RTI Health Solutions

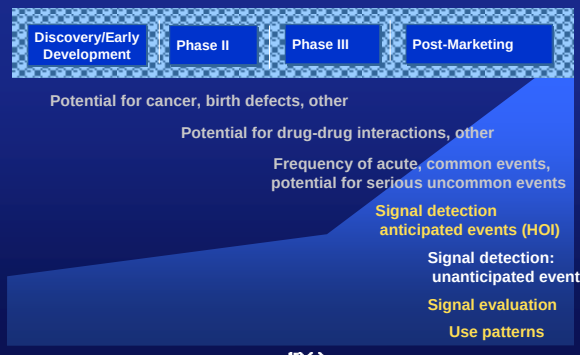
What Data Infrastructure Will Achieve FDA's Vision of Active Surveillance in the Near and Longer Term?

Views from an epidemiologist

Elizabeth B. Andrews, PhD, MPH
VP, Pharmacoepidemiology and Risk Management

LEADING RESEARCH…
MEASURES THAT COUNT

Emergence and evaluation of signals across the therapeutic lifecycle



What do we want to know from active surveillance?

- **Incidence of events**
  - Serious acute events in new users (e.g. liver failure, MI)
  - Serious events in users of inpatient medications
  - Serious events in long-term use (e.g. fracture)
  - Serious events of long latency (e.g., lymphoma, breast cancer)
  - Pregnancy outcomes, esp. birth defects
- **Incidence of events among important patient subgroups**
  - Pediatrics
  - Elderly
- **Comparative risks**
  - Compared to background population
  - Compared to other therapeutic choices
- **Predictors of increased risk**
  - Smoking, BMI, OTC use, co-morbidities, genetic markers

What do we want to know from active surveillance?

- **Utilization**
  - Distinguish chronic vs. intermittent use
  - Identify abuse and diversion (+consequences)
  - Describe prescribing within and outside of intended use
  - Characterize prescribers
  - Describe prescribing with other medications and other conditions
- **Other (potentially not knowable)**
  - Inter-generational effects
  - Small increases in risk in very rare outcomes
  - Small-medium increases in risk in small subgroups

Complementary Views

- "Le mieux est l'ennemi du bien." Voltaire's *Dictionnaire Philosophique* (1764)
- "The best is the enemy of good."
- But – don't overlook the gaps for the long term

Missing pieces from most claims databases

- Patient-reported information (smoking, BMI, family history)
- Lab tests and results (available in some)
- Inpatient medications
- OTC medications
- Long-term utilization
- Long-term follow-up for outcomes
- Poor specificity in many outcomes (e.g., birth defects, tumor type)
- Practice patterns outside of "covered care"

Missing pieces from EMR

- Patient-reported information (smoking?, BMI?, family history)
- Inpatient medications
- OTC medications
- Services outside of the EMR base

What else is missing?

- Linkage across information sources (e.g., EMR and Cancer registries)
- Validated algorithms for (all?) the health outcomes of interest
  - Validated algorithms in large databases (EMR, claims)
  - Repository of methods to facilitate their reuse and increase efficiency of future studies
- Methodological concerns
  - Confounding by indication
  - Unmeasured confounders
  - Incomplete data
  - Reverse causation
  - Compliance (i.e. information errors stemming from using prescription data)

Near-term

- **Facilitate: "ad hoc requests for targeted surveillance of selected events that emerge at some time after initial marketing, and routine surveillance of newly marketed products for a specified number of targeted events of interest. "**

Longer-term

- **Linkage with other databases (e.g., cancer registries)**
- **Improve data coverage and continuity**
- **Improve quality of data - EA wish list**
  - Distinguish between actual diagnosis and rule out
  - Capture LMP!
- **Supplement with patient-reported information**
- **Improve methods for identifying outcomes and making comparisons**

Response Panel

Kenneth D. Mandl, MD, MPH

Director, Intelligent Health Laboratory
Children's Hospital Informatics Program
Children's Hospital Boston

Center for Biomedical Informatics
Harvard Medical School

Should test some assumptions across projects

- ✓ Little or no exchange of person-level data is needed
  - ☞ What are the use cases really?
  - ☞ Will the preponderance of cases require central data aggregation or not? Line list data or not?
  - ☞ How does comparative effectiveness play in here?
- ✓ There are no options to exchange rich data because of the business interests of stakeholders
  - ☞ $48B elephant in the room—should these activities be clinically linked in some way? (DARTnet successes in a provider centered model; other models . . .)
  - ☞ Personal health model

Types of data

- Claims
- EHR
- Personally controlled health records
- "Registry" (meticulously entered data)

PATIENT AS INTEGRATOR OF DISTRIBUTED DATA

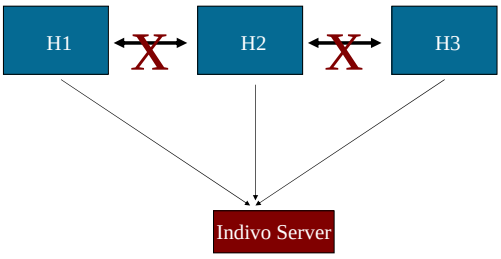
Children's Hospital Informatics Program | Harvard Medical School | HST

In 1994 we observed that institutions rarely share data

- Proprietary
- Perceived competition
- Privacy
- Health Insurance Portability and Accountability Act
- No dedicated resources to do so

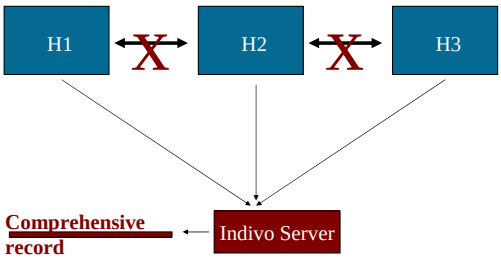
Children's Hospital Informatics Program | Harvard Medical School | HST

What if we gave patients a tool to request their records electronically?



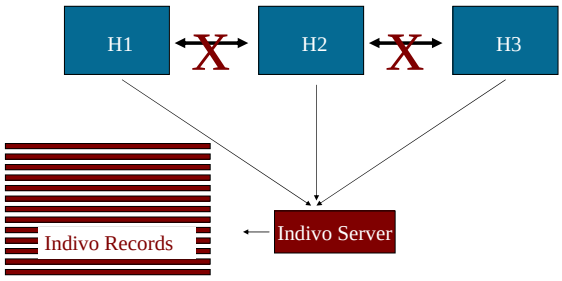
Children's Hospital Informatics Program | Harvard Medical School | HST

And create a personal health record



Children's Hospital Informatics Program | Harvard Medical School | HST

The collection of these records is a population health database



Our original statement on personal control

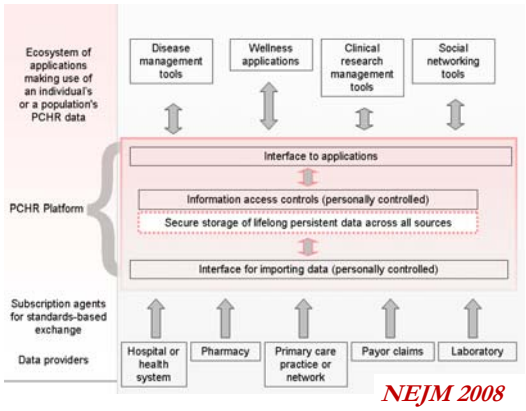
- A PCHR stored all of an individual's medical history in a container with:
  - ✓ patient control
  - ✓ interoperability
  - ✓ open standards

Information in practice

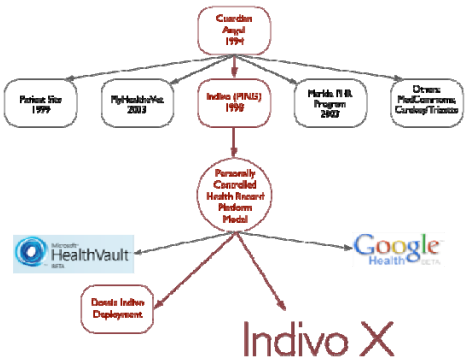
Public standards and patients' control: how to keep electronic medical records accessible but private

Kenneth D Mandl, Peter Szolovits, Isaac S Kohane

BMJ 2001;322:283–7



NEJM 2008



PCHR vendors and users create large accessible populations for public health study and intervention

The NEW ENGLAND JOURNAL of MEDICINE

Tectonic Shifts in the Health Information Economy

Kenneth D. Mandl, M.D., M.P.H., and Isaac S. Kohane, M.D., Ph.D.

In a recent shift in the health information landscape, large corporations are seeking an integral and transformative role in the management of health care information. The mechanism by which this transformation is likely to take place is through the creation of computer platforms that will enable patients to manage health data in personally controlled health records (PCHRs). Two types of large corporations are involved. Technology companies such as Google and Microsoft see

at one hospital, a visit to an emergency department at another hospital, and test results at a laboratory. She logs into her hosted PCHR at a secure Web site. Since she has established subscriptions to automatic updates from each of these clinical entities, her PCHR is current with copies of those data.

The PCHR enables the patient to authorize access to information (views or even copies of the record) to others, including clinical providers,

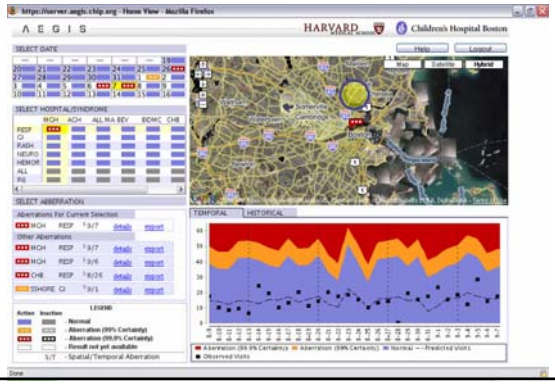
New England Journal of Medicine, April 2008

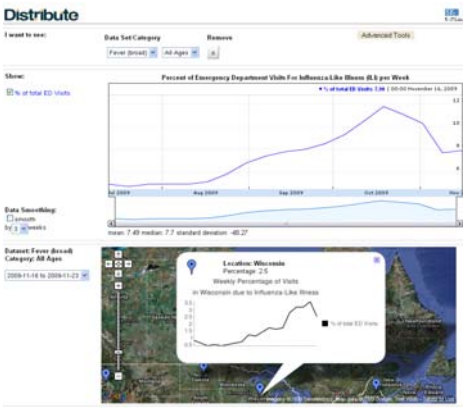
High potential for information altruism

- ✓ Surveyed experienced PCHR users about willingness to share information from record for population health and public health
  - ☞34% users "very agreeable"
  - ☞35% "moderately agreeable"
  - ☞21% "somewhat agreeable" to sharing for population health monitoring
  - ☞After more than one year of exposure to a pilot system, ONLY 9% report they are "not agreeable" to sharing

JMIR (submitted)

Live data





Journal of the American Medical Informatics Association Volume 14 Number 6 Nov / Dec 2007

Research Paper

The Value of Patient Self-report for Disease Surveillance

Florence T. Bourgeois, MD, MPH, Stephen C. Porter, MD, MPH, Clarissa Valim, ScD, MD, Tiffany Jackson, BS, E. Francis Cook, ScD, Kenneth D. Mandl, MD, MPH

Abstract **Objective:** To determine the accuracy of self-reported information from patients and families for use in a disease surveillance system.

Design: Patients and their parents presenting to the emergency department (ED) waiting room of an urban, tertiary care children's hospital were asked to use a Self-Report Tool, which consisted of a questionnaire asking questions related to the subjects' current illness.

Measurements: The sensitivity and specificity of three data sources for assigning patients to disease categories was measured: the ED chief complaint, physician diagnostic coding, and the completed Self-Report Tool. The gold standard metric for comparison was a medical record abstraction.

Results: A total of 936 subjects were enrolled. Compared to ED chief complaints, the Self-Report Tool was more

Table 3 ▪ ED Chief Complaint, Diagnostic Code, and Self-Report Tool Classification*

	ED Chief Complaint		Diagnostic Code		Self-Report Tool	
Disease category	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %
Fever	48 (130/272)	98 (634/645)	21 (50/243)	99 (576/577)	91 (253/278)	88 (579/658)
Respiratory	42 (112/269)	97 (628/648)	52 (130/252)	99 (561/568)	88 (239/273)	84 (555/663)
Gastrointestinal	78 (171/218)	96 (670/699)	75 (129/172)	99 (640/648)	86 (190/221)	92 (656/715)
Dermatologic	37 (22/60)	99 (853/857)	41 (21/52)	99 (765/768)	82 (49/60)	87 (762/876)
Neurological	60 (26/43)	99 (867/874)	67 (20/30)	99 (788/790)	77 (33/43)	90 (801/893)
Allergic	50 (5/10)	99 (904/907)	0 (0/8)	99 (811/812)	90 (9/10)	99 (916/926)
Injury	84 (161/192)	98 (712/725)	91 (172/188)	97 (615/632)	97 (190/196)	96 (708/740)
Other	62 (102/164)	87 (652/753)	75 (112/150)	79 (531/670)	71 (122/172)	84 (644/764)

*ED chief complaint available for 917/936 (98%) of subjects; diagnostic code available for 820/936 (88%) of subjects; Self-Report Tool available for 936 (100%) of subjects.

Personal health data

- Patient reported outcomes
  - ✓ Adverse effects
  - ✓ Efficacy endpoints
  - ✓ Adherence
  - ✓ Satisfaction
  - ✓ Quality of life
- Patient reported data
  - ✓ Over the counter meds
  - ✓ Complimentary/alternative meds

ADAPTING A BIOSURVEILLANCE DISTRIBUTED APPROACH

National Health Information Network Demo

JAMIA Editorial Comments

Editorial ■

Lessons Learned from the Shared Pathology Informatics Network (SPIN): A Scalable Network for Translational Research and Public Health

MICHAEL J. BECICH, MD, PhD

■ J Am Med Inform Assoc. 2007;14:534–535. DOI 10.1197/jamia.M2477.

The article by McMurry et al. in the current issue of JAMIA describes an innovative architecture to support National ing agreements . . .".[1] This work includes successful development, integration, and support for de-identification

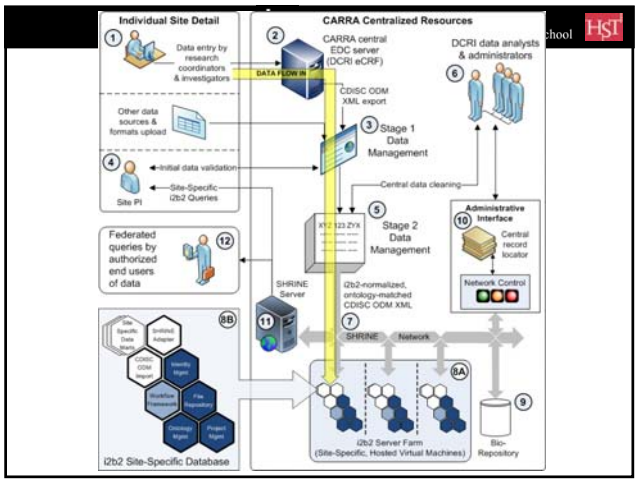
$8M GO Grant

- ARRA NIH funding for "Grand Opportunities"
- Based on the I2B2 system
- 60 Pediatric Rheumatology sites (children's hospitals)
- A live, evolving registry (meticulously entered data)
- Enables light sip and deep drink

One Solution: The i2b2 Model

Deliverables

- Turnkey 60-site i2b2/SHRINE Data Warehouse with 'Parsimonious' Core Data Set (light sip)
- Permissioned, secure query via SHRINE for detailed high quality data (deep drink)



ALTERNATIVE MODEL

"These medicines all taste pretty good—let's approve them."

careevolution
HEALTHCARE TECHNOLOGY

Distributed Data Networks for Active Medical Product Surveillance

Vik Kheterpal, MD
Principal
CareEvolution, Inc.

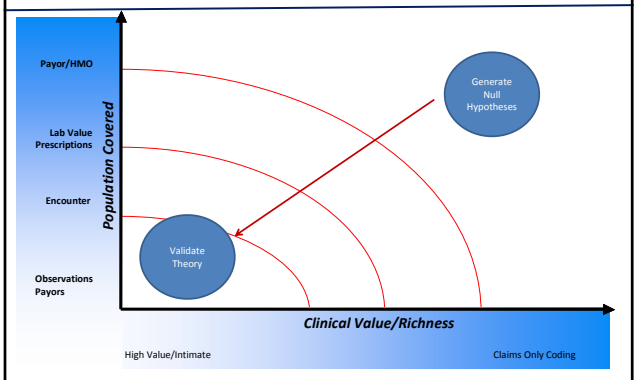
Some Observations – 100,000 foot view

Current Experiences	"Barriers to Scale"
"Distributed"… foregone conclusion?	Non Technical
Policy/Privacy Pressures	Recruitment of data "owners"
Technological Pleasant Surprise	Skepticism/Protectionism
People don't want to give up "their" data	Defining data needs
Common Themes	Fragmentation ("weight") for clinical
Convener Function	Technical
"ETL"	Last mile problem to EHR
Data Stays Locally	Data cleansing
EHR Independent – Middle Layer	Data Standardization
Common Data Model	Analytic Processing Readiness
Hybrid Query/Analytic Methods	How to scale down/out

Is the Runway Long Enough? This is not a one-time event.

Non Technical	Technical
Is asking for data truly unprecedented?	Last mile effort – need leverage
Claims justification	Each effort is purpose built
Public Health Requirement	Subset of data accessed
Governance structure re-invention	Middle Mile - monolithic
Tertiary Use	Specific for each "question"
Ownership	Yet common needs
Requestor Harmonization	LOINC
JCAHO, HEDIS… now CORE	RxNorm
How many efforts can providers engage in	Diagnoses/complaints
Ignores Fast Moving Landscape	Puritanical Approach
Provider – HITECH "meaningful use"	Must have normalized data
Quality Reporting requirements etc.	Common Data Model

Data Type and Provider Continuum

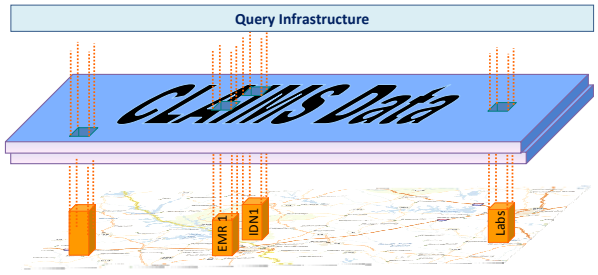


Broad But Shallow (claims) to Jump Start the Effort

Network Services to provide a public utility to participate

Standards based connectors to provide on/off ramps to Hwy

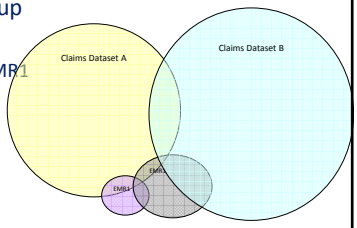
Web Services to allow all to participate



5

The Need for Linking

- Core Value Proposition
  - Subject Duplicity
  - Longitudinal Follow up
    - Acute Events IN CDa
    - Prescription Hx in EMR1
    - ER in CDb
- Issue
  - Privacy Concern
  - Business Issues



Privacy Protecting Linking

1. Just as the Bank does not know the contents of the safety deposit box, with Crypto-RLS you can provide linking without even KNOWING the contents
2. No risk in managing entire regional population
3. No clinical data centralization
4. Protects from
   - Internal threats
   - Disgruntled employees
   - External hacks
   - Inadvertent loss (theft, backup distribution)
5. Web Services provide a "catcher-pitcher" handoff

Separate the Two Questions:

"Where are records for Patient X"

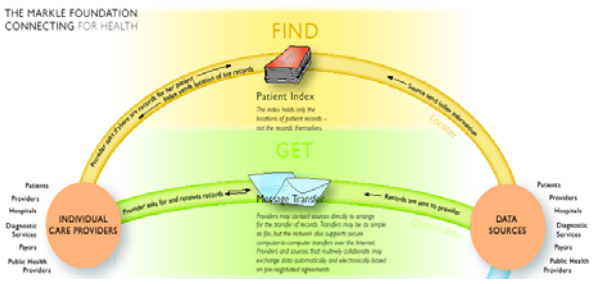
"How can I get Them?"

Follow CfH, HITSP/CCHIT, & NHIN Standards – so the **n**etwork wll be interoperable, privacy protecting, and scaleable…for other uses

Courtesy: Connecting for Health
Markle Foundation

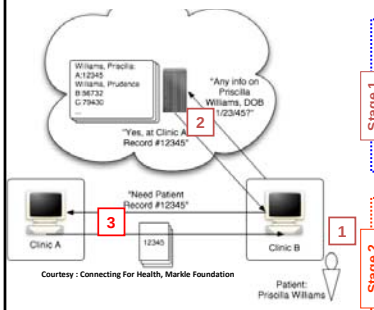
8

Separate Clinical Data from The "Patient Locator" aka "Record Locator Service/RLS"



Courtesy: Connecting for Health
Markle Foundation

Two Phased Query for Data



Discover Location of Records

Stage 1

Patient arrives in Clinic B

Clinic B "pings" the record locator service (RLS)

RLS returns list of locations of records

Retrieve the Record

Stage 2

Clinic B contact peer organizations and retrieves clinical records

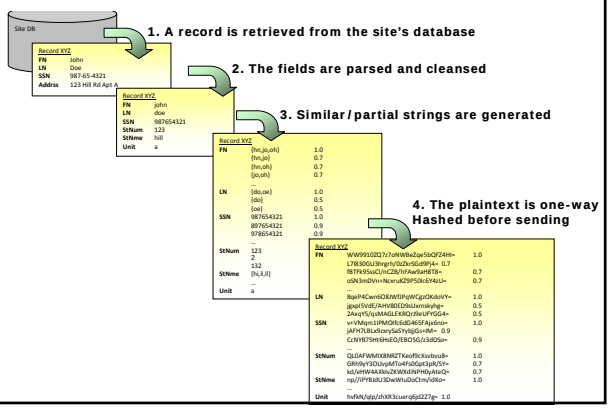
Within Each Site Adapter

Records at participating institutions with identifiers (First Name, Last Name, SSN, DOB, …)

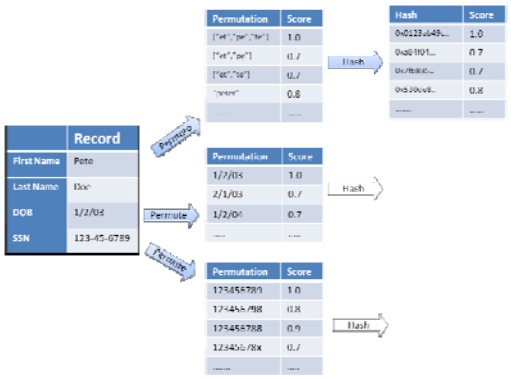
1) Standardizer Pass (delimiters, junk values, spaces)
2) Create permutations of identifiers with associated similarity score (NYSIS, Soundex, transpositions, Nicknames, …)
3) One-way hash the permutations
4) Persist hashed permutations in the BRI
5) Find other records in the BRI with matching permutations.
6) Link to other records based on identifier similarities
7) Group linked records into Patients

Shared Across Sites

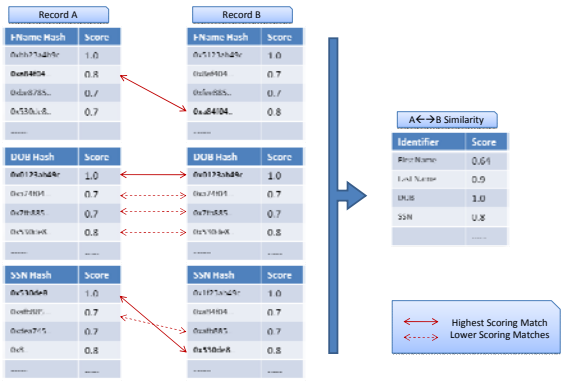
Details



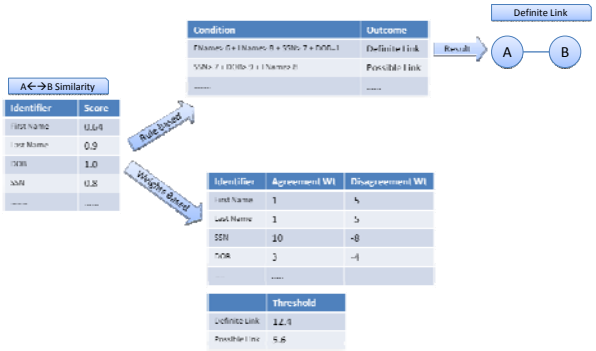
Permutation and Hashing



Similarity Calculation



Link Creation



Multicenter Perioperative Outcomes Group

http://mpog.med.umich.edu

MPOG Members

- University of Michigan (Coordinating center) -- Centricity
- Childrens Hospital of Philadelphia – Compurecord
- University of Pennsylvania -- DocuSys
- University of California at San Francisco -- Picis
- Columbia University – Compurecord
- Mt Sinai NY - Compurecord
- Massachusetts General Hospital -- iMDSoft
- Vanderbilt University -- Acuitec
- Thomas Jefferson University – Drager
- Univ of Miami -- Picis
- 25 others

"Applications" completed

- Death database lookup
  - Application to convert SSN to one-way hashed code
  - Updated to use SHA-2 hashing algorithm
  - Pings central MPOG server for vital status
  - 3rd party tool, to be licensed
- RxNorm categorization of preop medications
  - Maps to key med categories: beta blocker, insulin, oral hypoglycemic, ACE / ARB
  - Occurs centralized after upload
  - 3rd party tool, to be licensed

Recommendations

- Spawn distributed "blindfolded" linkage pilots
- Study scale of the "overlap" and "missed" signal problem without record linkage
- Utilize early experience to inform a multi-year roadmap for Active Surveillance
- Determine a data type to research question stratification model