Medical Product Safety Surveillance Research in Multi-site Settings

Application of "Big Data" to Pediatric Safety Studies Meeting

Silver Springs, MD
September 18-19, 2017

Jeffrey Brown, PhD

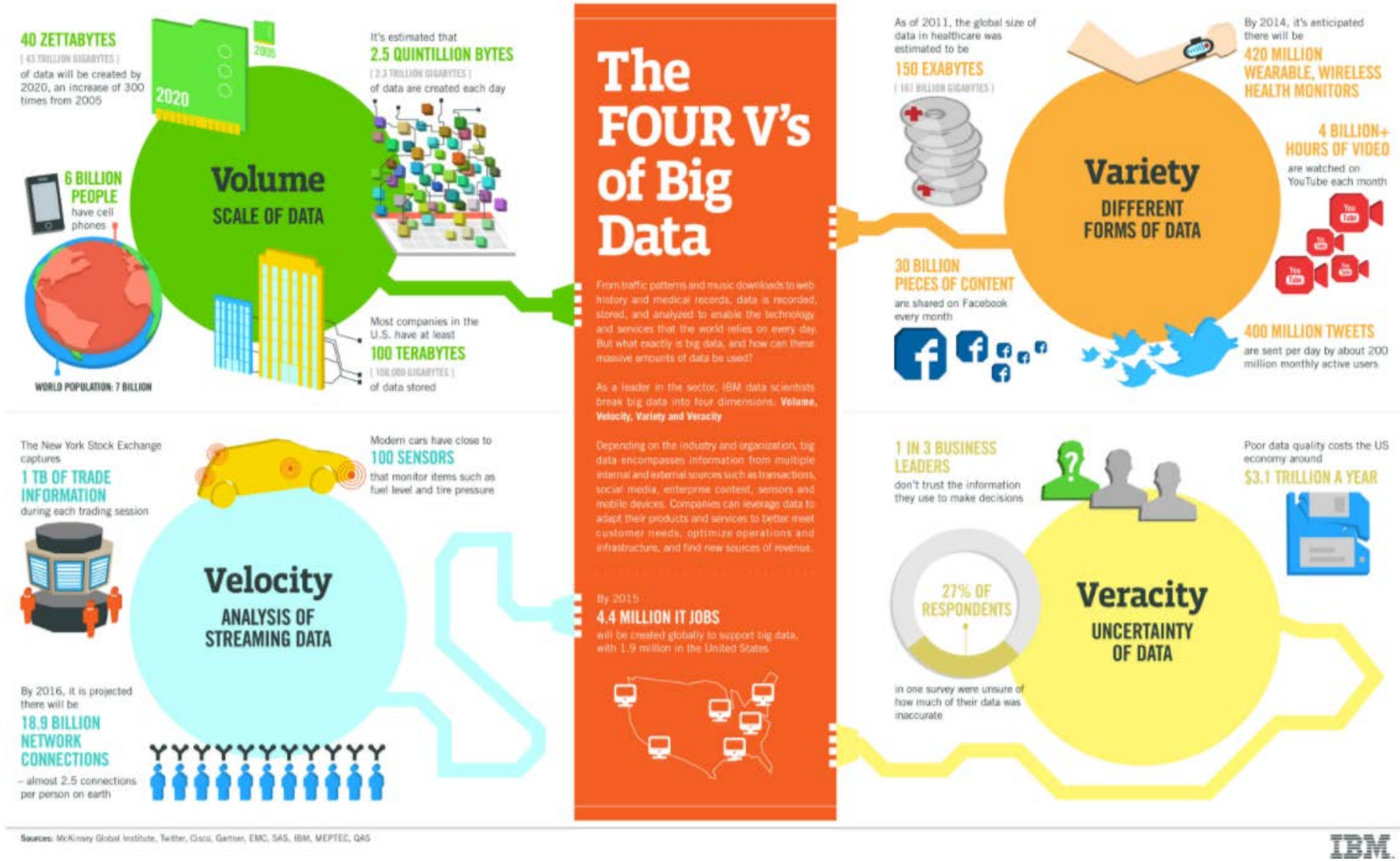
DEPARTMENT OF POPULATION MEDICINE
HARVARD MEDICAL SCHOOL
Harvard Pilgrim Health Care Institute

Disclosures

I am an employee of Harvard Pilgrim Health Care Institute.

I currently receive funding from FDA, NIH, PCORI, BBCIC, IMEDS, GSK, and Pfizer.

I am the inventor of PopMedNet, an open source software application to support distributed health data networks.



http://www.ibmbigdatahub.com/infographic/four-vs-big-data

Sentinel Infrastructure

The NEW ENGLAND JOURNAL of MEDICINE

HOME | ARTICLES & MULTIMEDIA | ISSUES | SPECIALTIES & TOPICS | FOR AUTHORS | CME

Perspective

Mini-Sentinel and Regulatory Science — Big Data Rendered Fit and Functional

Bruce M. Psaty, M.D., Ph.D., and Alasdair M. Breckenridge, M.D.

N Engl J Med 2014; 370:2165-2167 | June 5, 2014 | DOI: 10.1056/NEJMp1401664

Share:

Article | References | Citing Articles (31) | Metrics

In medicine, "big data" come in many forms. With the financial incentives provided by Medicare and Medicaid for the "meaningful use" of electronic health records (EHRs), the quantity of electronic medical data has expanded rapidly. Simultaneously, genomewide association studies funded by the National Heart, Lung, and Blood Institute have produced data sets with millions of genetic variants for each participant, encouraged the development of consortia with hundreds of thousands of study participants, and resulted in discoveries about the genetic origins of human health and disease.

Sentinel's charge

Assess the use, safety, and effectiveness of regulated medical products by using electronic healthcare data plus other resources

Create data, informatics, and methodologic capabilities to support these activities

Speedily!

What does "Big Data" Offer?

- **Breadth** – large numbers of individuals get us closer to the underlying source population – *potential reduction in selection bias*?
- **Depth** – increasing amount of data on each individual increases the chance that we will have measures of likely confounders – *potential reduction in information bias*?
- **Diversity** – different types of data offer the potential to "cross check" findings for any particular data source – *potential to enhance control for residual bias and/or improve generalizability*?

What is needed to generate actionable evidence?

- Adequate data
  - Medical Product Exposure
  - Health Outcomes of Interest
  - Confounders
- Appropriate method
- To answer the question of interest
- To a satisfactory level of precision

What is unique about Big Data for pediatrics?

- Age by itself is not a barrier (if date of birth is known)
- Issues around exposure to medical products during pregnancy and birth outcomes
  - Complex to link moms and babies to asses birth outcomes
  - Health plan data challenges in days after birth
    - Coded for the mom or baby?
- Unique patterns of care?
  - Critical information dispersed (no data source has a clear view)
    - Hospital, pediatrician, insurer, birth registry, vaccine registry
  - Do kids see more specialists leading to more data dispersion?
  - Care at school?
- Regulatory constraints/ research with minors

What is unique about Big Data for pediatrics?

- Age by itself is not a barrier (if date of birth is known)
- Issues around exposure to medical products during pregnancy and birth outcomes
  - Complex to link moms and babies to asses birth outcomes
  - Health plan data challenges in days after birth
    - Coded for the mom or baby?
- Unique patterns of care?
  - Critical information dispersed (no data source has a clear
    - Hospital, pediatrician, insurer, birth registry, vaccine registry
  - Do kids see more specialists leading to more data dispersion?
  - Care at school?
- Regulatory constraints/ research with minors

Volume

Veracity

Variety / Velocity

Variety

Sentinel partner organizations

Lead – HPHC Institute

DEPARTMENT OF POPULATION MEDICINE
HARVARD MEDICAL SCHOOL
Harvard Pilgrim Health Care Institute

Data and scientific partners

HealthCore Anthem
VANDERBILT SCHOOL OF MEDICINE
health care systems research network
MASSACHUSETTS
HUMANA
HCA Hospital Corporation of America
OPTUM
KAISER PERMANENTE
aetna

Scientific partners

QUINTILES
DukeMedicine
Penn Medicine
UAB
DEPARTMENT OF MEDICINE BRIGHAM AND WOMEN'S HOSPITAL HARVARD MEDICAL SCHOOL
AHIP America's Health Insurance Plans
UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH
HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH
UF College of Pharmacy UNIVERSITY of FLORIDA
UIC
THE UNIVERSITY OF IOWA COLLEGE OF PUBLIC HEALTH
RUTGERS

An ideal distributed network should…

- Accommodate many data holders' data
- Incorporate new kinds of data as they become available
- Maximize local control of data and uses
- Minimize data exchange
- Include local experts in study design and interpretation
- Allow a study protocol to be implemented identically and efficiently across the network
- Support standardized, reusable components
- Generate actionable information

An ideal distributed network should…

- Accommodate many data holders' data
- Incorporate new kinds of data as they become available
- Maximize local control of data and uses
- Minimize data exchange
- Include local experts in study design and interpretation
- **Allow a study protocol to be implemented identically and efficiently across the network**
- **Support standardized, reusable components**
- **Generate actionable information**

These needs lead to a common data model

- Standard data structure allows
  - Partners to execute identical distributed programs locally
  - Development of reusable tools
- Based on Guiding Principles
- Focused on most relevant data domains

Sentinel CDM Oversight

- A strong coordinating center manages the data partner network that actively participates in the creation, implementation, updating, maintenance, enhancement, and use of the Sentinel CDM (SCDM)
- The SCDM …requires that data comparable in **format and meaning** are stored at all sites

For evidence generation and decision making:

Big Data needs Big Curation and Big Expertise

Platelet count units of measure (easier said than done)

Veracity

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	

Raebel MA, Haynes K, Woodworth TS, Saylor G, Cavagnaro E, Coughlin KO, Curtis LH, Weiner MG, Archdeacon P, and **Brown JS.** Electronic Clinical Laboratory Test Results Data Tables: Lessons from Mini-Sentinel. Pharmacoepidemiol Drug Saf. 2014 Feb;23(6):609-18.

Observed result units for HbA1c (easier said than done)

Veracity

*Glycosylated hemoglobin (HbA1c) original result units**

%	%T.HGB	% TL HGB	% HGB
HEMOGLOBIN	%T.Hgb	% OF TOTAL	PERCENT
U	%T.Hgb	% of Hgb	Percent
%HB	% NGSP	% of total	HbA1c%
% OF T	%NGSP	%THb	%HbA1c
%AIC	% TOTAL HGB	%NGSP	% A1C
MG/DL	G/DL	mmol/mol†	Blank
% A1C	% A1c	%Hb	g/dL
NULL	%THb		

Raebel MA, Haynes K, Woodworth TS, Saylor G, Cavagnaro E, Coughlin KO, Curtis LH, Weiner MG, Archdeacon P, and **Brown JS.** Electronic Clinical Laboratory Test Results Data Tables: Lessons from Mini-Sentinel. Pharmacoepidemiol Drug Saf. 2014 Feb;23(6):609-18.

Select Data Model Guiding Principles

- The SCDM is able to incorporate new data types and data elements as future needs indicate
- The SCDM design is transparent, intuitive, well documented and easily understood by analysts, investigators, and stakeholders. It is easy to use by experienced analysts and investigators; special skills or knowledge beyond those commonly found among pharmacoepidemiologist and analysts should not be necessary.
- The SCDM enables interoperability with appropriate evolving healthcare coding standards
- The SCDM captures values found in the source data; any mapping to standard vocabularies are transparent
- Calculated variables should not be stored in the SCDM

SCDM key considerations

- Inclusion of a variable does not imply completeness
- Completeness may vary by source and over time
- Availability of data in the source system does not mean it is usable for research
  - **Especially in a multi-site environment**
- Maintaining standardization is an ongoing and iterative process

For evidence generation and decision making:

Big Data needs Big Curation and Big Expertise

Sentinel Data Partners

aetna℠

OPTUM™

health care systems research network

HealthCore® Anthem®

HCA
Hospital Corporation of America℠

VANDERBILT SCHOOL OF MEDICINE

MASSACHUSETTS

HUMANA®

KAISER PERMANENTE®

CMS.gov
Centers for Medicare & Medicaid Services

Starting 2017

Numerous data elements are available

Demographics and Medical Encounters

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 date(s)	Service date(s)
Drug coverage	Sex	National drug code (NDC)	Encounter ID	Encounter ID	Encounter ID
Medical coverage	ZIP code	Days supply	Encounter type & provider	Encounter type & provider	Encounter type & provider
Medical record availability	Etc.	Amount dispensed	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 and time
Test type, immediacy & location	Height and 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 and time	Administration start and end date and time
Encounter ID	Encounter ID
National Drug Code (NDC)	Transfusion administration ID
Route	Transfusion product code
Dose	Blood Type
Etc.	Etc.

Sentinel distributed database*

- Populations with well-defined person-time for which most medically-attended events are known
- 425 million person-years of observation time
- 43 million people currently accruing new data
- 5.9 billion pharmacy dispensings
- 7.2 billion unique medical encounters
- 42 million people with at least one laboratory test result

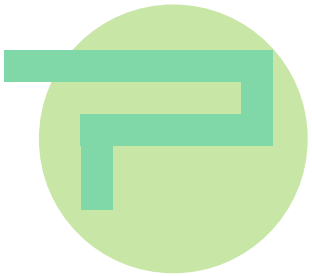
https://www.sentinelinitiative.org/sentinel/snapshot-database-statistics

* As of January 2017

Three ways to address questions

Rapid Analyses

Custom Programs



- Analysis as specified
- Custom inputs, custom output
- Longer execution

Routine Analytic Framework (RAF)

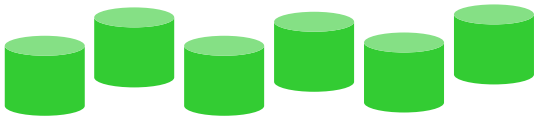
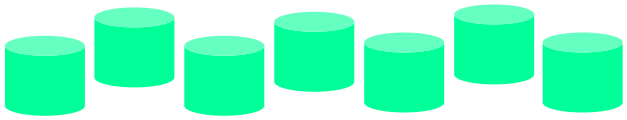
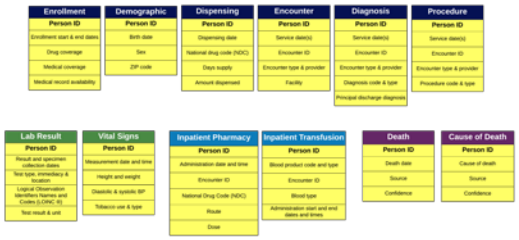
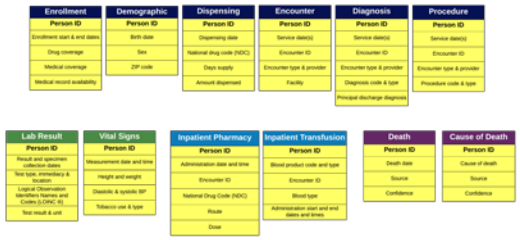
- Off-the-shelf query "templates"
- Standard inputs, standard output
- Quick execution

RAF + custom code

- Hybrid approach: custom code leveraging RAF
- Standard inputs, custom output

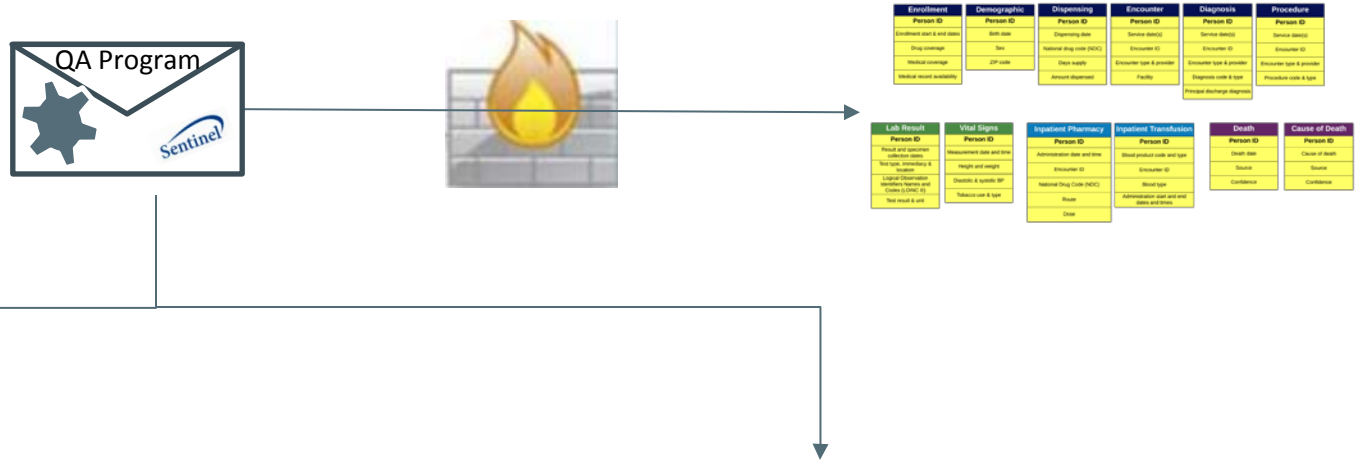
Rapid Response Requires Robust Data Quality Assurance – In Advance of Its Use

The database is dynamic – updates overwrite the preceding data!

	Data Delivery 1	Data Delivery 2
Data Partner Source Database Structure		
		
Transformed database in Sentinel CDM Format		
Timeframe of Data in Database		

The quality assurance process

Send a standard QA program to check DP's data in waiting



Compliance Checks

Level 1: Completeness, validity, accuracy

Level 2: Cross-variable and cross-table integrity

Judgment Call Checks

Level 3: Trends: consistency

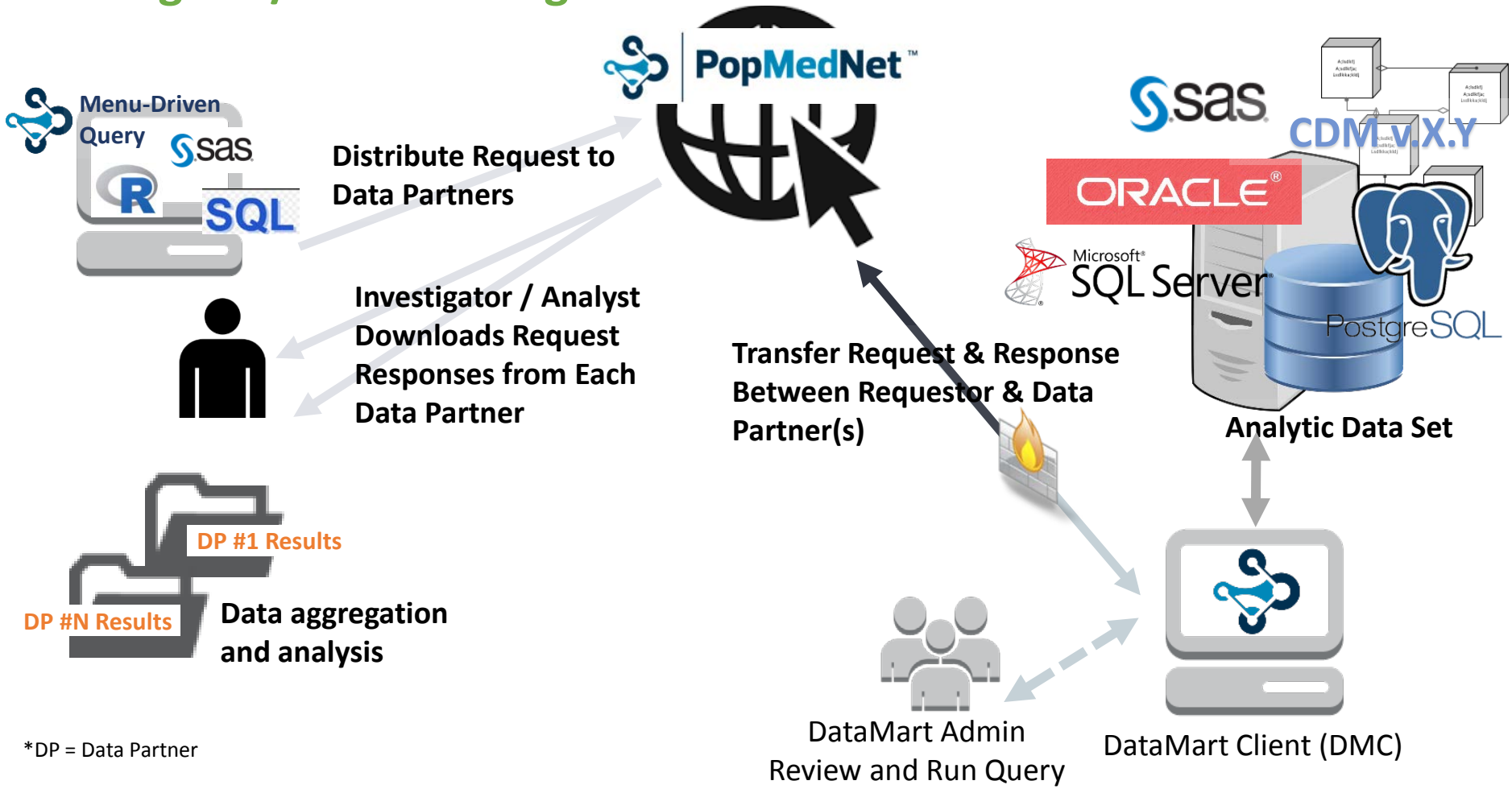
Level 4: Logical: plausibility, convergence

Sentinel quality assurance statistics

- The QA team (six people) reviews ~50 data updates per year from 17 Data Partners
- Since 1/1/2016, the dataset has needed to be re-refreshed and QA package re-run 16 times to fix an issue
- In the latest data deliveries from the 5 largest DPs, 25 checks required DP follow-up
  - 22 of the 25 were Level 3 checks

Investigator/Coordinating Center

Data Partner N



*DP = Data Partner

Examples of Sentinel studies

- Rotavirus and intussusception
- Mother-infant cohort to monitor vaccine safety during pregnancy
- Length of enrollment among adolescents
- Medication use during pregnancy
  - Use of antiemetic drugs
  - TDAP
- Blood transfusion during pregnancy
- Mobile App: collect data from patients

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Intussusception Risk after Rotavirus Vaccination in U.S. Infants

W. Katherine Yih, Ph.D., M.P.H., Tracy A. Lieu, M.D., M.P.H., Martin Kulldorff, Ph.D., David Martin, M.D., M.P.H., Cheryl N. McMahill-Walraven, M.S.W., Ph.D., Richard Platt, M.D., Nandini Selvam, Ph.D., M.P.H., Mano Selvan, Ph.D., Grace M. Lee, M.D., M.P.H., and Michael Nguyen, M.D.

Yih, N Engl J Med. 2014;370:503

FDA Releases Final Study Results of a Mini-Sentinel Postlicensure Observational Study of Rotavirus Vaccines and Intussusception

SHARE | TWEET | LINKEDIN | PIN IT | EMAIL | PRINT

FDA Safety Communication — June 13, 2013

FDA Releases Final Study Results of a Mini-Sentinel Postlicensure Observational Study of Rotavirus Vaccines and Intussusception

FDA Approves Required Revised Labeling for RotaTeq Based on the Study Results

Developing a mother-infant cohort in Sentinel's PRISM Program as a resource to monitor the safety of vaccine use during pregnancy

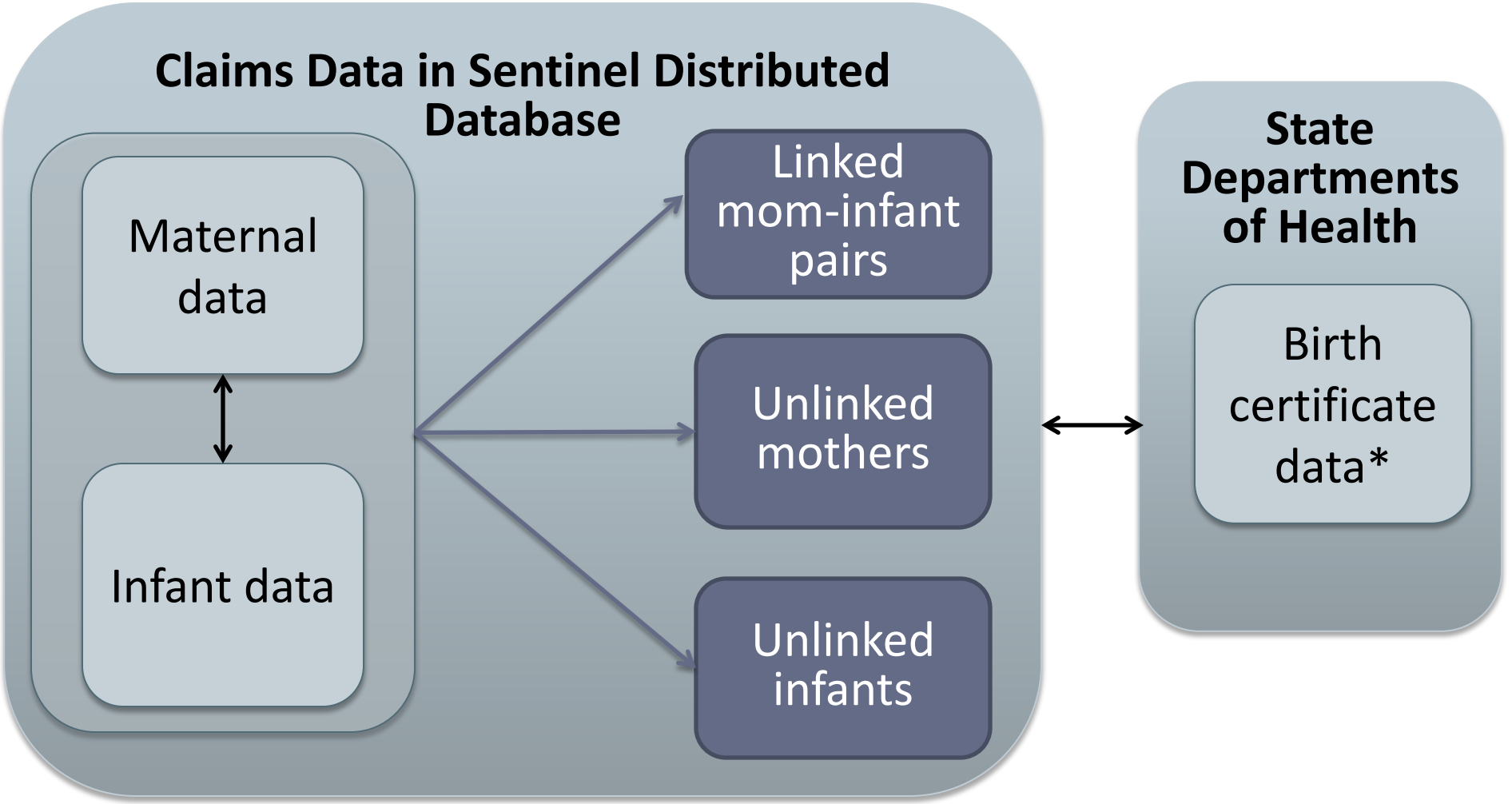
Alison Kawai[1], ScD, Susan Andrade[2], ScD, Robert Rosofsky[3], MA, Lauren Zichittella[1], MPH , Katherine Haffenreffer[1], BS, Cheryl Walraven[4], PhD, MSW, Kevin Haynes[5], PharmD, MSCE, Mano Selvan[6], PhD, Anita M. Loughlin[7], PhD, Azadeh Shoaibi[8], PhD, MS, MHS, Steven Anderson[8], PhD, MPP, Grace Lee[1,9], MD, MPH

[1] Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute; [2] Meyers Primary Care Institute; [3] Health Information Systems Consulting; [4] Aetna Inc.; [5] HealthCore Inc.; [6] Comprehensive Health Insights Inc; [7] OptumInsight Inc; [8] Center for Biologics Evaluation and Research, Food and Drug Administration; [9] Boston Children's Hospital

33rd International Conference on Pharmacoepidemiology & Therapeutic Risk Management; August 26-30, 2017; Palais des congrès de Montreal Montreal, Canada

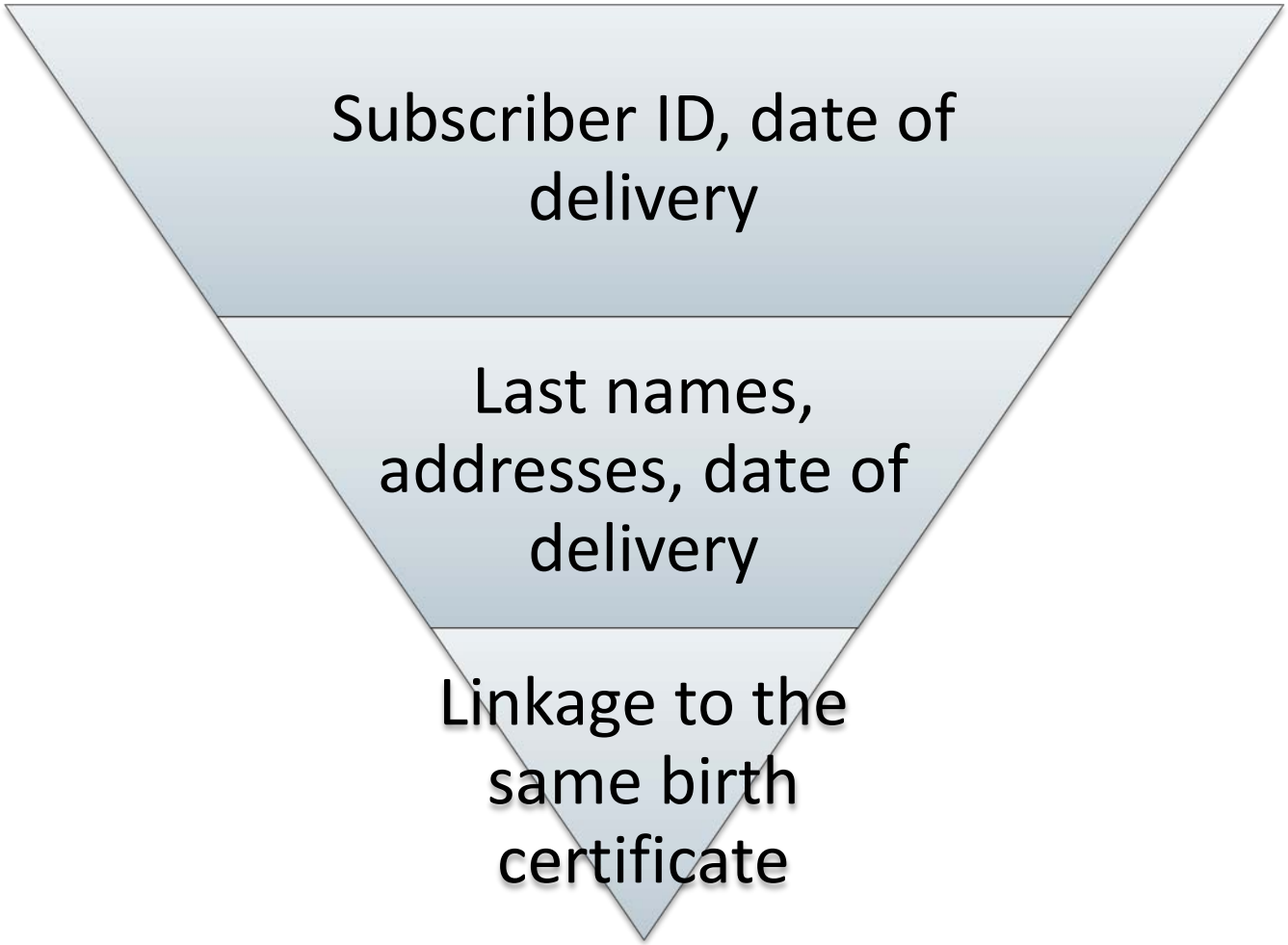
Objective

- To develop capabilities to assess **infant outcomes** following **maternal vaccination** within Sentinel's vaccine safety system
  - Post-licensure Rapid Immunization Safety Monitoring Program (PRISM)
- To develop a **mother-infant cohort**
- To develop and validate a **claims-based gestational age algorithm** within the mother-infant cohort

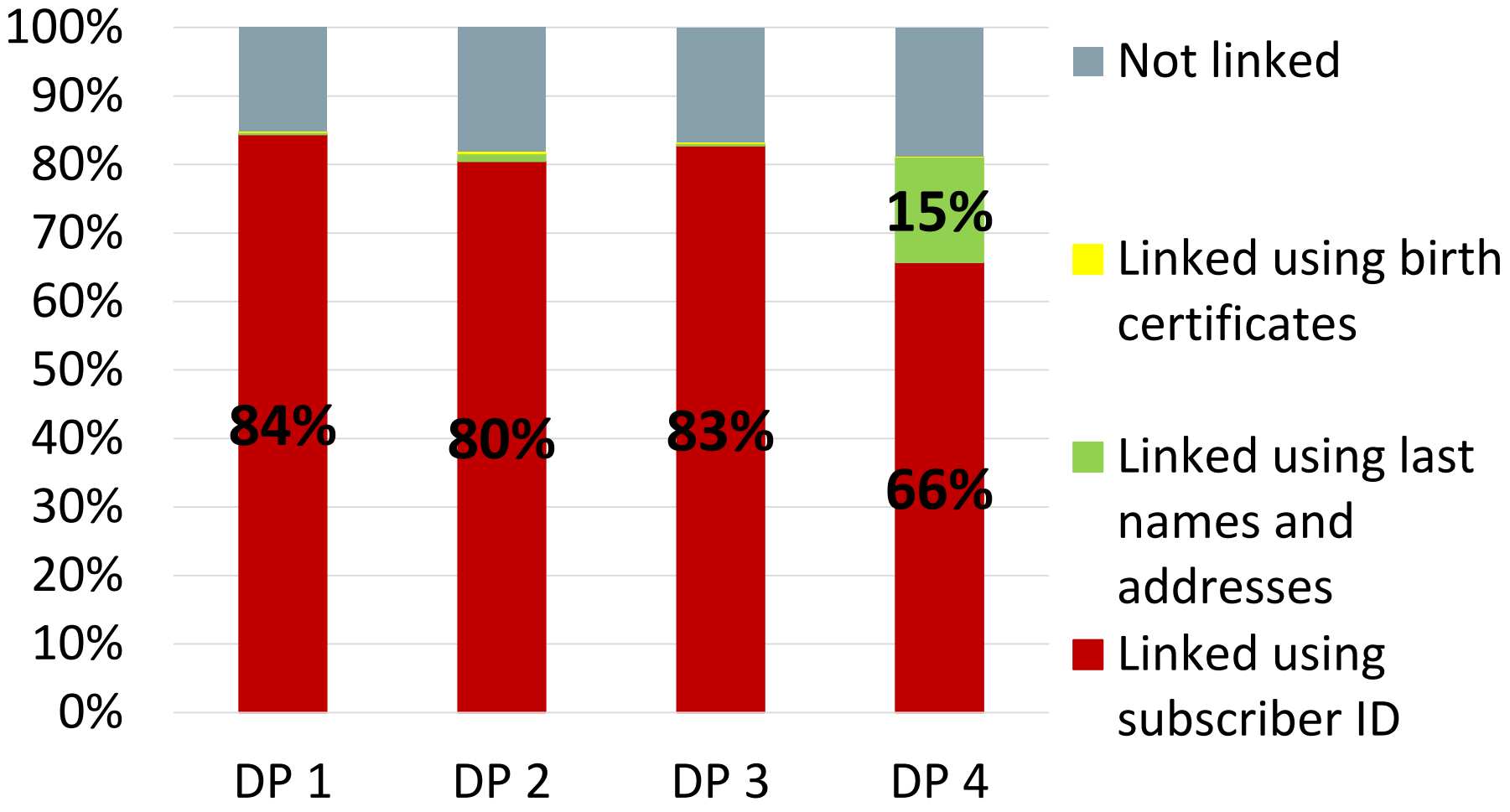


*Birth certificates available for 9 states

Methods to link deliveries to infants

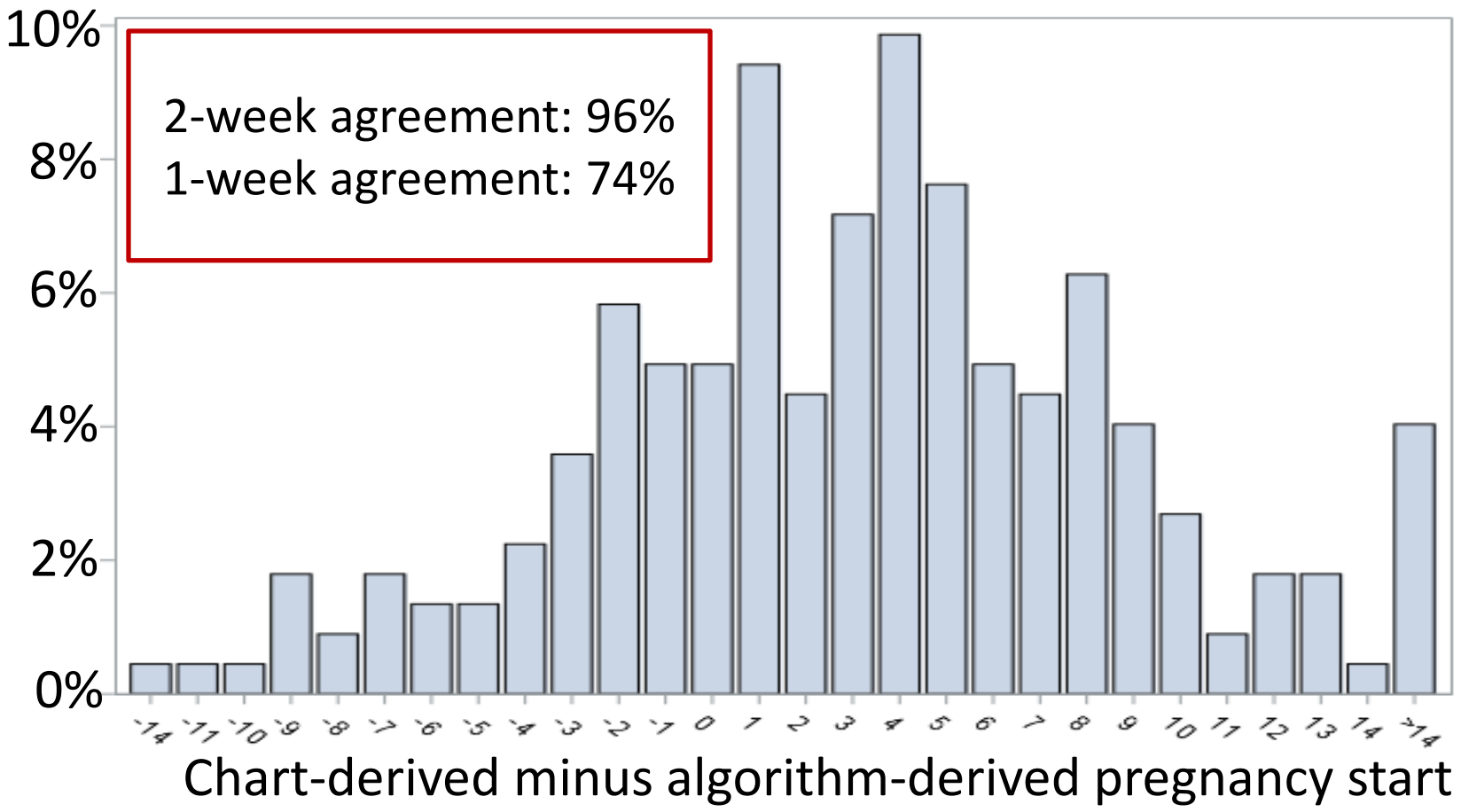


Percent deliveries linked to infants (N=651,607)



Validation of pregnancy start algorithm*
N=223 mother-infant pairs

Veracity



*A total of 313 mother-infant pairs were chart-reviewed

Conclusions

- Successfully linked mothers to infants in 4 large Sentinel Data Partners
- Demonstrated the validity of a claims-based algorithm for pregnancy start
- Supports the feasibility of assessing infant outcomes following maternal vaccination exposures
- Further validation of electronic data elements is needed

Length of enrollment after HPV vaccination

- **Big data needs enough data for study needs**
- 1.94 million new users with 1 year pre exposure and 6 months post exposure enrollment
- 927,000 with 1 year pre exposure and 2 years post exposure enrollment

Longitudinality…

	New HPV Users
Minimum of 365 Days of Enrollment Before Index	
6-Month Enrollment Span After Index	1,940,014
12-Month Enrollment Span After Index	1,558,125
18-Month Enrollment Span After Index	1,178,460
24-Month Enrollment Span After Index	927,484
36-Month Enrollment Span After Index	569,552

https://www.sentinelinitiative.org/drugs/assessments/length-enrollment-among-adolescents

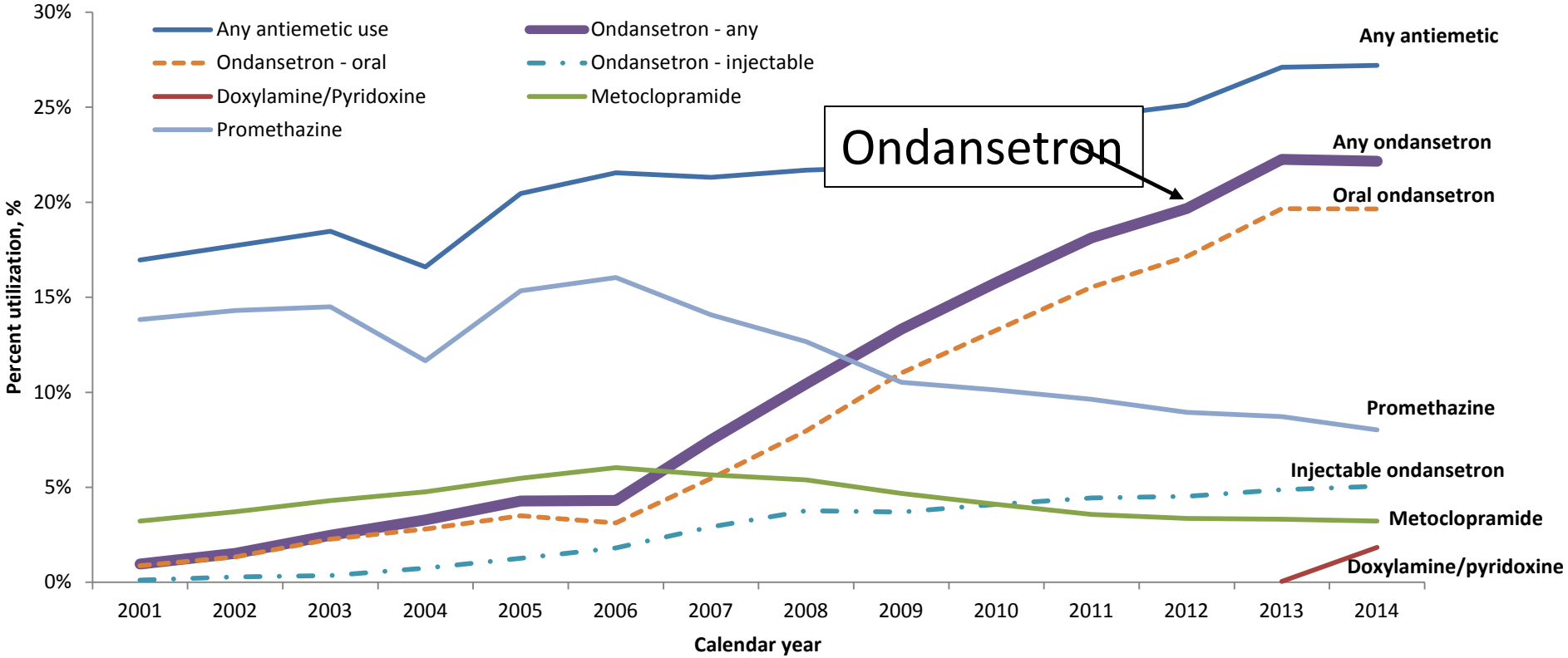
PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2017; **26**: 592–596
Published online 21 February 2017 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.4185

BRIEF REPORT

Antiemetic use among pregnant women in the United States: the escalating use of ondansetron

Lockwood G. Taylor[1]*, Steven T. Bird[1], Leyla Sahin[1], Melissa S. Tassinari[1], Patty Greene[1], Marsha E. Reichman[1], Susan E. Andrade[2], Katherine Haffenreffer[3] and Sengwee Toh[3]

Use of antiemetic drugs among live birth pregnancies in the Sentinel Distributed Database, 2001-2014[a,b]



[a] Dashed lines for oral and injection ondansetron form represent a portion of all total ondansetron use as shown by the solid purple line. Summation of oral and injection utilization sums to greater than total ondansetron use since some women received both products.

[b] Not all Mini-Sentinel data partners contributed data for the entire study period

Taylor. Pharmacoepidemiology and Drug Safety 2017;26:592

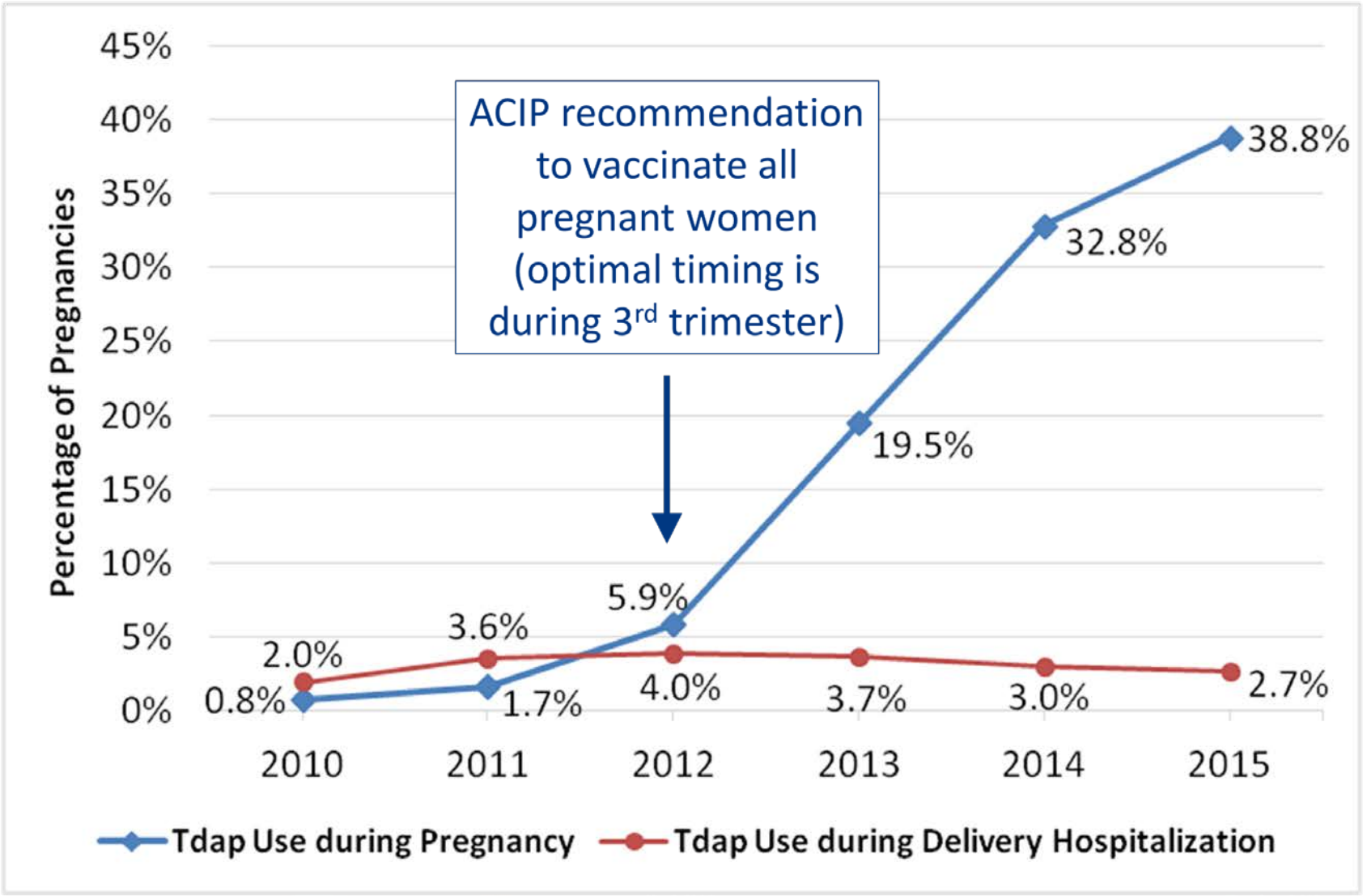
Trends of Tetanus, Diphtheria, and Acellular Pertussis (Tdap) Vaccination during Pregnancy in the Sentinel System

Genna Panucci, SM[1], Kinnera Chada[2], PhD, Hector Izurieta, MD, MPH[2], Azadeh Shoaibi, PhD, MS, MHS[2], Maria Said, MD, MHS[2], Richard Forshee, PhD[2], Joyce Obidi, PhD[2], Andrew Petrone, MPH[1], Noelle Cocoros, DSc, MPH[1], Tiffany Woodworth, MPH[1], Alison Kawai, ScD, SM[1]
[1]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA; [2]Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA

- In 2011, the Advisory Committee on Immunization Practices (ACIP) recommended that unvaccinated women receive Tetanus, Diphtheria, and Acellular Pertussis (Tdap) vaccination during pregnancy to protect infants from pertussis
- In 2012, the recommendation was expanded to include all pregnant women

33rd International Conference on Pharmacoepidemiology & Therapeutic Risk Management; August 26-30, 2017; Palais des congrès de Montreal Montreal, Canada

Figure 2. Tdap Vaccinations during Pregnancy or Delivery by Year



Blood transfusion during pregnancy

- Need for rapid assessment of frequency of transfusion during pregnancy
- Identified 1,946,032 deliveries from 2008-2015 (~8% of U.S. deliveries)
- 21,048 (1.1%) pregnancies had blood transfusion
- Aggregate report across 15 data partners completed within 3 working days of final specification

www.sentinelinitiative.org/vaccines-blood-biologics/assessments/blood-transfusions

Engaging Patients in Evidence Generation

Mobile App Project

Funded by a grant from the Patient Centered Outcomes Research Trust Fund which is overseen by the Office of the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services

Mobile App Study Team

FDA

- David Martin (PI)

Patient Partners

- Kacie Washington
- Karen Byeman

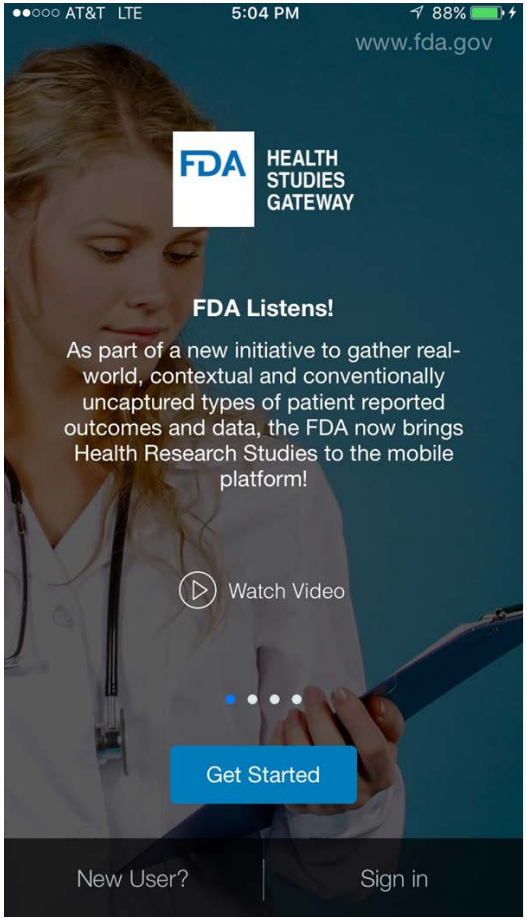
Harvard Pilgrim

- Juliane Reynolds
- Zac Wyner
- Chayim Herzig-Marx

KP Washington

- Sascha Dublin
- Pedja Klasnja
- Linda Kiel
- Catherine Lim
- Deryn Haug
- Ladia Albers-Junkans
- Several testers

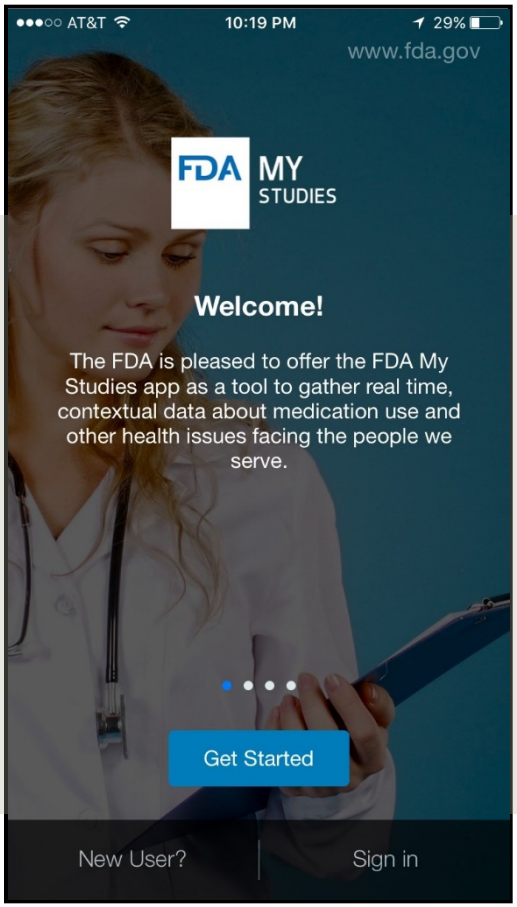
FDA Health Studies Gateway

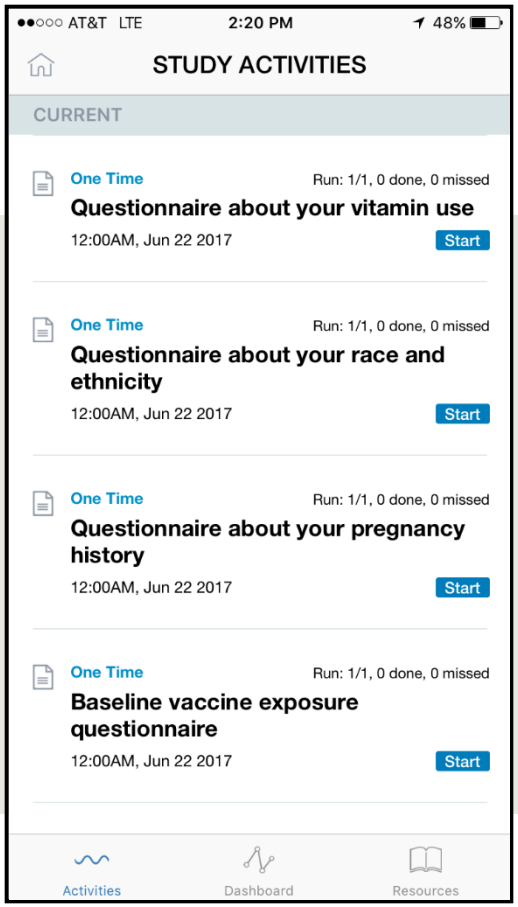


- First effort to link patient-reported data from a mobile platform to the Sentinel Infrastructure
- Study Mobile apps built using Apple ResearchKit and ResearchStack (Android)
- **Initial use case will be medication safety during pregnancy**
- Collaborators include Harvard Pilgrim Healthcare Institute, Kaiser Permanente Washington, LabKey, Boston Technology Corporation, and University of California San Diego

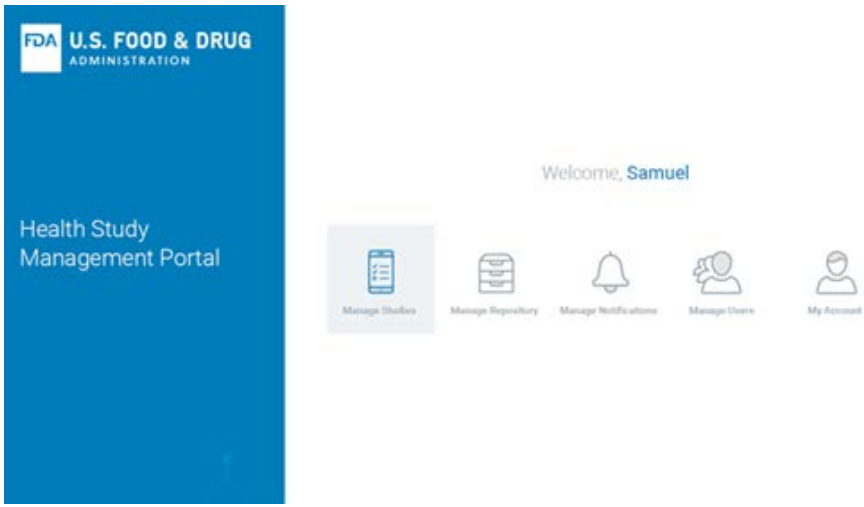
Note: App is not currently active. Wireframes are samples and will be altered before launch.

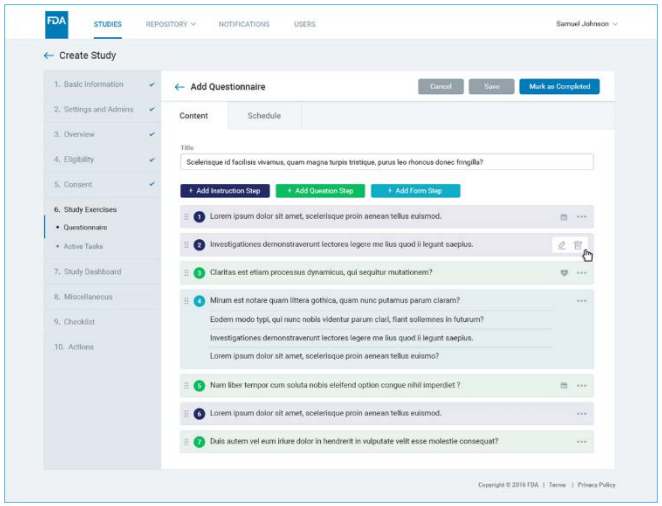
Screenshots from App





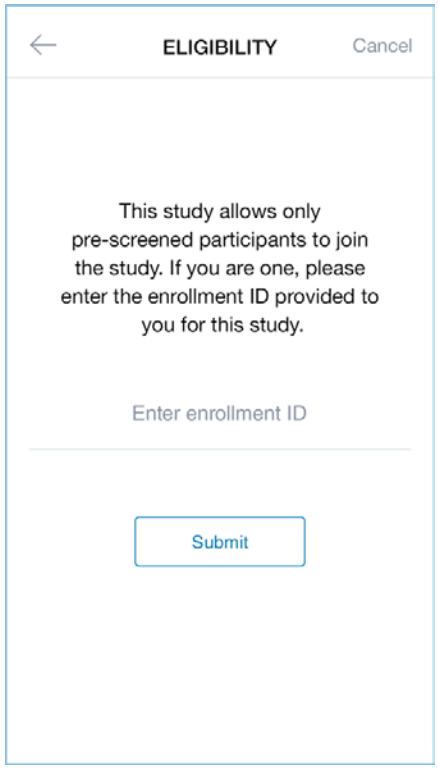
Create

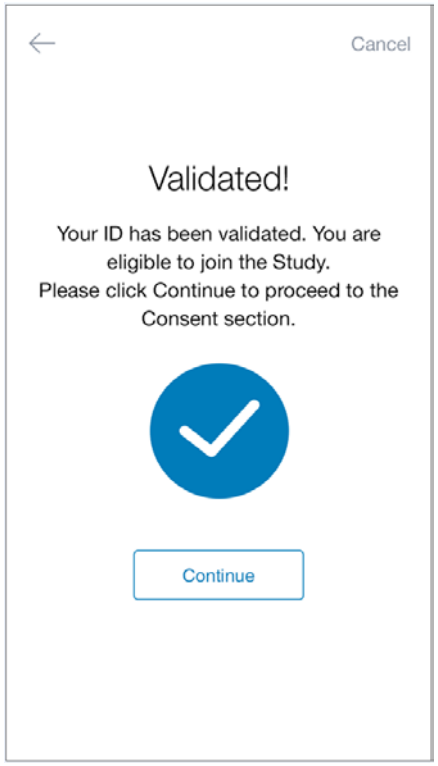


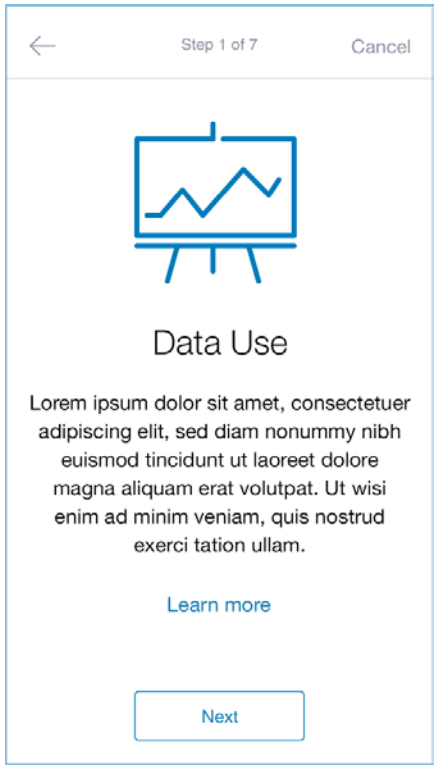


- Configure Study Elements (including questions and active tasks)
- Create patient enrollment tokens and map them to patient IDs

Enroll

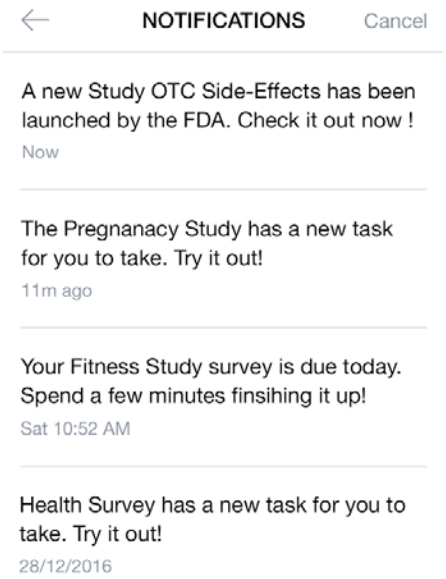


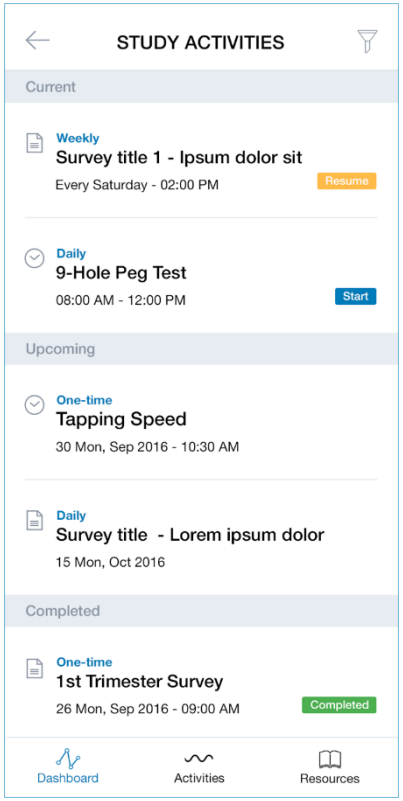


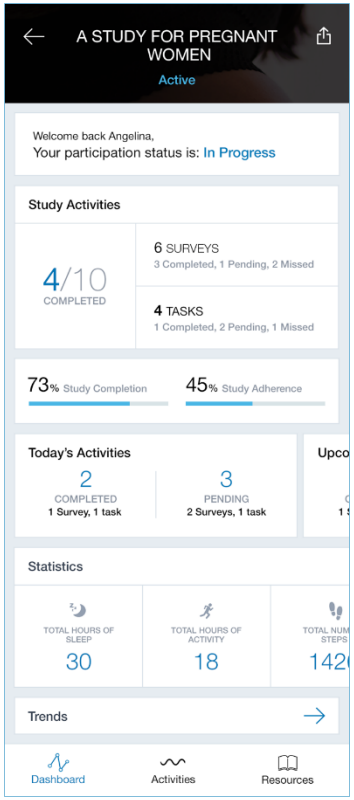


- Select a cohort and distribute enrollment tokens
- Participants download the app in iOS or Android app stores
- Informed consent via the app

Engage

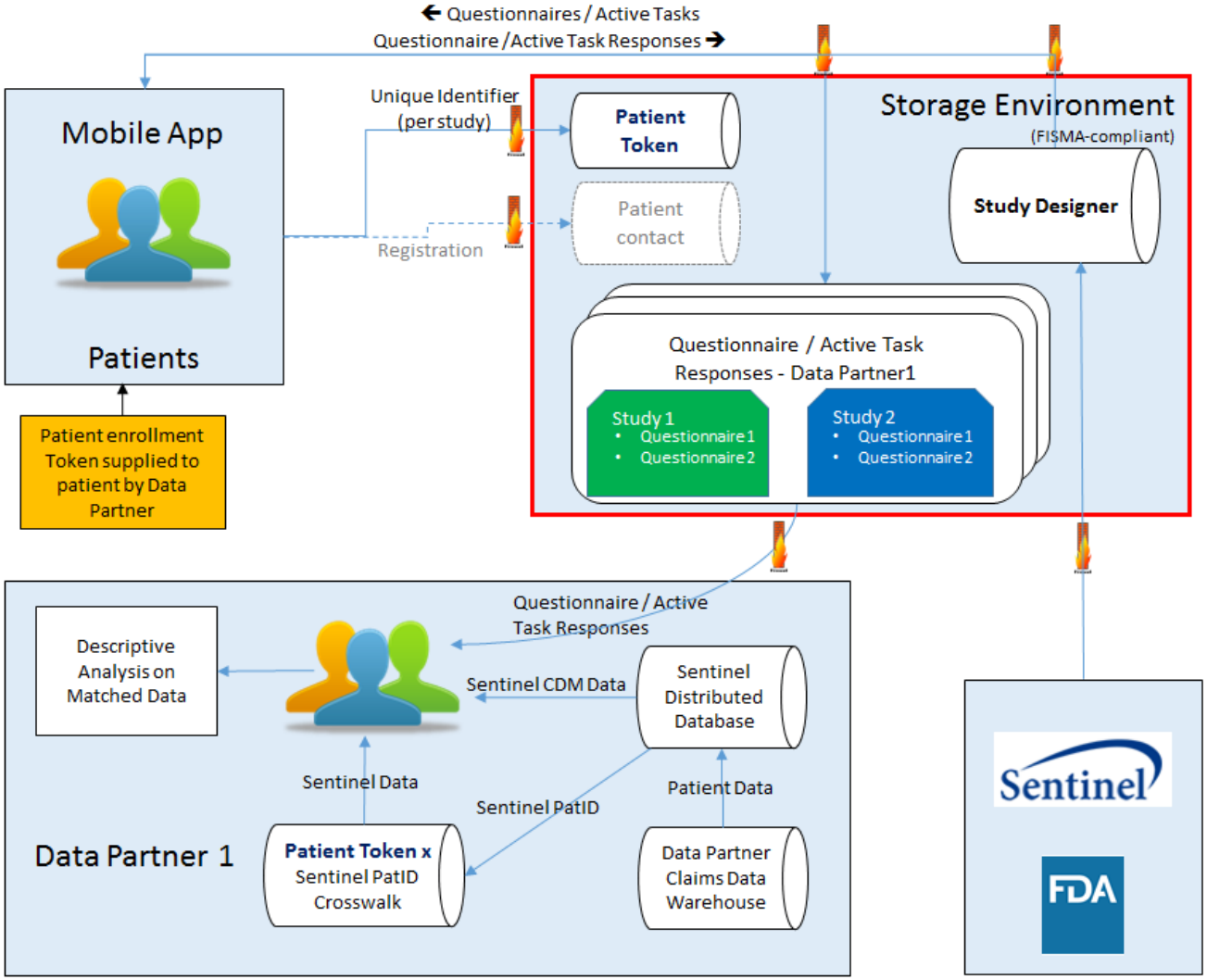






- Data collected directly from patients (eg, due date, pregnancy start date)
- Participants respond when they choose within the study schedule
- Study Dashboard displays progress as well as highlights from data collection

Link Primary and Secondary Data



What's next?

- Incorporation of the mom-baby linked data for routine analyses
- NLP and other approaches to obtain critical data elements difficult to extract or not available in source data (veracity)
  - Pregnancy start
  - Family history
  - Treatment regimens
  - Disease progression
  - Radiologic findings
  - Demographics
  - Test results
- Methods to improve veracity
- Better tools to enable use of dispersed data (variety)
  - Horizontally and vertically partitioned distributed regression
  - Efficient patient finding and linkage
- Approaches for high velocity data (eg, inpatient, social media)
- **Application of research methods to Big Data**

Thank You