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

1

MOVING FROM DATA TO EVIDENCE

Key terms from the FDA perspective

— Robert M. Califf, M.D., Commissioner of the U.S. Food and Drug Administration

- **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

http://blogs.fda.gov/fdavoice/index.php/2015/12/what-we-mean-when-we-talk-about-data/

2

BIG DATA AND SUFFICIENCY

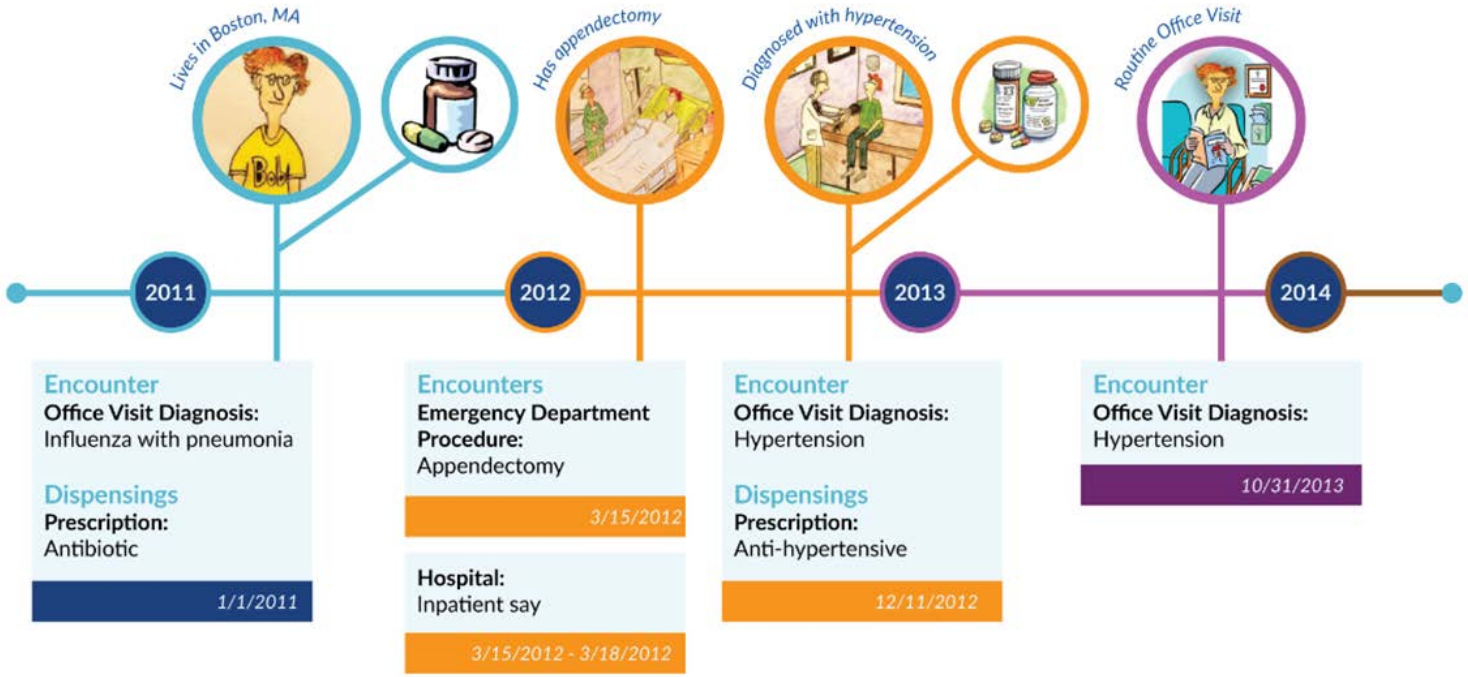
What does “Big Data” Offer?

- **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

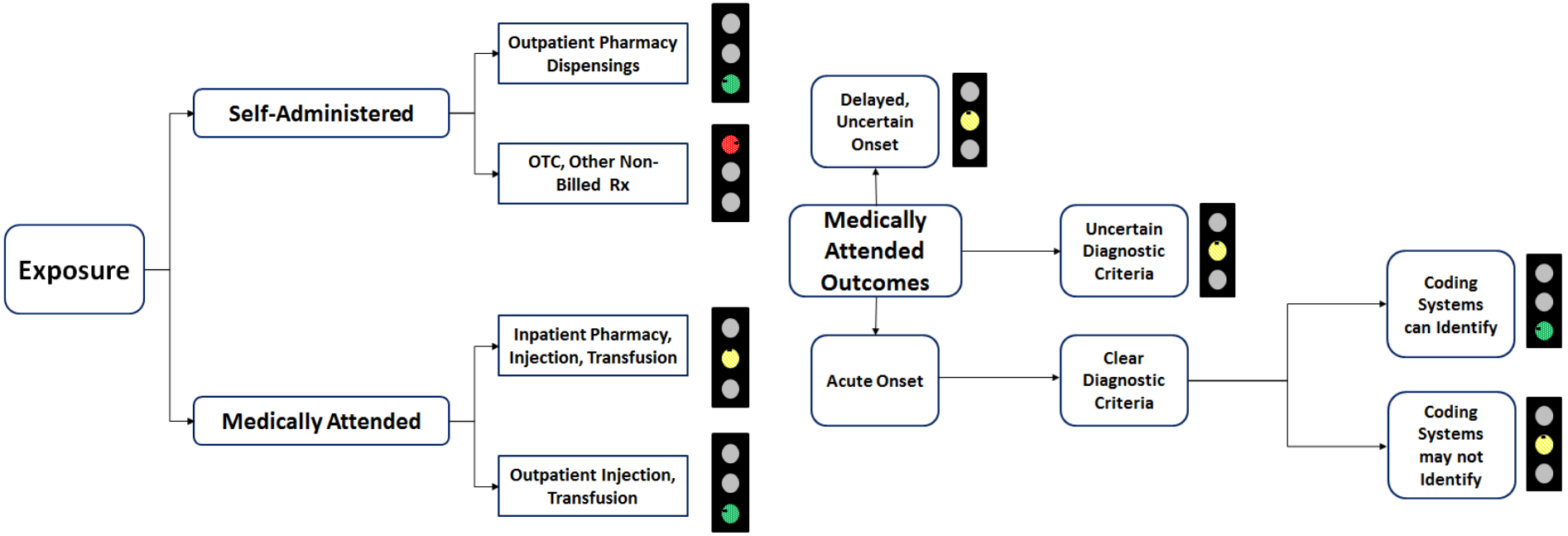
Administrative Data from Health Plans



- 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



Graphics developed by Jeffrey S. Brown, PhD and Kevin Haynes, PharmD, MSCE

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

HCA

Hospital Corporation of America℠

- 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

3

EXPANDING THE DEPTH AND DIVERSITY OF BIG DATA

Data Linkage

Administrative

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.

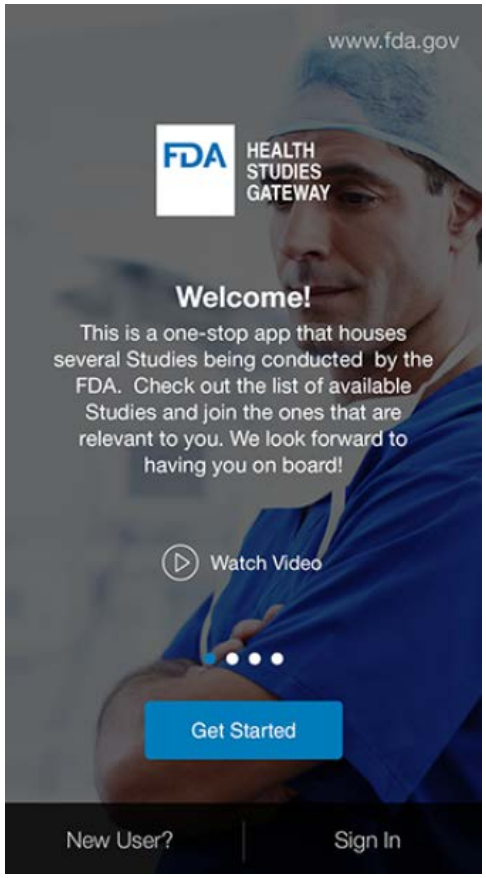
Clinical

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.

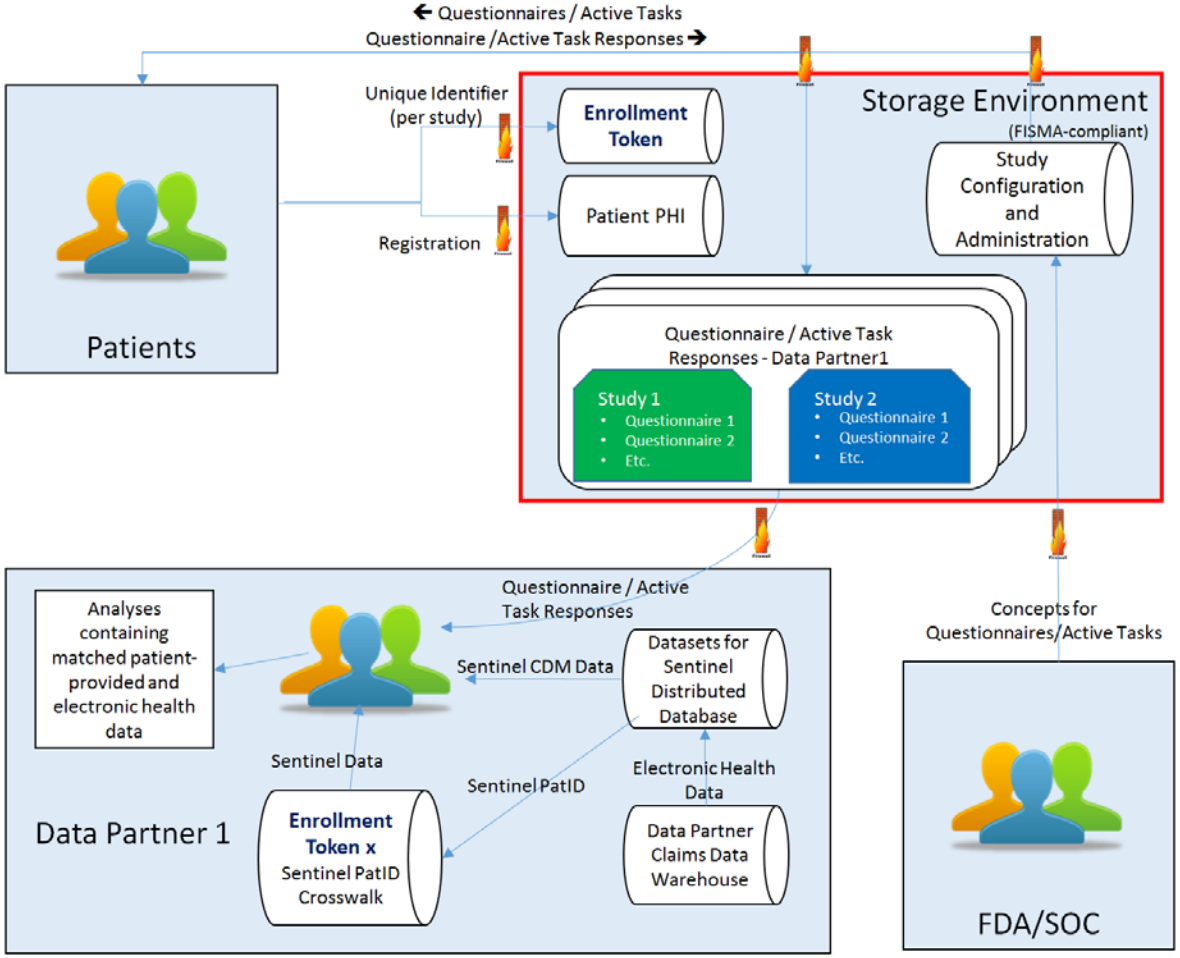
Incorporating information from patients



- 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



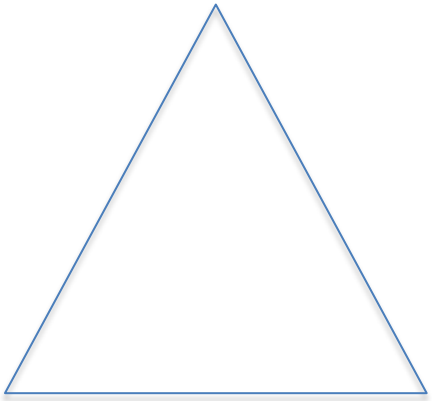
Note: Schematic representation will change as development continues

4

GENERATING EVIDENCE

Tradeoffs

Rapid



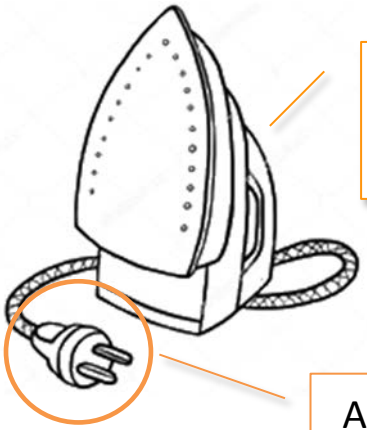
Unbiased and free from measurement error

Inexpensive in terms of staff time and financial resources

Optimizing Evidence Generation Tradeoffs

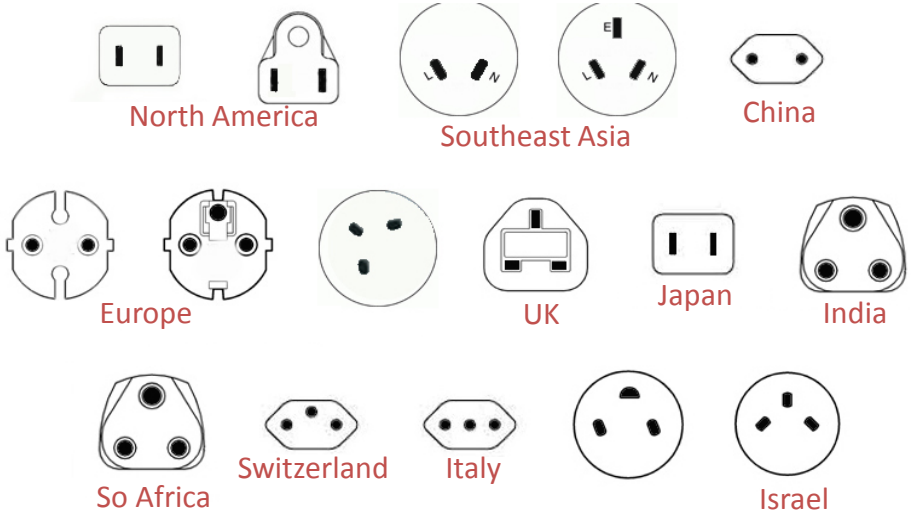
- Establish partnerships and build capacity suitable for broad-based evidence generation
- Focus on core data elements and less complex use cases and then expand
- Automate and/or Repurpose processes and concepts when possible
- Use the most parsimonious approach that will still meet regulatory decision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

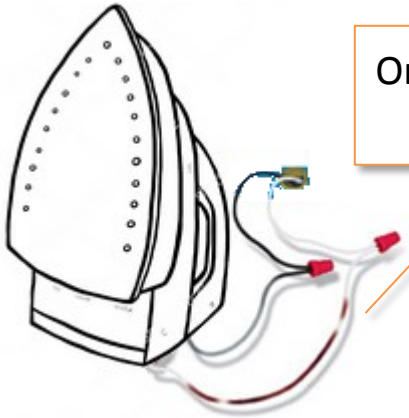
Single Study/Custom Code Approach



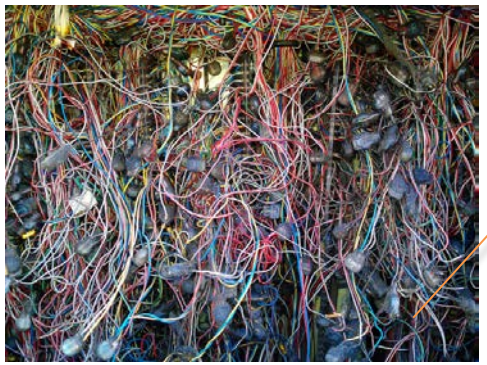
Analytical question: Drug Utilization patterns over time

Application to data





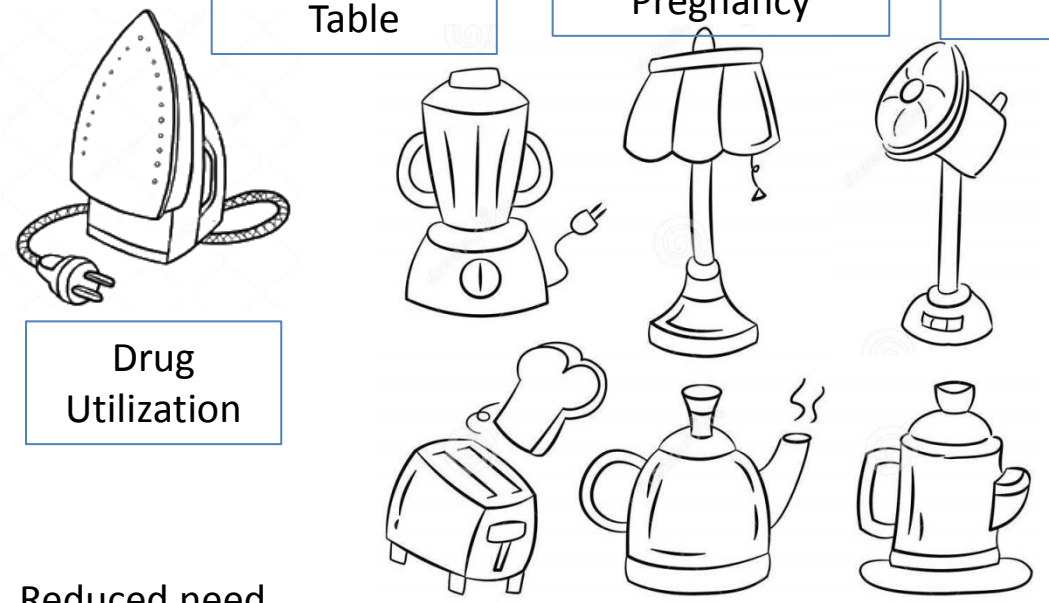
One SAS or R script for each study



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

Analogy developed by Christian Reich, MD, PhD

Standardized Data and Analytics



Summary Table

Utilization in Pregnancy

Pre/Post Exposure Evaluation

Drug Utilization

Concomitant Utilization

SCRI analysis

Propensity score matched analysis

Data Partner 1 DP2 DP3

DP4 DP5 DP6 DP7

DP8 DP9 DP10 DP11

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

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

Common Data Model standardizes format for distributed analyses

Analogy 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.

- Description of drug utilization or health outcome patterns
- Incidence of health outcomes after exposures
- Inferential analyses comparing the incidence of health outcomes among exposed patients or person time to unexposed patients or person time

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.

5

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)

5

PUBLIC ACCESS TO A HIGH QUALITY EVIDENCE GENERATION SYSTEM

Reagan-Udall Foundation
FOR THE FOOD AND DRUG ADMINISTRATION

FDA

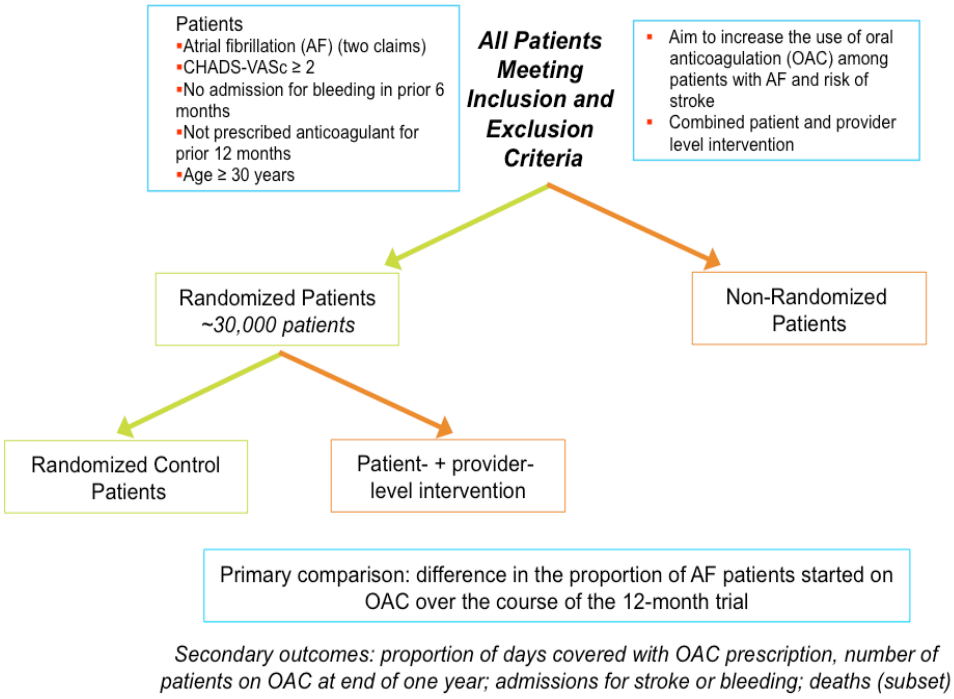
The Role of the Reagan-Udall Foundation for the FDA

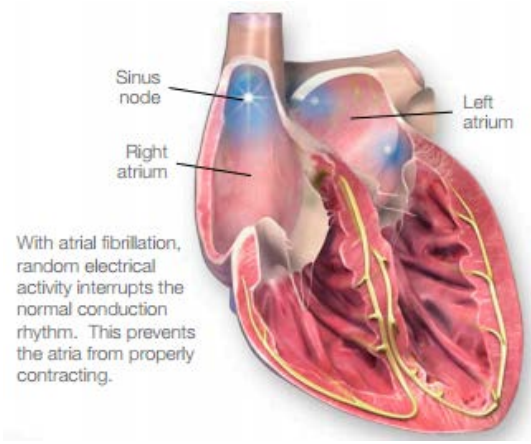
- The organization established by the United States Congress to provide a framework for public private partnerships intended to advance regulatory science on behalf of the agency
- RUF is establishing a distributed database modeled on the Sentinel system and provides governance so private-sector entities 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

5

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

American Heart Association image http://www.heart.org/idc/groups/heart-public/@wcm/@hcm/documents/downloadable/ucm_300294.pdf

6

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


FDA
U.S. FOOD & DRUG
ADMINISTRATION