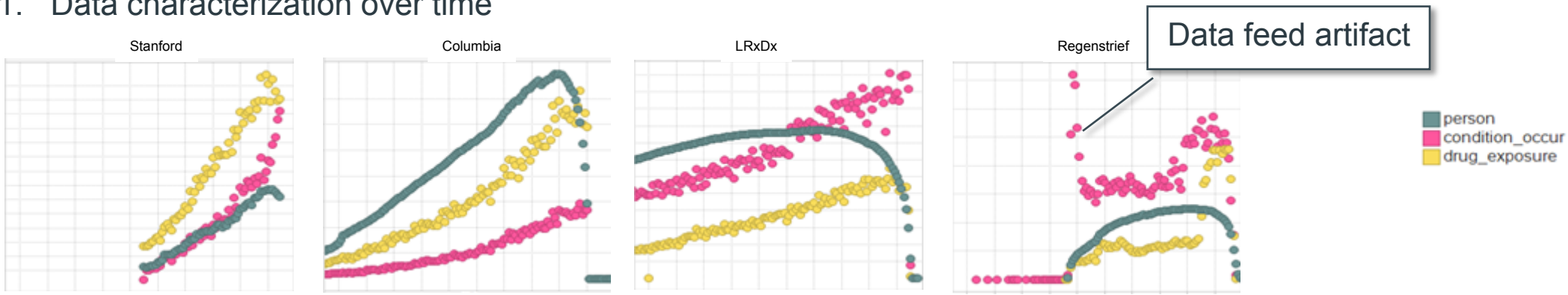


Manual Processing Quality Review with Tools – Business Rules



Data feed artifacts need to be detected and fixed

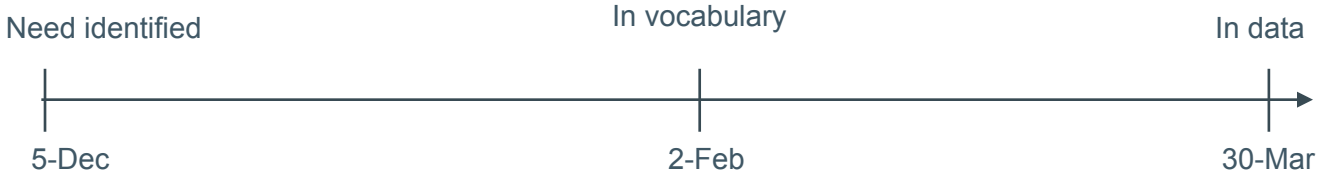
1. Data characterization over time



2. Codes and data feed gaps

Data Partner	Present	Initially Missing	Fixed
Columbia	ICD-PX	HCPCS, CPT4	CPT4 2-Mar
Stanford	ICD-PX,CPT	HCPCS	2-Mar
Regenstrief	ICD-PX,CPT, HCPCS	EHR and claim feed	2-Mar
LRxDx	CPT4, ICD-PX	ICD-PX without dot	underway
Hospital	CPT4, ICD-PX	HCPCS	25-Jan
AmbEMR	CPT4, order text	ICD-PX, HCPCS	25-Jan

3. Correction – ISBT-128 codes from Blood Banks



Data Source	Cases
Columbia	171,336
Regenstrief	303,752
Stanford	271,187

Software Validation

OHDSI Tools – ATLAS and ARACHNE

- Unit testing – tests a functional unit within a tool
- Code profiling – identifies code inefficiencies, including possible vulnerabilities
- Continuou
  environm
- Automate
  broken ar
- Manual te
  perform n
- Security a

ARACHNE
ATLAS

http://forums.ohdsi.org/t/software-validity-and-meeting-regulatory-requirements/3438

OHDSI

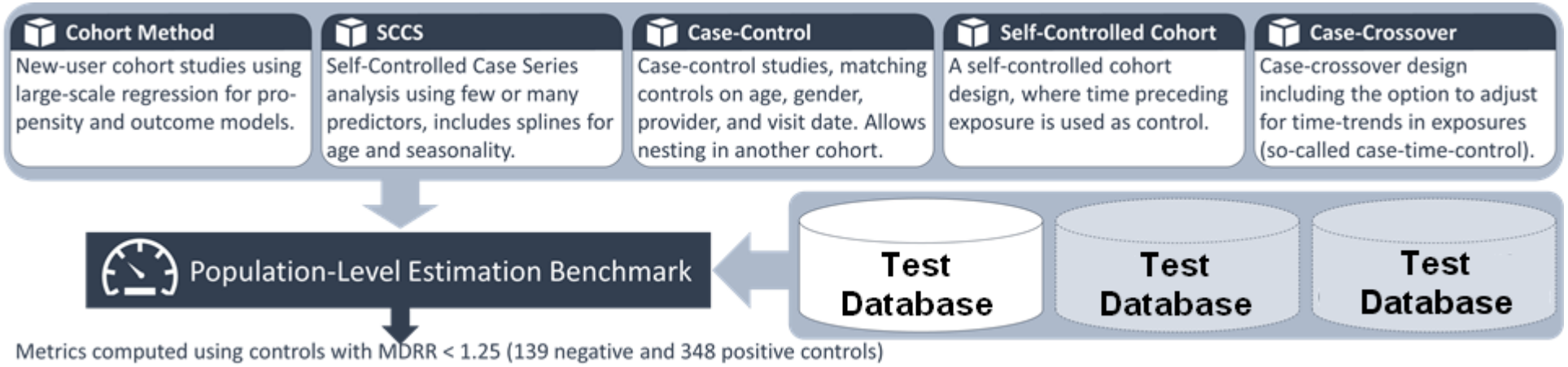
Software validity and meeting regulatory requirements

Developers

schuemie Martijn Schuemie Oct '17

Whenever we perform an observational study, one important consideration is the validity of our analysis software; Does our analysis code do what it is supposed to do? Although we have gone to great lengths to ensure the validity of the OHDSI Methods Library, we haven't done a very good job of documenting what we

Method Validation



Method	Analysis choices	AUC	Coverage of 95% CI	Mean precision	MSE	Type 1 error	Type 2 error	Missing
Case-control	Matching on age and gender, 2 controls per case	0.92	0.12	1812.92	0.6	0.81	0.01	0.01
Case-control	Matching on age and gender, 10 controls per case	0.91	0.1	3303.4	0.58	0.84	0.01	0.01
Case-control	Matching on age and gender, nesting in indication, 2 controls per case	0.9	0.3	1344.33	0.48	0.64	0.04	0.01
Case-control	Matching on age and gender, nesting in indication, 10 controls per case	0.91	0.25	2189.06	0.55	0.7	0.03	0.01
Case-crossover	Simple case-crossover	0.85	0.35	486.51	0.76	0.7	0.07	0
Case-crossover	Nested case-crossover	0.85	0.43	284.12	1.34	0.59	0.11	0
Case-crossover	Nested case-time-control, matching on age and gender	0.82	0.61	117.27	1.5	0.44	0.19	0.01
Cohort method	No matching, simple outcome model	0.76	0.42	131.74	1.17	0.49	0.18	0.04
Cohort method	Matching plus simple outcome model	0.82	0.61	85.66	0.58	0.26	0.23	0.11
Cohort method	Stratification plus stratified outcome model	0.86	0.68	104.05	1.46	0.19	0.23	0.06
Cohort method	Matching plus stratified outcome model	0.8	0.82	39.54	0.43	0.08	0.35	0.13
Cohort method	Matching plus full outcome model	0.77	0.86	25.22	0.42	0.01	0.54	0.49
SCCS	Simple SCCS	0.9	0.28	1958.69	0.45	0.71	0.02	0
SCCS	Using pre-exposure window	0.89	0.26	1871.1	0.48	0.75	0.03	0
SCCS	Using age and season	0.91	0.28	1913.83	0.45	0.7	0.01	0
SCCS	Using event-dependent observation	0.88	0.25	1906.17	0.5	0.7	0.02	0
SCCS	Using all other exposures	0.9	0.41	962.33	0.39	0.55	0.03	0
Self-controlled cohort	Length of exposure, index date in exposure window	0.9	0.32	1418.27	0.3	0.55	0.09	0.01
Self-controlled cohort	30 days of each exposure, index date in exposure window	0.91	0.52	466.84	0.08	0.49	0.11	0
Self-controlled cohort	Length of exposure, index date in exposure window, require full obs	0.91	0.34	1217.81	0.29	0.51	0.09	0.01
Self-controlled cohort	30 days of each exposure, index date in exposure window, require full obs	0.91	0.52	466.84	0.08	0.49	0.11	0
Self-controlled cohort	Length of exposure, index date ignored	0.94	0.36	1392.35	0.18	0.5	0.1	0.01
Self-controlled cohort	30 days of each exposure, index date ignored	0.93	0.55	438.31	0.09	0.26	0.14	0
Self-controlled cohort	Length of exposure, index date ignored, require full obs	0.94	0.39	1187.46	0.17	0.44	0.1	0.01
Self-controlled cohort	30 days of each exposure, index date ignored, require full obs	0.93	0.55	438.31	0.09	0.26	0.14	0

Summary

- Real World Data: QA responsibility with secondary use
- Quality = Data + Processes + Software/Methods
- Transparent and open approach needed for trust and reproducibility
- QA mechanisms: Tools for review, automated QA
- More work and research needed

Acknowledgement

- US FDA Center for Biologics Evaluation and Research (CBER) Office of Biostatistics and Epidemiology: CBER Sentinel Central Team
- OHDSI Collaborative
- Stanford University
- Regenstrief Institute
- Columbia University
- University of California Los Angeles (UCLA)
- Georgia Institute of Technology

PCORnet® Data Curation & Common Data Model

Keith Marsolo, PhD

On behalf of the PCORnet Coordinating Center's Distributed Research Network Operations Center (DRN OC)

pcornet®

The National Patient-Centered Clinical Research Network

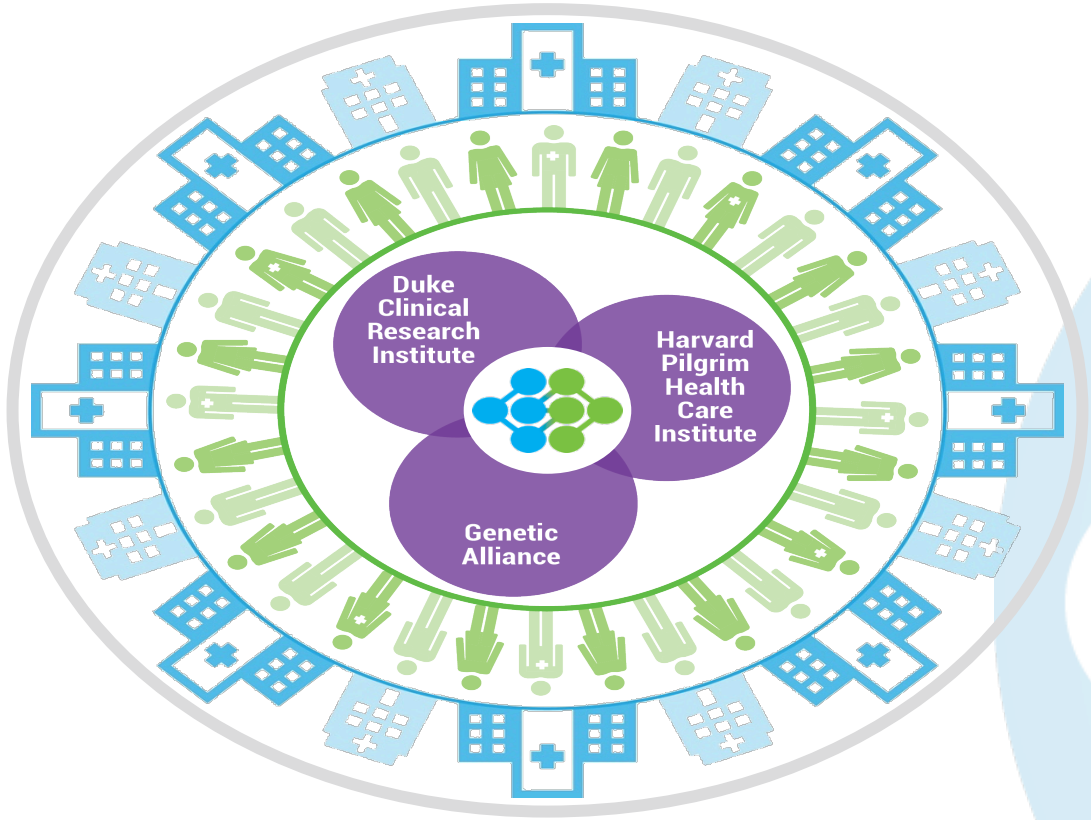
Disclosures

- I and my spouse/partner have no relevant relationships with commercial interests to disclose.

- *This work was supported through several Patient-Centered Outcomes Research Institute (PCORI) Program Awards (CC2-Duke-2016; ASP-1502-27079; OBS-1505-30699; OBS-1505-30683). All statements in this presentation, including its findings and conclusions, are solely those of the author and do not necessarily represent the views of PCORI, its Board of Governors or Methodology Committee.*

pcornet®

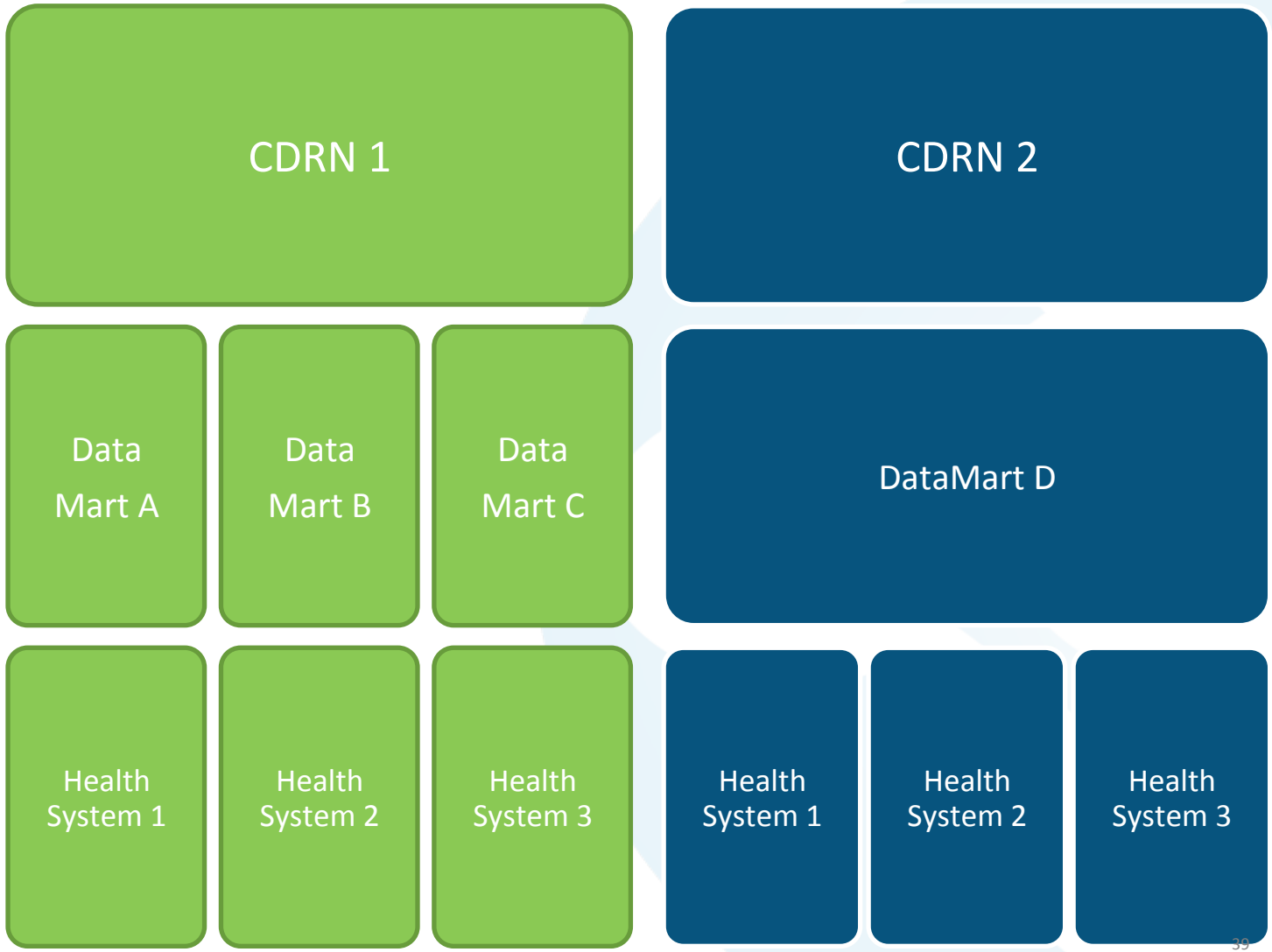
PCORnet® embodies a "network of networks" that harnesses the power of partnerships



Patient-Powered Research Networks (**PPRNs**) + Clinical Data Research Networks (**CDRNs**) + Health Plan Research Networks (**HPRNs**) + Coordinating Center = A national infrastructure for people-centered clinical research

PCORnet Terminology

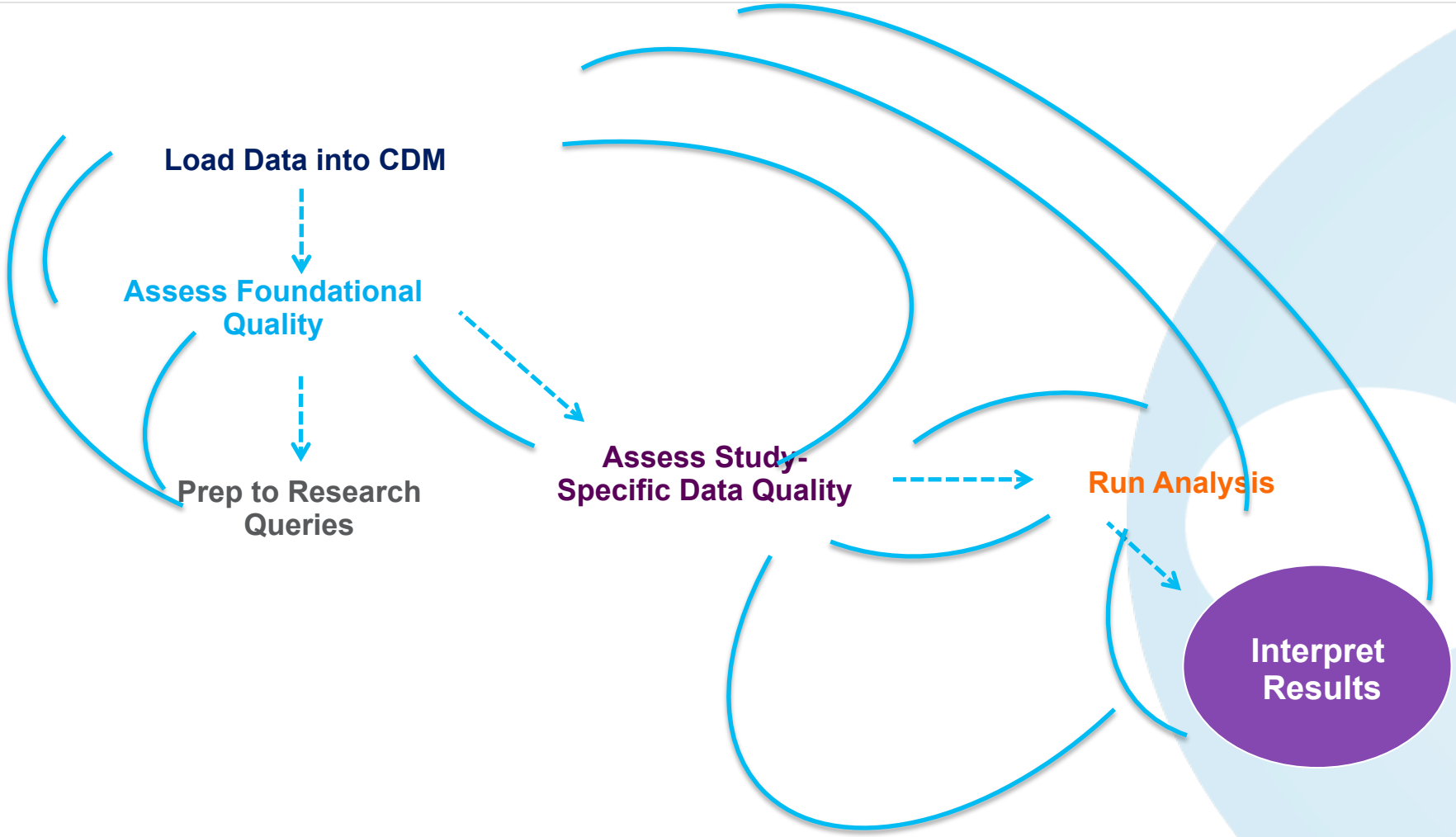
- Networks can consist of 1 or more DataMarts
- DataMarts can have 1 or more sites / health systems contribute data
- DataMarts are the unit of query
- Current PCORnet stats:
  - ~110 Health Systems / Health Plans
  - 80 DataMarts (also known as network partners)



pcornet®

Queries within PCORnet®

Learning within PCORnet®



Variation when loading the CDM

Network partners often have to make decisions on how to map their source data to the CDM

SITE 1
Social Work Visit
Allied Health
Office Visit
Nurse Visit
Procedure Visit
Employee Health
Vascular Lab
Sleep Study Visit
Social Work Visit

SITE 2
Office Visit
Specimen
Postpartum Visit
Clinical Support
Initial Prenatal

SITE 3
Home Care Visit
Office Visit
Therapy Visit
Orders Only
Cardiology Testing
Hospital Encounter

Common Data Model
Ambulatory Visit (AV)
Emergency Department (ED)
ED Admit to Inpatient (EI)
Inpatient Hospital (IP)
Non-Acute Inst. Stay (IS)
Observation Stay (OS)
Institutional Consult (IC)
Other Ambulatory (OA)
Other (OT)
Unknown (UN)
No Information (NI)

Reality is even more complicated (encounter types from one EHR)

REGISTRATION
EMPTY
LAB REQUISITION
INITIAL CONSULT
ANTI-COAG VISIT
PROCEDURE VISIT
OFFICE VISIT
CONSENT FORM
SCREENING FORM
EXTERNAL HOSPITAL ADMISSION
LETTER (OUT)
REFILL
IMMUNIZATION
HISTORY
RESEARCH ENCOUNTER
REFERRAL
ORDERS ONLY
RX REFILL AUTHORIZE
MEDS ONLY (WEB)
MEDS VOID (WEB)
RESOLUTE PROFESSIONAL BILLING HOSPITAL PROF FEE
EPISODE CHANGES
ANCILLARY ORDERS
PHARMACY VISIT
BPA
ROUTINE PRENATAL
INITIAL PRENATAL
OPHTH OFFICE VISIT
ABSTRACT
WALK-IN
TREATMENT PLAN
ALLIED HEALTH
NURSE ONLY
SOCIAL WORK
NUTRITION
PHYSICAL THERAPY
OCCUPATIONAL THERAPY
SPEECH THERAPY
RESPIRATORY THERAPY
CASE MANAGEMENT
EDUCATION
SURGICAL H&P
CLINICAL SUPPORT
MEDS ONLY / E - PRESCRIBE
PFT ONLY
TRANSPLANT PRE-EVALUATION
TRANSPLANT EVALUATION
TRANSPLANT FOLLOW-UP
TRANSPLANT RESULTS ENTRY
IMMUNOTHERAPY
ALLERGY TESTING
SPECIMEN COLLECTION
AUTO RELEASE ORDERS
URODYNAMIC TESTING
PRE-NATAL
CONSULT CHECKLIST
BOWEL MANAGEMENT
CARE CONFERENCE
INTAKE/TRIAGE
VNS REPROGRAM/SHUTOFF
CLINICAL NOTE
GENETICS
PASTORAL
THERAPY VISIT
INTAKE - NEW PATIENT
HIM SCANS
PRE-VISIT PLANNING
TRANSCRIBED ORDERS
SCHOOL TEACHER/INTERVENTION
CHILD LIFE
THERAPY PROGRESS SUMMARY
BRONCHOSCOPY REQUEST
HEMONC SOCIAL WORK
AUD CONSULT
OPH CONSULT
ALG CONSULT
UROLOGY COMPLEX INTAKE

EEG
EXERCISE
CARDIOLOGY TESTING
PUMP/CGM INITIATION ORDERS
MED TAPER SCHEDULE
GENETIC COUNSELOR
NEONATOLOGY TESTING
CARE CONFERENCE - PATIENT/FAMILY PRESENT
HOME VISIT - PALLIATIVE CARE
ABUSE REPORTING
CARE COORDINATOR
SPECIAL NEEDS SUMMARY
EARLY INTERVENTION
HI NEURODEVELOPMENTAL CLINIC TRACKING
INFUSION ORDERS
ENT CLINIC VISITS
FEES/VOICE
HEPATOBLASTOMA LIVER TRANSPLANT FOLLOW UP
PRE-ADOPTION ENCOUNTER
EB PLANNING
FEES CLINIC
VPI - ENT/SPEECH
INTAKE
HVMC PLANNING
PRE-OP PHYSICAL
PLAN OF CARE
ENT INPATIENT VISIT
HOSPITAL TO HOSPITAL TRANSFER
DEVELOPMENTAL TESTING
BIOETHICS CONSULT
ENDO STIM TESTING
HIM INTERFACE CREATED
SURGICAL SITE INFECTION
DERM PATCH TESTING
INTAKE CONSULT
ADEC INTAKE
CPST-PSY ENCOUNTER
ECONSULT TELEMEDICINE
ROADMAP
HOSPITAL ENCOUNTER
UPDATE
PCP/CLINIC CHANGE
WAIT LIST
CLERICAL ORDERS
MOTHER BABY LINK
LACTATION ENCOUNTER
CANCELED
APPOINTMENT
SURGERY
ANESTHESIA
ANESTHESIA EVENT
UNMERGE
HEALTH MAINTENANCE LETTER
PATIENT EMAIL
E-VISIT
MOBILE ORDER ONLY
QUESTIONNAIRE SERIES SUBMISSION
PATIENT OUTREACH
CONTACT MOVED
NURSE TRIAGE
E-CONSULT
E-CONSULT COMMUNITY ORDER
TELEMEDICINE
EXTERNAL CONTACT
OPHTH EXAM
HOSPICE ADMISSION
HOME HEALTH ADMISSION
HOME CARE VISIT
HOME CARE UPDATE
PATIENT WEB UPDATE
COMMUNITY ORDERS
COMMITTEE REVIEW
POST MORTEM DOCUMENTATION
BILLING ENCOUNTER
HOSPITAL
CONFIDENTIAL
OPH TESTING
EDUCATOR
VOICE CLINIC
TELEPHONE

Reducing variation with CDM Implementation Guidance

- Created to address instances where there is ambiguity in the CDM specification:
  - CDM is silent on the issue – *what to do if date of death is completely unknown?*
  - Unexpected complexity in source data – *how to separate race & ethnicity if captured in a single field?*

ENCOUNTER Table Implementation Guidance

Guidance

- Each ENCOUNTERID will generally reflect a unique combination of PATID, ADMIT_DATE, PROVIDERID and ENC_TYPE.
- Every diagnosis and procedure recorded during the encounter should have a separate record in the DIAGNOSIS or PROCEDURES Tables.
- Multiple visits to the **same** provider on the same day may be considered one encounter, especially if defined by a reimbursement basis; if so, the ENCOUNTER record should be associated with all diagnoses and procedures that were recorded during those visits.
- Visits to **different** providers for different encounter types on the same day, however, such as a physician appointment that leads to a hospitalization, would generally correspond to multiple encounters within the ENCOUNTER table.
- Rollback or voided transactions and other adjustments should be processed before populating this table.
- Although "Expired" is represented in both DISCHARGE_DISPOSITION and DISCHARGE_STATUS, this overlap represents the reality that both fields are captured in hospital data systems but with variation in how each field is populated.
- Do not include scheduled encounters.
- Partners should ensure that "administrative" encounters (e.g., e-mail, phone, documentation-only), are coded to the appropriate encounter type, which is typically "OA" for outpatient visits.

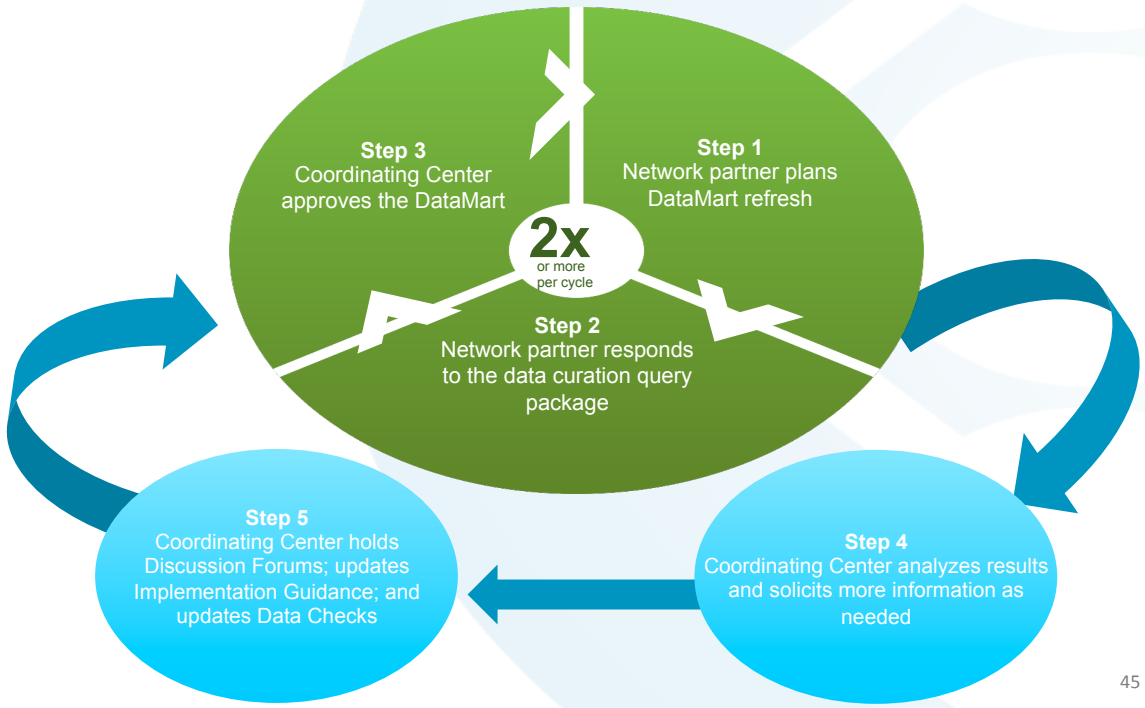
DEMOGRAPHIC Table Specification

Field Name	*RDBMS Data Type*	*SAS Data Type*	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*	*Data Element Provenance*	*Field-Level Implementation Guidance*
HISPANIC	RDBMS Text(2)	SAS Char(2)	Y=Yes N=No R=Refuse to answer NI=No information UN=Unknown OT=Other	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.	MSCDM v4.0 with modified field size and value set Compatible with "OMB Hispanic Ethnicity" (Hispanic or Latino, Not Hispanic or Latino)	**Populating RACE and HISPANIC if race and ethnicity are not captured separately within the source system** (e.g., "Hispanic or Latino" is included as a selection under Race) - for patients with a known race (e.g., Race is something other than "Hispanic or Latino", partners should set HISPANIC to "OT" and RACE to the appropriate race code. For patients who are listed as having a race of "Hispanic," partners should set HISPANIC to "Y" and RACE to "OT". In this situation, the combined race/ethnicity field is treated as known field capturing values for both race and ethnicity, which is why the preference is to use "OT" instead of "NI".

pcornet®

Assessing data quality – Foundational Data Curation

- Purpose
  - Evaluate data quality and fitness-for-use across a broad research portfolio
  - Generate meaningful, actionable information for network partners, investigators and other stakeholders
- Resources
  - Data quality checks
  - Data curation query packages
  - Analyses and reports
  - Discussion Forums



pcornet®

Data Curation Cycles: Our Journey So Far

Aspect	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
Start date	January 2016	November 2016	July 2017	January 2018	July 2018
CDM version	V3.0	V3.0	V3.1	V3.1	V4.1
CDM tables	7 (DEMOGRAPHIC, DIAGNOSIS, ENROLLMENT, ENCOUNTER, HARVEST, PROCEDURES, VITAL)	11 (added DISPENSING, PRESCRIBING, LAB_RESULT_CM and DEATH)	15 (added CONDITION, PRO_CM, DEATH_CAUSE and PCORNET_TRIAL)	15	18 (added MED_ADMIN, PROVIDER, OBS_CLIN (partial))
Distributed queries [3]	Diagnostic Query [1] Data Curation Query	Data Curation Query	Data Curation Query	Data Curation Query	Data Curation Query
Self-service queries	None	Diagnostic Query [1] Code Errors [2]	Diagnostic Query [1] Code Errors [2]	Diagnostic Query [1] Code Errors [2]	Diagnostic Query [1] Code Errors [2]
Annotated Data Dictionary	Excel spreadsheets	REDCap database	REDCap database	REDCap database	REDCap database
Data Quality Checks [4]	13 data checks 498 measures	20 data checks *(7 new, 9 revised)* 587 measures	26 data checks *(6 new, 8 revised)* 644 measures	27 data checks *(1 new, 5 revised)* 654 measures	31 data checks *(4 new, 13 revised)* 1144 measures
Analyses and Investigations	One-on-one discussions with DataMart teams	Network-wide Discussion Forums; DataMart-specific feedback	Network-wide Discussion Forums; DataMart-specific feedback	Network-wide Discussion Forums; DataMart-specific feedback	Network-wide Discussion Forums; DataMart-specific feedback

1. Evaluates table and field-level conformance with the CDM
2. Detects potential errors in diagnosis, procedure, lab, and Rx codes based on heuristics such as field length and presence of alphanumeric characters
3. Available at https://github.com/PCORnet-DRN-OC/PCORnet-Data-Curation
4. Available at http://pcornet.org/pcornet-data/

Cycle 5 Data Checks

Category	Type	Check	Description	Changes from v4
Data Model Conformance	Required	DC 1.01	Required tables are not present	Added MED_ADMIN, OBS_CLIN, OBS_GEN, and PROVIDER
	Required	DC 1.02	Expected tables are not populated	None
	Required	DC 1.03	Required fields are not present	Added RAW fields and new fields
	Required	DC 1.04	Fields do not conform to data model specifications for data type, length, or name.	Added new fields
	Required	DC 1.05	Tables have primary key definition errors	Added MED_ADMIN, OBS_CLIN, and PROVIDER
	Required	DC 1.06	Fields contain values outside of data model specifications	Added new fields
	Required	DC 1.07	Fields have non-permissible missing values	Added new fields and removed DIAGNOSIS.ENCOUNTERID and PROCEDURES.ENCOUNTERID
	Required	DC 1.08	Tables contain orphan PATIDs	Added MED_ADMIN and OBS_CLIN
	Required	DC 1.09	Tables contain orphan ENCOUNTERIDs	Reclassified from Investigative to Required; changed from a 5% to a 0% threshold; added MED_ADMIN and OBS_CLIN
	Required	DC 1.10	Replication errors between the ENCOUNTER, PROCEDURES and DIAGNOSIS tables	None
	Required	DC 1.11	More than 5% of encounters are assigned to more than one patient	Reclassified from Investigative to Required
	Required	DC 1.12	Tables contain orphan PROVIDERIDs	New
Data Plausibility	Investigative	DC 2.01	More than 5% of records have future dates	Added new fields
	Investigative	DC 2.02	More than 10% of records fall into the lowest or highest categories of age, height, weight, diastolic blood pressure, systolic blood pressure, or dispensed days supply	None
	Investigative	DC 2.03	More than 5% of patients have illogical date relationships	Added new fields
	Investigative	DC 2.04	The average number of encounters per visit is > 2.0 for inpatient (IP), emergency department (ED), or ED to inpatient (EI) encounters	None
	Investigative	DC 2.05	More than 5% of results for selected laboratory tests do not have the appropriate specimen source	Added new value set
	Investigative	DC 2.06	Median lab result values for selected tests are statistical outliers	New
	Investigative	DC 2.07	The average number of principal diagnoses per encounter is above threshold [2.0 for inpatient (IP) and ED to inpatient (EI))	New
Data Completeness	Investigative	DC 3.01	The average number of diagnoses records with known diagnosis types per encounter is below threshold [1.0 for ambulatory (AV), inpatient (IP), emergency department (ED), or ED to inpatient (EI) encounters]	None
	Investigative	DC 3.02	The average number of procedure records with known procedure types per encounter is below threshold [0.75 for ambulatory (AV) encounters, 0.75 for emergency department (ED) encounters, 1.00 for ED to inpatient (EI) encounters, and 1.00 for inpatient (IP) encounters]	None
	Investigative	DC 3.03	More than 10% of records have missing or unknown values for the following fields: BIRTH_DATE, SEX, DISCHARGE_DISPOSITION (IP/EI encounters only), DISCHARGE_DATE (IP/EI encounters only), PX_DATE, RX_ORDER_DATE, DISPENSE_SUP, DX_ORIGIN, PX_SOURCE, VITAL_SOURCE, DEATH_SOURCE, CONDITION_SOURCE, RX_SOURCE, MEDADMIN_SOURCE, DIAGNOSIS.ENCOUNTERID, or PROCEDURES.ENCOUNTERID	Added new fields
	Required	DC 3.04	Less than 50% of patients with encounters have DIAGNOSIS records	None
	Required	DC 3.05	Less than 50% of patients with encounters have PROCEDURES records	None
	Investigative	DC 3.06	More than 10% of IP (inpatient) or ED to inpatient (EI) encounters with any diagnosis don't have a principal diagnosis	None
	Investigative	DC 3.07	Encounters, diagnoses, or procedures in an ambulatory (AV), emergency department (ED), ED to inpatient (EI), or inpatient (IP) setting are less than 75% complete three months prior to the current month	None
	Investigative	DC 3.08	Less than 80% of prescribing orders are mapped to a RXNORM_CUI which fully specifies the ingredient, strength and dose form	None
	Investigative	DC 3.09	Less than 80% of laboratory results are mapped to LAB_LOINC	None
	Investigative	DC 3.10	Less than 80% of quantitative results for tests mapped to LAB_LOINC fully specify the normal range	None
	Investigative	DC 3.11	Vital, prescribing, or laboratory records are less than 75% complete three months prior to the current month	None
	Investigative	DC 3.12	Less than 80% of quantitative results for tests mapped to LAB_LOINC fully specify the SPECIMEN_SOURCE and RESULT_UNIT	New

Empirical Data Curation Report

Table IIIB. Records With Extreme Values

This table supports Data Check 2.02 (more than 10% of records fall into the lowest or highest categories of age, height, weight, diastolic blood pressure, systolic blood pressure, or dispensed days supply). A high percentage of records in these categories may signal incorrect measurement units. Exceptions for blood pressure measures are expected for pediatric populations. Data check exceptions are highlighted in blue and should be investigated and explained in the ETL ADD.

		Data Check Parameters			Records with values in the lowest category		Records with values in the highest category			
Table	Field	Low	High	Records	N	%	N	%	Median	Source table
VITAL	DIASTOLIC	<40 mgHg	>120 mgHg	41,883,101	4,546,498	10.9	9,674	0.0	n/a	VIT_L3_DIASTOLIC
VITAL	SYSTOLIC	<40 mgHg	>210 mgHg	41,883,101	46,753	0.1	1,752	0.0	n/a	VIT_L3_SYSTOLIC

Table IVI. Lab Data Completeness

This table shows the level of data completeness for LAB_RESULT_CM records and supports Data Check 3.09 (less than 80% of laboratory results are mapped to LAB_LOINC) and Data Check 3.10 (less than 80% of quantitative results for tests mapped to LAB_LOINC fully specify the normal range). Data check exceptions occur if the percentage is <80% or the numerator is 0. The data check exception threshold is high in order to better understand the inherent limitations and opportunities for improvement in these data. Exceptions are highlighted in blue and should be investigated and explained in the ETL ADD.

Description	Criteria	Numerator	Denominator	Percentage	Source table
Number of distinct LAB_LOINCs		471			LAB_L3_LOINC
Percentage of results mapped to a known LAB_LOINC	LAB_LOINC is not null	59,427,917	59,427,917	100.00	LAB_L3_RECORDC; LAB_L3_N
Percentage of results mapped to a known LAB_LOINC with a known result	LAB_LOINC is not null and (RESULT_NUM is not null and RESULT_MODIFIER is not null) or RESULT_QUAL is in ("BORDERLINE", "POSITIVE", "NEGATIVE" or "UNDETERMINED")	51,305,569	59,427,917	86.33	LAB_L3_RECORDC
Number of quantitative results for tests mapped to LAB_LOINC	LAB_LOINC is not null and RESULT_NUM is not null and RESULT_MODIFIER is not null	51,305,569			LAB_L3_RECORDC
Percentage of quantitative results for tests mapped to LAB_LOINC which fully specify the normal range.	LAB_LOINC is not null and RESULT_NUM is not null and RESULT_MODIFIER is not null and NORM_MODIFIER_LOW, NORM_RANGE_LOW, NORM_MODIFIER_HIGH, and NORM_RANGE_HIGH are all populated per CDM specifications.**	41,193,033	51,305,569	80.29	LAB_L3_RECORDC

Cycle 4 Discussion Forum Schedule

- March 5 – General overview of Cycle 4 findings
- March 12 – Exploratory analyses (e.g., unmatched codes, potential duplication of records) & overview of Data Curation Lab Groups
- March 19 – Identification of lab mapping errors through outlier detection
- March 26 – Medication mapping issues

Study-specific data characterization

- Assess data on the intended cohort related to study aims
- Ensure that outcomes / variables of interest are available & complete
- Determine whether partners actually have enough data / patients to participate
- Requires upfront investment, but can save significant time overall

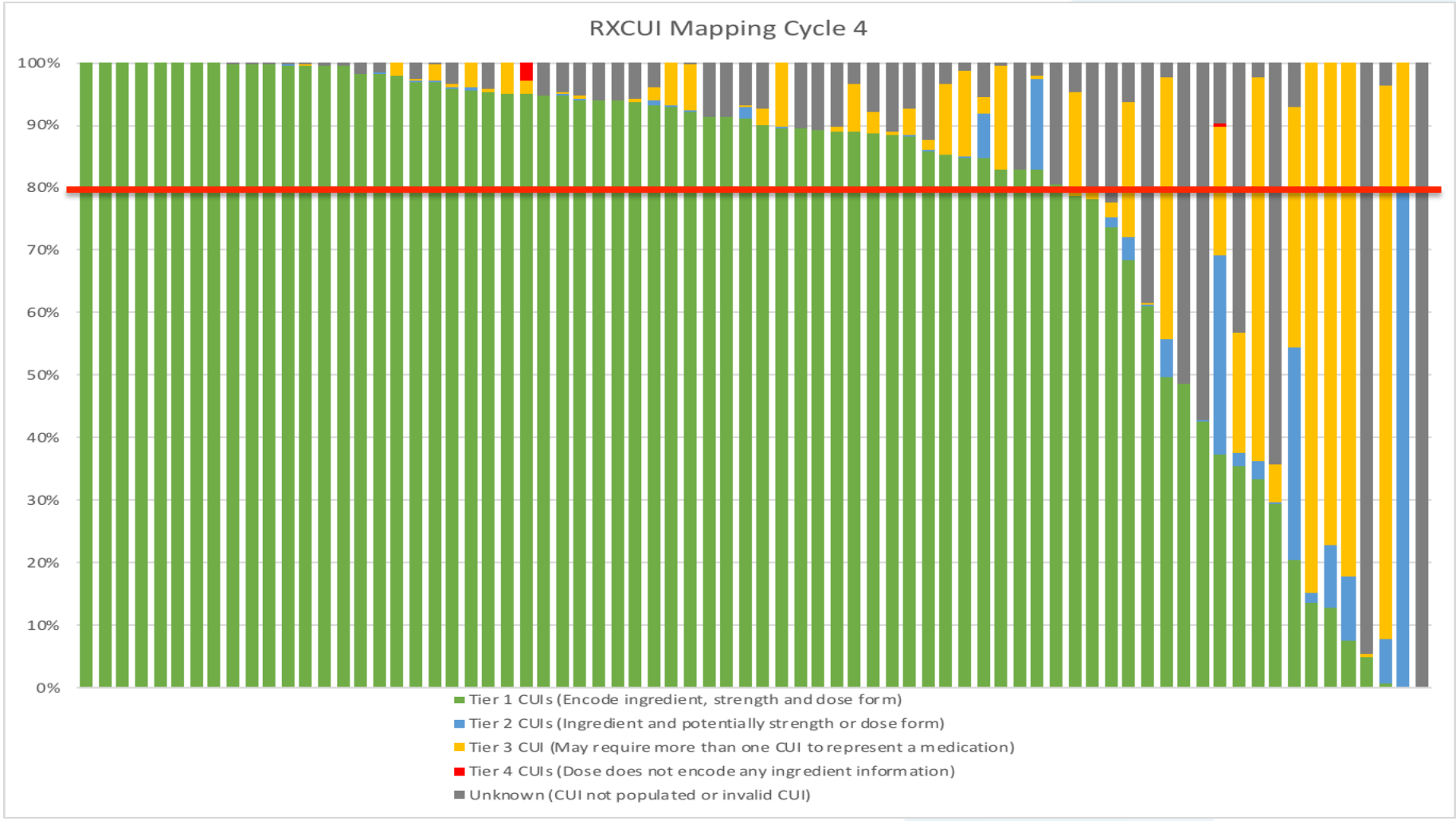
Antibiotics study example

- Study Aims: To evaluate the comparative effects of different types, timing, and amount of antibiotics prescribed during the first 2 years of life on:
  - Body mass index and risk of obesity at 5 and 10 years
  - Growth trajectories from infancy onwards
- Conducted study-specific data characterization to assess site eligibility / suitability of prescribing data to support study
- Sample findings
  - Days supply – highly missing
  - Start date minus end date – low percent missing – **very different from the global measure**
  - RxNorm – **variability in how partners mapped to RxNorm**
- Critical to overall success of the study

pcornet®

Study findings influencing data curation – medication coding

- Information about the medication ingredient, strength, and dose form is needed for many studies
- Implementation Guidance developed to establish the preferred mapping strategy
- Data Curation added a data check to measure adherence to the guidance



Study findings influencing data curation – data latency

- Knowing when to expect CDM data to be complete is essential for many study activities
- The ADAPTABLE* study team used data curation results to evaluate data latency and establish censoring dates
- Data curation added a data check to measure data latency and completeness

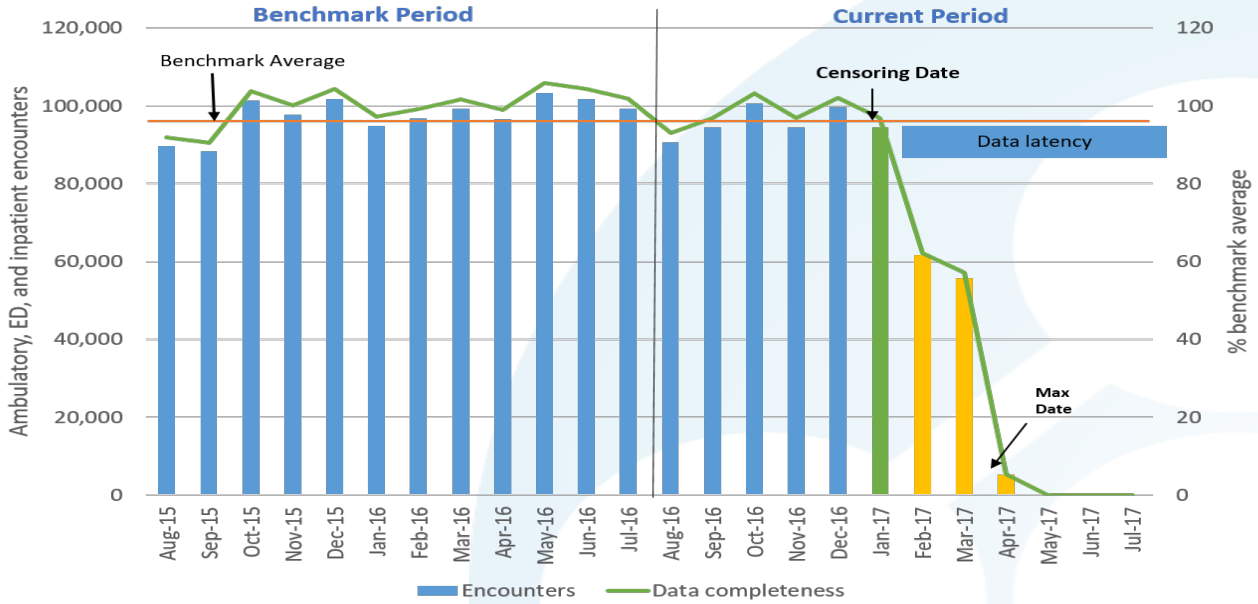


Table IVG. Data Latency and Completeness of Vital, Prescribing, and Lab Data, Past 2 Years

This table includes VITAL, PRESCRIBING, and LAB_RESULT_CM data from the most recent 24 month period; month -0 is the month the data curation query was run. Data completeness is determined by comparing the actual volume to the expected volume in each month. Expected volume is determined by taking the average volume during the benchmark period of months -12 to month -23. Data completeness is reported as a percentage of the benchmark average. Temporal differences may be affected by data availability, ETL processes, date shifting, secular trends, and/or changes in data provenance.
These data support Data Check 3.11 (vital, prescribing, or laboratory records are less than 75% complete three months prior to the current month). Data check exceptions occur if the month -3 result is <75% of the benchmark average or 0 records. Data check exceptions are highlighted in blue. Data check exceptions and unexpected results should be investigated and explained in the ETL ADD.

	Vitals		Prescriptions		Labs	
Month	Records	Percent of benchmark average	Records	Percent of benchmark average	Records	Percent of benchmark average
Month -0	60,980	9.8	16,015	13.0	82,977	13.0
Month -1	495,533	79.4	118,617	96.3	583,263	91.3
Month -2	560,362	89.7	121,318	98.5	604,813	94.7

*theaspirinstudy.org

Conclusions

- Support of the CDM and data curation requires multi-disciplinary teams at network partners & coordinating center
  - Database developers
  - EHR subject matter experts
  - Statistical analysts
- PCORnet is first network of this size to curate domains like laboratory results and medication orders
  - While data are messy, they are improving
  - Allow for more rapid study execution in the future

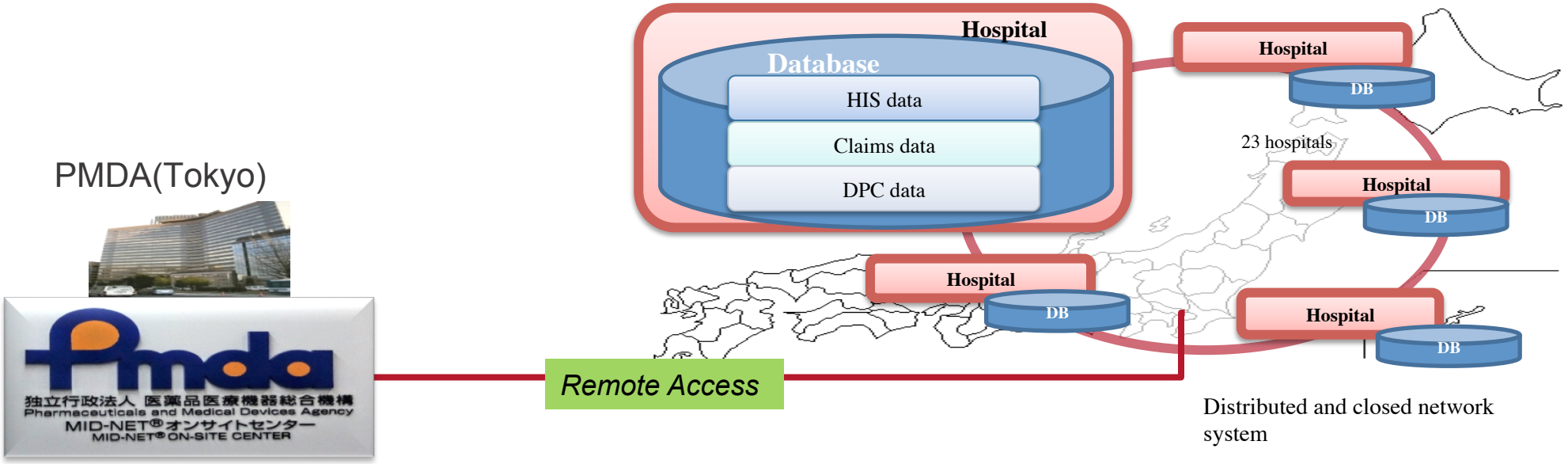
Data Quality Management of MID-NET®

Dr Yoshiaki Uyama
Director, Office of Medical Informatics and Epidemiology
Pharmaceuticals and Medical Devices Agency (PMDA)

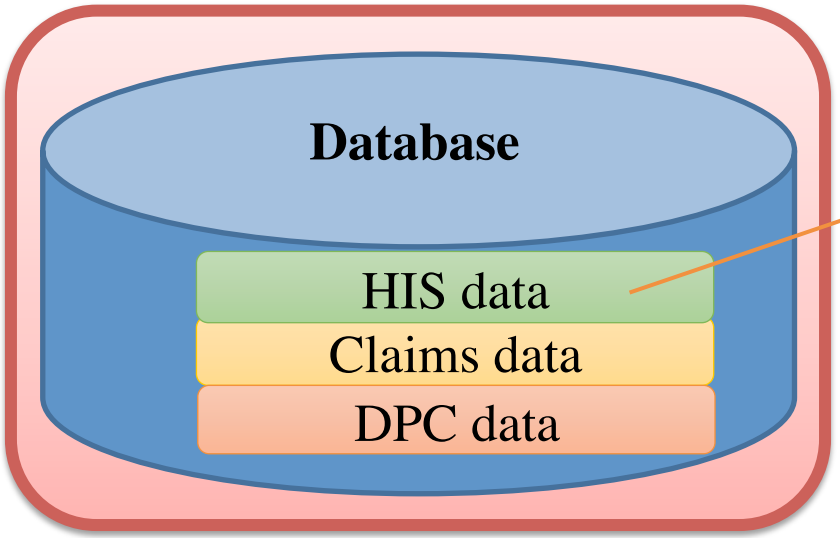
What is MID-NET® ?

MID-NET®
Medical Information Database Network

- The Medical Information Database Network in Japan for a real-time assessment of drug safety (currently >4M patients).
  - The project was started in 2011
- PMDA has led the project for establishing an integrated real time EMRs database with high quality



Common Data Model of the MID-NET®



Example of standard codes

Contents	Standard Code
Disease	ICD-10
Drug	YJ, HOT9 (JP specific codes)
Laboratory test	JLAC10 (JP specific codes)

HIS data

- Patient identifying data
- Medical examination history data

(including admission , discharge data)

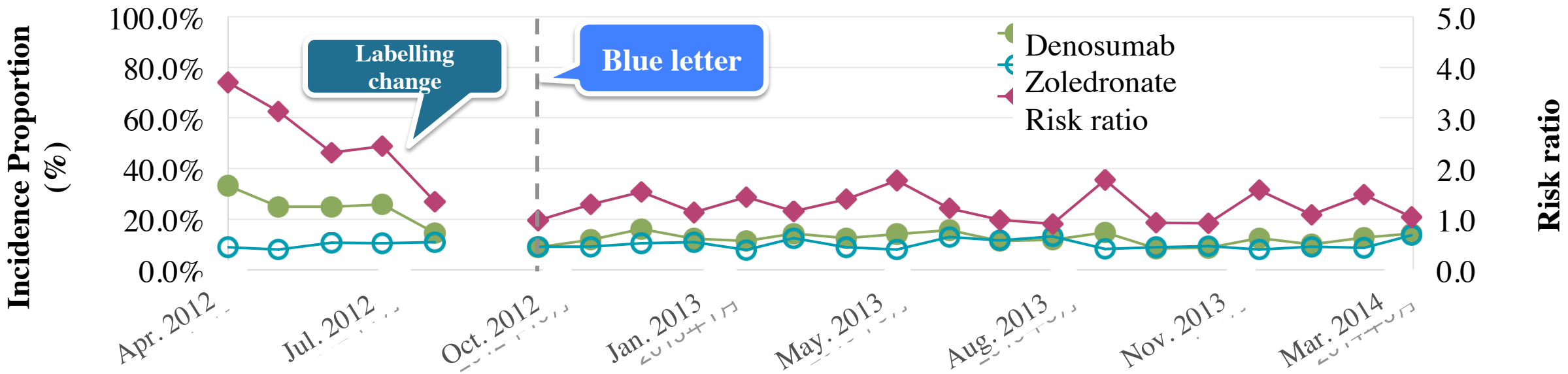
- Disease order data
- Discharge summary data
- Prescription order/compiled data
- Injection order/compiled data
- Laboratory test data
- Radiographic inspection data
- Physiological laboratory data
- Therapeutic drug monitoring data
- Bacteriological test data

Example: MID-NET® pilot
denosumab and severe hypocalcemia

Pilot study
Unpublished data

■**Objective**
✓To examine impacts of label change and warning letter in clinical practice for the risk of hypocalcemia associated with denosmab

Monthly transition of the incidence of
hypocalcemia (adjusted serum calcium conc.< 8.5mg/dL)



・Calculate the incidence of hypocalcemia during 28 days from a prescription date.
・Perform segment regression analysis based on the incidence of hypocalcemia / month.

Importance of Data Quality Management

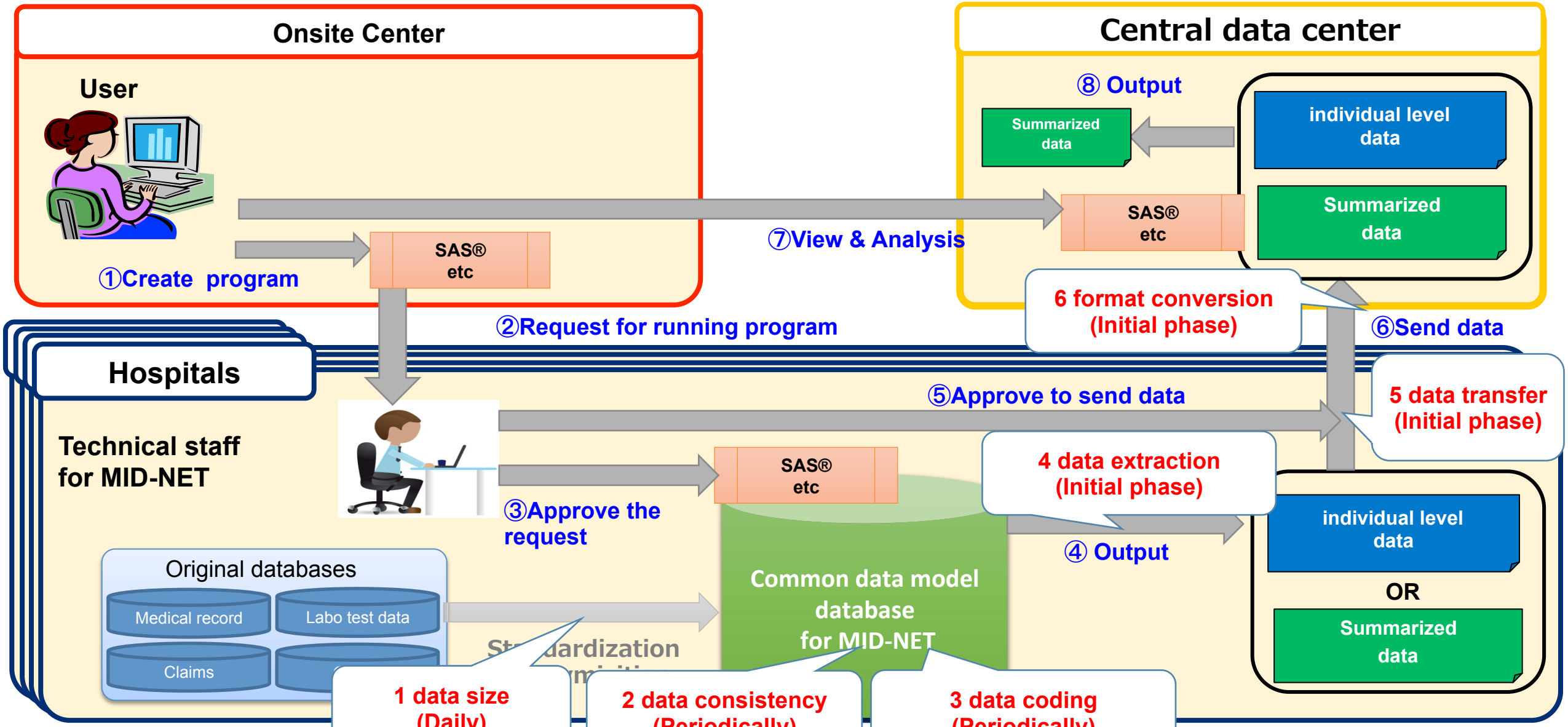
Reliable Data	X	Inappropriate analysis	=	Uninterpretable results
Unreliable Data	X	Appropriate analysis	=	Uninterpretable results
Reliable Data	X	Appropriate analysis	=	Interpretable results

High data quality as well as appropriate analysis are pre-requisite in utilizing real world data for providing scientifically interpretable results

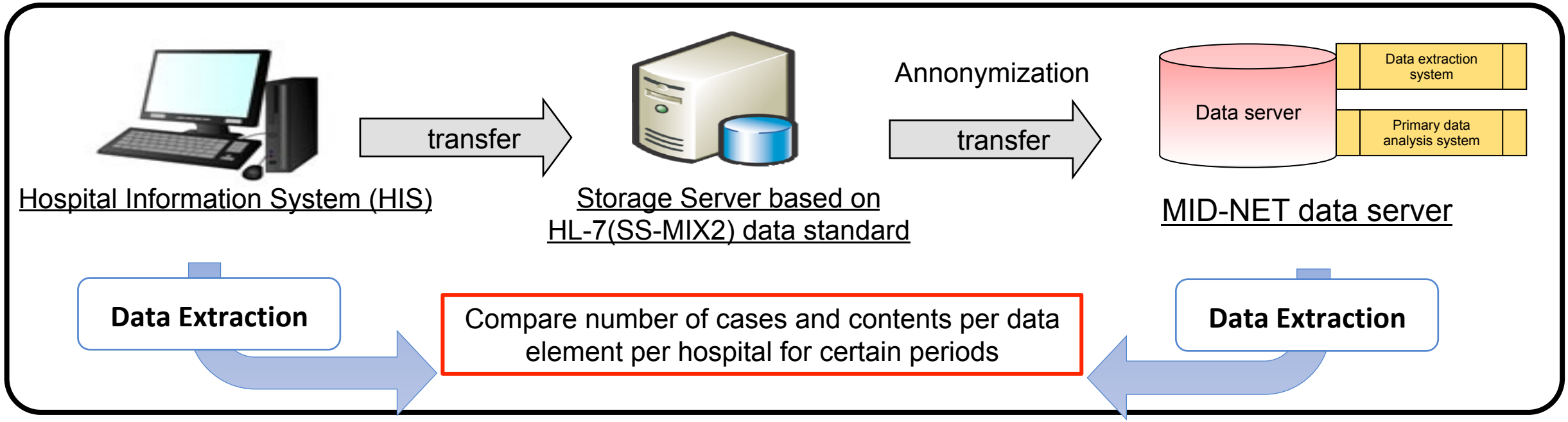
Data Quality Management of MID-NET®

- Daily management
  - Daily monitoring trends of data size sent to the MID-NET®
    - If marked changes are observed, necessary measures are taken
- Periodical management
  - Consistency check between the original data (Hospital data) and MID-NET® data
  - Updating data coding tables (standardized codes for diseases, products, lab. tests etc.)

Major points managed for data quality in the MID-NET®



Example: Data Consistency Check



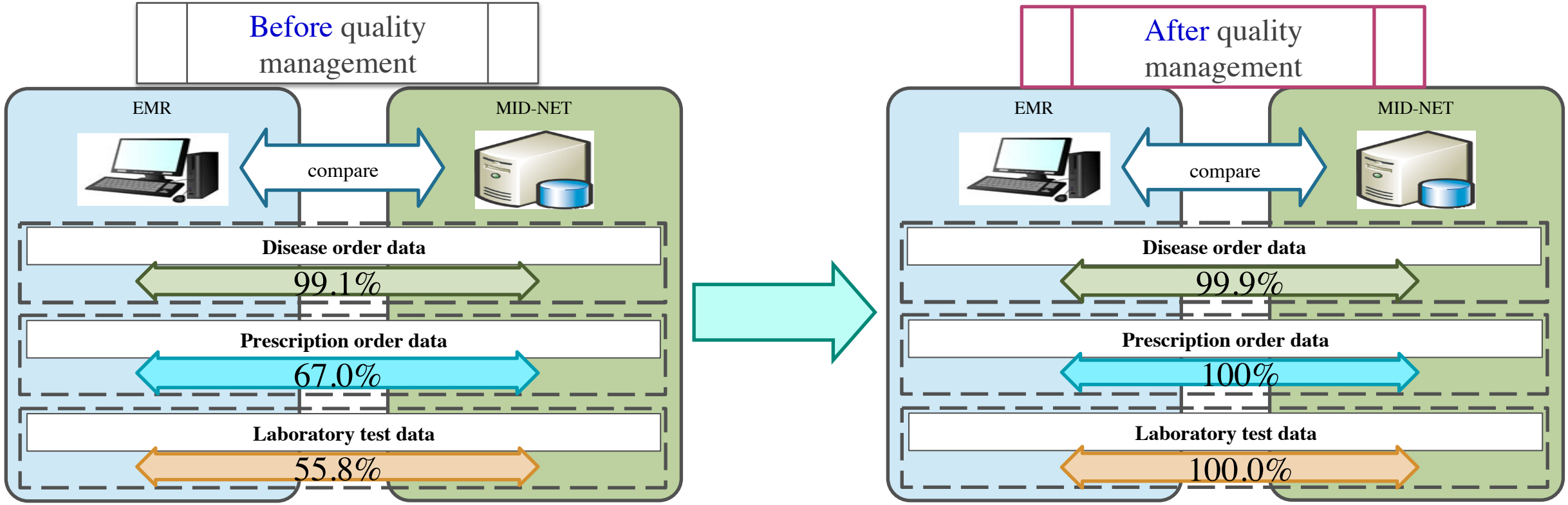
Examples of data inconsistency

- Lack of a unit
- Difference in a place of data storage among sites etc.
  e.g.; single dose, daily dose vs total dose

At the beginning, approximately hundreds of issues per site were identified for further investigation or consideration

MID-NET®: data consistency with the original data

PMDA has worked with cooperative hospitals for assuring data quality of MID-NET®.



Standardized data coding process
-Example; Laboratory test-

- Confirming appropriateness of a code for individual laboratory test by checking a distribution of laboratory test results (Approximately 200 tests)

Distribution of laboratory test results among hospitals

Confirmed

Further investigation were conducted in case of different distributions for understanding a reason and identifying an appropriate code

Examples of available laboratory test

ALT, AST, BUN, K, Creatinine, LDH, Gamma-GT, Cl, ALP, MCHC, MCH, Uric Acid, cGFR, TG, Cholesterol, Amylase, Blood Glucose, LDL-C, Inorganic Phosphate, HDL-C, PT-INR, HbA1c, PT, APTT, CEA, Fe, FT4, IgG, TSH, Sedimentation rate, RPR, IgM, HbA1c(NGSP), TPHA, AFP, Ferritin, Hb, Reticulocyte, Blood Gases(TCO_2), Blood Gases(pH),etc

Advantages and Limitation of MID-NET®

Advantages

- Various kinds of data including laboratory test results
- High data quality (daily and periodical check)
- Real-time data update (every 1-4 weeks)

Limitations

- May be not enough sample size (currently 4M)
- No linkage of a patient among hospitals
- Need to consider data generalizability due to limited cooperative organizations (mainly mid-large size hospitals like University hospitals)

Lessons Learned

- Points to establish a reliable and valuable database
  - Data quality management with routine monitoring
    - In addition to the daily monitoring, consistency between data stored in the database and original data (EMRs) should be checked and confirmed periodically
    - Data coding process should be standardized among all sites
  - Deep understanding regarding real situations in a site for sending data
    - Appropriate measures can only be taken with the deep understanding
  - Strong collaborations among all relevant organizations (hospitals, IT companies, academia, operating center, regulatory agency etc.)

MIHARI
Medical Information for Risk Assessment Initiative

MID-NET®
Medical Information Database Network

- **PMDA web site**
  http://www.pmda.go.jp/english/index.html
- **E-mail:**
  uyama-yoshiaki@pmda.go.jp

Thank you very much for your kind attention !!

CANADIAN NETWORK FOR OBSERVATIONAL
DRUG EFFECT STUDIES (CNODES)

Quality Assurance Processes in CNODES

Kristian B. Filion PhD FAHA

Assistant Professor and William Dawson Scholar
Departments of Medicine and of Epidemiology, Biostatistics, and Occupational Health, McGill University

Disclosures

- Salary support award from the *Fonds de recherche Québec – santé* (FRQS; Quebec Foundation for Health Research)
- William Dawson Scholar award from McGill University
- Research grants from Canadian Institutes of Health Research
- No conflicts 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

CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

*Nominated Principal Investigator

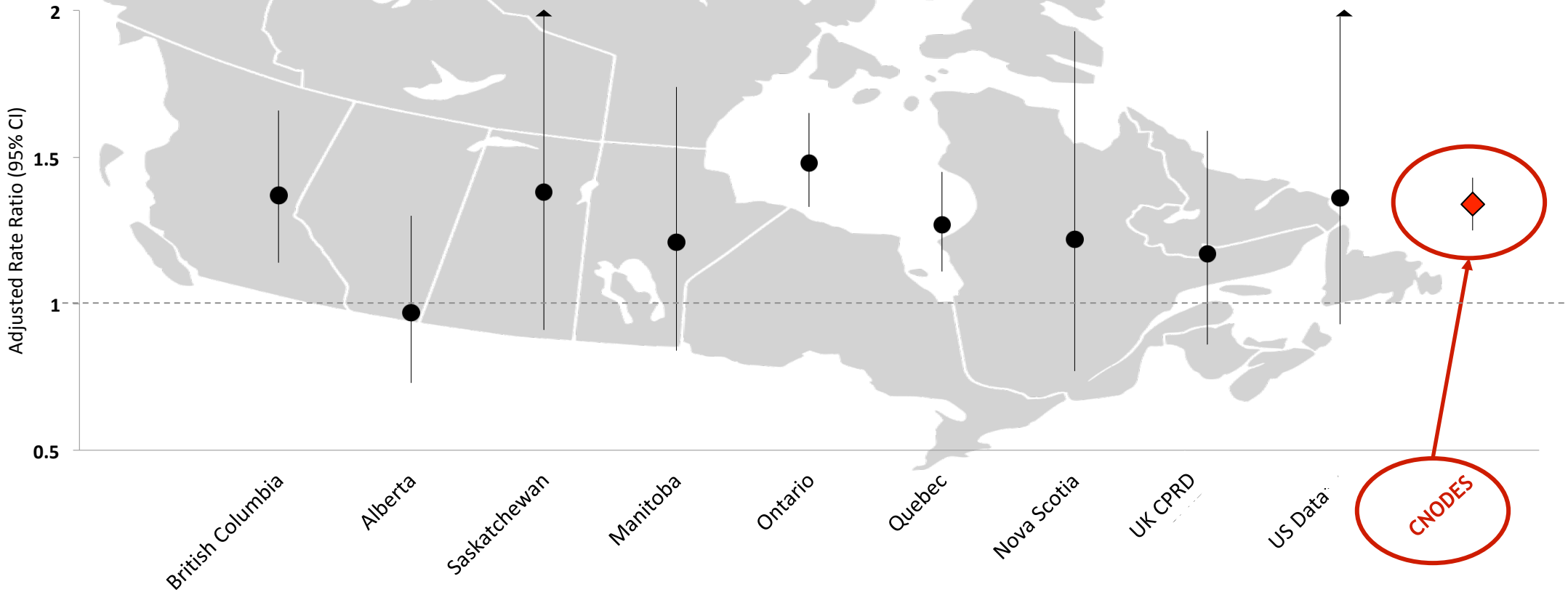
CNODES at a glance

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

Data sources

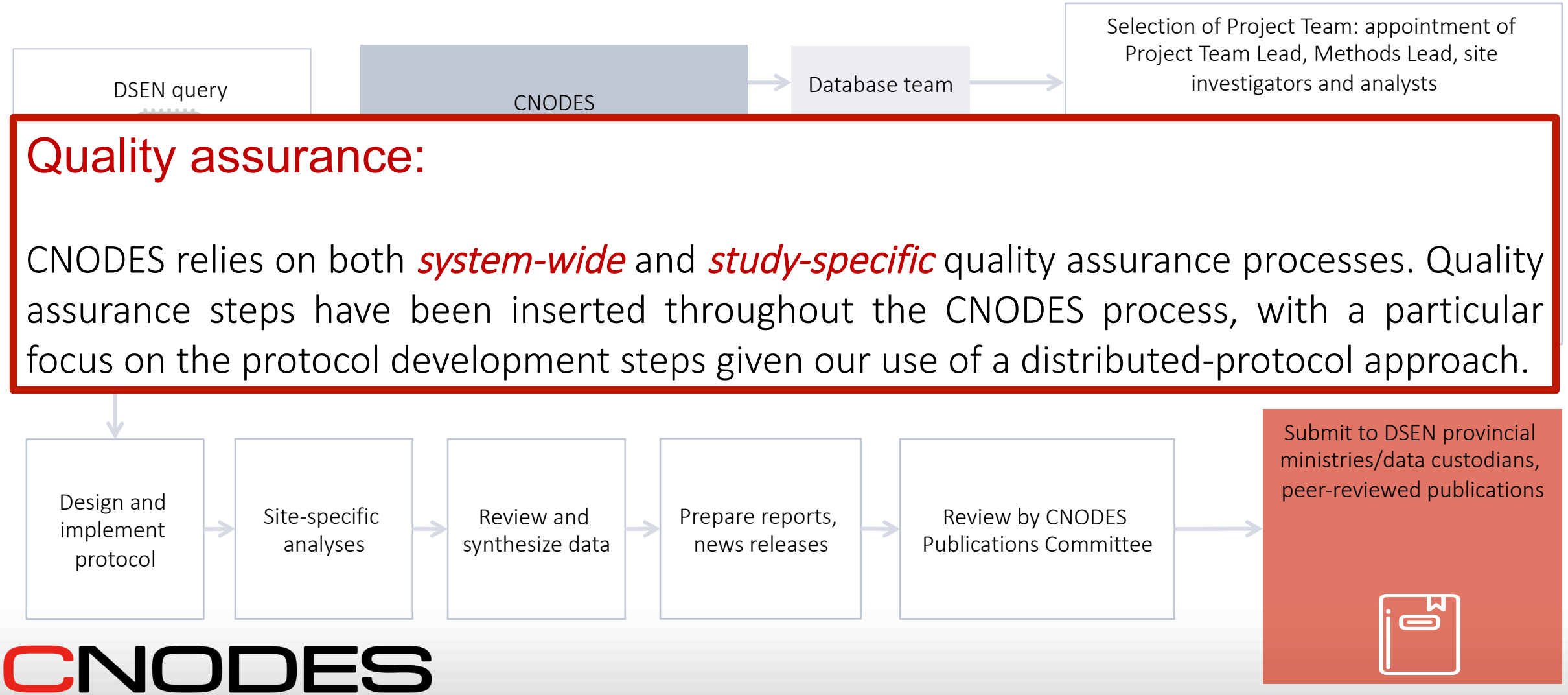
Data from across Canada

Example from a CNODES study examining the association between statin potency and acute kidney injury (Dormuth et al. 2013), using data from the provinces below and two international databases
(point estimate of relative risk with 95% confidence interval).



The CNODES process

From query submission to project completion and knowledge translation



CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

CNODES: Key steps in distributed-protocol approach

1. *Scientific Protocol*

Overview document describing study objectives, suitable for ethics review

2. *Statistical Analysis Plan (SAP)*

Detailed technical document describing the methodology for implementation

3. Phased implementation
   - *Phase I:* perform descriptive analyses, drug utilization
   - *Phase II:* detailed safety analyses and sensitivity analyses

CNODES policies and procedures

- Several policies and procedures have been developed to ensure that projects are carried out similarly by project team members across the country:

Policies and Tools	Description
Analyst Toolbox	Collection of coding and procedures for analysts
Project Guide	Describes in detail each step and role of a CNODES research project
Protocol Development Guide	Documents the process to standardize and facilitate the timely development of study protocols
Publications Policy	Describes the proper acknowledgement and attribution of authorship
Conflict of Interest Policy	Outlines practices to ensure that research is rigorous, transparent and free of undeclared conflicts of interest
Knowledge Translation (KT) Messaging	Details the process for developing KT and communicating with stakeholders

CNODES policies and procedures

Improve quality by minimizing bias and increasing reproducibility

- Registration of study protocols (transparency)
- Pre-specification of all variables and analyses
- Advanced study design and analytic methods (e.g., high-dimensional propensity score analysis, new user designs, highly restricted cohorts)
- Site-specific results deposited blind to those from other sites
- Independent review and synthesis of results

Case study #1

Gut

ORIGINAL ARTICLE

Proton pump inhibitors and the risk of hospitalisation for community-acquired pneumonia: replicated cohort studies with meta-analysis

Kristian B Filion,[1] Dan Chateau,[2] Laura E Targownik,[3] Andrea Gershon,[4] Madeleine Durand,[5] Hala Tamim,[6] Gary F Teare,[7] Pietro Ravani,[8] Pierre Ernst,[1] Colin R Dormuth,[9] the CNODES Investigators

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2013-304738).

For numbered affiliations see end of article.

Correspondence to
Dr Kristian B Filion,
Division of Clinical
Epidemiology, McGill

ABSTRACT

Objective Previous observational studies suggest that the use of proton pump inhibitors (PPIs) may increase the risk of hospitalisation for community-acquired pneumonia (HCAP). However, the potential presence of confounding and protopathic biases limits the conclusions that can be drawn from these studies. Our objective was, therefore, to examine the risk of HCAP with PPIs prescribed prophylactically in new users of non-steroidal anti-inflammatory drugs (NSAIDs).

Significance of this study

What is already known on this subject?

- Previous observational studies and their meta-analysis have found that proton pump inhibitors are associated with an increased risk of community-acquired pneumonia.
- Potential confounding by gastroesophageal

CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

Methods

7 databases

- Alberta, Manitoba, Ontario, Quebec, Nova Scotia, CPRD, MarketScan

Study population

- New users of non-steroidal anti-inflammatory drugs (NSAIDs)

Outcome:

- Hospitalization for community-acquired pneumonia

Exposure:

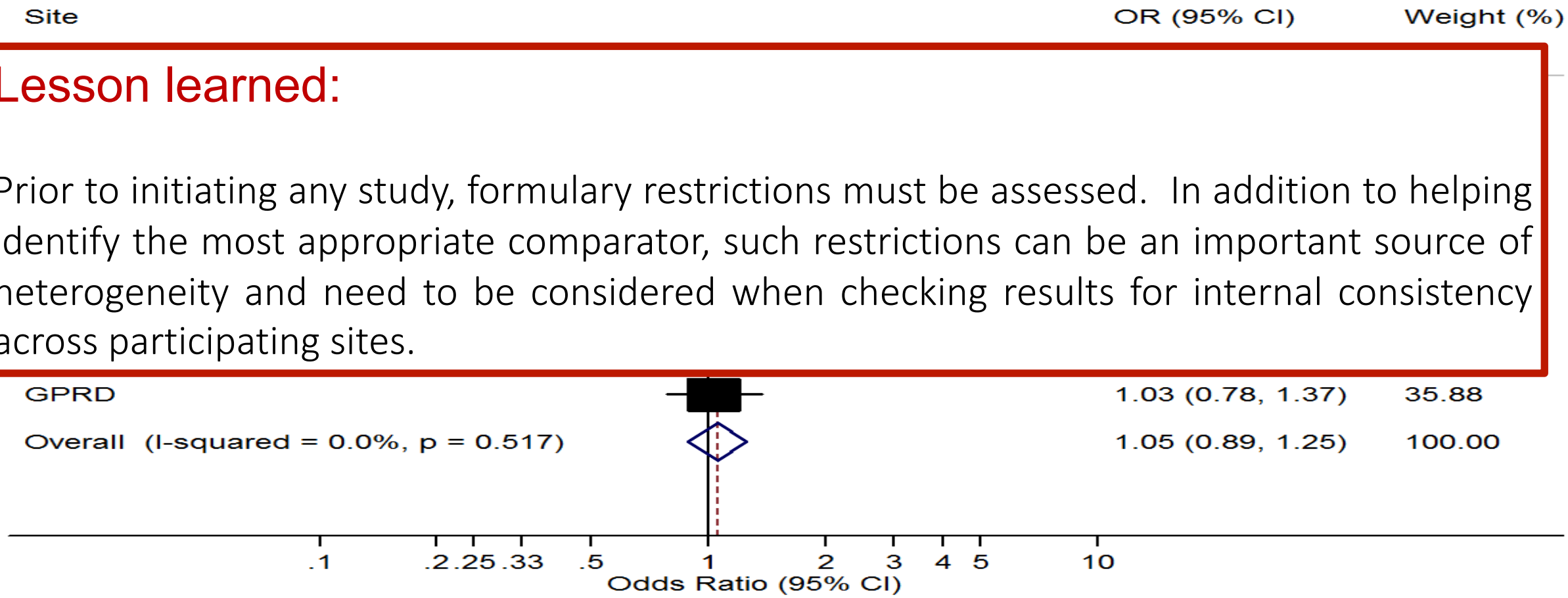
- New PPI on the same day as NSAID prescription vs no PPI

Statistical analysis

- Intention-to-treat analysis
- Follow-up = 6 months
- Logistic regression with high-dimensional propensity scores (HDPS)

Filion et al. Gut 2014.

PPIs and HCAP



Lesson learned:

Prior to initiating any study, formulary restrictions must be assessed. In addition to helping identify the most appropriate comparator, such restrictions can be an important source of heterogeneity and need to be considered when checking results for internal consistency across participating sites.

Filion et al. Gut 2014.

Case study #2

British Journal of Clinical Pharmacology

Br J Clin Pharmacol (2016) **82** 461–472 461

DRUG SAFETY

Ventricular tachyarrhythmia and sudden cardiac death with domperidone use in Parkinson's disease

Correspondence Dr Christel Renoux, Centre for Clinical Epidemiology, Lady Davis Research Institute, Jewish General Hospital, 3755 Cote Ste-Catherine, Montreal, Quebec H3T 1E2, Canada. Tel.: +1 (514) 340 - 8222 ext 4561; Fax: +1 (514) 340 - 7564; E-mail: christel.renoux@mcgill.ca

Received 27 January 2016; **revised** 15 March 2016; **accepted** 2 April 2016

Christel Renoux[1,2,3], Sophie Dell'Aniello[1], Paul Khairy[4], Connie Marras[5], Shawn Bugden[6], Tanvir Chowdhury Turin[7], Lucie Blais[8], Hala Tamim[9,10], Charity Evans[11], Russell Steele[1,12], Colin Dormuth[13], Pierre Ernst[1,14] and the Canadian Network for Observational Drug Effect Studies (CNODES) investigators*

CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

Quality assurance

- Nested case-control study: 7 Canadian provinces and CPRD
- Important heterogeneity identified:
  - Incidence rates of VT/SCD ranged from 19.8 (BC) to 53.4 (Quebec) per 10,000 person-

Lesson learned:

Local variability in coding and its precision needs to be considered when developing study protocols and interpreting study results.

Identifying sources of database heterogeneity and testing their impact on study findings through empirical and simulation studies can strengthen the design and analysis of network data.

- Quebec: rarely recorded secondary discharge diagnoses, contributed to higher rate of SCD

Doyle CM (Submitted).

Case study #3

BMJ

BMJ 2014;348:g3244 doi: 10.1136/bmj.g3244 (Published 29 May 2014)

RESEARCH

Higher potency statins and the risk of new diabetes: multicentre, observational study of administrative databases

Colin R Dormuth *assistant professor*[1], Kristian B Filion *assistant professor*[2], J Michael Paterson *scientist*[3], Matthew T James *assistant professor*[4], Gary F Teare *director of measurement and analysis*[5], Colette B Raymond *research scientist*[6], Elham Rahme *associate professor*[7], Hala Tamim *associate professor*[8], Lorraine Lipscombe *adjunct scientist*[3], for the Canadian Network for Observational Drug Effect Studies (CNODES) Investigators

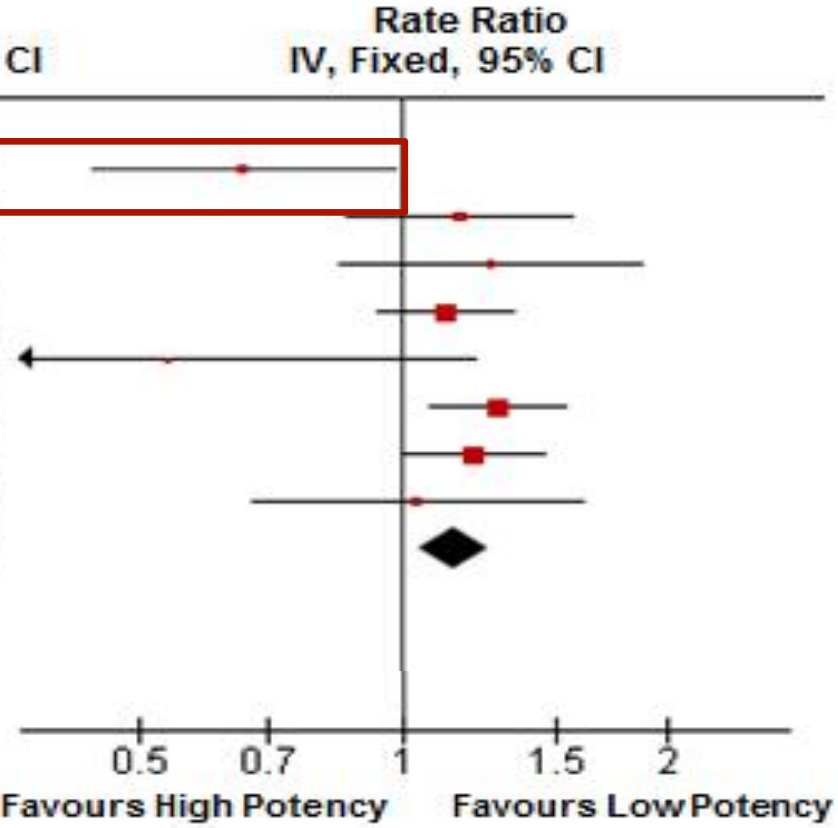
CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

High vs low potency statin and new diabetes

Subgroup	Lower Potency Cases	Lower Potency Controls	Higher Potency Cases	Higher Potency Controls	Weight	Rate Ratio IV, Fixed, 95% CI
1.25.1 Diabetes Within 2 Years of Therapy						
Alberta	68	531	90	944	5.2%	0.66 [0.44, 0.98]
CPRD	103	1,064	247	2,266	9.2%	1.17 [0.87, 1.57]
Manitoba	47	447	170	1,514	5.2%	1.27 [0.85, 1.88]
Marketscan	180	1,853	502	4,652	25.3%	1.12 [0.94, 1.34]
Nova Scotia	18	125	23	216	1.3%	0.54 [0.24, 1.21]
Ontario	236	2,658	675	6,196	26.5%	1.29 [1.08, 1.53]
Quebec	260	2,775	507	4,681	23.1%	1.21 [1.00, 1.46]
Saskatchewan	42	378	188	1,585	4.3%	1.04 [0.67, 1.61]
Subtotal	954	9,831	2,402	22,054	**100.0%**	**1.15 [1.05, 1.26]**

Heterogeneity: $Chi^2 = 13.32$, $df = 7$ ($P = 0.06$); $I^2 = 47\%$

Test for overall effect: $Z = 3.00$ ($P = 0.003$)



CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

Dormuth et al. BMJ 2014.

Quality assurance

- Following Steering Committee review:
  - SAS programs were verified locally by two analysts

Lesson learned:

The heterogeneity observed in this study is consistent with other studies that have shown that unexpected findings can sometimes be explained by differences in data structure or capture, confounding due to different local conditions, and and/or chance. This highlights the importance of replication, a key strength of CNODES.

remained.

Conclusions

- CNODES has adapted *system-wide* quality assurance processes as well as *study-specific* quality assurance procedures.
- With our use of a *distributed protocol* approach, much of our attention has focused *on protocol development* and *internal consistency* across sites, while using external information where possible.
- A key issue is the need for *local expertise*; our approach ensures that the individuals who know the data source best are those applying the protocol to it.
- Ultimately, quality assurance is the responsibility of the *entire research team*.

Thank you

Visit us at www.cnodes.ca

CNODES

CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

kristian.filion@mcgill.ca

CIHR IRSC