Predictive Analytics

Perspective for Health TechNet

Joseph Bormel, MD, MPH
July 20, 2018

Copyright 2018
Joseph Bormel, MD, MPH

Predictive Analytics Underlies All Societal Trends

Societal

INTERNET TRENDS 2018

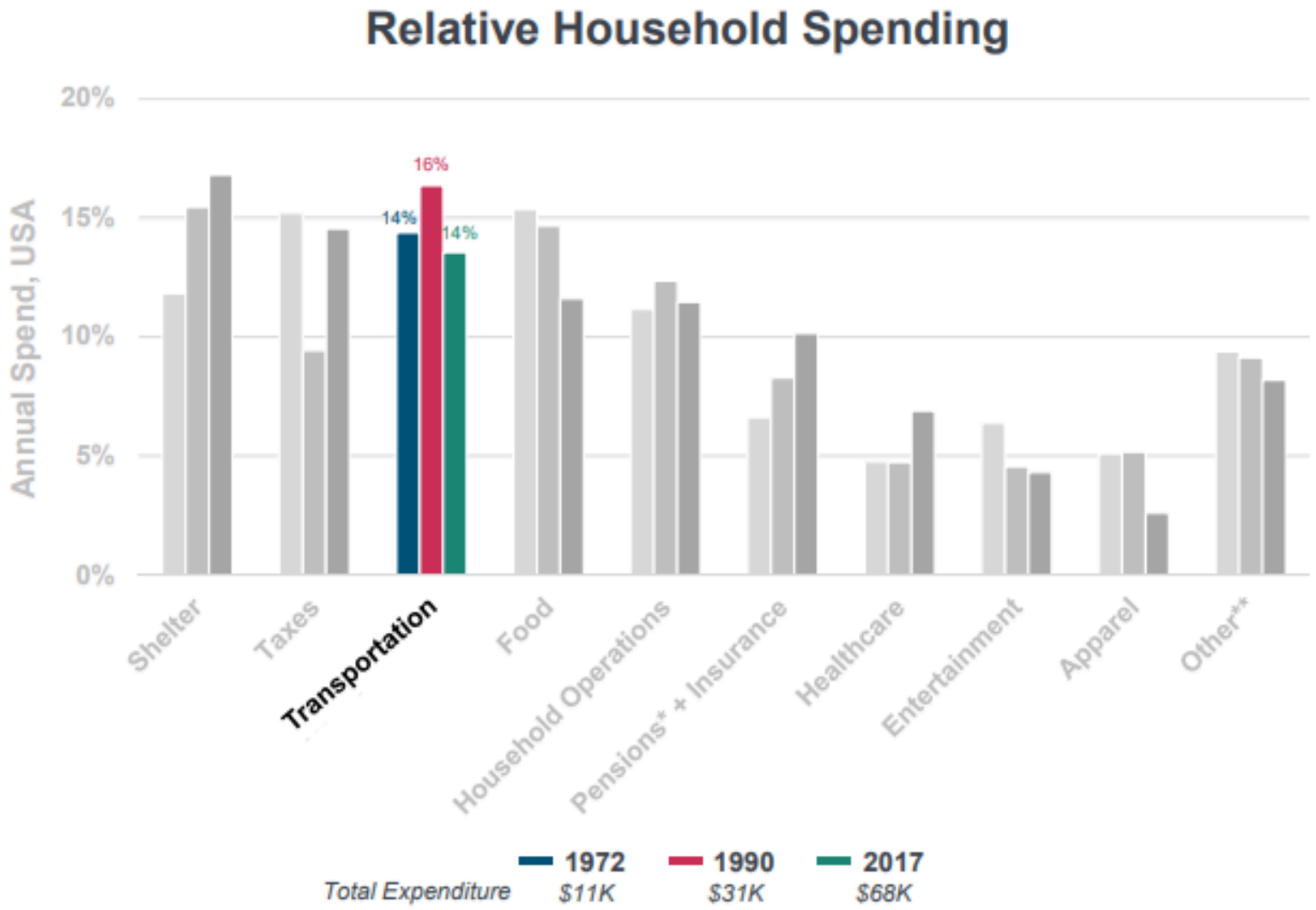
Mary Meeker
May 30 @ Code 2018

KLEINER PERKINS

https://youtu.be/HdjcdZqODoE
http://www.kpcb.com/internet-trends

Copyright 2018
Joseph Bormel, MD, MPH

Transportation as % of Household Spending = 14% vs. 14% (1972)... #3 Segment of $ Spending Behind Shelter + Taxes



KLEINER PERKINS
2018
INTERNET TRENDS

Source: USA Bureau of Labor Statistics (BLS), Consumer Expenditure Survey. *Pensions + Insurance includes deductions for private retirement accounts, social security, and life insurance. **Other includes education and miscellaneous other expenses. Note: Results based on Surveys of American Urban & Rural Households (Families & Single Consumers). 1972 data reflects non-annual survey conducted by BLS + Census Bureau to adjust CPI. 1990 and 2017 Data Based on Annual Survey performed by BLS + Census Bureau. Healthcare costs include insurance, drugs, out-of-pocket medical expenses, etc.. 2017 = mid-year figures.

http://www.kpcb.com/internet-trends

Copyright 2018
Joseph Bormel, MD, MPH

Agenda

- Analytics in Service of Better Health
- Authorized Uses
- Dark Matter Issues
- Opportunities for Surveillance

".. Answer safety questions faster"

The Reagan-Udall Foundation has formed a public-private partnership to provide access for private-sector entities, such as regulated industry, academic institutions, and non-profit organizations, to a system based on the Food and Drug Administration's (FDA) **Sentinel Initiative**. This collaboration works with selected Sentinel data partners and the Harvard Pilgrim Healthcare Institute, functioning as the Analytic or Coordinating Center, to facilitate the analyses of medical product safety evaluations.

http://reaganudall.org/innovation-medical-evidence-development-and-surveillance

Copyright 2018
Joseph Bormel, MD, MPH

Analytics in Service of Better Health

Product/Service Market vs Right

Transparency (Cost, Quality, Access)

Science / Learning system

Goals and Incentives

Social Determinants (SDoH)

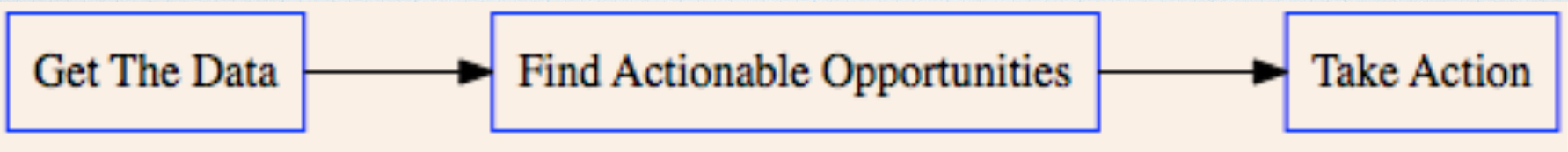
About social determinants of health

The social determinants of health are the conditions in which people are born, grow, live, work and age. These circumstances are shaped by the distribution of money, power and resources at global, national and local levels. The social determinants of health are mostly responsible for health inequities - the unfair and avoidable differences in health status seen within and between countries.

http://www.who.int/social_determinants/sdh_definition/en/

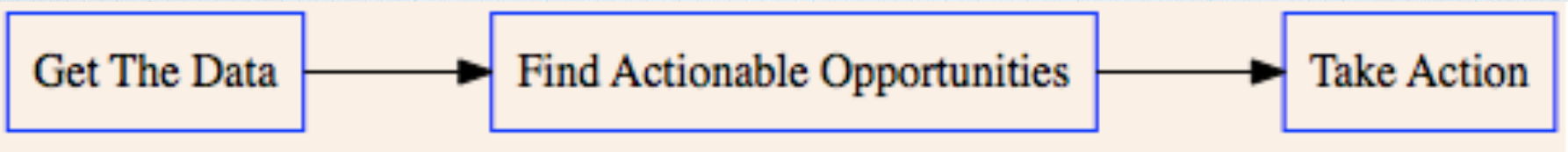
Copyright 2018
Joseph Bormel, MD, MPH

Successful Initiatives



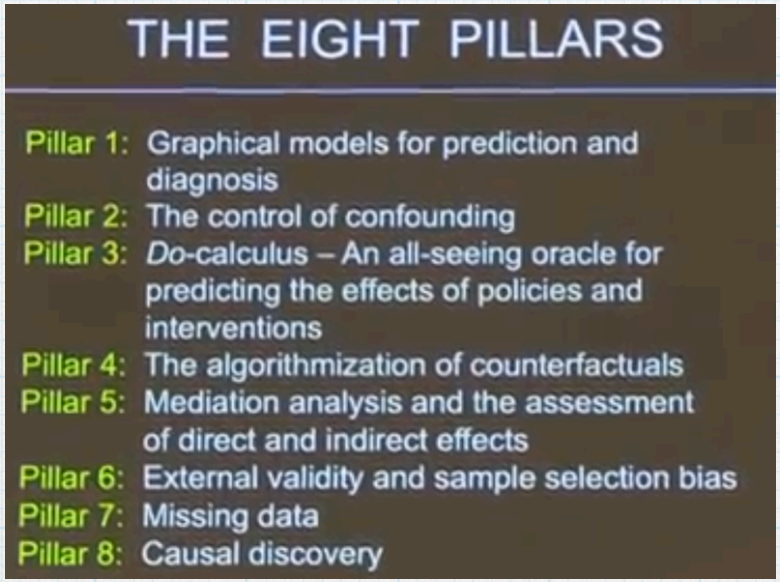
Copyright 2018
Joseph Bormel, MD, MPH

Successful Initiatives



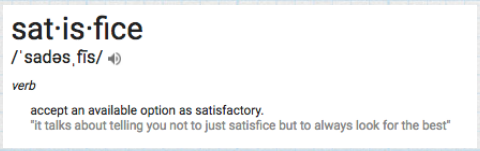
May not exist
May not be accessible
Cost to capture (time, ease)
Competing tasks
  to capturing the patient's story
How do providers
  choose what to record?
Satisfice ("good enough")

See **Data Limitations** slide



The **five rights** include:
the **right** information,
to the **right** person,
in the **right** intervention format,
through the **right** channel,
at the **right** time in workflow.

See "**Taking Action**" slide



See Judea Pearl paper:
https://spp.ucr.edu/wce2017/Papers/eight_pillars_of.pdf

https://healthit.ahrq.gov/ahrq-funded-projects/current-health-it-priorities/clinical-decision-support-cds/chapter-1-approaching-clinical-decision/section-2-overview-cds-five-rights

Copyright 2018
Joseph Bormel, MD, MPH

Taking Action

requires workflow integration and designed orchestration dialogues

2012-AMIA2012-kawamoto-cds copy

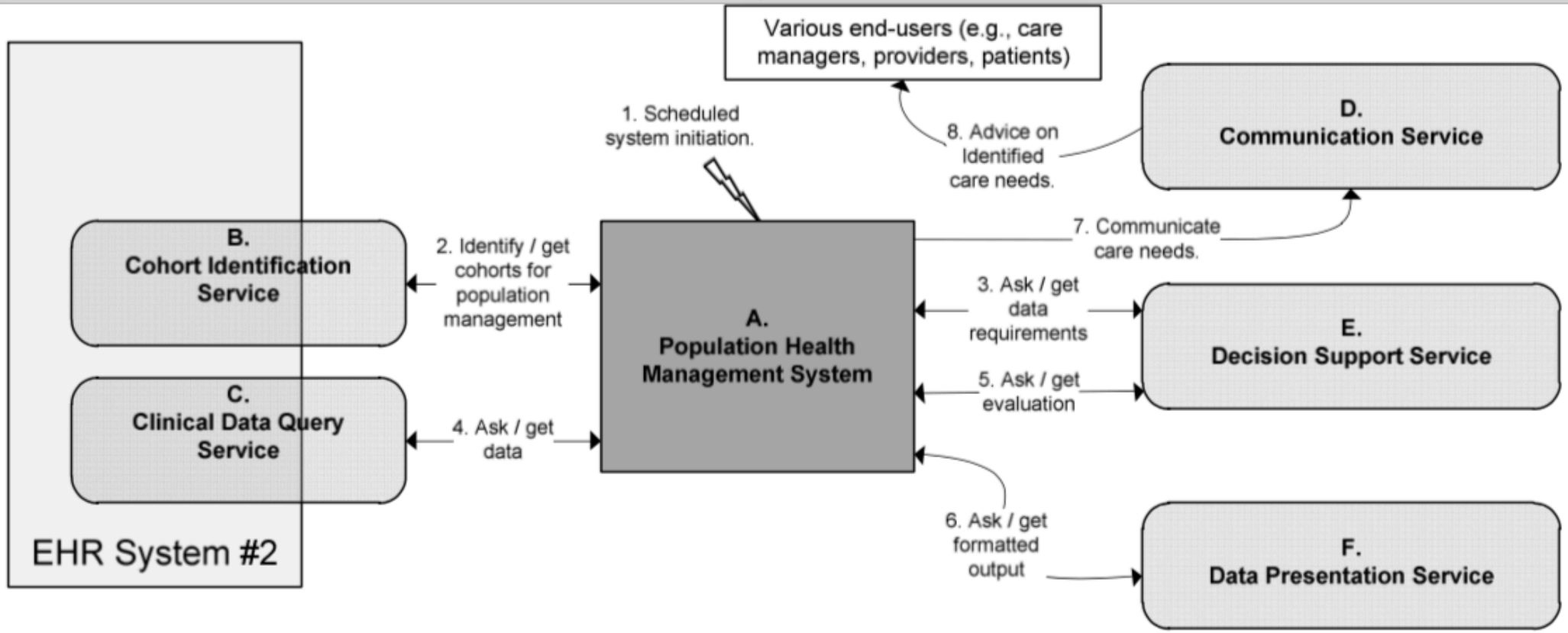


Figure 4. Example of a CIS-consuming SOA for CDS.

From AMIA2012 - Clinical Information System Services and Capabilities Desired for Scalable, Standards-Based, Service-oriented Decision Support: Consensus Assessment of the Health Level 7 Clinical Decision Support Work Group

Kensaku Kawamoto, MD, PhD,[1] Jason Jacobs,[1] Brandon M. Welch, MS,[1]
Vojtech Huser, MD, PhD,[2] Marilyn D. Paterno, MBI,[3] Guilherme Del Fiol, MD, PhD,[1] David Shields,[1] Howard R. Strasberg, MD, MS,[4] Peter J. Haug, MD,[5] Zhijing Liu, PhD,[6] Robert A. Jenders, MD, MS,[7] David W. Rowed, MBBS,[8] Daryl Chertcoff,[9]
Karsten Fehre, DI,[10,11] Klaus-Peter Adlassnig, PhD, MSc,[10,11] A. Clayton Curtis, MD, PhD[12]

Copyright 2018
Joseph Bormel, MD, MPH

Common to make design mistakes

Cannot infer Prevalence from Visits

Need rates by SDoH

Reliability of fatal overdose data >> non-fatal

WNL = “We Never Looked”

Rates of ED visits per 100,000 population involving illicit drugs, 2010

Drugs (1)	Rate of ED visits per 100,000 population (2)
Total ED visits, illicit drugs (3)	**378.5**
Cocaine	157.8
Heroin	72.6
Cannabinoids	152.2
Marijuana	149.0
Synthetic cannabinoids	3.7
Amphetamines/methamphetamine	44.6
Amphetamines	16.7
Methamphetamine	30.7
MDMA (Ecstasy)	7.1
GHB	0.6
Flunitrazepam (Rohypnol)	0.2
Ketamine	0.3
LSD	1.2
PCP	17.3
Misc. hallucinogens	2.0
Inhalants	3.7
Combinations not tabulated above	2.0

(1) The classification of drugs used in DAWN is derived from the Multum *Lexicon*, © 2011 Lexi-Comp DAWN's unique requirements (2010). The Multum Licensing Agreement governing use of the Lexicon http://www.samhsa.gov/data/dawn/MultumLicenseAgreement.pdf.
(2) All rates are ED visits per 100,000 population. Estimates of ED visits are based on a representative hour EDs. Population estimates are drawn from the set of United States Resident Population Estimates Census Bureau.
(3) ED visits often involve multiple drugs. Such visits will appear multiple times in this table (e.g., a vi of visits by drug will be greater than the total, and the sum of percentages by drug will be greater than 1
NOTE: CI = confidence interval. RSE = relative standard error.
SOURCE: Center for Behavioral Health Statistics and Quality, SAMHSA, Drug Abuse Warning Netw

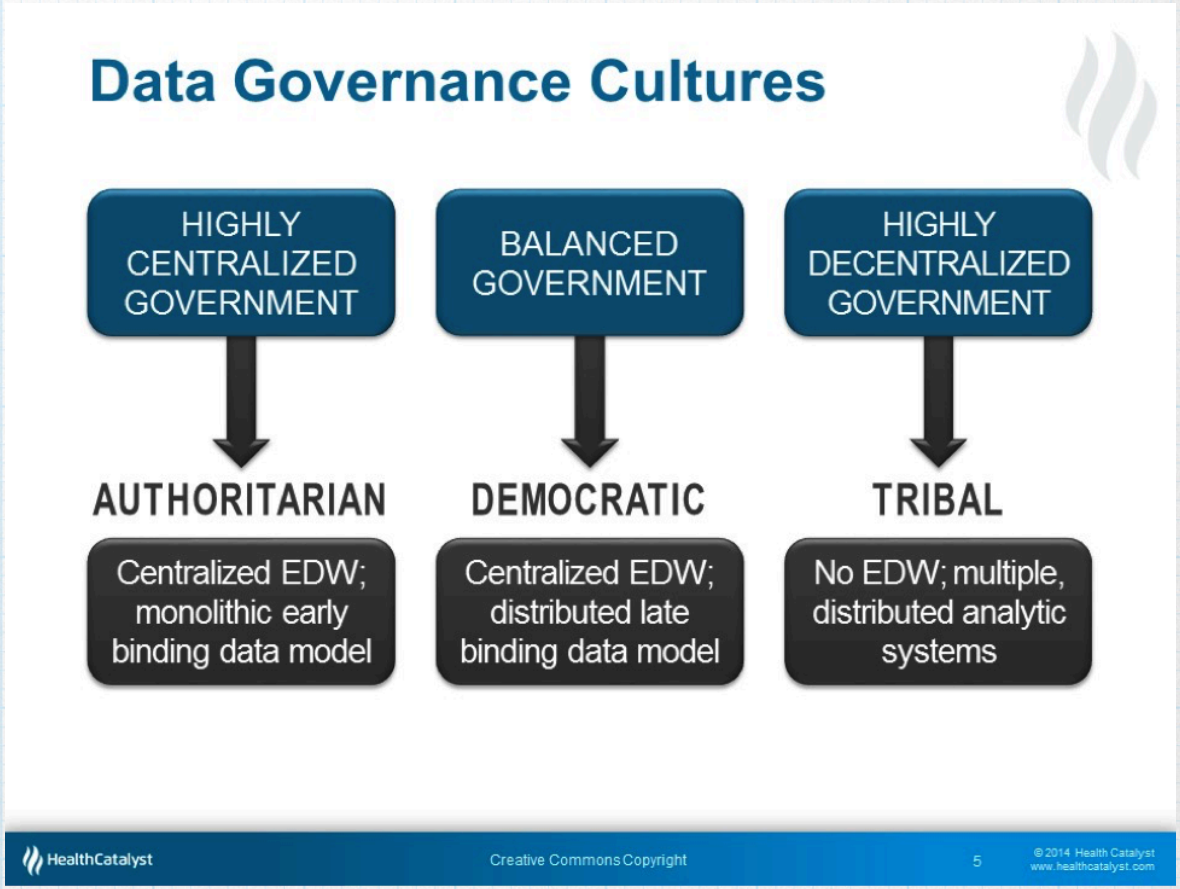
Examples of opiates include:

- Heroin.
- Morphine.
- Oxycodone (trade names include: OxyContin and Percocet).
- Hydrocodone (trade names include: Vicodin and Lortab).
- Codeine.
- Fentanyl.

Copyright 2018
Joseph Bormel, MD, MPH

Authorized Uses

Data use, privacy, confidentiality, **trust** vary widely based on model



Copyright 2018
Joseph Bormel, MD, MPH

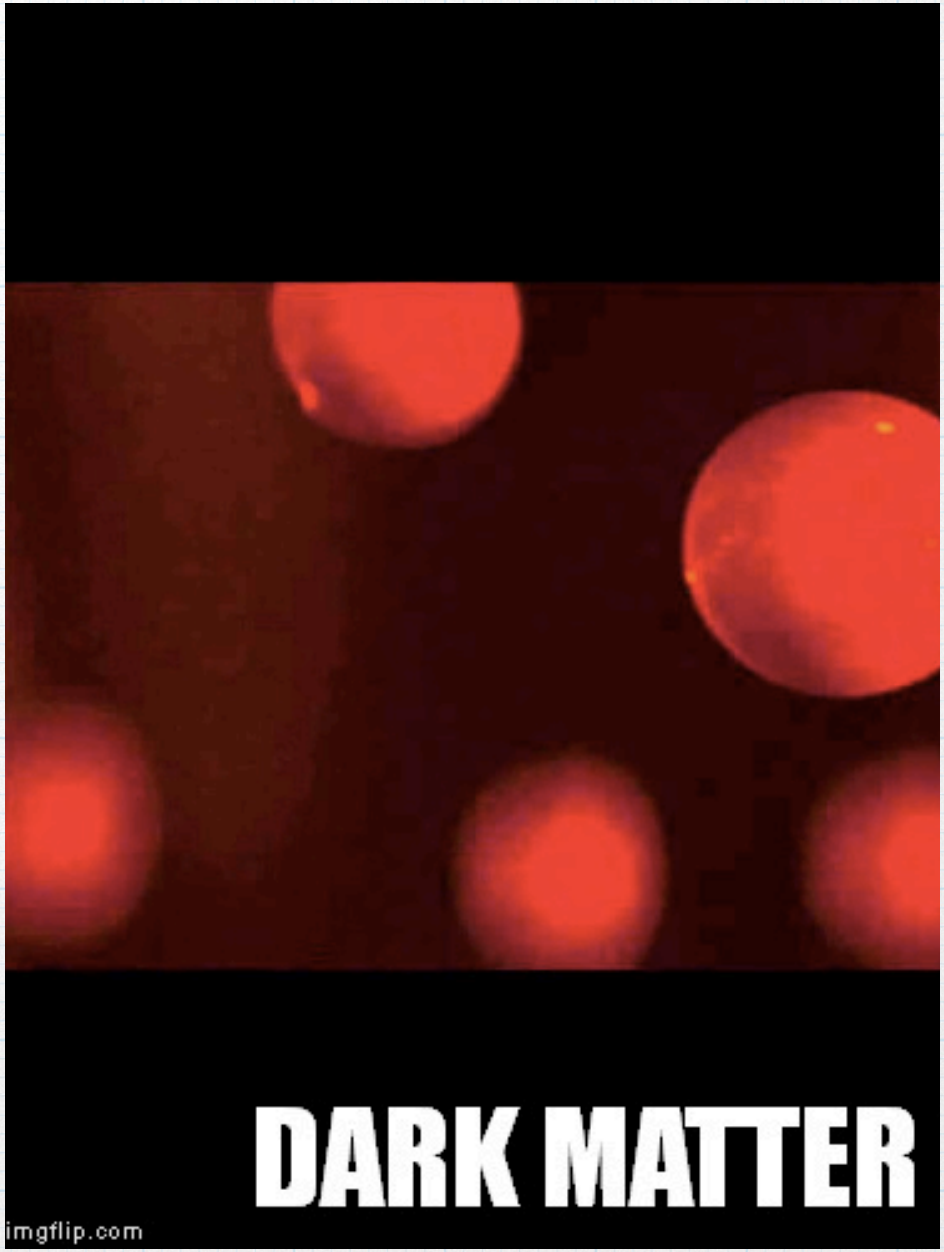
Authorities

	Public Health	Research	Survey	Commercial
Patient Consent and IRB	No	Yes	Implied (voluntary)	No Inaccessible Opt-Out Consent
Collection informed by Hypothesis	No (exploratory)	Yes (confirmatory)	Yes	No Late-binding with massive aggregation / inference*
Duty	dissemination including regulatory; PHA, CIPSEA	Publication	Publication	Proprietary only
Cost Effectiveness	inexpensive (massive resuse)	expensive (often one-off)	depends on design	Yes
Scale	exhaustive depending on data partners	inclusion criteria and availability/recruitment	respondents rate	all consumers on the grid

Cambridge Analytics and Facebook practice
https://www.theatlantic.com/technology/archive/2018/06/what-we-know-about-facebooks-latest-data-scandal/561992/

Copyright 2018
Joseph Bormel, MD, MPH

Dark Matter Issues



Copyright 2018
Joseph Bormel, MD, MPH

Harvard Business Review

DATA

If Your Data Is Bad, Your Machine Learning Tools Are Useless

by Thomas C. Redman

APRIL 02, 2018

ALAN SCHEIN PHOTOGRAPHY/GETTY IMAGES

First, clarify your objectives and assess whether you have the right data to support these objectives.

Second, build plenty of time to execute data quality fundamentals into your overall project plan.

Third, maintain an audit trail as you prepare the training data.

Fourth, charge a specific individual (or team) with responsibility for data quality as you turn your model loose.

Copyright 2018
Joseph Bormel, MD, MPH

Dark Matter Issues

- Predictive analytics (including AI) often presupposes that in the available data is signal that can portent the future.

- Clinical and administrative (claims) data is always incomplete. Data is collection is not free. People are pragmatic. Recording observations is often unsafe.

- Adequate analytics would need to surface the darkness. For example, a gap analysis between ideal model data and what is present in our data would be needed to assess what is known and unknown with any confidence.

- For example, in patients suspected of possible opioid use disorder, what is their relevant history (e.g. use) and social determinate of health (e.g. homelessness, transportation, employment, diet, loneliness, substance use in the residence, incarceration in extended setting).

- Some essential data will almost always be dark, such as intention behind acts. (e.g. Why was that drug given or taken?)

"Roughly 80 percent of the mass of the universe is made up of material that scientists cannot directly observe. Known as dark matter, this bizarre ingredient does not emit light or energy."

https://www.space.com/20930-dark-matter.html

"Dark" is synonymous with "invisible".

https://en.wikipedia.org/wiki/Dark_matter

Copyright 2018
Joseph Bormel, MD, MPH

Seductive Nature of Free Text

* Analyzing free text **can reveal** references to problems such as diseases, medications, allergies, anatomy, medical procedures, and history.
* It is tempting to believe that this information is **necessary and sufficient** to characterize the patient for diagnostic and therapeutic purposes.
* In practice, natural language processing is used to **make sure that we are not losing available information.** That's very important.
* It is rare that an adequacy analysis is performed. When they are performed, embarrassing issues of data integrity are surfaced which can threaten the initiative.

Copyright 2018
Joseph Bormel, MD, MPH

Data Capture Issues

- What gets **said** by patient?
- What gets **documented** by provider?
  - What was heard?
  - How hard is it to document. (Does that change with SMART on FHIR, or other)
  - How safe is it to document (e.g. liability, pre-existing conditions, employment, legal citizenship, insurance coverage associated with suicide, etc)
- **How** is it said?
  - e.g. “Suicide Attempt” versus “Intentional Overdose”
- Other important factors
  - **Reputation** Index of the information source
  - How **aligned** is information with models needed for purpose (e.g. Opioid Crisis)
  - **Jurisdictional** Issues (Federal, State, County requirements, definitions, even sales tax)
  - Meets needs of those documenting in terms of bidirectional, community (**internal use**)

Copyright 2018
Joseph Bormel, MD, MPH

Design Limitations

using the Drug Abuse Warning Network as a typical example

* There was little public documentation of DAWN's **data quality**, and a few published studies raised concerns that measurement error in DAWN was substantial and systemic.
* Because DAWN monitors **episodes**, not **individuals**, data are misinterpreted if analysts view them as **prevalence** measures.
* There is an **uncertain**, and perhaps **inconsistent**, relationship between the number of DAWN episodes and the true level of drug abuse problems in a given area.
* **Delays in reporting** DAWN data undermined its potential utility as an early warning mechanism for emerging drug problems.
* DAWN data were often **assumed to represent heavy or chronic drug users**, although the validity of that assumption has never been assessed. (What was sampled?)

  [In sum, there were problems with timeliness, quality, representation and sharing of data.]

From: 2002-08 Drug Abuse Warning Network-- Development of a New Design - orig 10.1.1.444.3232 .pdf, pages 11,12

Copyright 2018
Joseph Bormel, MD, MPH

Re-Designs are Common

Check for updates

PUBLIC HEALTH Reports

Enhancing Surveillance Systems

The Evolution of BioSense: Lessons Learned and Future Directions

Public Health Reports
2017, Vol. 132(Supplement 1) 7S-11S
© 2017, Association of Schools and Programs of Public Health
All rights reserved.
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0033354917706954
journals.sagepub.com/home/phr
SAGE

Deborah W. Gould, PhD[1], David Walker, MPH[1], and Paula W. Yoon, ScD, MPH[1]

Abstract

The BioSense program was launched in 2003 with the aim of establishing a nationwide integrated public health surveillance system for early detection and assessment of potential bioterrorism-related illness. The program has matured over the years from an initial Centers for Disease Control and Prevention–centric program to one focused on building syndromic surveillance capacity at the state and local level. The uses of syndromic surveillance have also evolved from an early focus on alerts for bioterrorism-related illness to situational awareness and response, to various hazardous events and disease outbreaks. Future development of BioSense (now the National Syndromic Surveillance Program) includes, in the short term, a focus on data quality with an emphasis on stability, consistency, and reliability and, in the long term, increased capacity and innovation, new data sources and system functionality, and exploration of emerging technologies and analytics.

Keywords

Biosense, syndromic surveillance, emergency department data, public health surveillance

be reporting these data to BioSense. To speed up data acquisition, CDC established servers in large metropolitan hospitals, and data were sent directly to BioSense. By 2007, only about 10% of US EDs were reporting, and BioSense stakeholders became concerned with the slow rate of adding hospitals and increasing the coverage of EDs across the country. State and local public health officials expressed concern about CDC recruiting data sources directly rather than working through the public health departments. Additionally, because syndromic surveillance was a new and untested methodology, state and local health department epidemiologists raised questions about data quality, validity, and utility, and members of the US Congress expressed concerns about the number of false alerts and the cost and effectiveness of BioSense as an early warning system for bioterrorism.[7,8]

2017-07-10-pub Deborah W. Gould, PhD, pJB - The Evolution of BioSense- Lessons Learned and Future Directions- 0033354917706954.pdf

Copyright 2018
Joseph Bormel, MD, MPH

Re-Designs are Common

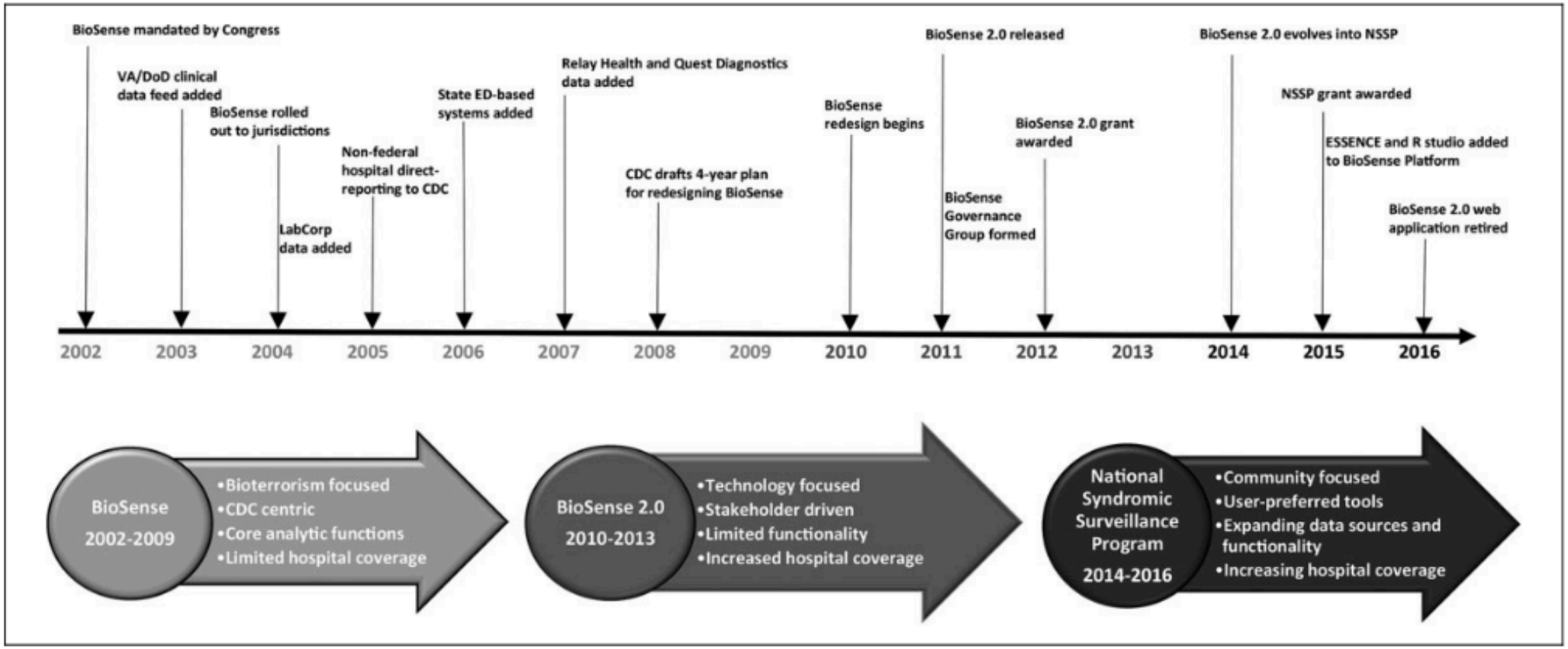
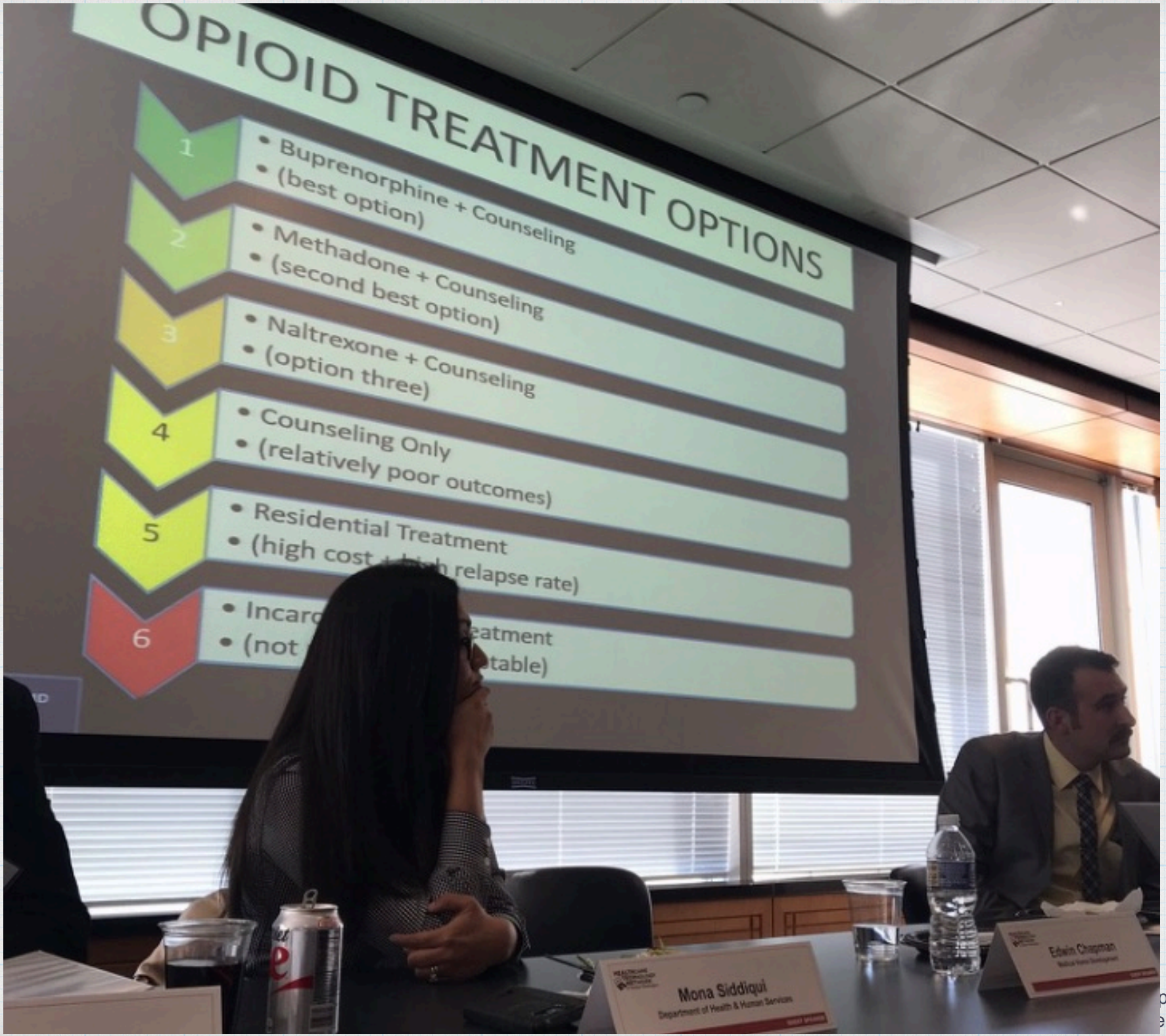


Figure. Evolution of the Centers for Disease Control and Prevention (CDC) BioSense public health surveillance system, 2002-2016. Abbreviations: ED, emergency department; ESSENCE, Electronic Surveillance System for the Early Notification of Community-Based Epidemics; NSSP, National Syndromic Surveillance Program; VA/DoD, Department of Veterans Affairs/Department of Defense.

ref: 2017-07-10-pub Deborah W. Gould, PhD et al - The Evolution of BioSense- Lessons Learned and Future Directions- 0033354917706954

Copyright 2018
Joseph Bormel, MD, MPH



opyright 2018
el, MD, MPH

Opportunities for Surveillance

https://www.sentinelinitiative.org/sentinel-system-story

https://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/UCM587649.pdf

Copyright 2018
Joseph Bormel, MD, MPH

What is the Sentinel System?

One of the FDA's biggest jobs is to make sure drugs, vaccines, and medical devices are safe. FDA wants to know if patients get bad side effects from these products. To make it faster and easier to learn about problems, FDA created a special program called the Sentinel System.

How the Sentinel System Works

Sentinel System's 3 important parts

- Information: The system looks at billing claims and patient records.
- Expert Team: Sentinel works with scientists, doctors and computer experts.
- Computer Programs: They study large groups of patients who take the same medicine, or use the same device.

Personal privacy

- No one at FDA or the Sentinel Operations Center has access to your name, address, or any other information that identifies you.
- For more information, visit sentinelinitiative.org.

Sentinel asks questions like:

- How many patients take the same drug?
- How many patients are getting bad side effects (swelling, bleeding, etc.)?
- Are side effects more common after taking one drug than after another drug that treats the same problem?

How does FDA use the information?

- FDA can choose to collect more information.
- FDA can provide updated safety information for patients and providers.
- If you have concerns about your own medical products, please contact your doctor.

https://www.sentinelinitiative.org/sentinel-system-story

Copyright 2018
Joseph Bormel, MD, MPH

Sentinel uses a **distributed data approach** in which **Data Partners** maintain physical and operational control over electronic data in their existing environments.

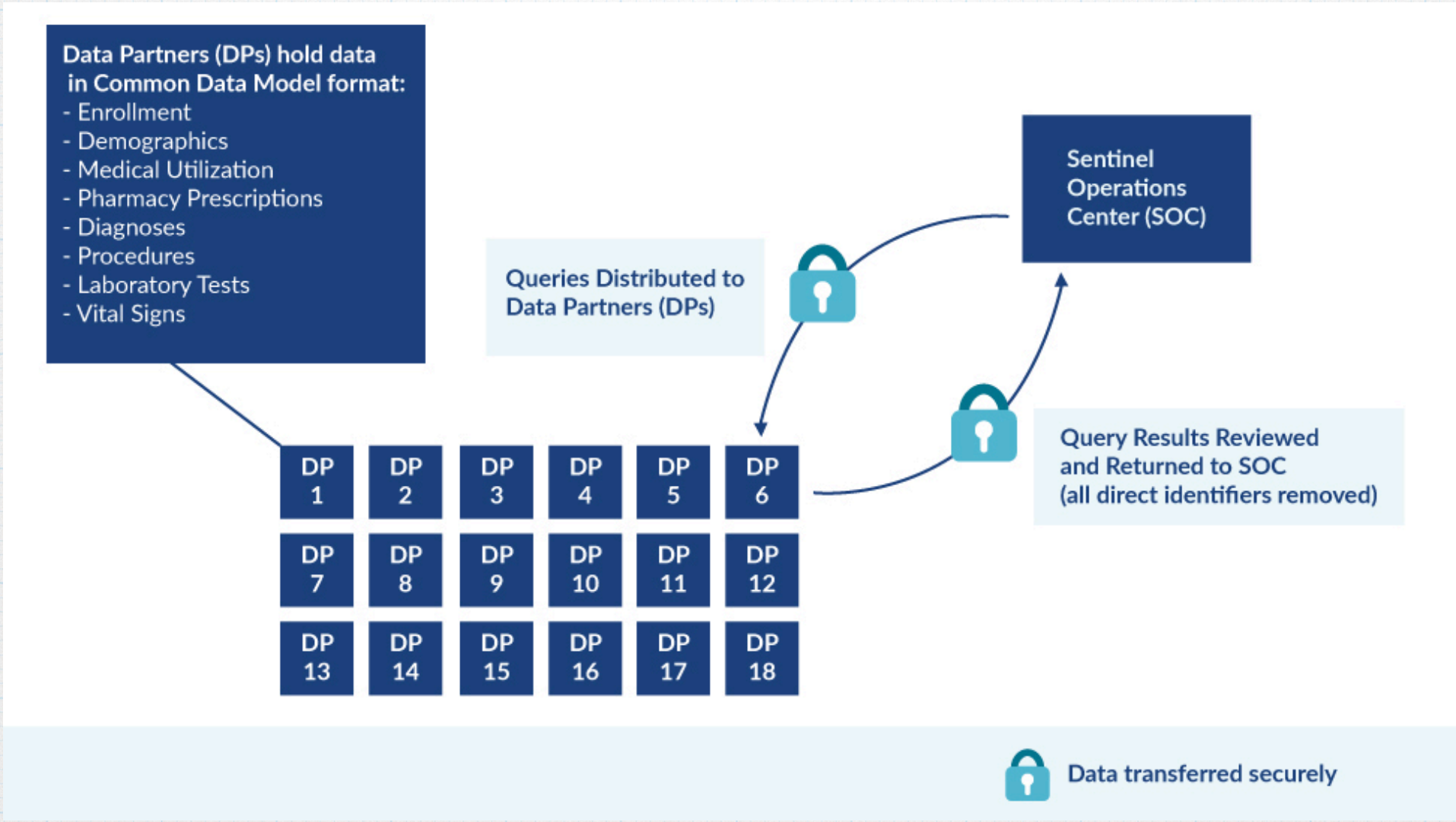
The distributed approach is achieved by using a standardized data structure referred to as the **Sentinel Common Data Model**.

Data Partners transform their data locally according to the Common Data Model, which enables them to execute standardized computer programs that **run identically at each Data Partner site**.

Data Partners are able to **review the results of the queries before sending them back** to the Sentinel Operations Center. Queries are distributed and results are returned through a secure portal in order to preserve privacy. **The combined collection of datasets** across all Data Partners is known as the Sentinel Distributed Database (SDD). The figure below (next slide) illustrates Sentinel's distributed data approach.

Copyright 2018
Joseph Bormel, MD, MPH

Distributed Database and Common Data Model



https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model

Copyright 2018
Joseph Bormel, MD, MPH

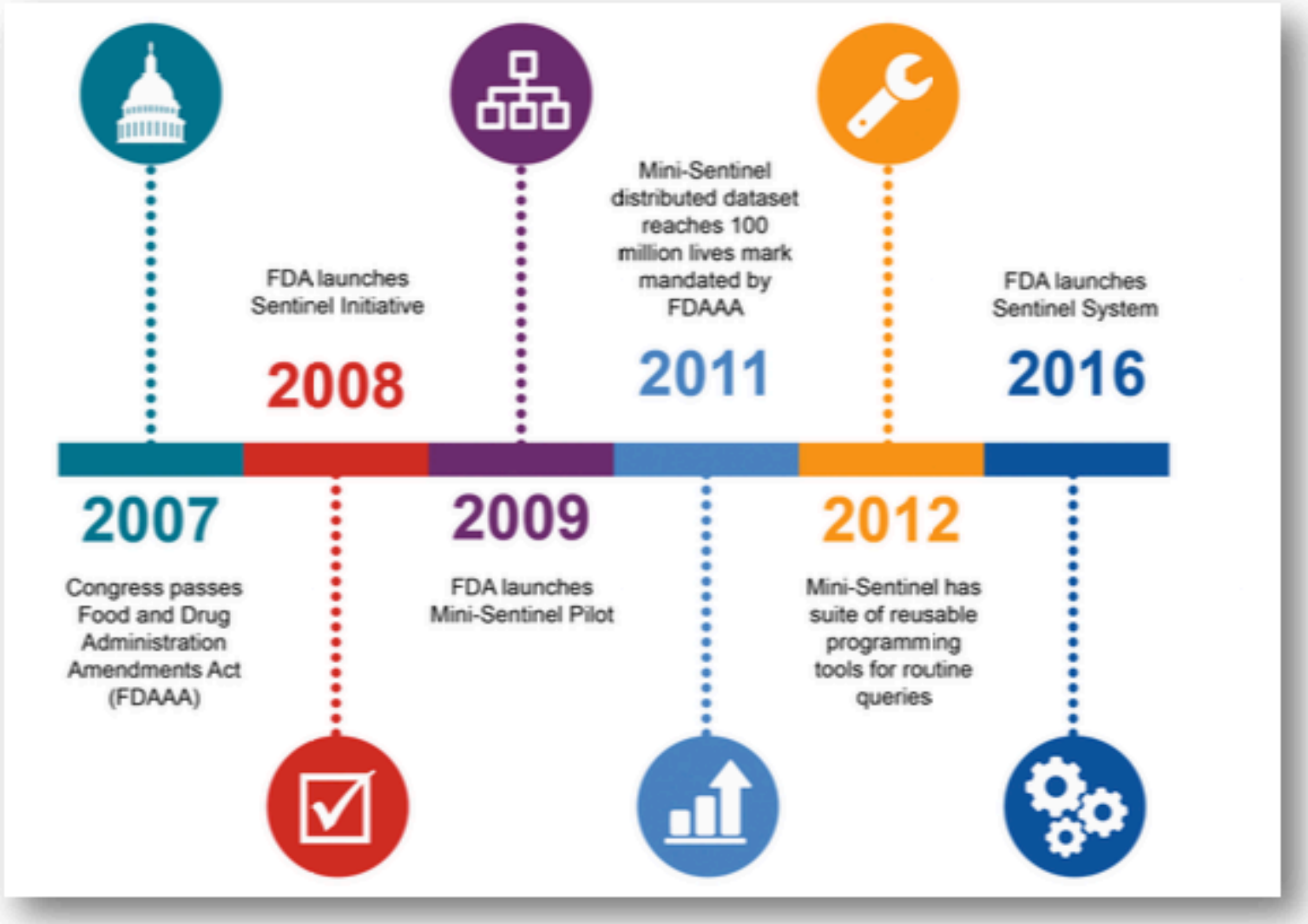
Data Partners

1. Aetna Informatics
2. Blue Cross Blue Shield of Massachusetts
3. Department of Population Health Sciences, Duke University School of Medicine
4. Harvard Pilgrim Health Care Institute
5. HealthCore, Inc. Government & Academic Research
6. HealthPartners Institute
7. Hospital Corporation of America
8. Humana, Inc., Comprehensive Health Insights
9. Kaiser Permanente Colorado Institute for Health Research
10. Kaiser Permanente Center for Health Research Hawaii
11. Kaiser Foundation Health Plan of the Mid-Atlantic States, Inc.
12. Kaiser Permanente Northern California, Division of Research
13. Kaiser Permanente Northwest Center for Health Research
14. Kaiser Permanente Washington Health Research Institute
15. Marshfield Clinic Research Institute
16. Meyers Primary Care Institute, a joint endeavor of Fallon Community Health Plan
17. Optum
18. Vanderbilt University School of Medicine, Department of Health Policy

https://www.sentinelinitiative.org/sentinel/data/data-partners

Copyright 2018
Joseph Bormel, MD, MPH

Timeline



https://www.sentinelinitiative.org/background

15

Copyright 2018
Joseph Bormel, MD, MPH

Access

- "… in the unique position of being able to offer industry, academia, and researchers access to a system similar to Sentinel for evaluating safety signals, implementing post-market studies, and assessing the impact of risk management actions."
- https://www.sentinelinitiative.org/sentinel/reagan-udall-foundation-and-imeds
- Data Partners find tremendous value in the internal use of their data and the Sentinel toolkit

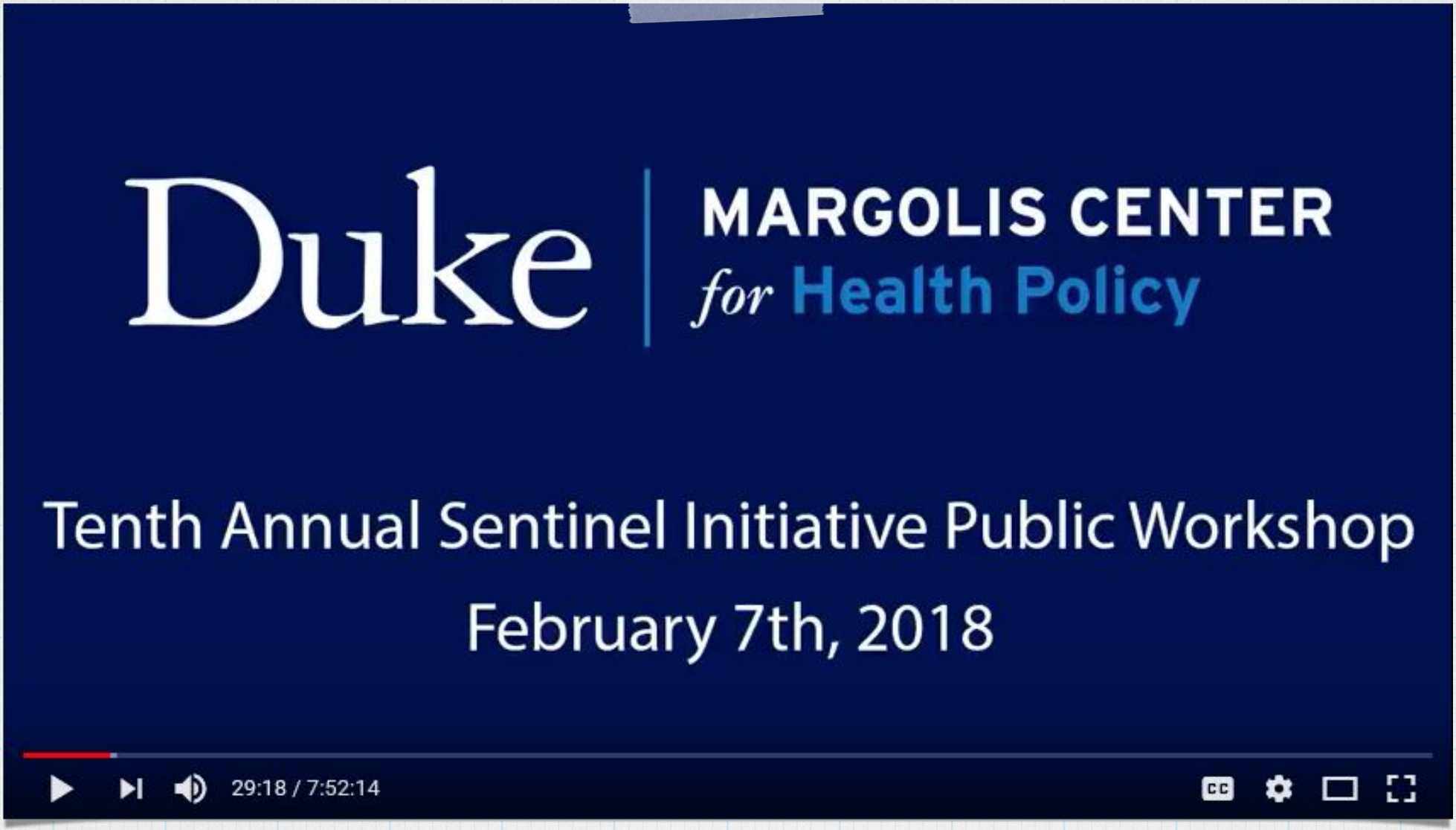
Reagan-Udall Foundation and IMEDS

REAGAN-UDALL FOUNDATION FOR THE FOOD AND DRUG ADMINISTRATION

The Innovation in Medical Evidence Development and Surveillance (IMEDS) program offered by the Reagan-Udall Foundation for the Food and Drug Administration (FDA) supports the FDA's vision of providing the Sentinel Initiative as a broader resource for public health and medical evidence generation.

By partnering with the Harvard Pilgrim Health Care Institute as the Analytic Center and select Data Partners contributing data to the Distributed Database, the Foundation is in the unique position of being able to offer industry, academia, and researchers access to a system similar to Sentinel for evaluating safety signals, implementing post-market studies, and assessing the impact of risk management actions. Learn more about the IMEDS program at the Reagan-Udall Foundation website and the FDA Voice Blog.

Copyright 2018
Joseph Bormel, MD, MPH



https://www.youtube.com/watch?v=r2axuu9swgA

Copyright 2018
Joseph Bormel, MD, MPH

Glossary, partial

Active Risk Identification and Analysis (ARIA): The U.S. Food and Drug Administration's (FDA) active post-market risk identification and analysis system, which is comprised of pre-defined, parameterized, reusable routine querying tools, combined with the electronic data in the Sentinel Common Data Model. Because ARIA uses parameterized tools and a trusted multi-site distributed database that undergoes continuous quality checks and refreshes, safety analyses can be done more efficiently to conduct medical product safety surveillance to fulfill the mandate in the FDA Amendments Act of 2007.

Sentinel Collaborating Institutions: A network of Data and Academic Partners that work with the FDA and Sentinel Coordinating Center to provide access to both healthcare data and scientific, technical, and organizational expertise.

Sentinel Coordinating Center: The Sentinel Coordinating Center includes the Sentinel Operations Center (SOC), comprised of the Applied Surveillance, Scientific Systems, and Administration Divisions housed at the Harvard Pilgrim Health Care Institute (HPHCI), and advisory groups. Both the Sentinel Coordinating Center and the SOC are led by the Sentinel Principal Investigator at HPHCI.

Sentinel Data Partners: Data Partners in the Sentinel System include a diverse group of organizations including academic medical centers, healthcare systems, and health insurance companies. Sentinel Data Partners maintain physical and operational control over electronic data in their existing environments.

Sentinel Infrastructure: The underlying data infrastructure created to enable analysis within the Sentinel System. The Sentinel Infrastructure involves: 1) a distributed data approach in which Data Partners maintain physical and operational control over electronic data in their existing environments; and 2) a Common Data Model consisting of standardized administrative and clinical information across Data Partners. The Sentinel Infrastructure has the potential to allow analysis of the data for other purposes besides safety for the FDA or those outside the FDA.

Sentinel Initiative: A multi-year effort beginning in 2008 to create a national electronic system for monitoring the performance of FDA-regulated medical products to improve the FDA's ability to identify and assess medical product safety issues.

Sentinel System: An active surveillance system that uses routine querying tools and pre-existing electronic healthcare data from multiple sources to monitor the safety of regulated medical products. Subcomponents of the Sentinel System include: ARIA, PRISM, BloodSCAN and STAT.

https://healthpolicy.duke.edu/sites/default/files/atoms/files/2018_sentinel_initiative_workshop_terms_final_1000_hours.pdf

Copyright 2018
Joseph Bormel, MD, MPH

Questions

Copyright 2018
Joseph Bormel, MD, MPH