Distributed Data Networks: Lessons from the HMO Research Network and Other National Health Plans

Distributed Data Networks for Active Medical Product Surveillance: Expert Think Tank

Jeffrey Brown, PhD

November 23, 2009

Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute

Our experience has taught us…

- Little or no exchange of person-level data is needed to answer many safety, effectiveness, and quality questions
- A relatively small subset of items in electronic health data systems can answer most safety, effectiveness, and quality questions
- Data holders do not like having their data outside their control
- Concentrate analytics in a single team
- Minimize impact on health plan operations

Implication of those experiences

- Data holders maintain physical possession of their person level data
- Data holders control the uses of their data
- Computer programs should run at multiple sites without modification

Data Sources to Meet Needs

- Enrollment: dates and type of coverage
- Demographics
- Claims – inpatient, outpatient
- Pharmacy dispensing
- Electronic medical records
- Access to full text inpatient and outpatient records
- Linkage to selected external registries, e.g., birth, death certificates, immunization

HMO Research Network (HMORN)

- 15 health plans across the US and Israel
- Different delivery systems
  - Insurers, medical groups, integrated delivery systems
- Different data systems
  - Claims, EMR, labs, upgrade schedules
- Different corporate SOPs, IRBs, beliefs
- Collaborate to conduct public health research

Data Standardization: Virtual Data Warehouse

- Created a *common data model* covering 10 priority data areas
- Uses existing coding standards (ICD-9, NDC, HCPCS)
- Relevant items transformed to common data model for entire population (Extract-Transform-Load [ETL])
  - Stored as SAS datasets
- Data remain at sites; no centralized data
- Checking of data quality and completeness via distributed programs
  - Within- and across-sites

Data Completeness

- **Enrollment data allows identification of defined populations at risk during specific periods**
  - Able to identify inception cohorts
- **Exposure granularity**
  - All care settings: NDCs, HCPCS, and ICD-9 procedure codes
  - Within EMR: vital signs, test results, lot numbers, etc.
- **Claims allow ascertainment of all care, regardless of setting**
  - **The absence of a claim implies no event occurred**
- **Ability to review medical charts**

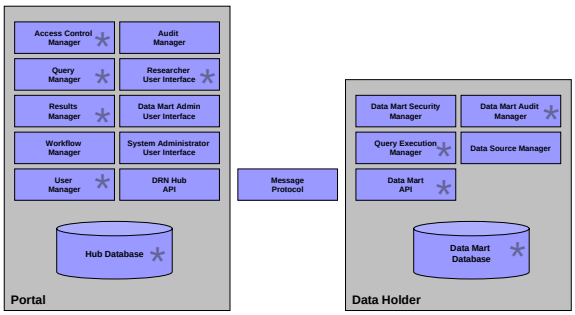
Data Areas and Plans

- Enrollment
- Demographic
- Utilization
- Pharmacy
- Census
- Cancer registry
- Death date/cause
- Vital signs
- Laboratory results
- Provider characteristics
- Planned:
  - Infusion
  - Prescribing
  - Radiology findings
  - Pathology findings
  - Inpatient details
  - NLP clinical note extraction

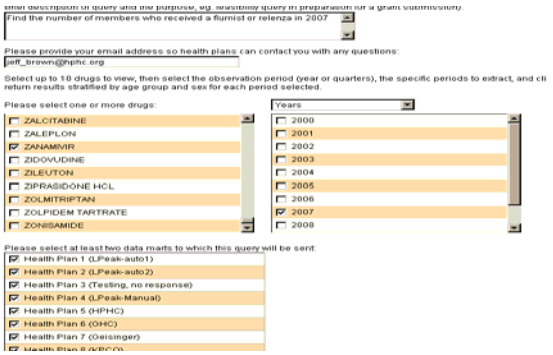
Distributing Queries & Returns: Current

- Queries (as SAS programs) are distributed via e-mail or shared portal (e.g., collaboration website)
- Results returned as appropriate for data (secure e-mail, secure FTP, certified mail)
  - Results typically contain no Protected Health Information

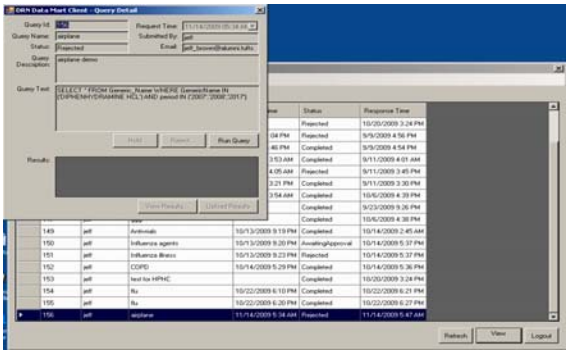
Distributing Queries & Returns: In development



HMORN DRN Portal Screenshot



HMORN DRN Desktop QEM Screenshot



Distribution: Minimal Automation Preferred Initial Approach

- Practical approach with our health plans' social, regulatory, and business environment
  - Lowers barriers to acceptance and implementation
  - Small IT footprint and limited risk
  - Focus on things we do well: data manipulation
  - Minimize need for extensive database expertise and ongoing maintenance/management of complex data structures
- Design allows automation of any step via role based access control
  - Ex.1: Require manual execution if submitted by a, b, or c
  - Ex. 2: Allow automated execution of all queries from x, y, and z
    - Unless topic is mental health

Query Execution Responsibilities

- Currently, HMORN plans may distribute programs to each other directly or via coordinating center
- Under planned architecture, queries distributed by an authorized user via portal
- Sites always control execution
  - Potential for fine grained control of automation

Security and Privacy

- Plans have complete control over all uses of their data
- Plans ensure their own data sharing procedures are followed
- Plans approve all data transmissions
- Portal will use standard web security

Scalability

- **Data sources**
  - Common data model can be extended (e.g., genetic data)
  - For very large data sets, only subsets need be converted to common model
- **Planned Network infrastructure**
  - Small IT footprint at sites
  - Most software and development resources focused on the portal
    - Authorizations, access controls, query management, etc
  - A limited number of authorized users – hundreds, not tens of thousands

Key challenges

- Common data model development and enhancement
  - Ensuring adherence to definitions during transformation
  - Ensuring data quality
- Maintaining a practical and pragmatic approach versus the waiting for the perfect solution
- Governance, governance, governance
  - Prioritization of limited resources
  - Access rights
  - Etc

Next steps

- Continue development and roll-out of new architecture
- Extend model to new partners

Distributed Data Networks: Lessons from the HMO Research Network and Other National Health Plans

Distributed Data Networks for Active Medical Product Surveillance: Expert Think Tank

Jeffrey Brown, PhD

November 23, 2009

Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute

Overview of the Vaccine Safety Datalink (VSD) and its Distributed Data Model

November 23, 2009

James Baggs, PhD
for the Vaccine Safety Datalink Project
Immunization Safety Office
Centers for Disease Control and Prevention

Vaccine Safety Datalink (VSD): Background

- Established in 1990, the VSD is a collaborative project among CDC and 8 managed care organizations (MCOs)
- Collects medical care and vaccination data on more than 9.5 million members annually (3.1% of the US population)
- Allows for planned immunization safety studies as well as timely investigations arising from
  - hypotheses from medical literature and pre-licensure
  - reports to the Vaccine Adverse Event Reporting System (VAERS)
  - changes in immunization schedules, or the introduction of new vaccines
- Since 2006, conducts routine Rapid Cycle Analysis (RCA) of newly licensed and approved vaccines or modifications to existing vaccine recommendations

VSD Sites: 2009

VSD: Strategic Priorities

1. Evaluate the safety of newly licensed vaccines
2. Evaluate the safety of new vaccine recommendations for existing vaccines
3. Evaluate clinical disorders following immunizations
4. Assess vaccine safety in special high risk populations
5. Develop and evaluate methodologies for vaccine safety assessment

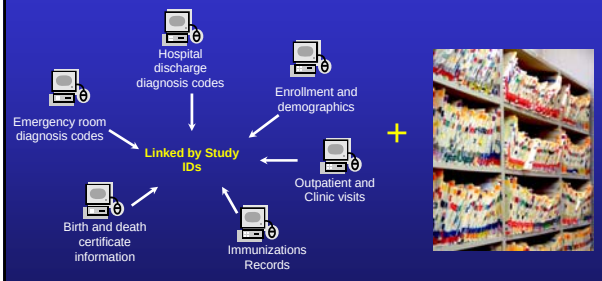
VSD Data

- Automated computerized data derived from electronic data sources and data warehouses at MCOs as well as additional outside data sources
- VSD data:
  - Compiled annually to create “cycle files”
- Cycle files organized by a standardized data dictionary
  - Contains data on:
    - Demographics
    - Health plan enrollment
    - Vaccinations (including lot #, manufacturer, location)
    - Hospitalizations and emergency room visits
    - Outpatient clinic visits and urgent care visits
    - Procedure codes
    - Mortality data
    - Additional birth information (e.g. birth weight) when available
- Stored as SAS datasets

VSD Studies

- Computerized data has limitations
- In addition to computerized data, VSD studies often employ additional data sources
  - Medical chart review
  - Survey
  - Additional computerized data sources such as pharmacy data, laboratory data, or radiology data

VSD Annual Cycle Files + Chart Review



The Dynamic Data Files

- To meet the changing needs of the VSD:
  - Restructuring annual cycle files was undertaken in 2005
  - We enhanced the infrastructure to capture near real time VSD event data:
    - vaccinations, hospitalizations, emergency room visits, outpatient and clinic visits, MCO enrollment data, and certain demographic data
- The newly developed files are now referred to as the VSD “Dynamic Data Files (DDF)”

The Dynamic Data Files

- Each MCO captures event based VSD data in near real time
- The annual cycle data files continue to be created
- DDF data follows VSD data dictionary
- For studies using the DDF, data are accessed on a weekly basis by CDC for analysis and/or extraction of necessary data

How VSD uses the DDFs

- VSD Studies
- Monitoring uptake of vaccines
- Rapid Cycle Analysis Studies began in early 2006:
  - Meningococcal Conjugate (Menactra®)
  - Rotavirus (Rotateq® and Rotarix®)
  - MMRV (Proquad®)
  - Tdap (ADACEL® and BOOSTRIX®)
  - HPV (Gardasil®)
  - Seasonal Influenza Vaccines
  - H1N1 Influenza Vaccines
  - DTaP-IPV (Kinrix ®)
  - DTaP-IPV/Hib (Pentacel ®)

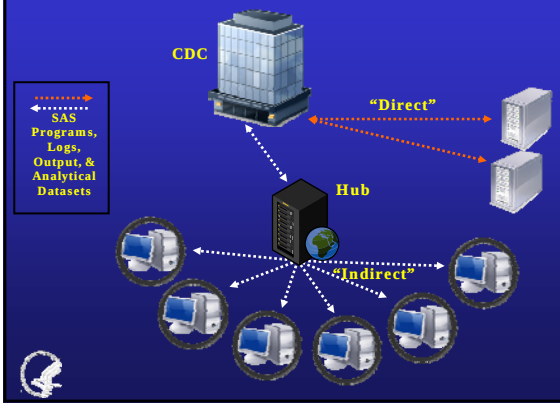
The VSD Uses a Distributed Data Model (DDM)

The Distributed Data Model (DDM) is a system that allows all individual level standardized VSD data (Cycle Files) to reside at the MCO, rather than be transferred to the CDC

The DDM:

- Maintains confidentiality and ownership of VSD sites data
- Utilizes encrypted and secure methods (SSH2/SAS Secure)
- Limits access to IRB approved data required for specific studies
- Allows for simultaneous multi-site processing
- All SAS programs are submitted through the DDM by CDC/VSD data analysts
- Timely research is possible because of rapid turn-around of submitted SAS programs

The VSD Distributed Data Model



The VSD Research Process

- Concept Sheet
  - Feasibility Assessment
- Proposal
  - Development and revision
  - Formal review by VSD project
  - Approval by CDC VSD team leader
- Request MCO participation
- IRB approvals and HIPAA data use agreements
- Data collection and statistical analysis
- Manuscript preparation and review by VSD project
- CDC and Site Clearance
- Dissemination of results
- Archival for data sharing program

VSD DM Processes

- Data Management workgroup
  - Meets 1-2 times per month via conference call
  - 1-2 meetings per year
  - Input gathered from investigators and project managers
  - Workgroup makes decisions regarding standardized data dictionary, DDM changes, other data issues
- Site data managers approve SAS programs for all IRB approved studies
- Sites monitor DDM activity
- All programs submitted via CDC analysts
- Specific coding guidelines

Key Challenges and Next Steps

- Always updating the DDM
  - Upload
  - Additional standardized macros
- Continual updates to the standardized data dictionary
  - Height/Weight data
  - Temperature
- Greater efficiency in SAS program approval process
  - Using the VSD website
- Managing increasing amounts of data
- Developing personnel infrastructure
  - DDM programmers

VSD At A Glance

- Has published over 100 articles
- Is currently conducting approximately 70 studies
- Over 19 million individuals included in VSD data files
- Over 85.5 million vaccine doses
- DDM fully operation since early 2004
- Since 2006, an average of 35 “jobs” per month per site submitted to the DDM
- Is in the process or has completed 10 active surveillance studies of new licensed vaccine in a post market setting
  - Several presentations to the Advisory Committee on Immunization Practices (ACIP)

The Summary of the VSD

- The VSD project uses a Distributed Data Model (DDM)
  - The DDM is an innovative model that ideally meets the needs of the VSD to conduct secure multi site studies while maintaining patient confidentiality
- Cycle Files (core data files) are still created on an annual basis
- Parallel datasets called the Dynamic Data Files (DDF) have been created and are updated weekly
- Along with retrospective studies, VSD now conducts near real time surveillance of potential vaccine associated adverse events (RCA)

VSD Investigators and Collaborators - Partial List

Centers For Disease Control, VSD team
- James Baggs, PhD
- Julianne Gee, MPH
- Natalie McCarthy, MPH
- Eric Weintraub, MPH

Kaiser Permanente of No. California (NCK), Oakland CA
- Roger Baxter, MD
- Nicky Klein, MD, PhD
- Ned Lewis

Northwest Kaiser Permanente (NWK), Portland OR
- Allison Naleway, PhD
- John Mullooly, PhD
- Karen Riedlinger
- Lois Drew

Harvard Pilgrim /Harv. Vanguard (HAR) Boston, MA
- Tracy Lieu, MD, MPH
- Richard Platt, MD, MSc
- Katherine Yih, PhD, MPH
- Richard Fox
- Grace Lee, MD, MPH
- Sharon Greene, MD, MPH

Group Health Cooperative (GHC), Seattle WA
- Lisa Jackson, MD, MPH
- Jennifer Nelson, PhD
- Lora Bounds

Marshfield Clinic Rsch. Foundation (MFC) Marshfield WI
- Edward Belongia, MD
- James Donahue, MD
- Nick Berger

Health Partners Rsch Foundation (HPM) Minneapolis MN
- Jim Nordin, MD
- Amy Butani

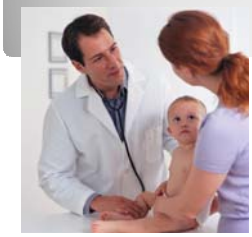
Kaiser Permanente of Colorado (KPC) Denver, CO
- Simon Hambidge, MD, PhD
- Jason Glanz, MS, PhD
- David McClure, PhD
- Matt Daley, MD
- David Ryerson

So. California Kaiser Permanente (SCK), CA Los Angeles, CA
- Steven Jacobson, MD, PhD
- Wansu Chen, MS
- Sungching Glenn, MS

Sites include > 125 staff working on VSD

Thank You!

DARTNet:
A New Model for Translational, Effectiveness and Safety Research



Wilson D. Pace, MD, FAAFP

Goals of this introduction

1. National context
2. Overview of DARTNet
3. Clinical context
4. Safety, CER and QI
5. Limitations
6. Next steps

National context

- Health care value assessment critical
- Improved methods for CER and safety monitoring of therapeutic activities
- New models of data acquisition
- Systems that combine claims data with clinical data and data directly collected from patients will add value
- Systems that can drive quality while providing new approaches to CER will be acceptable to clinicians

What is DARTNet?

- Public/private partnership - 4 primary partners (Growing # academic partners)
  - University of Colorado Department of Family Medicine and School of Pharmacy
  - American Academy of Family Physicians
    - National Research Network
    - The Robert Graham Policy Center
  - Clinical Integration Networks of America (CINA)
  - The University of Utah's Center for High Performance Computing

What is DARTNet?

- DARTNet is a federated network of electronic health record data from multiple organizations
  - Supports bi-directional electronic communication with these practices/ providers and patients
  - Facilitates data collection/ aggregation using multiple constructs
    - Point of care from office staff/providers/patients
    - Ancillary data to the PCMH – fulfillment data, claims data, patient entered data

DARTNet's Mission

- DARTNet's Mission is to explore how currently available EHR data can be used to supplement data from other datasets and sources to answer questions concerning the safety and effectiveness of medications and medical devices while improving the quality of the care provided by member organizations

DARTNet's Aims

- **Support the Concept of the Patient Centered Medical Home**
- **Enhance the State of the Art in Effectiveness**
- **Research**
- **Advance practice-based research capabilities**
- **Enhance HIT capabilities within ambulatory care**

How does DARTNet work?

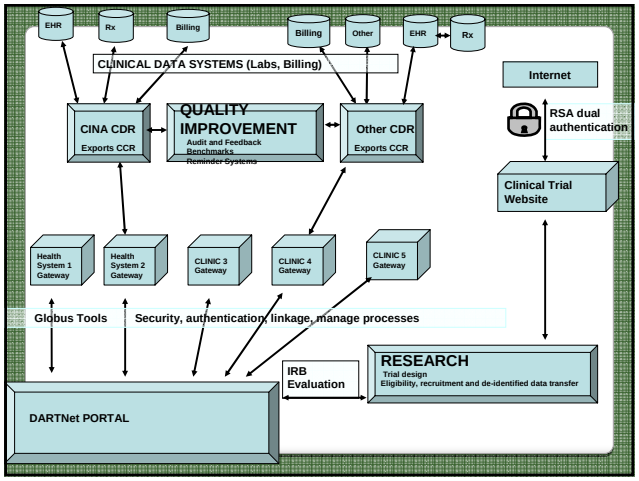
- **Step 1 -** Capture, code, & standardize data (ETL)
  - **Step 2 –** Database for query/ research secure Web-portal
  - **Step 3 –** Knowledge generated to inform and fuel clinical quality improvement

Data management overview

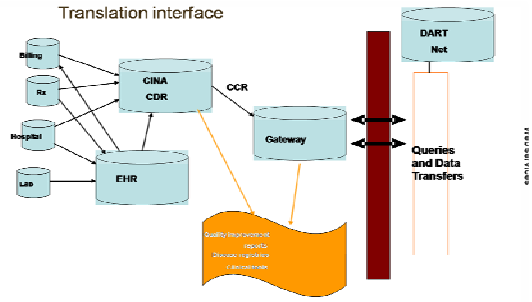
- Data stays locally
- Standardized locally with retention of original format for both:
  - oQuality checks
  - oRecoding in future
- Each organization retains control of patient level data
- Local processing allows expansion

Technical overview

- True distributed database
- EHR independent
- Data standardization middle layer tied to clinical decision support
- Distributed queries using Globus tools
- Exploring alternative data collection approaches
- Exploring multiple data sources



Single practice perspective



Security

- OGSA-DAI to Gateway connection
  - IP to IP specific
  - DNS registry reverse hand shake
- Three factor security at login
- User functions limited by role
- Query functions limited by type
  - Aggregate will return to system
  - Patient level requires local activation

Data Standardization Model

- Data element – Glucose measurement
  - Original data element tag
    - Glucose, fast
    - Glucose, rnd
  - Data concept - Blood glucose measurement - 33747003
  - Specific data elements
    - Fasting glucose measurement – 52302001
    - Random glucose measurement - 73128004

Data Standardization

- Original data location and label retained as well as value where applicable
- Each data element then mapped to a data concept and specific data label
- SNOMED-CT used for labs, history, procedures, allergies, vital signs
- RxNorm used for drugs with NDC retained
- Fulfillment data – can batch be attained?
- ICD-9 CM used for diagnoses

Point of Care Data

- Algorithm driven
- Current models Dx driven
- Could be drug driven, lab driven, reason for visit driven
- Pilot identified order of magnitude greater number of hypoglycemic events
- Many associated with OTC supplement use
- CA-MRSA study tracking clinical decision making from EHR triggers

Multi-faceted Research

- Data mining for traditional OCER
  - Drug fulfillment data critical
- Enhanced clinical data for OCER
- Information to inform studies
  - Eligibility criteria
  - Incidence and prevalence data
- Best practices research
- PBRN interventional trials

Learning Community

- Learning Community Activities
  - Benchmarking reports
  - Practice facilitation
  - Linkages (self-initiated and facilitated)
    - Website, Listserv, E-newsletter
  - Webinars
    - Best practices, case studies, how-to-workshops
  - Periodic face-to-face conference

Future Challenges

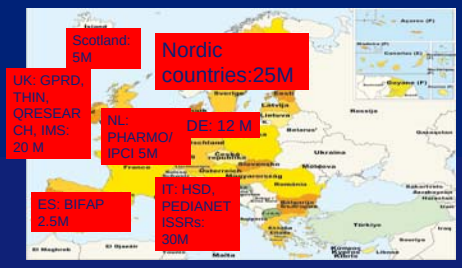
- Grow to reasonable size
- Find adequate infrastructure to support learning community
- Integrate various approaches to POC data collection
- Improve claims and fulfillment data
- Include the patient's voice

Erasmus MC
University Medical Center Rotterdam

A European perspective to distributed networks for drug and vaccine safety, EU-ADR, SOS and VAESCO II as examples



Healthcare databases in the EU



Total EU population: 456 M
databases on more than 100 M available

Erasmus MC

Legal basis for combining data in EU

- Directive 95/46/EC regulates the processing of personal data and the free movement of personal data (including health care) -> implemented in all countries.
- Principle: personal data may not be processed
  - Scientific purposes are an exception
  - However transparency is required (except when this is impossible)
    - Use of coded data in large databases is possible
- Each country may have different implementation of directive
  - Needs to be explored
  - Processing rules depend on country where the data are (also after they have been sent across borders)
- Each database has own ethical framework and procedures for processing data, these need to be satisfied as well

Erasmus MC

Working models for combining data

1. Combination of source databases
2. Combination of raw data cuts
3. Combination of person-level anonymous query results
4. Combination of aggregated data (multiple persons)
5. Common protocol: combination of model coefficients
6. Meta-analysis of different studies

Experience in EC Framework programme

- FP-6 TEDDY: local elaboration of data, with own statistical expertise according to common protocol
  CUMBERSOME
- FP 7 approaches: Distributed data networks
  Started with EU-ADR
  Now implemented in EU-ADR, SOS, ARITMO
  VAESCO

Erasmus MC

The EU-ADR Project

(formerly known as ALERT)

Exploring and **U**nderstanding **A**dverse **D**rug **R**eactions by Integrative Mining of Clinical Records and Biomedical Knowledge

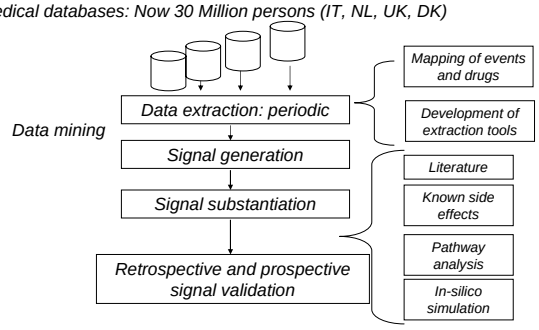
http://euadr-project.org/

Started february 2008

EU-ADR IS A RESEARCH AND DEVELOPMENT PROJECT FUNDED BY THE INFORMATION AND COMMUNICATION TECHNOLOGIES (ICT) AREA OF THE EUROPEAN COMMISSION UNDER THE VII FRAMEWORK PROGRAMME

EU-ADR concept

Medical databases: Now 30 Million persons (IT, NL, UK, DK)



www.euadr-project.org

Type of databases currently involved

Electronic medical	Administrative
• IPCI (NL)	• PHARMO (NL)
• QRESEARCH (UK)	• Aarhus (DK)
• PEDIANET (IT)	• ARS: Tuscany (IT)
• HSD (IT)	• UNIMIB: Lombardy (IT)

All database population-based, capture all events/exposure

Databases

SIMG	1,600,000	general population, all ages, medical record all Rx, all Dx, labs
Pedianet	160,000	general population, children, medical record all Rx, all Dx, labs
Lombardy	9,000,000	General population, all ages, all reimbursed dispensings, hospitalizations
Tuscany	4,000,000	General population, all ages, all reimbursed dispensings, hospitalizations, deaths, exemptions
IPCI	1,000,000	General population, all ages, all Rx, all Dx, labs
PHARMO	3,000,000	Population based, all ages, all reimbursed dispensings, hospitalizations, in hospital
Denmark	3,000,000	General population, all ages, all reimbursed dispensings, procedures, hospitalizations other registries
QRESEARCH	11,2000,000	General population, all ages, all Rx, all Dx, labs

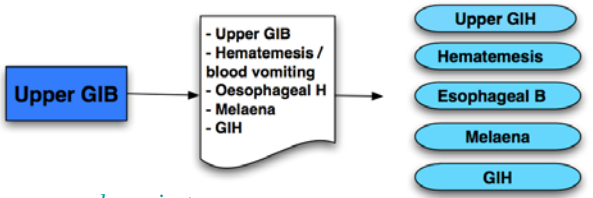
Structure & Standardization

Harmonization process

- **Pre-data extraction**
  1. Terminology mapping
- **Data extraction**
  2. Local use of dedicated software (Jerboa)
- **Post-data extraction**
  3. analysis of queries for event data extraction
     - consensus via discussion in case of disagreement

Workflow of Terminology mapping

1. Event
2. Event Description Form
3. Search for UMLS concepts



www.euadr-project.org

From: Paul Avillach, Frantz Thiessard, Bordeaux

C0017181		Gastrointestinal Hemorrhage
ICD10	K92.2	Gastrointestinal haemorrhage, unspecified
ICD9CM	578	Gastrointestinal haemorrhage
ICD9CM	578.9	GASTROINTEST HEMORR NOS
ICD9CM	578.9	Haemorrhage of gastrointestinal tract, unspecified
ICPC2ICD10DUT	MTHU011567	bloeding; gastro-intestinaal
ICPC2ICD10DUT	MTHU011669	bloeding; maagdarmkanaal
ICPC2ICD10DUT	MTHU030509	gastro-intestinaal; bloeding
ICPC2ICD10DUT	MTHU047008	maagdarmkanaal; bloeding
ICPC2ICD10ENG	MTHU011567	haemorrhage; gastrointestinal
ICPC2ICD10ENG	MTHU011669	haemorrhage; gastrointestinal tract
ICPC2ICD10ENG	MTHU030509	gastrointestinal; hemorrhage
ICPC2ICD10ENG	MTHU030540	gastrostaxis
ICPC2ICD10ENG	MTHU047008	gastrointestinal tract; hemorrhage
ICPC2P	D15001	Bleeding:gastrointestinal
ICPC2P	D15001	gastrointestinal bleeding
RCD	J68..	Gastrointestinal bleed
RCD	J68..	Gastrointestinal haemorrhage

alert

Data extraction

eu-adr

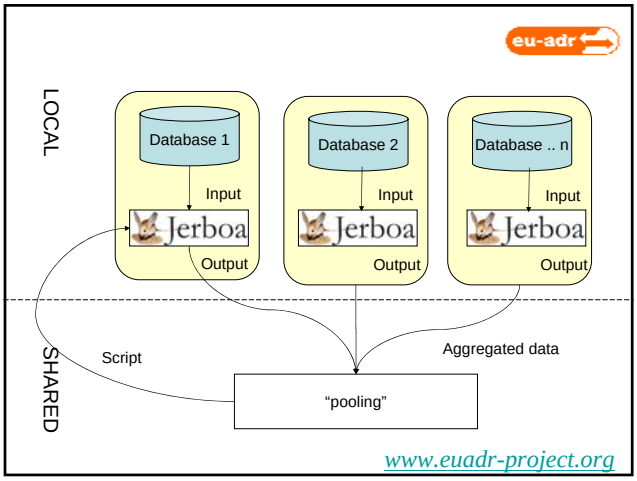
- Distributed network approach for obtaining the data
- No analytical capabilities or common software in all sites
- Software developed in JAVA on purpose:
- Runs on common data model

Jerboa

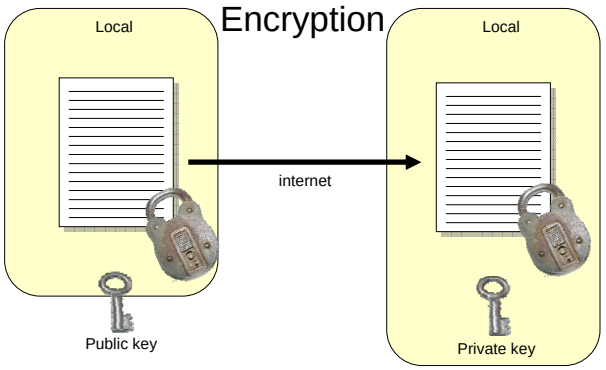
Common data model

Population	Drugs	Events	Vaccines
PatientID	PatientID	PatientID	PatientID
Startfol	DateRx	Date	Date
Endfol	ATC	Typevent	Typevacc
DOB	Dose		Lot
Gender			..

Jerboa



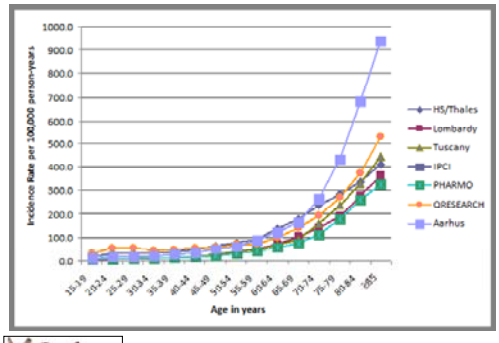
Safety



Comparison and benchmarking of the rates

		CRUDE IR per 100, 000 person-yrs		
COUNTRY	DATABASE	Not harmonized	Post -Terminology mapping	After consensus
ITA	Pedianet	16.6	16.0 (*14.5*)	16.0 (*14.5*)
	Health Search	126.4	109.3 (*65.3*)	109.3 (*65.3*)
	Lombardy	45.2	84.0 (45.8)	52.5 (*29.1*)
	Tuscany	80.3	71.8 (*32.2*)	71.8 (*32.2*)
NL	IPCI	65.4	61.0 (*44.2*)	61.0 (*44.2*)
	PHARMO	48.3	39.0 (*25.3*)	39.0 (*25.3*)
UK	Qresearch	85.6	83.4 (*59.5*)	83.4 (*59.5*)
DK	Aarhus UH	85.7	108.6 (66.9)	87.6 (*54.5*)
	TOTAL		66.9	66.9

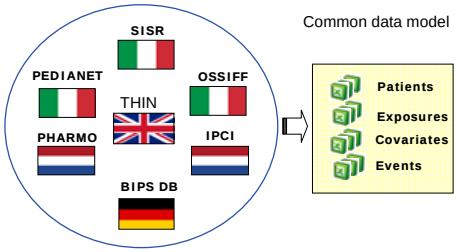
Quality control: Comparison of incidence rates (UGIB)



Jerboa

Other EC projects using distributed data approaches

Databases: GI/CV Safety of NSAIDs



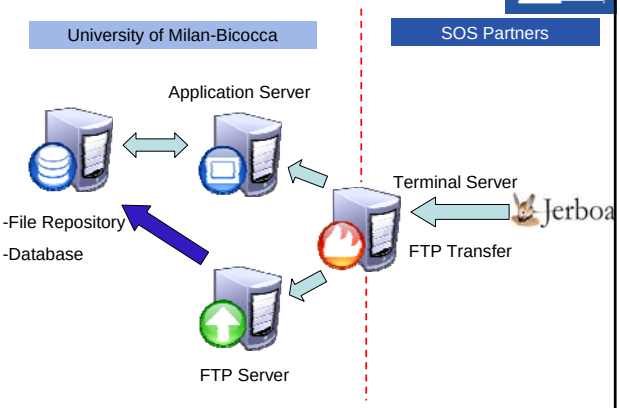
Common data model

- Patients
- Exposures
- Covariates
- Events

Total around 30 million source population

Period: 1999-2008

Hardware Configuration



Example 2 : VAESCO (vaccine safety)

- Background rates on outcomes: 9 countries (NO, SW, FI, DK, UK, NL, DE, IT, ES)
- Terminology mapping with UMLS (ICD-9, 10, ICPC, READ): free text: 12 conditions
- Common data model
- Jerboa for calculations, with/without censoring

Mapping terminologies

ENCO **Encephalitis (myelitis) other causes (unknown)**
READ-GPRD A130300, A136.00, A136000, A136z00, A361.00, AD00.00, F03..00, F031.00, F032011, F033.00, F033000, F033011, F033111, F033400, F033411, F033z00, F034.00, F035.00, F035z00, F036.00, F036100, F036200, F036z00, F03y.00, F03z.00, F212.00, A413.00, A413.11, F03X.00, Fyu0500, A560100, F03..00, F03..11, F03y.11, F03y.12, F03z.11, F286.15, F286.16
ICD-9: 323.51, 323.61, 323.62, 323.7, 323.81, 323.9, 049.8, 036.1, 046.3, 330.8
ICD-10: G36 (excluding G36.0), G04
ICPI: N71

ENCV **Encephalitis viral causes**
READ-GPRD: A4y0.00, A4zy000, A4zy100, A4zy200, A4zy500, A4zz.11, A520.00, A531400, A543.00, A550.00, A62..00, A620.00, A621.00, A622.00, A623.00, A624.00, A62y.11, A63..00, A630.00, A63y000, A722.00, AE20.00, Ayu8B00, F030.00, F030200, F030211, F030400, F030411, F030500, F030511, F030611, F030800, F030900, F030911, F030A00, F030z00, F035000, F035011, F035100, ZV73512
ICD-9: 323.4, 136.2, 056.01, 130.0, 054.3, 049.0, 094.1, 072.2, 013.0, 062.4, 323.0, 049.9, 062.8, 062.9,062, 045.0
ICD-10: A86, A83
ICPI: N70

Key challenges

1) Terminology mapping: how many events/conditions
2) Defining output files from queries that are ‘anonymous’ according to all countries and are flexible enough for analysis
3) Lack of analytical capability in various databases
4) Databases from private and public sources
5) Disparity
6) Differentiating between heterogeneity due to misclassification (lack of information) and true heterogeneity

Next steps

- Inclusion of vaccines
- More databases being ‘connected’
- Validation of events

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

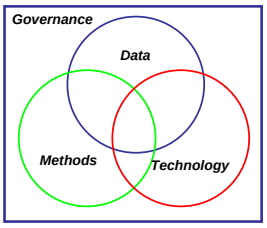
FOUNDATION FOR THE National Institutes of Health

Structure and Function of Distributed Networks for Active Surveillance: Experience in the Observational Medical Outcomes Partnership

Patrick Ryan
OMOP Research Investigator
23 November 2009

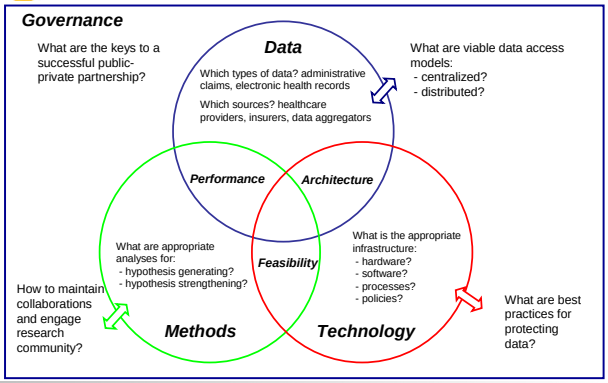
Observational Medical Outcomes Partnership

A public-private partnership to serve the public health by testing whether multi-source observational data can improve our ability to assess drug safety and benefits.



- Assess the appropriate technology and data infrastructure required for systematic monitoring of observational data
- Develop and test the feasibility and performance of the analysis methods
- Evaluate required governance structures

Outstanding questions for active surveillance



OMOP Phases

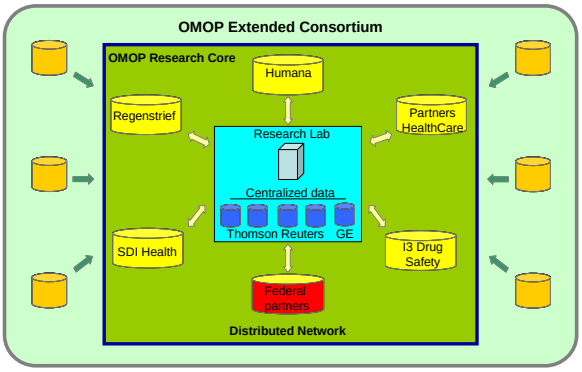
- **Phase 1:** FEASIBILITY OF DATA INFRASTRUCTURE (Feb – July 2009)
  - Establish a consistent framework to use across disparate observational data sources
  - Establish OMOP Research Community
- **Phase 2:** FEASIBILITY OF ANALYSES (Aug – Dec 2009)
  - Develop and test analysis methods within the OMOP Research Lab and other data environments
  - Establish standard data characterization procedures
  - Implement health outcomes of interest definitions
  - OMOP to facilitate comparisons across databases
- **Phase 3:** PERFORMANCE MEASUREMENTS (Jan – July 2010)
  - Evaluate performance of methods and data in identifying drug safety issues
  - OMOP to facilitate comparisons across databases
- **Phase 4:** UTILITY OF ANALYSES & PROCESS (July – Dec 2010)
  - Assess the effectiveness and usefulness of how the results and comparisons contribute to decision-making

OMOP data assessment: Provider willingness for data access models

	Organizations (n=21)	Total population (m)
Centralized model: Provide your data externally to load into the Central Research Core IT environment	7	297
Federated model: Facilitate OMOP researchers access to execute queries directly (through firewall)	4	252
Distributed CDM Model: OMOP queries run locally by your research staff	17	470
Distributed protocol model: Develop and run your own queries locally	19	413

Each access model would have access to over 250m lives in aggregate, indicating the FDAAA mandate of 100m persons is achievable under all alternative infrastructures without full participation of potential data sources

Testing data access models: OMOP data community

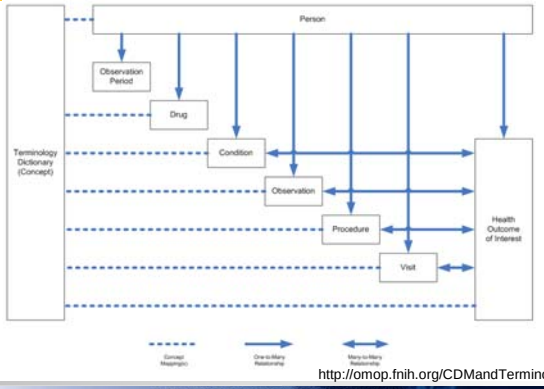


Common Data Model

- The common data model includes:
  - A single data schema that can be applied to disparate data types
  - Standardized terminologies
  - Consistent transformation for key data elements
- A common data model can:
  - Enable consistent and systematic application of analysis methods to produce comparable results across sources
  - Create a community to facilitate the sharing of tools and practices
  - Impose data quality standards
  - Create implementation efficiencies

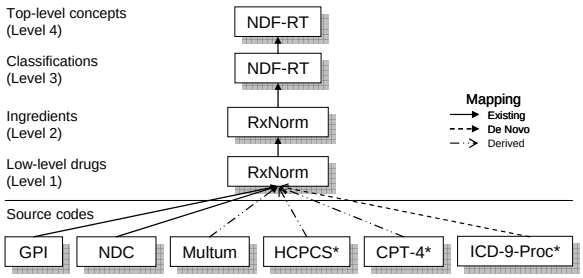
Common Data Model	
What We Are Doing	**What We Are Not Doing**
• Creating one model that could accommodate any relevant type of observational data • Facilitating comparison of analysis results across sources • Providing a conceptual model to allow researchers to develop analysis methods that are be portable across data sources	• Combining multiple datasets into one centralized database • Trying to force claims data into a EHR model or vice versa • Developing a graphical user interface to automatically create structured queries

Conceptual Schematic of OMOP Common Data Model



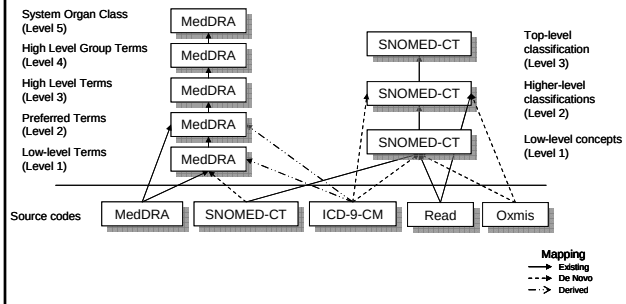
http://omop.fnih.org/CDMandTerminologies

Standardizing terminologies for drugs



http://omop.fnih.org/CDMandTerminologies

Standardizing terminologies for conditions

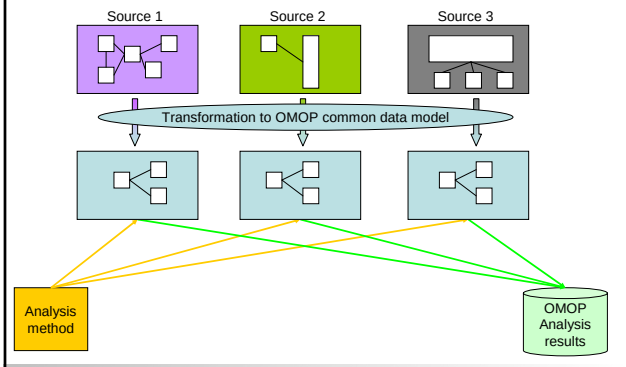


http://omop.fnih.org/CDMandTerminologies

Lessons about OMOP Common Data Model

- Common Data Model can accommodate disparate data types, including claims and EHR
  - All Research Core data sources have been successfully transformed into OMOP Common Data Model
  - All organizations completed CDM ETL in ~3 months
  - Upcoming research: evaluate consistency and time to execution of methods when applied to CDM vs. raw data
- CDM can be built to minimize information loss
  - OMOP only explicitly removed financial information from scope of model, but model could be expanded as needs arise
  - Largest source of information loss is in analysis, most methods don't take advantage of the wealth of available information
- Common challenges:
  - Value in developing standard processes in centralized environment, but distributed organizations may have different technology infrastructure
  - Standardizing terminologies (drugs, conditions, procedures, laboratory results) requires mapping
  - Conflict of static data for research and analysis vs. continuous data feeds for primary data collection purposes (reimbursement, care)

Role of common data model in OMOP Analysis process

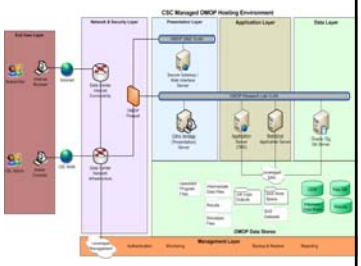


A standardized process for evaluating data characteristics

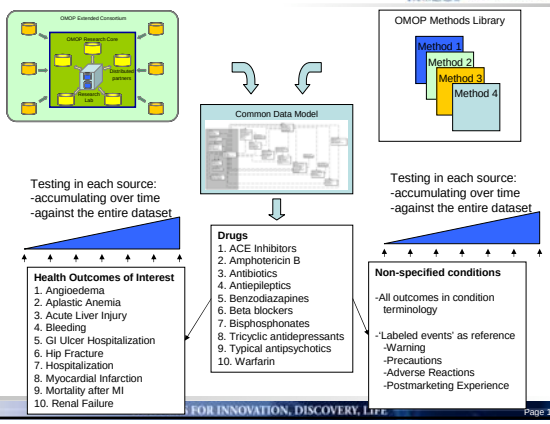
- OSCAR (Observational Source Characteristics Analysis Report) Provides a systematic approach for summarizing observational healthcare data stored in the OMOP common data model
  - Validation of transformation from raw data to OMOP common data model
  - Comparisons between data sources
  - Comparison of overall database to specific subpopulations of interest
  - Providing context for interpreting and analyzing findings of drug safety studies
- NATHAN (Natural History Analysis) provides structured descriptive statistics for a pre-specified cohort of interest
  - Exposed population (e.g. patients taking antibiotics)
  - Cases (e.g. patients with acute liver injury)
  - Exposed cases (e.g. patients taking antibiotics with acute liver injury)
- Standardized descriptive statistics provide meta-data to put drug-outcome assessments into appropriate context, and can facilitate comparisons across disparate sources within network

Infrastructure for a central coordinating center: OMOP Research Laboratory

- Objectives
  - Secure, protected access to all de-identified person-level data and summary results across data community
  - Infrastructure manage and facilitate analyses across large datasets
  - Environment to facilitate ongoing analysis research and development, distribution and retrieval across network
- Outstanding issues:
  - Hardware?
  - Software?
  - QA Process?
  - Governance?



OMOP research experiment workflow



Contact information

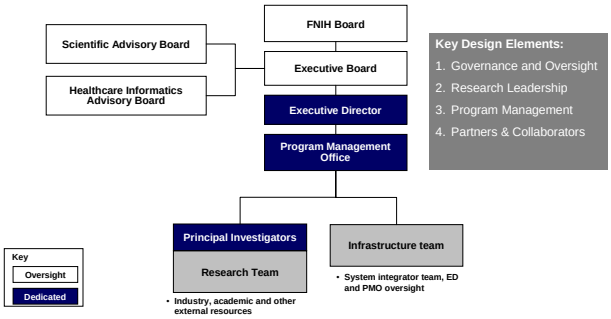
Patrick Ryan
Patrick.b.ryan@gsk.com

OMOP website: http://omop.fnih.org
OMOP Cup website: http://omopcup.orwik.com

Backups

Partnership Structure

Governance Provided by an Executive Board
Scientific and Informatics advisory boards inform decisions



- Industry, academic and other external resources
- System integrator team, ED and PMO oversight

Key Design Elements:

1. Governance and Oversight
2. Research Leadership
3. Program Management
4. Partners & Collaborators

Executive Board

A multi-stakeholder group, the OMOP Executive Board oversees the operation of the Partnership.

Janet Woodcock, MD
Director, Center for Drug Evaluation and Research, Food and Drug Administration
Chair, Observational Medical Outcomes Partnership Executive Board

Rebecca Burkholder
Vice President of Health Policy, The National Consumers League

Sherine Gabriel, MD, MSc
Professor of Medicine and Epidemiology, The Mayo Clinic

Cynthia Gilman, JD
Special Assistant to the President for Advancement of Cancer Research and Collaborative Partnerships, Henry Jackson Foundation

Jesse L. Goodman, MD, MPH
Chief Scientist and Deputy Commissioner for Science and Public Health (acting),
Food and Drug Administration

Ronald L. Krall, MD
Former Senior Vice President and Chief Medical Officer, GlaxoSmithKline

Richard Platt, MD, MSc
Professor and Chair of the Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care

Stephen Spielberg, MD, PhD
Marion Merrell Dow Chair in Pediatric Pharmocogenomics, Children's Mercy Hospital and Dean Emeritus, Dartmouth Medical School

Brian Strom, MD, MPH
George S. Pepper Professor of Public Health and Preventive Medicine; Professor of Biostatistics and Epidemiology, Medicine, and Pharmacology; Chair, Department of Biostatistics and Epidemiology; Director, Center for Clinical Epidemiology and Biostatistics; Vice Dean for Institutional Affairs, University of Pennsylvania School of Medicine
Senior Advisor to the Provost for Global Health Initiatives, University of Pennsylvania

David Wheadon, MD
Senior Vice President, Pharmaceutical Research and Manufacturers of America (PhRMA)

Research Investigators

The Principal Investigators (PIs) are the lead scientists for the OMOP project and guide and participate in the research across all four project phases

Marc Overhage, MD, PhD: Director, Medical Informatics and Research Scientist, Regenstrief Institute, Inc.; Regenstrief Professor of Medical Informatics, Indiana University School of Medicine,CEO; President of the Indiana Health Information Exchange

Paul Stang, PhD, FISPE: Senior Director, Epidemiology, Johnson & Johnson Pharmaceutical Research and Development

Abraham G. Hartzema PharmD, MSPH, PhD, FISPE: Professor and Eminent Scholar, Pharmaceutical Outcomes & Policy, Perry A. Foote Chair in Health Outcomes Research, University of Florida College of Pharmacy

Judy Racoosin, MD, MPH: Sentinel Initiative Scientific Lead, US Food and Drug Administration

Patrick Ryan: Manager Drug Development Sciences, GlaxoSmithKline R&D
OMOP Co-Investigator

Foundation for the NIH Program Staff

Thomas Scarnecchia, MS
Executive Director, OMOP

Emily Welebob, RN, MS
Senior Program Manager, Research

Christian Reich, MD, PhD
Senior Program Manager, Technology

OMOP Statistics and Programming Team

David Madigan, PhD
Columbia University
OMOP Methods Lead

Ivan Zorych, PhD
Columbia University

Cortney Hayflinger
Hayflinger Analytic Services

Mark Khayter
Ephir, Inc.

Ron Mantha
Ephir, Inc.

Carlos Alzola
Data Insights

Emmanuel Angel
Angelic Productions

Reed George
etera solutions

Eric Lantz
University of Wisconsin-Madison

Advisory Boards

A Scientific Advisory Board (SAB) will provide independent review of and expert input into the scientific aspects of OMOP's activities.

- Elizabeth Andrews, RTI Health Solutions
- Andrew Bate, Pfizer
- Jesse Berlin, Johnson & Johnson
- Robert Davis, Kaiser Permanente
- Steve Findlay, Consumer Union
- Sean Hennessy, University of Pennsylvania
- Mike Katz, FDA patient representative
- Allen Mitchell, Boston University
- David Page, University of Wisconsin
- Ken Rothman, RTI Health Solutions
- Judy Staffa, FDA
- Alec Walker, WHISCON

A Health Informatics Advisory Board (HIAB) will provide independent review and expert input into the OMOP's technology governance and project requirements related to privacy and security, terminology and coding, data and data models.

- Col. Kevin Abbott
- Jeff Brown, Harvard Medical School
- Stan Huff, Intermountain Healthcare
- Diane MacKinnon, IBM (retired)
- Ken Mandl, Harvard University
- Clem McDonald, National Library of Medicine
- David Memel, Klaipeda Consulting
- Joy Pritts, Georgetown University
- Rob Thwaites, United BioSource Corporation

Research Collaborators: Data and Infrastructure

as of 11/12/09

Organization	Team Leader	Activity
Computer Sciences Corporation	Dan Foltz	Research Lab
Department of Veterans Affairs Center for Medication Safety	Fran Cunningham, PharmD	Distributed Partner
GE Healthcare	Michael Lieberman, MD	Research Lab
i3 Drug Safety	Arnold Chan, M.D., Sc.D.	Distributed Partner
Indiana University - Regenstrief Institute	J. Marc Overhage, MD, PhD	Distributed Partner
Partners HealthCare System	Shawn Murphy, MD, PhD	Distributed Partner
ProSanos Corporation	Stephanie Reisinger	Simulated Data
SDI Health	Gregory Hess, MD, MBA, MSc	Distributed Partner
Thomson Reuters	Stella Chang, MPH	Research Lab
University of Miami-Humana Health Services Research Center	Vinit Nair, BS Pharm., MS, RPh	Distributed Partner

Research Collaborators: Methods

as of 11/12/09

Organization	Team Leader	Activity
Columbia University	David Madigan, PhD	Methods Lead
Eli Lilly and Company	Karin L. Benoit	Methods Partner
GPRD Group of the MHRA	John Parkinson, BSc, PhD	Methods Partner
Harvard Pilgrim Health Care Institute	Lingling Li, PhD	Methods Partner
Indiana University - Regenstrief Institute	Siu L. Hui, PhD	Methods Partner
M Alan Brookhart, PhD and SAS Institute	M. Alan Brookhart, PhD	Methods Partner
Merck Research Laboratories	Dr. A. Lawrence Gould	Methods Partner
ProSanos Corporation	Stephanie Reisinger	Methods Partner
Risk Benefit Statistics LLC	Robert L. (Bob) Obenchain, PhD, FASA	Methods Partner
RTI International	Suzanne L. West, MPH, PhD	HOI Library
Slone Epidemiology Center at Boston University	David Kaufman, ScD	Methods Partner
United BioSource Corporation	Matthew W. Reynolds, PhD	HOI Library
University of North Carolina at Chapel Hill	Stacie Dusetzina	HOI Library
University of Utah	Brian Sauer, PhD	Methods Partner
University of Wisconsin-Madison	David Page, PhD	Methods Partner
Uppsala Monitoring Center	Niklas Norén, PhD	Methods Partner