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
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.
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.

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 assess 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 assess 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
Sentinel partner organizations

DEPARTMENT OF POPULATION MEDICINE

Lead – HPHC Institute

HPHC Institute

Harvard Medical School

Harvard Pilgrim Health Care Institute

Data and scientific partners

HealthCore

Anthem

Vanderbilt School of Medicine

HCA

Hospital Corporation of America

OPTUM

Kaiser Permanente

Aetna

Scientific partners

PENN MEDICINE

UAB

DEPARTMENT OF MEDICINE

BRIGHAM AND WOMEN’S HOSPITAL

HARVARD MEDICAL SCHOOL

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

info@sentinelsystem.org
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)

<table>
<thead>
<tr>
<th>Blank</th>
<th>FL</th>
<th>TH/UL</th>
<th>X10(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>K/CMM</td>
<td>THOU/CMM</td>
<td>1000/UL</td>
</tr>
<tr>
<td>/100 W</td>
<td>k/cm</td>
<td>thou/cm</td>
<td>X10(3)/MCL</td>
</tr>
<tr>
<td>/CMM</td>
<td>K/CU MM</td>
<td>thou/mm3</td>
<td>X10(3)/UL</td>
</tr>
<tr>
<td>CMM</td>
<td>K/CUMM</td>
<td>THOU/UL</td>
<td>X10(6)/MCL</td>
</tr>
<tr>
<td>10 3 L</td>
<td>K/MCL</td>
<td>THOUS/CU.MM</td>
<td>X10*9/L</td>
</tr>
<tr>
<td>10X3 UL</td>
<td>K/UL</td>
<td>THOUS/MCL</td>
<td>X10E3/UL</td>
</tr>
<tr>
<td>10^3/UL</td>
<td>k/uL</td>
<td>THOUS/UL</td>
<td>X1000</td>
</tr>
<tr>
<td>10*3/uL</td>
<td>KU/L</td>
<td>Thou/uL</td>
<td>X10X3</td>
</tr>
<tr>
<td>10?3/uL</td>
<td>K/MM3</td>
<td>THOUSA</td>
<td>X10^3/UL</td>
</tr>
<tr>
<td>10E3/uL</td>
<td>K/mm3</td>
<td>THOUSAND</td>
<td>x10</td>
</tr>
<tr>
<td>10e3/uL</td>
<td>LB</td>
<td>THOUSAND/UL</td>
<td>X10E3/ul</td>
</tr>
<tr>
<td>10e9/L</td>
<td>PLATELET CO</td>
<td>U</td>
<td>X10E3</td>
</tr>
<tr>
<td>E9/L</td>
<td>T/CMM</td>
<td>X 10-3/UL</td>
<td>K/A?L</td>
</tr>
<tr>
<td>BIL/L</td>
<td>TH/MM3</td>
<td>X 10(3)/UL</td>
<td>K/B5L</td>
</tr>
<tr>
<td>bil/L</td>
<td>th/mm3</td>
<td>X10 3</td>
<td></td>
</tr>
</tbody>
</table>


info@sentinelsystem.org
Observed result units for HbA1c (easier said than done)

<table>
<thead>
<tr>
<th>Glycosylated hemoglobin (HbA1c) original result units*</th>
</tr>
</thead>
<tbody>
<tr>
<td>% HEMOGLOBIN U</td>
</tr>
<tr>
<td>%HB</td>
</tr>
<tr>
<td>% OF T</td>
</tr>
<tr>
<td>% AIC</td>
</tr>
<tr>
<td>NULL</td>
</tr>
</tbody>
</table>

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
Numerous data elements are available

### Demographics and Medical Encounters

<table>
<thead>
<tr>
<th>Enrollment</th>
<th>Demographic</th>
<th>Dispensing</th>
<th>Encounter</th>
<th>Diagnosis</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
</tr>
<tr>
<td>Enrollment start &amp; end dates</td>
<td>Birth date</td>
<td>Dispensing date</td>
<td>Service date(s)</td>
<td>Service date(s)</td>
<td>Service date(s)</td>
</tr>
<tr>
<td>Drug coverage</td>
<td>Sex</td>
<td>National drug code (NDC)</td>
<td>Encounter ID</td>
<td>Encounter ID</td>
<td>Encounter ID</td>
</tr>
<tr>
<td>Medical coverage</td>
<td>ZIP code</td>
<td>Days supply</td>
<td>Encounter type &amp; provider</td>
<td>Encounter type &amp; provider</td>
<td>Encounter type &amp; provider</td>
</tr>
<tr>
<td>Medical record availability</td>
<td>Etc.</td>
<td>Amount dispensed</td>
<td>Facility</td>
<td>Diagnosis code &amp; type</td>
<td>Procedure code &amp; type</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etc.</td>
<td>Etc.</td>
</tr>
</tbody>
</table>

### Clinical

<table>
<thead>
<tr>
<th>Lab Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Result and specimen collection dates</td>
</tr>
<tr>
<td>Test type, immediacy &amp; location</td>
</tr>
<tr>
<td>Logical Observation Identifiers Names and Codes (LOINC ®)</td>
</tr>
<tr>
<td>Test result &amp; unit</td>
</tr>
<tr>
<td>Etc.</td>
</tr>
</tbody>
</table>

### Vital Signs

<table>
<thead>
<tr>
<th>Vital Signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Measurement date and time</td>
</tr>
<tr>
<td>Height and weight</td>
</tr>
<tr>
<td>Diastolic &amp; systolic BP</td>
</tr>
<tr>
<td>Tobacco use &amp; type</td>
</tr>
<tr>
<td>Etc.</td>
</tr>
</tbody>
</table>

### Death

<table>
<thead>
<tr>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Confidence</td>
</tr>
</tbody>
</table>

### Cause of Death

<table>
<thead>
<tr>
<th>Cause of Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Source</td>
</tr>
</tbody>
</table>

### State Vaccine

<table>
<thead>
<tr>
<th>State Vaccine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Admission Type</td>
</tr>
<tr>
<td>Vaccine code &amp; type</td>
</tr>
<tr>
<td>Provider</td>
</tr>
</tbody>
</table>

### Inpatient Pharmacy

<table>
<thead>
<tr>
<th>Inpatient Pharmacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Person ID</td>
</tr>
<tr>
<td>Encouter ID</td>
</tr>
<tr>
<td>Route</td>
</tr>
<tr>
<td>Dose</td>
</tr>
<tr>
<td>Etc.</td>
</tr>
</tbody>
</table>
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

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 Analyses
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

Transformation Program

Data Delivery 1

Data Delivery 2

1/1/2000 1/1/2016

1/1/2000 4/1/2016

info@sentinelsystem.org
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.
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
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. McMahan-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.

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

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 Kawai¹, ScD, Susan Andrade², ScD, Robert Rosofsky³, MA, Lauren Zichittella¹, MPH, Katherine Haffenreffer¹, BS, Cheryl Walraven⁴, PhD, MSW, Kevin Haynes⁵, PharmD, MSCE, Mano Selvan⁶, PhD, Anita M. Loughlin⁷, PhD, Azadeh Shoaibi⁸, PhD, MS, MHS, Steven Anderson⁸, PhD, MPP, Grace Lee¹, 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
Claims Data in Sentinel Distributed Database

- Maternal data
- Infant data
  - Linked mom-infant pairs
  - Unlinked mothers
  - Unlinked infants

State Departments of Health
- Birth certificate data*

*Birth certificates available for 9 states
Methods to link deliveries to infants

- Subscriber ID, date of delivery
- Last names, addresses, date of delivery
- Linkage to the same birth certificate
Percent deliveries linked to infants (N=651,607)

- 84% linked using birth certificates
- 80% linked using last names and addresses
- 83% linked using subscriber ID
- 66% linked using birth certificates
- 15% not linked
Validation of pregnancy start algorithm*
N=223 mother-infant pairs

2-week agreement: 96%
1-week agreement: 74%

*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

<table>
<thead>
<tr>
<th>Duration After Index</th>
<th>New HPV Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum of 365 Days of Enrollment Before Index</td>
<td></td>
</tr>
<tr>
<td>6-Month Enrollment Span After Index</td>
<td>1,940,014</td>
</tr>
<tr>
<td>12-Month Enrollment Span After Index</td>
<td>1,558,125</td>
</tr>
<tr>
<td>18-Month Enrollment Span After Index</td>
<td>1,178,460</td>
</tr>
<tr>
<td>24-Month Enrollment Span After Index</td>
<td>927,484</td>
</tr>
<tr>
<td>36-Month Enrollment Span After Index</td>
<td>569,552</td>
</tr>
</tbody>
</table>

https://www.sentinelinitiative.org/drugs/assessments/length-enrollment-among-adolescents
Antiemetic use among pregnant women in the United States: the escalating use of ondansetron

Lockwood G. Taylor¹* ID, Steven T. Bird¹, Leyla Sahin¹, Melissa S. Tassinari¹, Patty Greene¹, Marsha E. Reichman¹, Susan E. Andrade², Katherine Haffenreffer³ and Sengwee Toh³
Use of antiemetic drugs among live birth pregnancies in the Sentinel Distributed Database, 2001-2014\textsuperscript{a,b}

- Ondansetron - oral
- Ondansetron - injectable
- Doxylamine/Pyridoxine
- Metoclopramide
- Promethazine

\textsuperscript{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.

\textsuperscript{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, SM1, Kinnera Chada2, PhD, Hector Izurieta, MD, MPH2, Azadeh Shoaibi, PhD, MS, MHS2, Maria Said, MD, MHS2, Richard Forshee, PhD2, Joyce Obidi, PhD2, Andrew Petrone, MPH1, Noelle Cocoros, DSc, MPH1, Tiffany Woodworth, MPH1, Alison Kawai, ScD, SM1

1Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA; 2Center 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.
Figure 2. Tdap Vaccinations during Pregnancy or Delivery by Year

ACIP recommendation to vaccinate all pregnant women (optimal timing is during 3rd trimester)
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
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.*
Welcome!

The FDA is pleased to offer the FDA My Studies app as a tool to gather real-time, contextual data about medication use and other health issues facing the people we serve.

Get Started

New User? | Sign in

---

**STUDY ACTIVITIES**

**CURRENT**

- **One Time**
  - **Run:** 1/1, 0 done, 0 missed
  - **Questionnaire about your vitamin use**
  - **Start**

- **One Time**
  - **Run:** 1/1, 0 done, 0 missed
  - **Questionnaire about your race and ethnicity**
  - **Start**

- **One Time**
  - **Run:** 1/1, 0 done, 0 missed
  - **Questionnaire about your pregnancy history**
  - **Start**

- **One Time**
  - **Run:** 1/1, 0 done, 0 missed
  - **Baseline vaccine exposure questionnaire**
  - **Start**

---

**Congratulations! What is your due date?**

<table>
<thead>
<tr>
<th>Month</th>
<th>Day</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>2</td>
<td>2014</td>
</tr>
<tr>
<td>May</td>
<td>3</td>
<td>2015</td>
</tr>
<tr>
<td>June</td>
<td>4</td>
<td>2016</td>
</tr>
<tr>
<td>July</td>
<td>5</td>
<td>2017</td>
</tr>
<tr>
<td>August</td>
<td>6</td>
<td>2018</td>
</tr>
<tr>
<td>September</td>
<td>7</td>
<td>2019</td>
</tr>
<tr>
<td>October</td>
<td>8</td>
<td>2020</td>
</tr>
</tbody>
</table>

Next

Skip this question
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

- Mobile App
  - Patients
    - Patient enrollment
      - Token supplied to patient by Data Partner

- Storage Environment (FISMA-compliant)
  - Study Designer
    - Questionnaire/Active Task Responses - Data Partner 1
      - Study 1
        - Questionnaire 1
        - Questionnaire 2
      - Study 2
        - Questionnaire 1
        - Questionnaire 2

- Data Partner 1
  - Patient Token
  - Sentinel Patient ID (PatID)

- Sentinel
  - Sentinel Data
  - Sentinel Distributed Database
    - Sentinel CDM Data
    - Patient Data
  - Data Partner Claims Data Warehouse
  - Patient Data

- Descriptive Analysis on Matched 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