FDA Approaches to Analytical Challenges Posed by Big Data

David Martin, MD, MPH
Captain, US Public Health Service
Office of the Center Director
Center for Drug Evaluation and Research
Disclaimer and Disclosure

• The views expressed herein are those of the author and should not be construed as FDA’s views or policies

• The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services

• No conflicts of interest to disclose
MOVING FROM DATA TO EVIDENCE
Key terms from the FDA perspective

- **Data** are raw measurements

- **Information** is obtained from data combined with critical context about what is being measured

- **Evidence** is derived from the analysis of information
BIG DATA AND SUFFICIENCY
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 Sufficiency?

• Adequate data
  – Medical Product Exposure
  – Health Outcomes of Interest
  – Confounders

• Appropriate method

• To answer the question of interest

• To a satisfactory level of precision
Administrative Data from Health Plans

- Enrollment files are a source of demographic information as well as confirmation of person-time under observation.
- Claims files include all of the submitted, approved, and paid claims for services covered under medical and/or pharmacy benefits so they are a source of exposures and outcomes.

Graphic developed by Jeffrey S. Brown, PhD
Administrative Data Information Capture

Exposure

- Self-Administered
  - Outpatient Pharmacy Dispensings
  - OTC, Other Non-Billed Rx

- Medically Attended
  - Inpatient Pharmacy, Injection, Transfusion
  - Outpatient Injection, Transfusion

Delayed, Uncertain Onset

Medically Attended Outcomes

- Uncertain Diagnostic Criteria
  - Coding Systems can identify
- Clear Diagnostic Criteria
  - Coding Systems may not identify

Acute Onset
Electronic Health Records

• Positive attributes
  – Additional clinical detail that may relate to intermediate endpoints or add context for temporality
  – Access to Laboratory, Pathology, and Imaging results

• Challenges
  – Observation of person-time: A patient’s care may be documented in more than one Electronic Health Record if they seek care at different institutions or practices
  – Large amount of unstructured data – structured data might not substantially augment administrative data
  – Prescriptions vs. Dispensings
Electronic Health Records - Example

- Hospital Corporation of America (HCA) captures 4-5% of inpatient care in U.S.
  - Potential to provide FDA with visibility for temporal relationship between treatments and outcomes during a hospital episode
  - Cannot define cohorts based on information prior to or after the hospitalization episode
EXPANDING THE DEPTH AND DIVERSITY OF BIG DATA
# Data Linkage

## Administrative

<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>Etc.</td>
<td>Principal discharge diagnosis</td>
<td></td>
</tr>
</tbody>
</table>

## Clinical

<table>
<thead>
<tr>
<th>Lab Result</th>
<th>Vital Signs</th>
<th>Registry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
<td>Person ID</td>
<td>Death</td>
</tr>
<tr>
<td>Result and specimen collection dates</td>
<td>Measurement date and time</td>
<td>Person ID</td>
</tr>
<tr>
<td>Test type, immediacy &amp; location</td>
<td>Height and weight</td>
<td>Cause of Death</td>
</tr>
<tr>
<td>Logical Observation Identifiers Names and Codes (LOINC ®)</td>
<td>Diastolic &amp; systolic BP</td>
<td>Person ID</td>
</tr>
<tr>
<td>Test result &amp; unit</td>
<td>Tobacco use &amp; type</td>
<td>Death date</td>
</tr>
<tr>
<td>Etc.</td>
<td>Etc.</td>
<td>Source</td>
</tr>
</tbody>
</table>

## Registry

<table>
<thead>
<tr>
<th>Death</th>
<th>Cause of Death</th>
<th>State Vaccine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
</tr>
<tr>
<td>Death date</td>
<td>Cause of death</td>
<td>Vaccination date</td>
</tr>
<tr>
<td>Source</td>
<td>Source</td>
<td>Admission Type</td>
</tr>
<tr>
<td>Confidence</td>
<td>Confidence</td>
<td>Vaccine code &amp; type</td>
</tr>
<tr>
<td>Etc.</td>
<td>Etc.</td>
<td>Provider</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Etc.</td>
</tr>
</tbody>
</table>
Incorporating information from patients

- First effort to link patient-reported data from a mobile platform to the Sentinel Infrastructure
- Study Mobile apps built using Apple ResearchKit and ResearchStack
- Initial use case will be medication safety during pregnancy
- Participant engagement using notifications and dashboard with study-specific data visualizations
- Collaborators include Harvard Pilgrim Healthcare Institute, Group Health Research Institute, LabKey, Boston Technology Corporation, and University of California San Diego

Note: App is not currently active. App wireframe is a sample and visual design will change.
Linking Primary and Secondary Data

Note: Schematic representation will change as development continues
GENERATING EVIDENCE
Tradeoffs

Rapid

Unbiased and free from measurement error

Inexpensive in terms of staff time and financial resources
Optimizing Evidence Generation Tradeoffs

• Establish partnerships and build capacity suitable for broad-based evidence generation
• Focus on core data elements and less complex use cases and then expand
• Automate and/or Repurpose processes and concepts when possible
• Use the most parsimonious approach that will still meet regulatory decision making needs
Single Study/Custom Code Approach

Analytical question: Drug Utilization patterns over time

Application to data

One SAS or R script for each study

• Not scalable
• Expensive
• Slow
• Prohibitive to non-expert routine use

Analogy developed by Christian Reich, MD, PhD
Standardized Data and Analytics

Common Data Model standardizes format for distributed analyses

Additional modular programs enhance speed and provide a roadmap for clinical and epidemiologic reasoning

Drug Utilization

Reduced need for custom programming – the backbone becomes a “modular program”

Common Data Model

Analysis developed by Christian Reich, MD, PhD
Real world big data use cases for safety

- **Dabigatran and bleeding**
  - The FDA ascertained that bleeding rates associated with dabigatran, a new drug, were not significantly higher than bleeding rates associated with warfarin, an older drug, despite the large number of postmarket adverse event reports of serious and fatal bleeding events.

- **Olmesartan and sprue-like enteropathy**
  - The FDA confirmed results of case studies that demonstrated increased risk of sprue-like enteropathy (intestinal problems, including severe chronic diarrhea with substantial weight loss) with long-term olmesartan use, but it did not find class effects.

- **Rotavirus vaccine and intussusception**
  - The FDA identified that administration of rotavirus vaccine (Rotateq) led to an increased risk of intussusception (a serious intestinal condition), which was not detected during clinical trials before approval.

- **Influenza vaccine and febrile seizures**
  - The FDA found no increase in risk of febrile seizures in children after receiving vaccination with Fluzone.

Table appears in Ball, R et al., The FDA’s Sentinel Initiative – A Comprehensive Approach to Medical Product Surveillance. *Clinical Pharmacology & Therapeutics*, 0(0):1-4.
TRANSPARENCY – AN IMPORTANT ISSUE WITH BIG DATA
Special Considerations

• Investigators are rarely able to actually share data because use is licensed from data holders and the minimum necessary standard applies.

• Investigators translate clinical constructs into electronic health data specifications and finally into analytic software code – Reporting in publications is often abbreviated.

• Publication bias and “p-hacking”
FDA promotes best practices through Sentinel

- All queries and studies are publicly posted
- Protocols for customized studies are posted prior to execution of primary analyses
- All parameters used for specifications are posted with query results – Basic query software code is also posted
- These actions promote **Replicability** (similar findings with application of the same design and parameters to different large healthcare data sources) or **Reproducibility** (if the same large healthcare data sources are used)
PUBLIC ACCESS TO A HIGH QUALITY EVIDENCE GENERATION SYSTEM
The Role of the Reagan-Udall Foundation for the FDA

• The organization established by the United States Congress to provide a framework for public private partnerships intended to advance regulatory science on behalf of the agency

• RUF is establishing a distributed database modeled on the Sentinel system and provides governance so private-sector entities gain access with appropriate oversight and transparency
  – Sentinel data partners are invited to participate
  – The analytic/coordinating center utilized by the FDA through the Sentinel System also participates
  – Private sector entities may sponsor rapid queries or customized studies
  – Pilot project with Pfizer complete

www.fda.gov
INTEGRATION OF REAL WORLD CLINICAL CARE AND CLINICAL RESEARCH
IMPACT Afib

Implementation of a randomized controlled trial to improve treatment with oral anticoagulants in patients with atrial fibrillation

Collaborators include Harvard Pilgrim Healthcare Institute, Duke University Medical Center, and Healthcore

Patients
- Atrial fibrillation (AF) (two claims)
- CHADS-VASc ≥ 2
- No admission for bleeding in prior 6 months
- Not prescribed anticoagulant for prior 12 months
- Age ≥ 30 years

All Patients Meeting Inclusion and Exclusion Criteria
- Aim to increase the use of oral anticoagulation (OAC) among patients with AF and risk of stroke
- Combined patient and provider level intervention

Randomized Patients
~30,000 patients

Randomized Control Patients

Patient- + provider-level intervention

Primary comparison: difference in the proportion of AF patients started on OAC over the course of the 12-month trial

Secondary outcomes: proportion of days covered with OAC prescription, number of patients on OAC at end of one year; admissions for stroke or bleeding; deaths (subset)
FUTURE USES OF BIG DATA TO SUPPORT REAL WORLD EVIDENCE GENERATION FOR REGULATORY DECISIONS
Enhancing Use of Real World Evidence for Use in Regulatory Decision-Making

Opportunity:
As the ability to generate and use “real-world evidence” (RWE) continues to evolve and grow, it is important that FDA explore the possibilities of using this data to evaluate safety and effectiveness.

Proposed Approach:
• Conduct public workshops to gather input into topics related to the use of RWE for regulatory decision-making.
• Initiate appropriate activities (e.g. pilot studies or methodology development projects) to address key issues in the use of RWE for regulatory decision-making purposes.
• Publish draft guidance on how RWE can contribute to the assessment of safety and effectiveness in regulatory submissions (e.g. supplemental applications, postmarketing requirements).

http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm446608.htm