

Evaluation of Health Care Interventions and Big Data: **Review of Associated Data Issues**

Carl Asche, PhD

Center for Outcomes Research, University of Illinois College of Medicine

Data access challenges such as data ownership, security issues and costs often serve as barriers to data access.

Careful thought is required to fully realize the potential of 'Big Data' to draw accurate conclusions.

Available 'Big Data' includes an ever-growing repository of clinical, genetic, genomic, social, outcome, and claims information.¹

Opportunity to improve patient care through:

- (1) assessment of treatment effectiveness and
- (2) outcome prediction in close to real-time is upon us.

Challenges in fully realizing the potential of 'Big Data'.

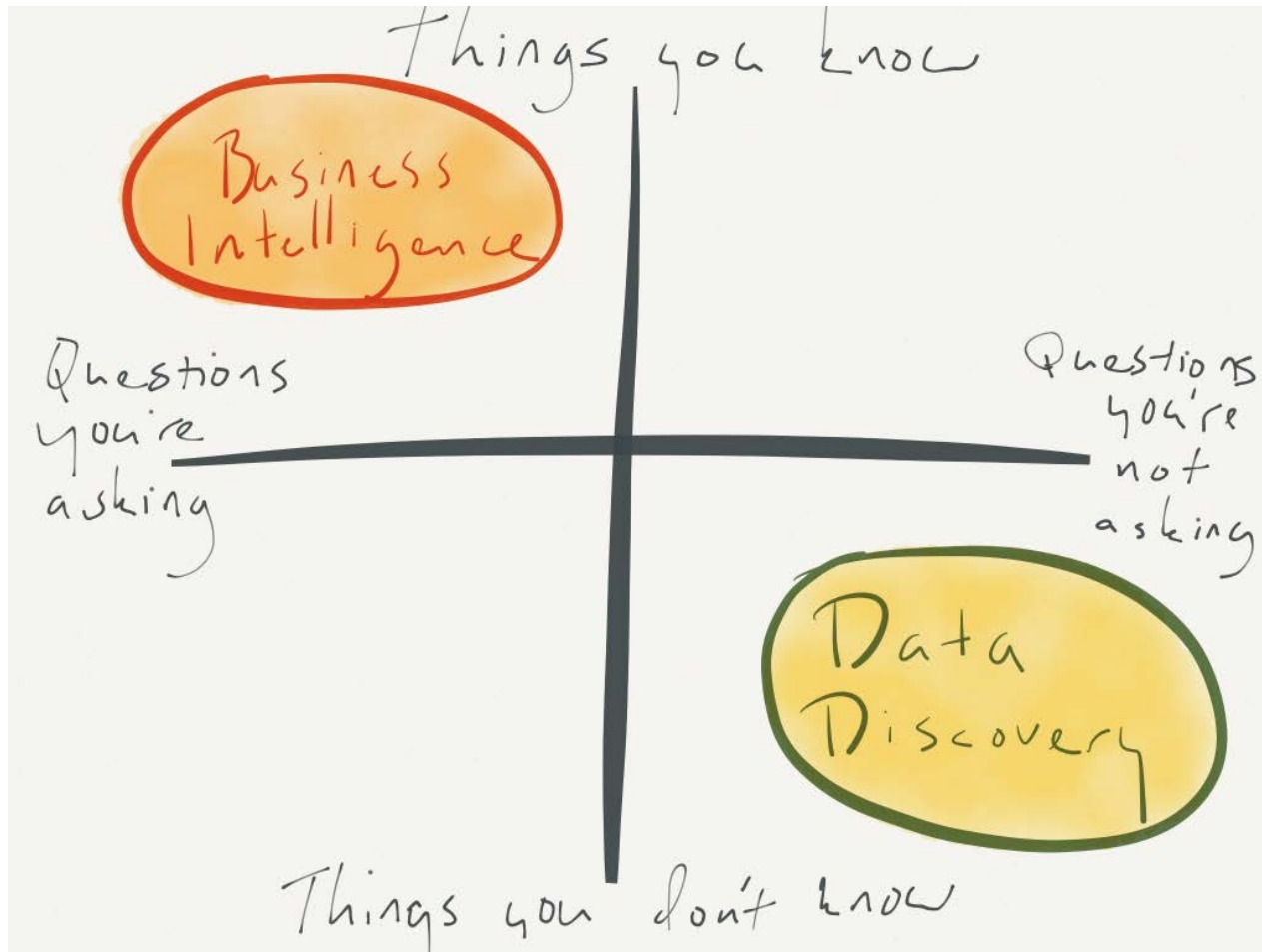
As unique multidisciplinary partnerships develop across industry, medicine and research, data issues specific to validation, linkage, and access must be addressed.

1. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health affairs (Project Hope)*. 2014 Jul;33(7):1123-31. PubMed PMID: 25006137.

The Issue of Data Validation & Linkage

- Lack of complete or linked information.
- Understand why the data exists and how this colors our interpretation.
- Establishing confidence in each data component's correctness and accuracy becomes paramount.

Issue: Big Data or Complete Information



Missing, Lost & Unavailable Data

Some data pieces are missing from the researcher's arsenal.

These key data points are not readily available for integration into larger data sets, limiting researchers' assessment of certain interventions and public health outcomes

Unavailable Data

- Key information on patient progress and outcomes.
- Most individuals over the age of 65 can be excluded from commercial insurance claims data
- Specialist care & outcomes can be omitted
- Data can be lost.

Reporting Bias

Terris et al.² model identifies six spheres where bias can arise, beginning with the patient's propensity to access services.

Additional areas of bias can include:

- Physician's propensity to detect, treat and record
- Community and system-based factors
- Factors associated with the processing and storage of information within the secondary database itself.²

2. Terris DD, Litaker DG, Koroukian SM. Health state information derived from secondary databases is affected by multiple sources of bias. *Journal of clinical epidemiology*. 2007 Jul;60(7):734-41. PubMed PMID: 17573990.

Unmeasured & Confounding

- Need to address unmeasured confounding when interpreting results.^{3,4}
- Current strategies to address confounding include:
 - Multiple imputation⁵
 - Propensity score calibration⁶
 - Use of external validation data.⁷

3. Schneeweiss S. Learning from big health care data. *The New England journal of medicine*. 2014 Jun 5;370(23):2161-3. PubMed PMID: 24897079.
4. Brookhart MA, Sturmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Medical care*. 2010 Jun;48(6 Suppl):S114-20. PubMed PMID: 20473199.
5. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ (Clinical research ed)*. 2009 Jun 29;338:b2393. PubMed PMID: 19564179.
6. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *American journal of epidemiology*. 2005 Aug 1;162(3):279-89. PubMed PMID: 15987725.
7. Sturmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Medical care*. 2007 Oct;45(10 Supl 2):S158-65. PubMed PMID: 17909375.

Strategies to Support Data Validation & Linkage

Expand Data Definitions

- Some variable definition changes reduce granularity, others can introduce standardization ripe for comparison
- Tools for accurately mapping ICD-9 to ICD-10 codes will be important for reducing bias in cohort discovery and commonly used comorbidity algorithms.
- Researchers such as Boyd et al. have begun creating tools on translating ICD-9 to ICD-10 and vice versa.⁸

8. ISPOR Digest of International Databases Working Group. Uses, Applications and Future Directions of the ISPOR Digest of International Databases. ISPOR 19th Annual European Congress Vienna, Austria. Available from: https://www.ispor.org/sigs/Digest_SIG-Forum_Vienna_2016.pdf

Match with Greater Precision

Evolving methodologies allow researchers to improve stringency of matching amongst observed patient populations thereby better simulating randomization.

Link Populations

To promote linkage of disparate data variables, stakeholders have invested in creating system-level data warehouses, with varying degrees of success.⁹

Key lessons regarding linkage can be learned from large clinical databases and registries across the globe:

- The Clinical Practice Research Database (CRPD)
- The Medical Expenditure Panel Survey (MEPS)
- National health care registries across Nordic.
- The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP).

Digest Databases

- Current success in large-scale data inventories has, to date, been more often associated with government-lead initiatives and organized as either administrative databases or disease/specialty specific clinical registries.¹⁰
- ISPOR Digest Working Group.⁸
- You can find the Oncology Databases here:

<https://www.ispor.org/OncologyORResources/SearchOologyResources.aspx>

8. ISPOR Digest of International Databases Working Group. Uses, Applications and Future Directions of the ISPOR Digest of International Databases. ISPOR 19th Annual European Congress Vienna, Austria. Available from:

https://www.ispor.org/sigs/Digest_SIG-Forum_Vienna_2016.pdf



The Issue of Data Access

Question of who has access to this information.

Key components of data access include:

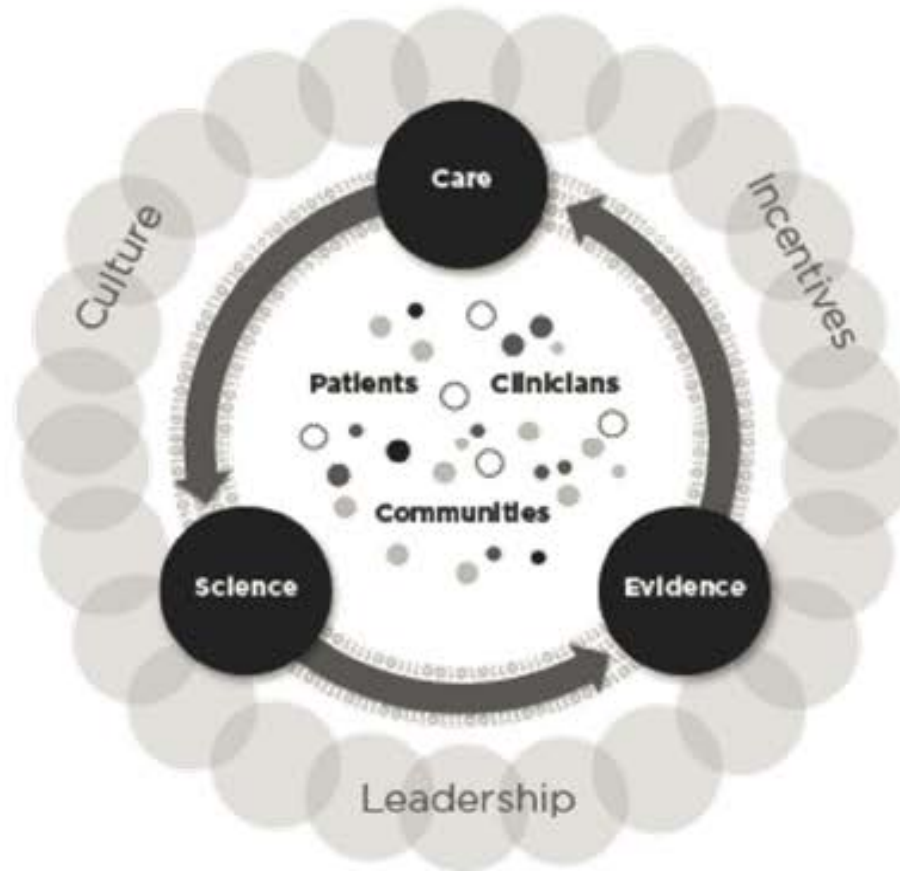
- Data Ownership
- Security & Privacy
- Cost & Availability

Emerging partnerships between public and private organizations may improve researchers' abilities to access data.

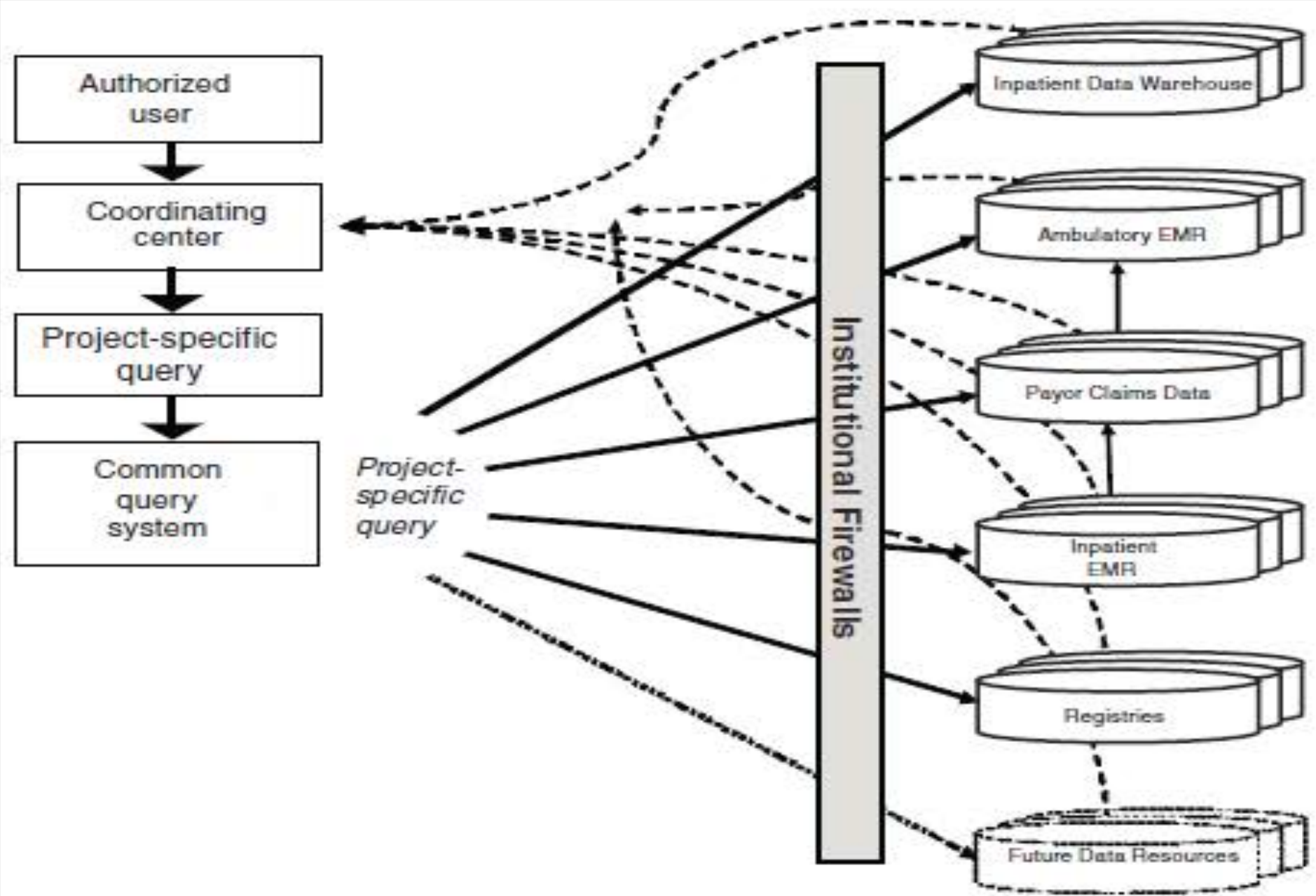
Strategies to Support Data Access

- Campaign for Public Support & Understanding
- Sustainably Fund Research
- Risk Assessment Plans
- Establish Distributed Data Networks

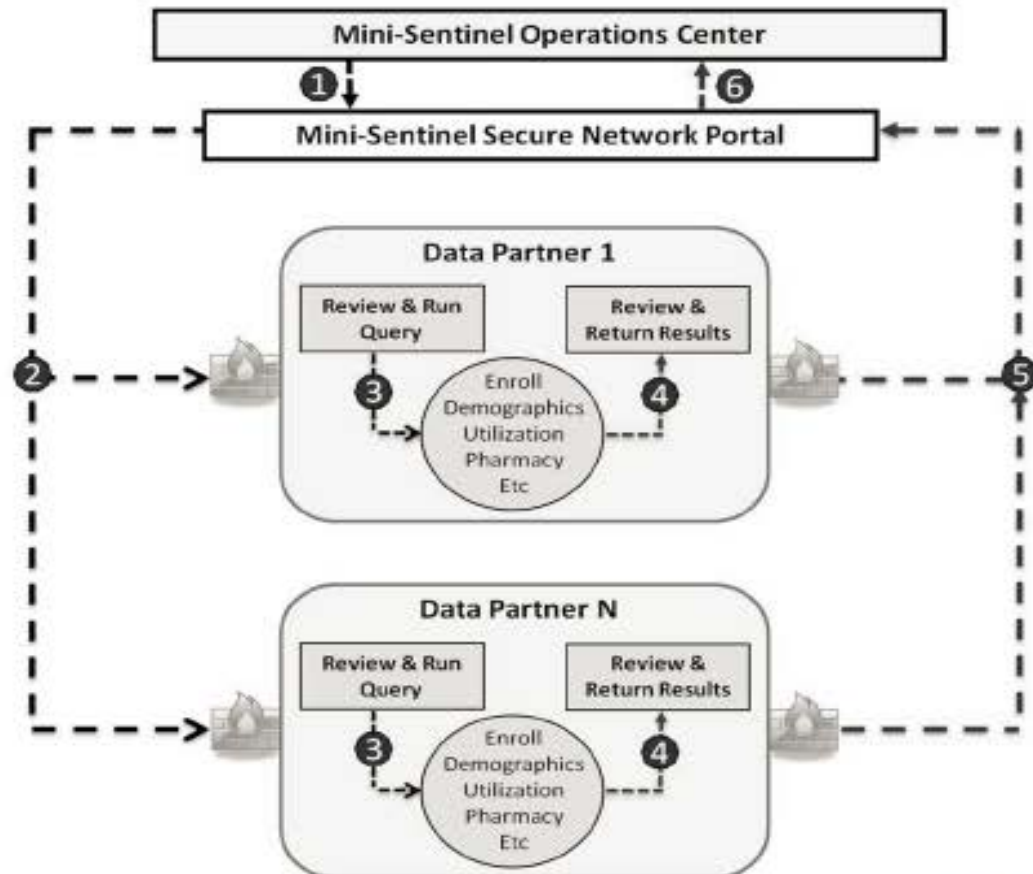
Vision: High-Value Care Driven by Rapidly Improving Science



Source: IOM, 2012



FDA Sentinel Initiative: Mini-Sentinel's Distributed Database



1- User creates and submits query (a computer program)

2- Data partners retrieve query

3- Data partners review and run query against their local data

4- Data partners review results

5- Data partners return results via secure network

6- Results are aggregated

Conclusion

Opportunities to accurately assess treatment effectiveness and outcome prediction... exciting development.

To fully realize the potential of these data sources and draw accurate conclusions, careful thought is necessary.

Solutions to data access challenges are still in flux, with data ownership, security needs, and cost often serving as barriers to data access.