Health Informatics

Lecture 1: Introduction

Samantha Kleinberg
samantha.kleinberg@stevens.edu

Why health informatics?

U.S.

Obama to Unveil Research Initiative to Develop Tailored Medical Treatments

By ROBERT PEAR JAN. 30, 2015

Email

Share

Tweet

Save

More

BIRDMAN
NOW PLAYING
GET TICKETS

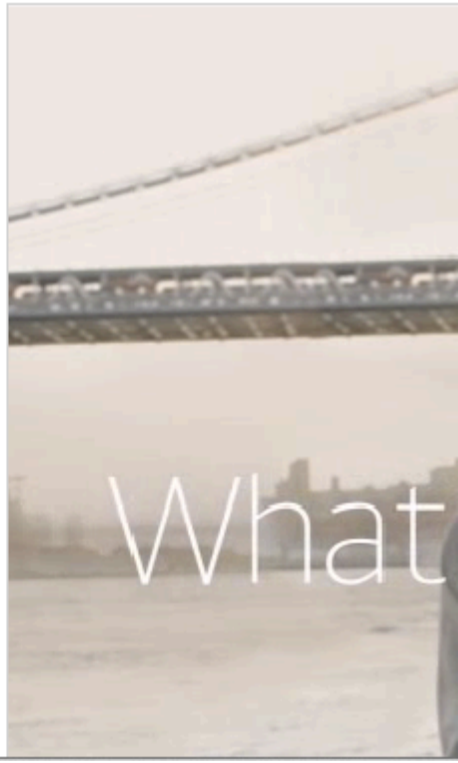
WASHINGT
biomedical r
one million A
tailored to ir
officials said

Federal officials described the project as a research consortium that would collect information from large numbers of people. The data could include medical records, laboratory test results, profiles of patients' genes, and information about their diet, tobacco use, lifestyle and environment.

Dr. Francis S. Collins, the director of the National Institutes of Health, said the studies would help doctors decide which treatments would work best for which patients.

White House officials said the "precision medicine initiative," also known as personalized or individualized medicine, would begin with a down payment of $215 million in the president's budget request for the fiscal year that starts on Oct. 1.

Mr. Obama plans to announce the details at a White House event on Friday with patients' advocates, researchers, and drug and biotechnology company



Why are you here?

What could we do with this?

- Who’s the user?
- What problem do they have?
- How can CS help?

But what do we need to know first?

- What's the data actually? Is it correct?
- Can we really apply ML/DM methods to it? What are the challenges?
- Can we share data freely? Why not?
- How do we preserve privacy

(and much more)

Syllabus

Key policies

1. No late work. There are few deadlines, but if the deadline's 11:59pm, work submitted at 12:01am isn't accepted

   (Why? Try submitting an NIH grant 2 minutes late!)

2. Plagiarism = F

   3X over the last year

What's plagiarism?

Presenting someone else's work as your own

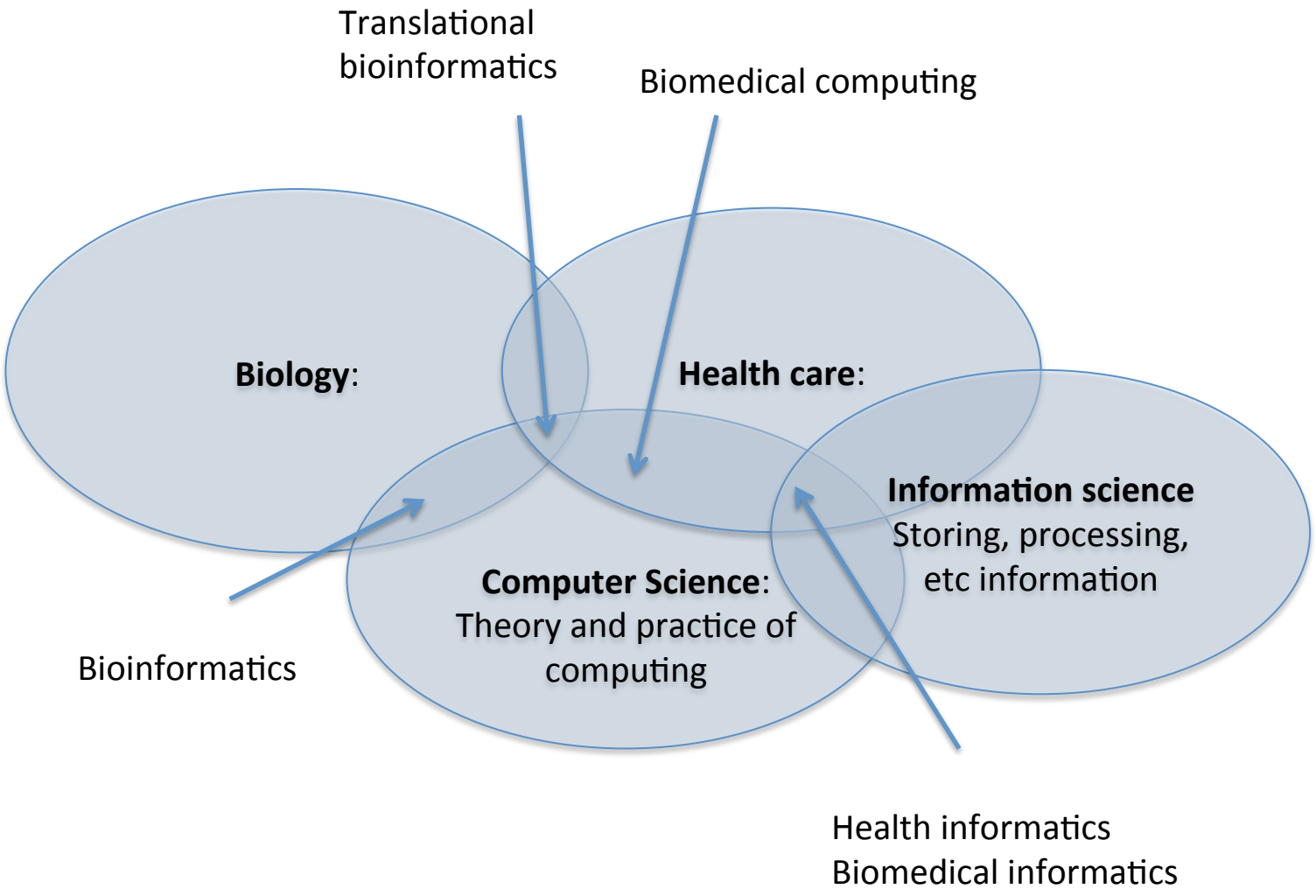
- Copying an entire paper
- Copying parts of other works, without attribution
  - E.g. quotes without citations
- Changing a few words, but keeping ideas and structure, without acknowledging source

Easy to avoid! Do your own work and acknowledge all sources. Quotes must be in quotes. Don't submit a collage.

If you're having trouble…

- Come to office hours! Monday 2-3 in Lieb 306 and by appointment
- Email me

What’s health informatics?

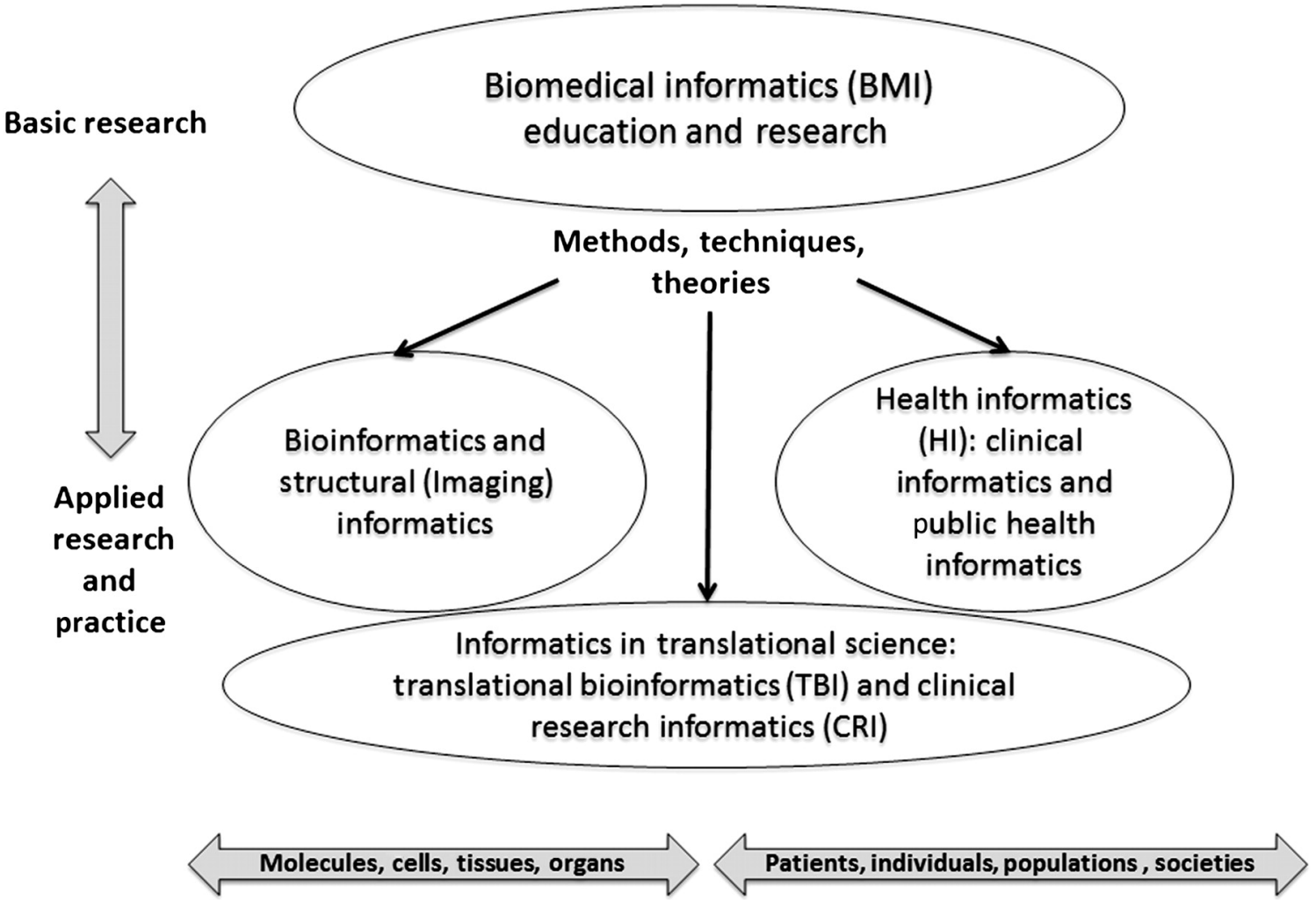

Translational bioinformatics
Biomedical computing
Biology:
Health care:
Information science
Storing, processing, etc information
Computer Science:
Theory and practice of computing
Bioinformatics
Health informatics
Biomedical informatics

Another definition

- **Definition**: Biomedical informatics (BMI) is the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, driven by efforts to improve human health.
- **Scope and breadth of discipline:** BMI investigates and supports reasoning, modeling, simulation, experimentation, and translation across the spectrum from molecules to individuals and to populations, from biological to social systems, bridging basic and clinical research and practice and the healthcare enterprise.
- **Theory and methodology:** BMI develops, studies, and applies theories, methods, and processes for the generation, storage, retrieval, use, management, and sharing of biomedical data, information, and knowledge.
- **Technological approach:** BMI builds on and contributes to computer, telecommunication, and information sciences and technologies, emphasizing their application in biomedicine.
- **Human and social context:** BMI, recognizing that people are the ultimate users of biomedical information, draws upon the social and behavioral sciences to inform the design and evaluation of technical solutions, policies, and the evolution of economic, ethical, social, educational, and organizational systems.

Kulikowski, C.A., Shortliffe, E.H., Currie, L.M., Elkin, P.L., Hunter, L.E., Johnson, T.R., Kalet, I.J., Lenert, L.A., Musen, M.A., Ozbolt, J.G., Smith, J.W., Tarczy-Hornoch, P.Z. & Williamson, J.J., 2012, AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline, *Journal of the American Medical Informatics Association*, pp. amiajnl-2012-001053

Biomedical informatics and its areas of application and practice, spanning the range from molecules to populations and society (adapted from Shortliffe and Blois3).



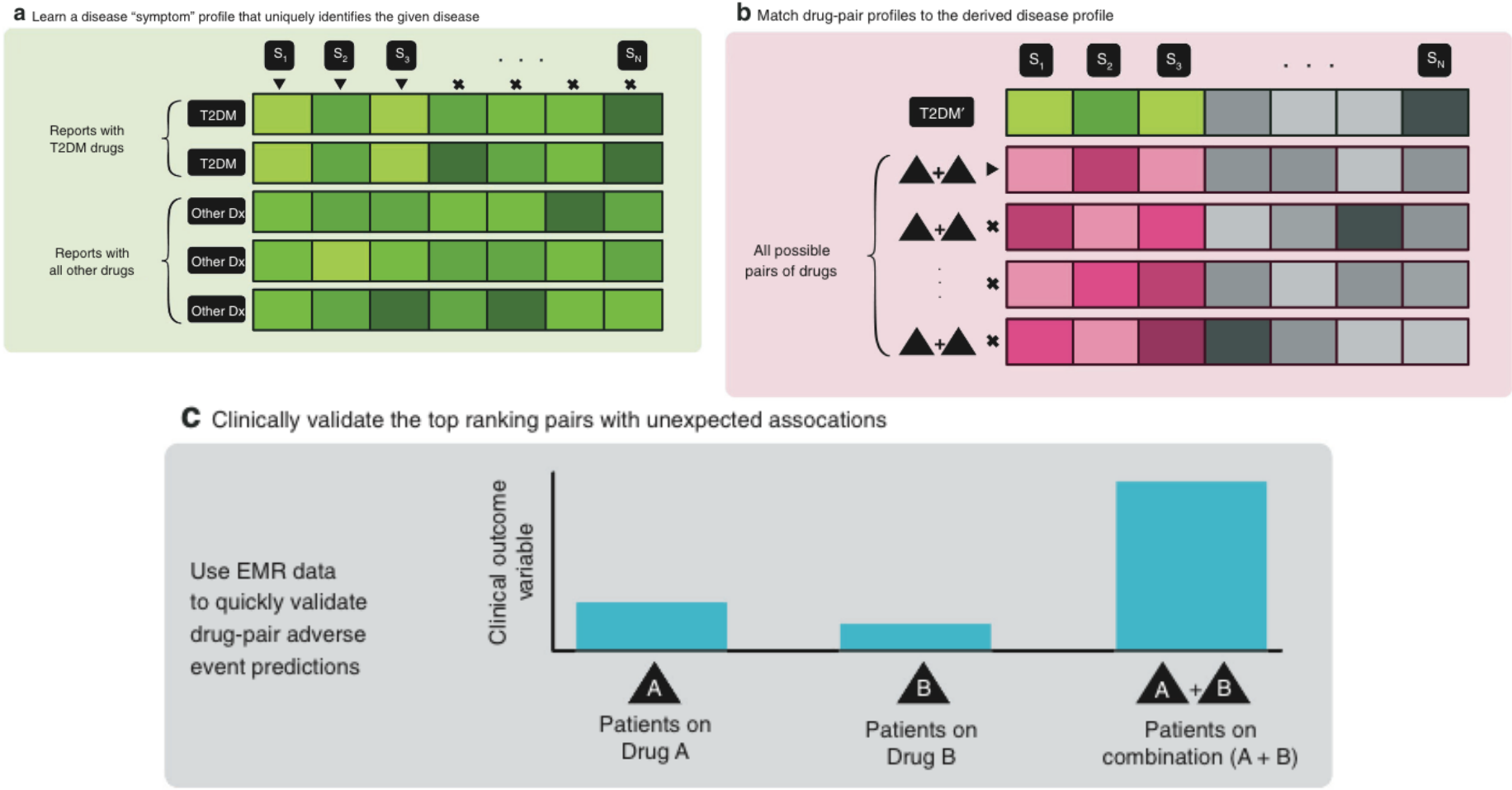
Kulikowski C A et al. J Am Med Inform Assoc doi:10.1136/amiajnl-2012-001053

Copyright © by the American Medical Informatics Association. All rights reserved.

Main subfields

- (Translational) Bioinformatics
- Public health informatics
- Healthcare/Clinical informatics
- Consumer/mHealth

Example: Translational bioinformatics



Tatonetti, N.P., Denny, J.C., Murphy, S.N., Fernald, G.H., Krishnan, G., Castro, V., Yue, P., Tsau, P.S., Kohane, I., Roden, D.M. & Altman, R., 2011, Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels, *Clinical Pharmacology & Therapeutics*, 90(1), pp. 133-42

Predicting cancer stage from online health forum text

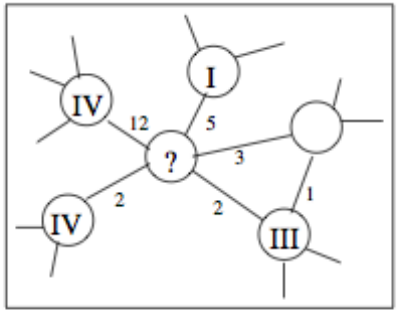


Figure 1: Nodes in the social network of forum member interaction.

Baseline				Text Based			
Stage	Precision	Recall	F	Stage	Precision	Recall	F
I	76.2	26.4	39.3	I	54.9	63.9	59.1
II	79.4	18.7	30.3	II	51.6	55.0	53.2
III	76.6	35.0	48.0	III	52.7	30.3	38.5
IV	76.4	50.7	60.9	IV	82.5	71.2	76.4
Network Based				Combined			
Stage	Precision	Recall	F	Stage	Precision	Recall	F
I	50.4	56.7	53.4	I	57.1	65.4	61.0
II	49.6	49.1	49.3	II	56.6	53.5	55.0
III	65.7	27.7	39.0	III	56.1	48.3	51.9
IV	59.3	83.7	69.4	IV	84.7	81.3	83.0

Table 3: Stage prediction results (Precision, Recall, and F-measure).

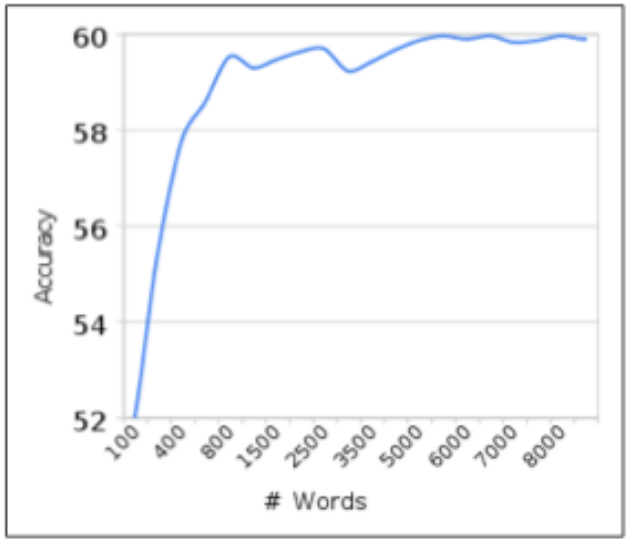
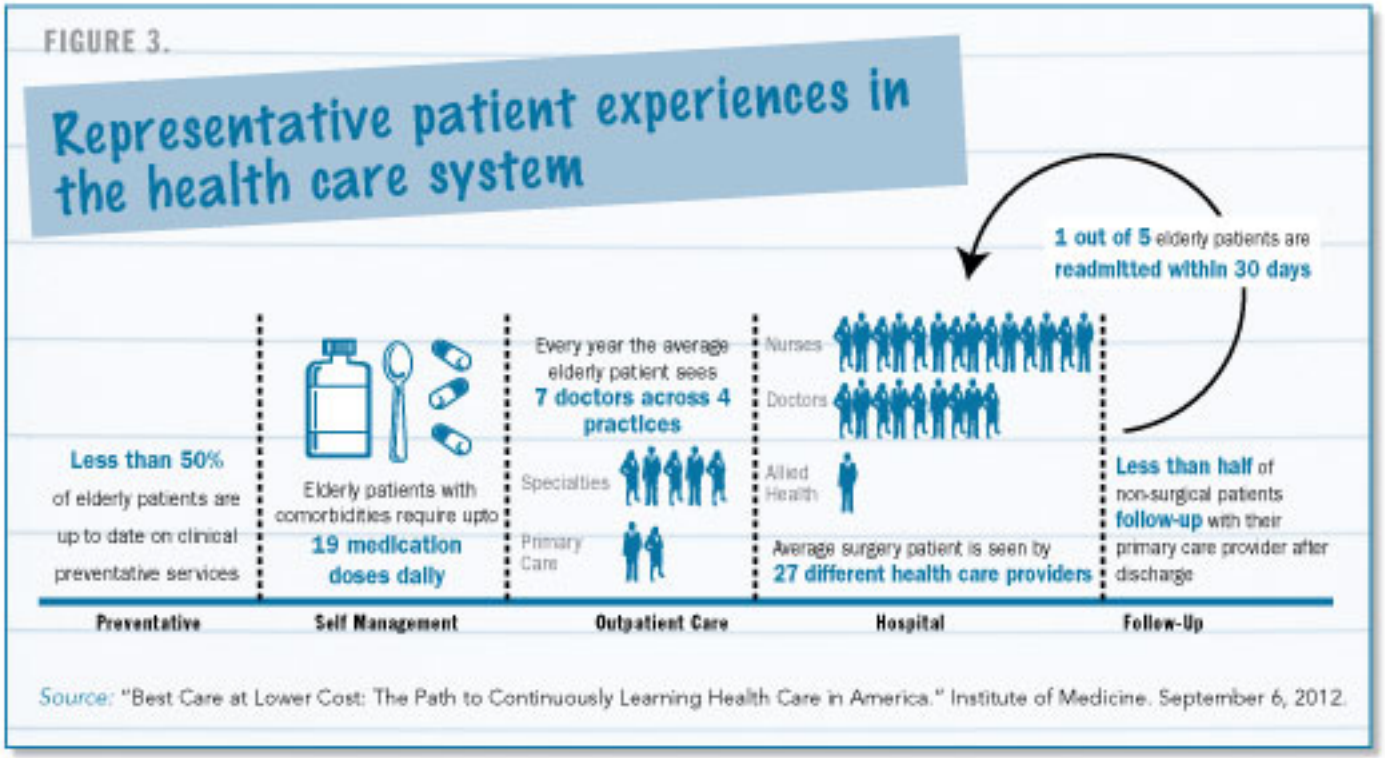


Figure 3: Overall text-based prediction accuracy against post history length.

Mukund Jha and Noemie Elhadad. Cancer stage prediction based on patient online discourse. In Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, pages 64–71, Uppsala, Sweden, July 2010

Clinical informatics



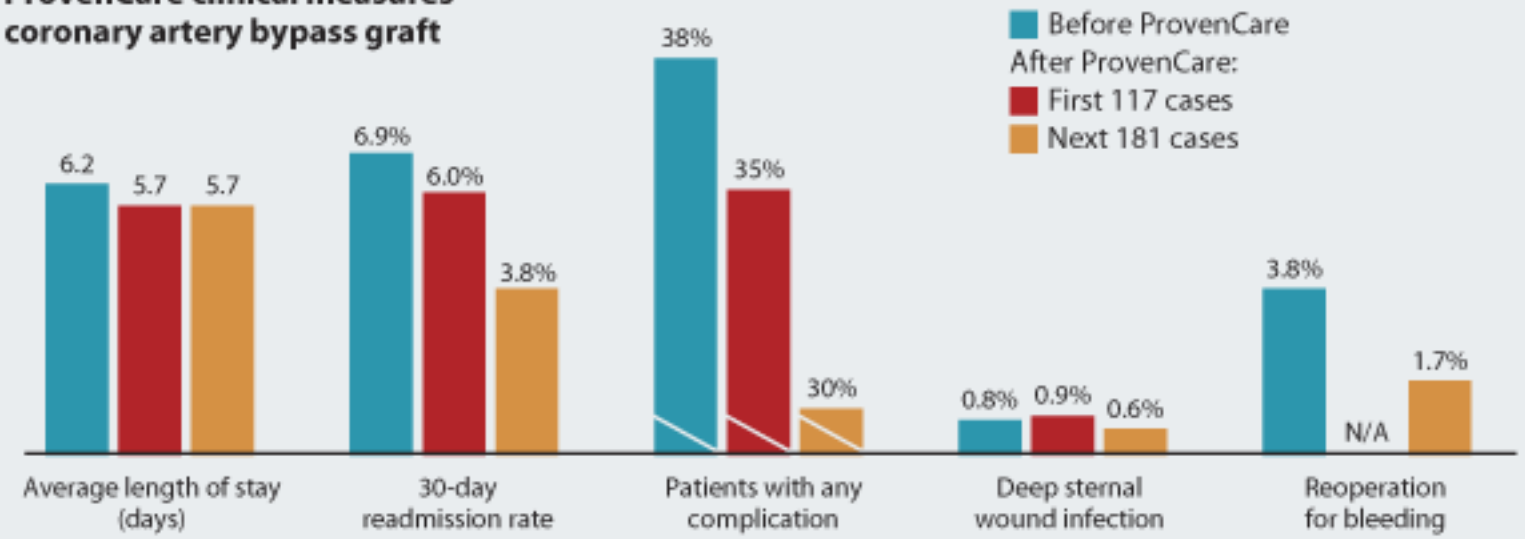
http://www.allhealth.org/sourcebookcontent.asp?CHID=181

ProvenCare by the Numbers

Heart program produces improved clinical measures

Geisinger Health Plan researchers compared selected clinical measures for elective coronary artery bypass graft patients before and after the implementation of the ProvenCare cardiac surgery program. The program went live three years ago. The first 117 cases after implementation showed performance improvements.The measures showed continued improvement after 181 cases, as Geisinger maintained an effort to standardize care for CABG patients.

ProvenCare clinical measures — coronary artery bypass graft



Source: Geisinger Health System, 2008.

Deep sternal wound infection	0.8%	0.6%	25%

http://www.managedcaremag.com/archives/0903/0903.geisinger.html

AMIA Themes

- **Biomedical Data Visualization**
- **Clinical Informatics:** findings related to the design, development, and implementation of state-of-the-art clinical systems
- **Clinical Research Informatics:** addressing the critical need for effective information management to address the many challenges facing clinical research
- **Clinical Workflow and Human Factors:** human factors aspects of clinical information system implementation and use that revolves around usability, workflow, and patient safety
- **Consumer Informatics and PHRs:** Personal Health Records (PHRs) and the consumer perspective in the use of health information science designed to improve patient engagement, medical outcomes, and the health care decision-making process
- **Mobile Health:** mHealth, Web 2.0, social media, telehealth/ telemedicine, and related topics
- **Data Interoperability and Information Exchange:** methods that organizations have undertaken to develop and implement various clinical data integration and exchange activities
- **Data Mining, NLP, Information Extraction Retrieval:** application of data mining, natural language processing, information extraction retrieval to all areas of biomedicine
- **Achieving Meaningful Use:** ways to promote the successful and effective development, implementation, and evaluation of Electronic Health Records as the nation works toward "meaningful use" of these systems
- **Global eHealth:** Global eHealth challenges
- **Informatics Education and Workforce Development:** effort to create a trained HIT workforce
- **Informatics in Health Professional Education:** application of information technology in health professional education
- **Interactive Systems:** human-computer interaction (HCI) research, compelling designs, or innovative interactive technologies
- **Policy and Ethical Issues:** unprecedented national HIT activity and ethical considerations posed as more practitioners and the public interface with these technologies.
- **Public Health Informatics and Biosurveillance:** disease detection, communications, workforce development, standards and interoperability, and best practices to combine the domains of health information science and technology with the practice and science of public health
- **Imaging Informatics:** intersection of imaging science, biomedical engineering and biomedical informatics,
- **Simulation and Modeling:** use of various computer-based simulation and modeling methodologies and tools as they can be applied within the field of biomedical informatics
- **Terminology and Standards Ontologies:** complex issues surrounding standard syntax, semantics, and pragmatics of design, development and use of various application-specific and general purpose clinical terminologies and ontologies
- **Translational Bioinformatics and Biomedicine:** opportunities in biomedical informatics that arise from the storage, retrieval, analysis, and dissemination of molecular and genomic information in a clinical setting context

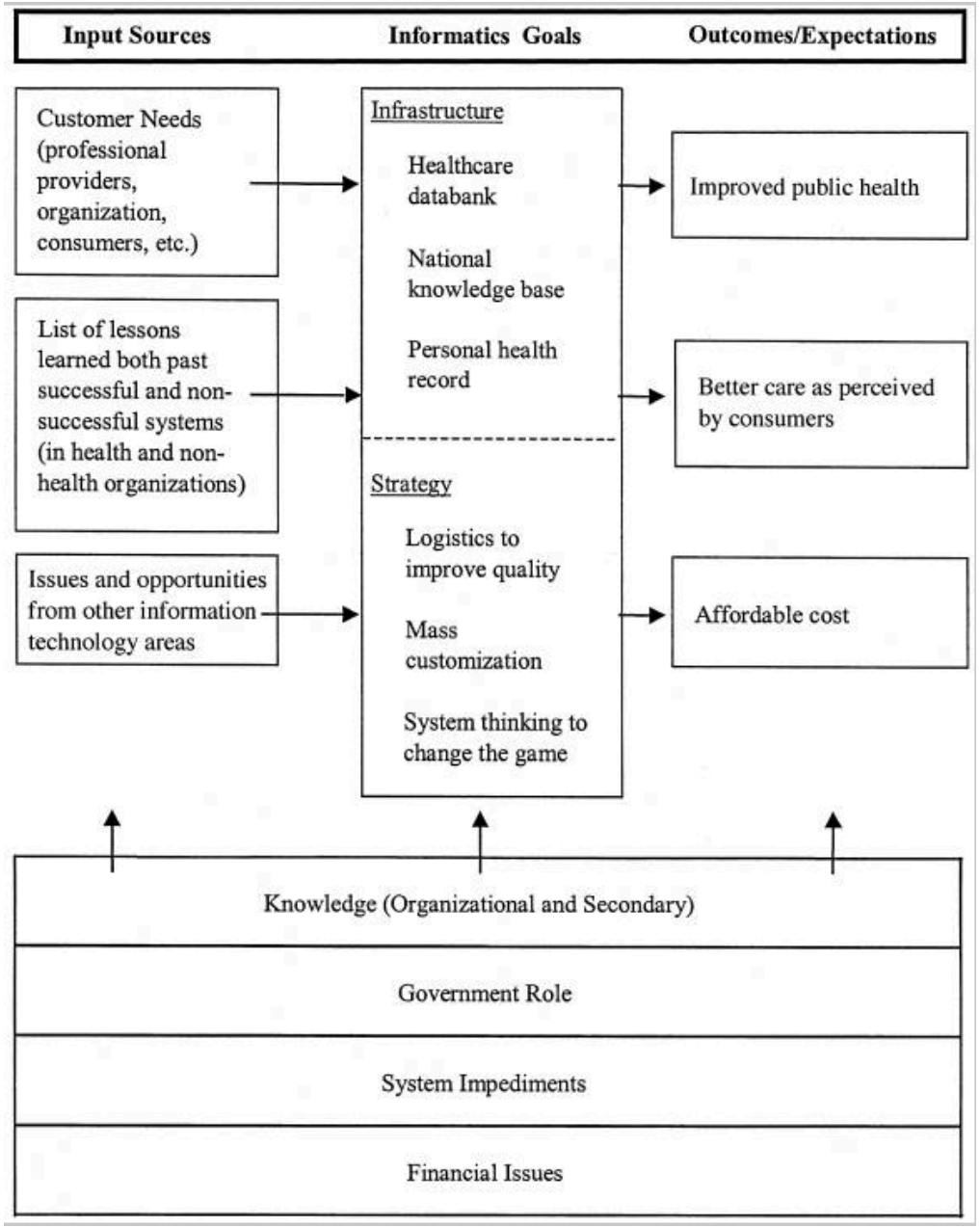
http://www.amia.org/amia2013/themes

Goals

Conceptual Value/Mission Model for Academic Health Centers

Value: Quality	Mission:		
	Patient Care	**Education**	**Research**
Right Thing (*Vision*)	■ Carriage trade	■ Specialists	■ Advancing and communicating knowledge
	■ Individualization	■ Family Practice Physicians	
	■ Mass customization	■ Academic Physicians	■ Outcome research
	■ Population-based medicine	■ Allied Health Professionals	
	■ Regional health model		
Right Way (*Which, How*)	■ Traditional medicine	■ Lecture	■ Laboratory-based research
	■ Complementary medicine	■ Problem-based learning	■ Department/discipline
	■ Evidence-based medicine	■ Small group	■ Interdisciplinary institutes
	■ Standards of care	■ Community-based learning	■ Academic health center
	■ Practice guidelines	■ Apprenticeships	
	■ Protocols		
	■ Learning organization		
	■ Person empowerment		
	■ Shared responsibility		
Right Time (*When*)	■ Acute care on demand	■ Undergraduate medical education	■ Undergraduate education
	■ Preventive medicine		■ Graduate education
	■ Public health education	■ Graduate medical education	
		■ AA/BA education	
		■ Just-in-time learning	
		■ Just-in-time training	
Right Person (*Who*)	■ Physician	■ Physicians	■ Faculty
	■ Integrated staff	■ Other faculty	■ Interdisciplinary team
	■ Skill mix		
	■ Patient		
	■ Consumer/public		

Stead, W.W. & Lorenzi, N.M., 1999, Health Informatics Linking Investment to Value, *Journal of the American Medical Informatics Association*, 6(5), pp. 341-8.



Stead, W.W. & Lorenzi, N.M., 1999, Health Informatics Linking Investment to Value, *Journal of the American Medical Informatics Association*, 6(5), pp. 341-8.

An Envisioned Cycle That Ties Patient Care with Knowledge Creation and Dissemination



Providers Caring for Patients

Electronic Health Records

Regional and National Public Health and Disease Registries

Biomedical and Clinical Research

Standards for Prevention and Treatment

Creation of Protocols, Guidelines, and Educational Materials

Information, Decision-Support, and Order-Entry Systems

A "Learning Healthcare System"

Slide by AMIA

Health IT

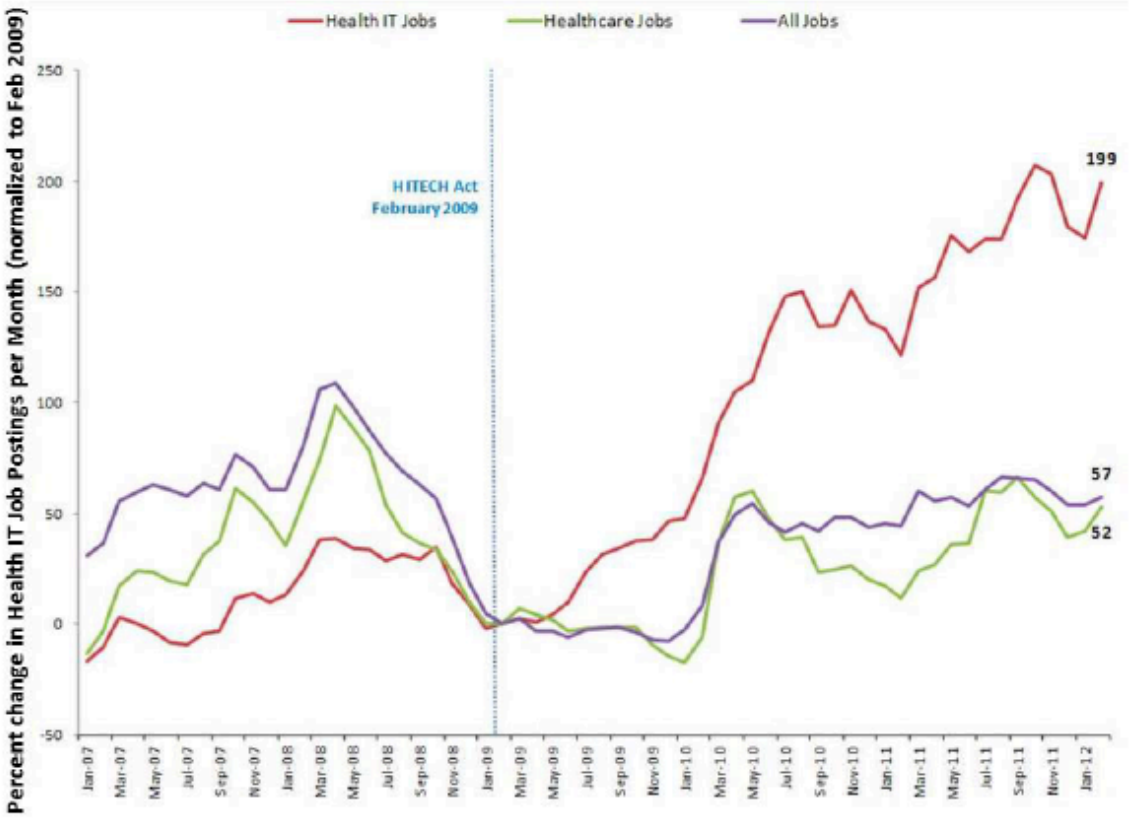
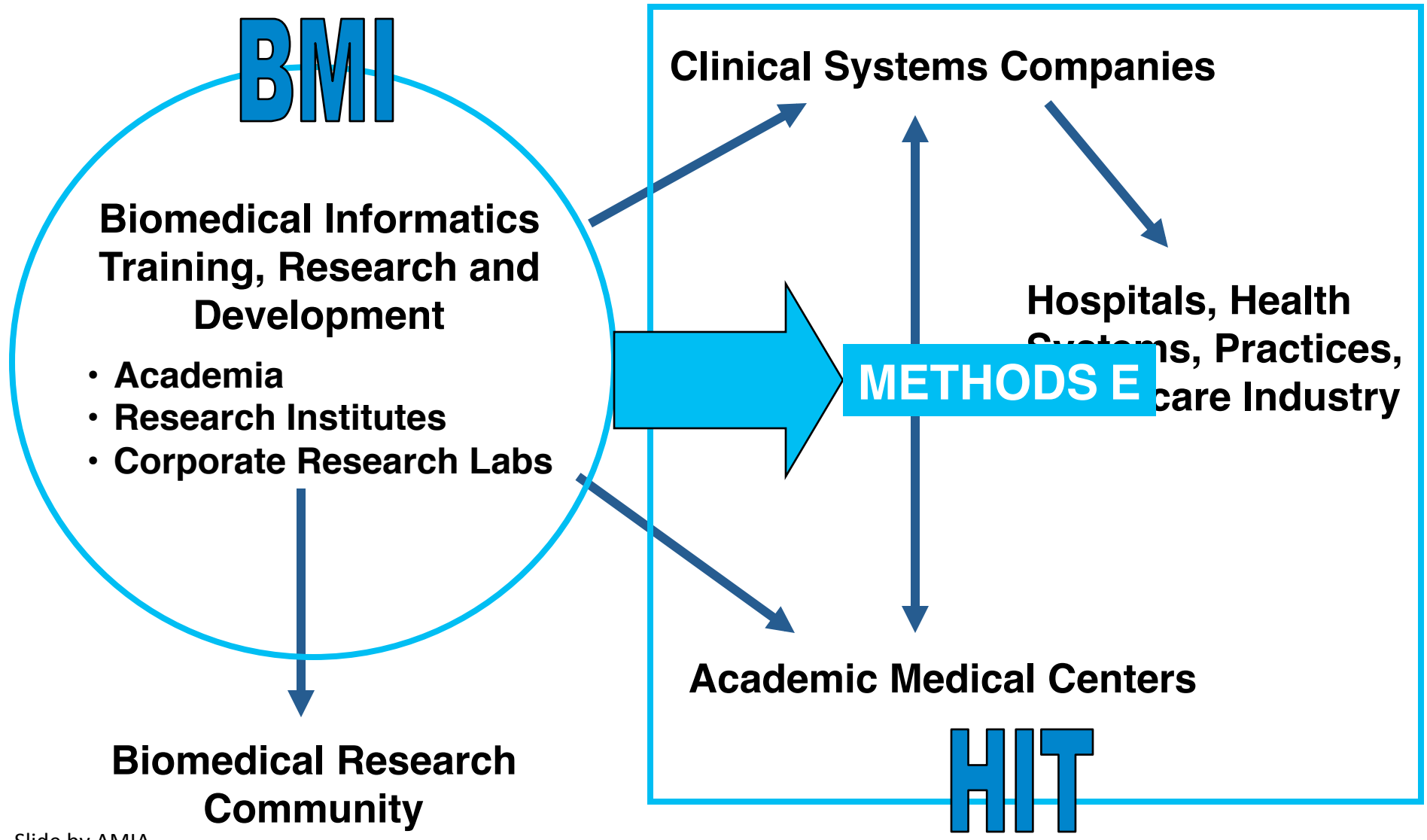


Figure 1.10 Percent change in online health IT job postings per month, relative to healthcare jobs and all jobs: normalized to February 2009 when ARRA passed. Source: ONC analysis of data from O'Reilly Job Data Mart, ONC Data Brief, No. 2, May 2012 [http://www.healthit.gov/sites/default/files/pdf/0512_ONCDataBrief2_JobPostings.pdf (Accessed 4/10/13)]

Shortliffe, E.H. & Cimino, J.J., 2013, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine,* Springer Verla

BMI and HIT



Slide by AMIA

LONGER TERM GOALS

Create a research cohort of **> 1 million American volunteers** who will share genetic data, biological samples, and diet/lifestyle information, all linked to their electronic health records if they choose.



Pioneer **a new model for doing science** that emphasizes **engaged participants, responsible data sharing, and privacy protection.**

Personalized medicine + informatics

- Using -omics in clinical care
- Decision support
- Find similar patients – treatments, outcomes

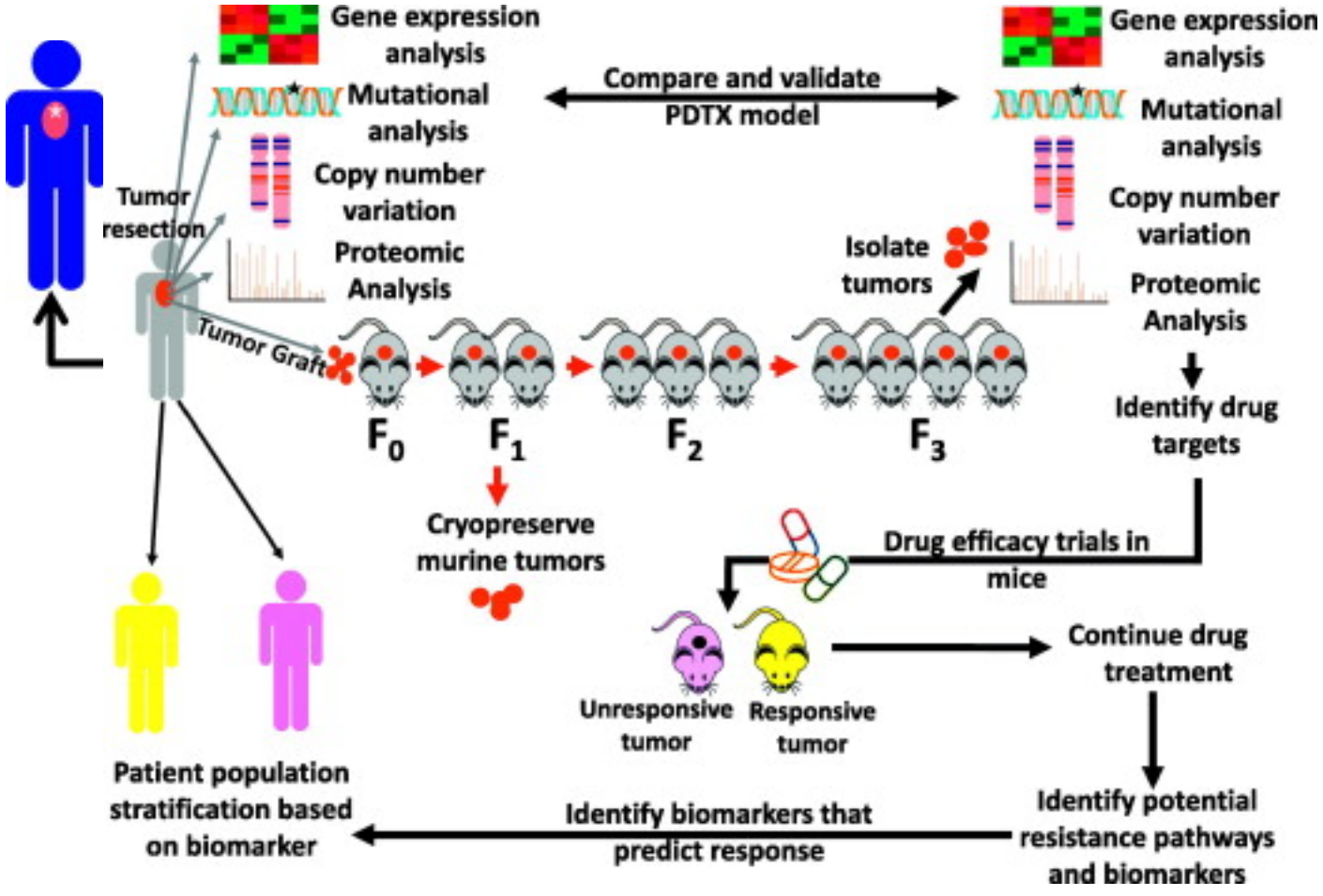
Mini-review

One mouse, one patient paradigm: New avatars of personalized cancer therapy

Prerna Malaney[a], Santo V. Nicosia[a], Vrushank Davé[a, b],

[a] Morsani College of Medicine, Department of Pathology and Cell Biology, Tampa, FL 33612, USA

[b] Department of Molecular Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA



Chronic disease

- Diabetes
- Obesity
- Asthma
- Epilepsy

…

Articles Register Sign in

glu

Glu is an active and diverse type 1 diabetes online community designed to accelerate research and amplify the collective voice of those living with T1D. **Register Now.**



Support:
Find and connect with others touched by T1D.

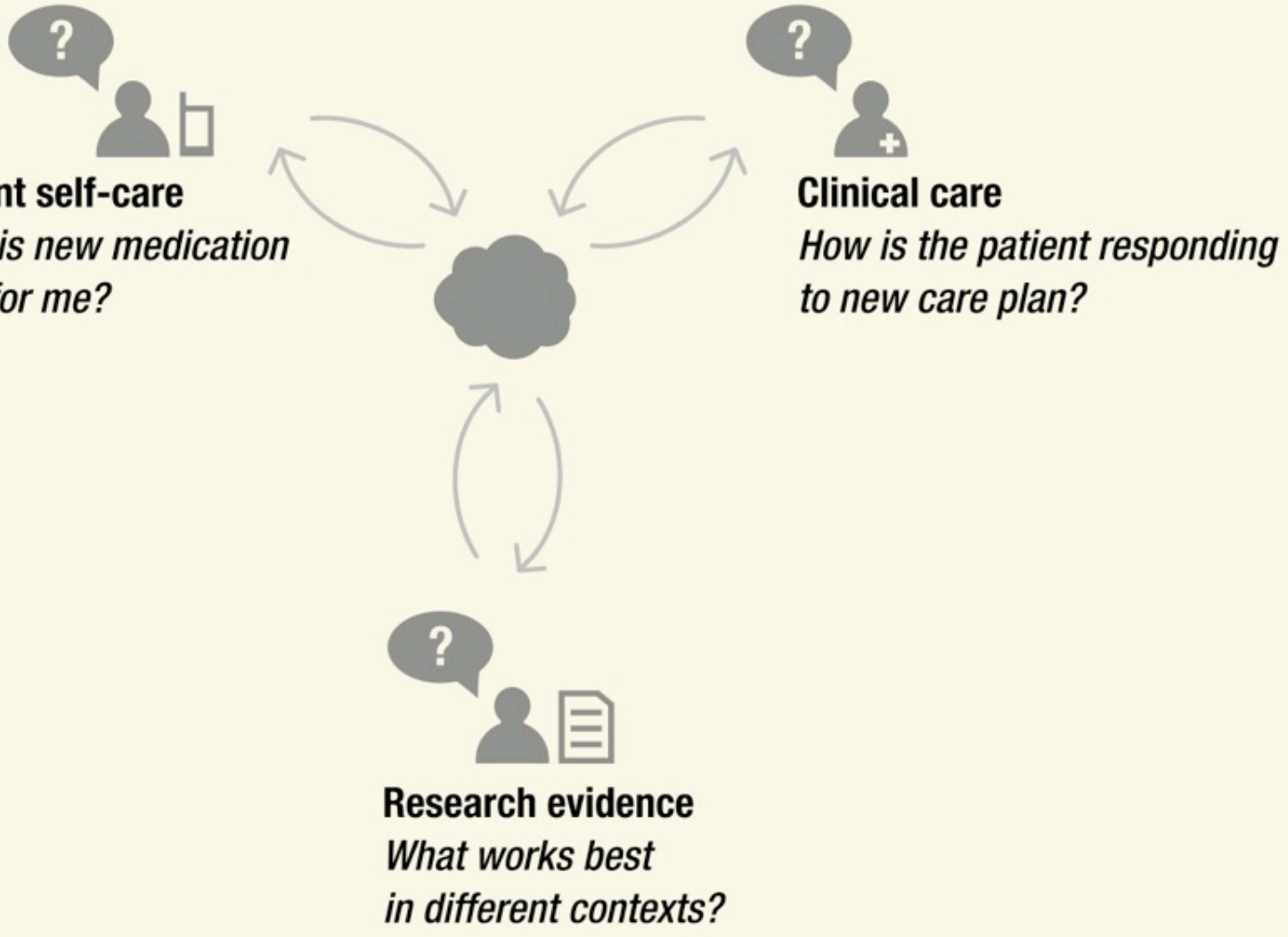


Empower:
Get involved as a citizen scientist to help accelerate research.

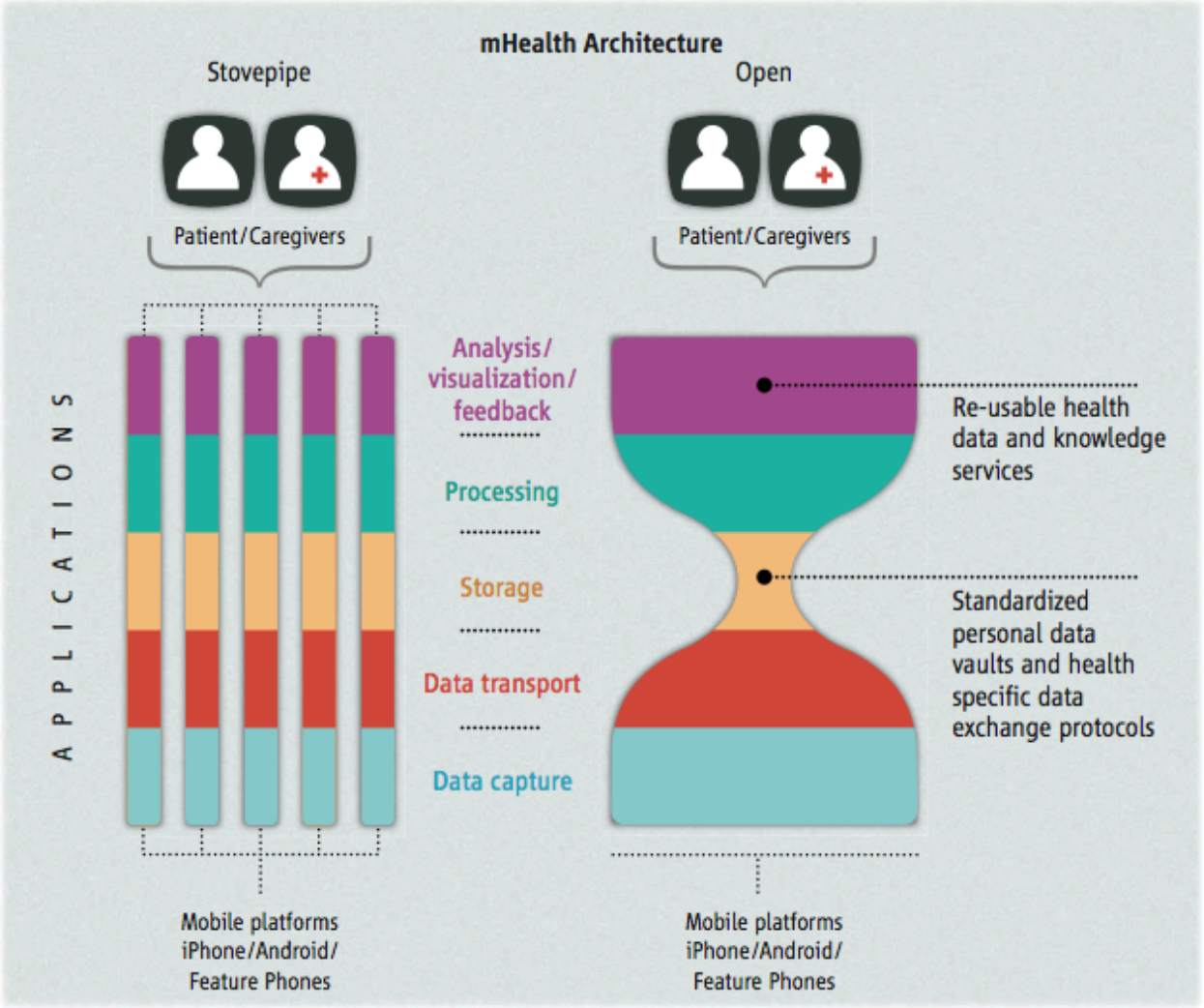


Educate:
Learn from discussions, articles, research, and from each other.

Essential data driven feedback loops



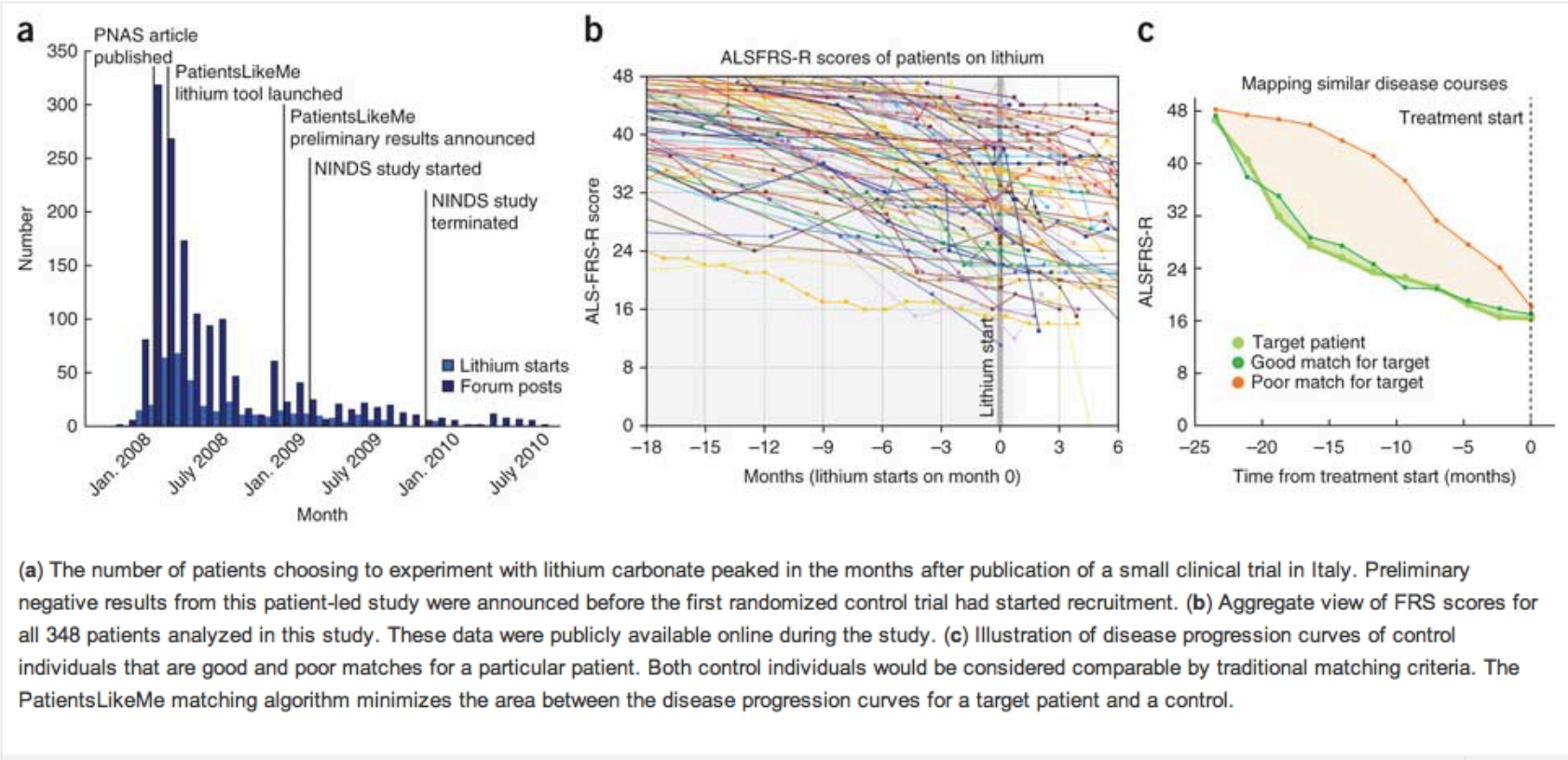
http://mobilizelabs.org/why-mhealth



mHealth architecture: Stovepipe versus Open. The narrow waist of the open hourglass will include at least health-specific syntactic and semantic data standards; patient identity standards; core data processing functions such as feature extraction and analytics; and data stores that allow for selective, patient-controlled sharing. Standards should be common with broader health IT standards whenever possible.

Estrin, D. & Sim, I., 2010, Open mHealth architecture: an engine for health care innovation, *Science(Washington)*, 330(6005), pp. 759-60

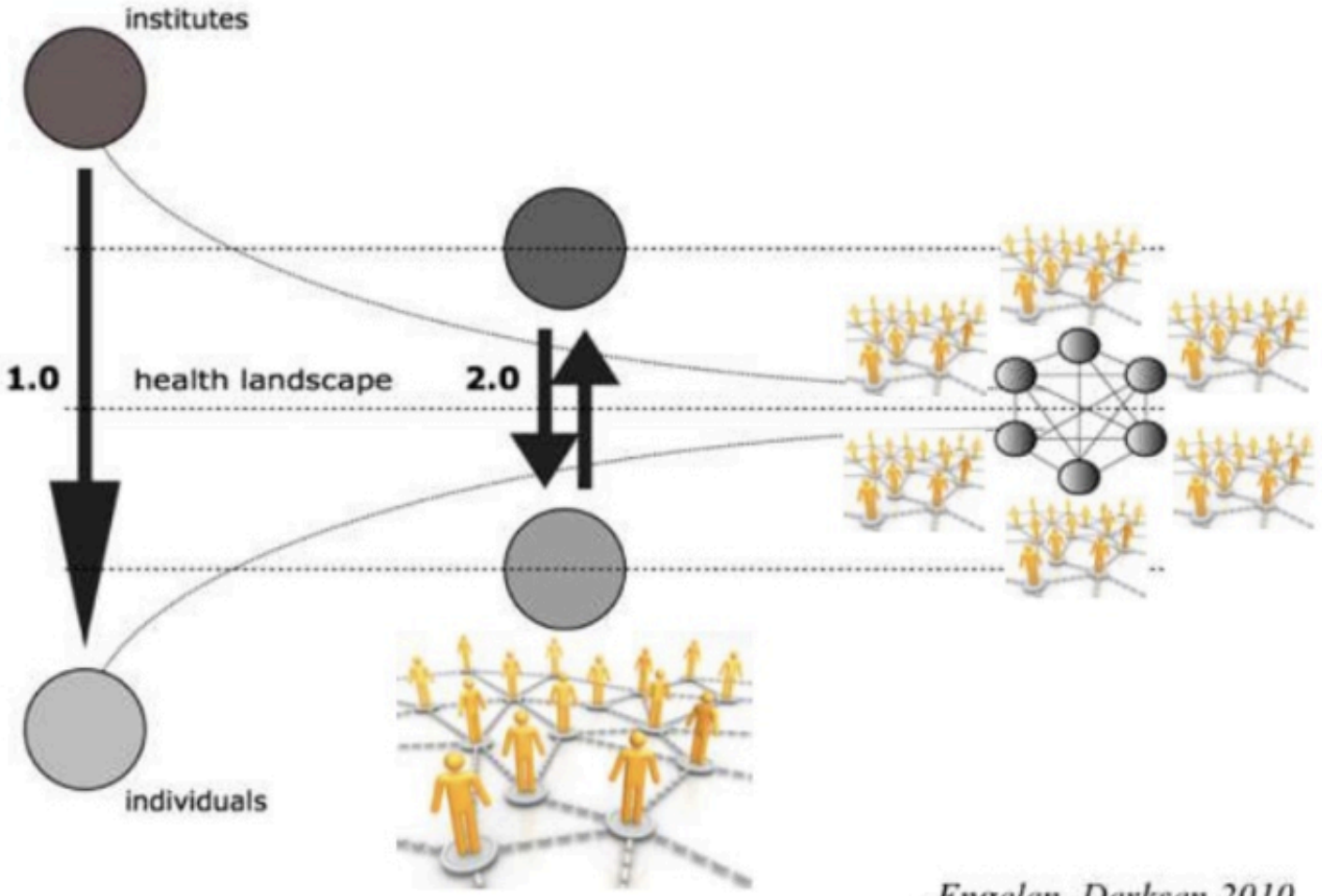
Citizen science



(**a**) The number of patients choosing to experiment with lithium carbonate peaked in the months after publication of a small clinical trial in Italy. Preliminary negative results from this patient-led study were announced before the first randomized control trial had started recruitment. (**b**) Aggregate view of FRS scores for all 348 patients analyzed in this study. These data were publicly available online during the study. (**c**) Illustration of disease progression curves of control individuals that are good and poor matches for a particular patient. Both control individuals would be considered comparable by traditional matching criteria. The PatientsLikeMe matching algorithm minimizes the area between the disease progression curves for a target patient and a control.

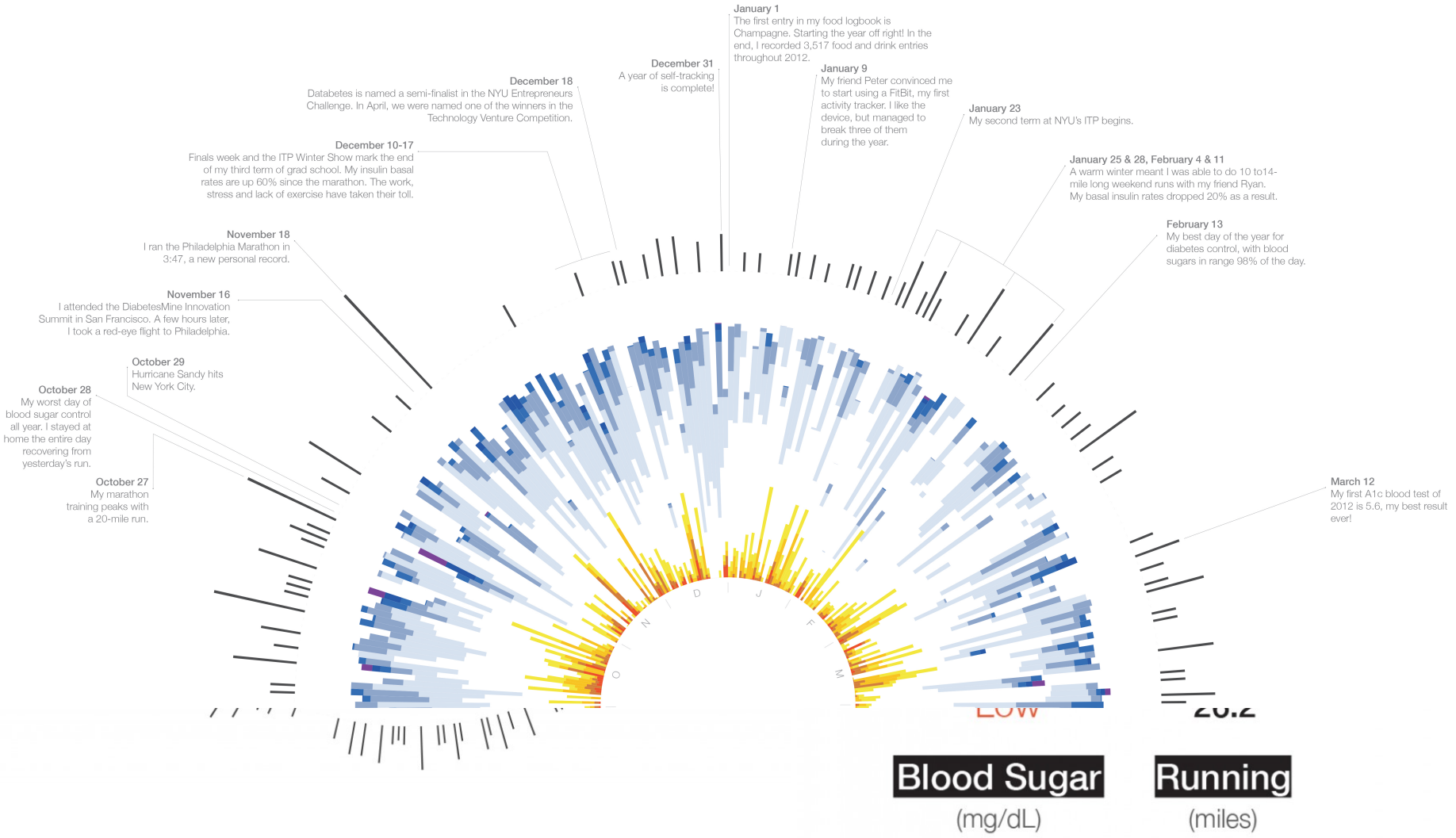
Wicks, P., Vaughan, T.E., Massagli, M.P. & Heywood, J., 2011, Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm, *Nature biotechnology*, 29(5), pp. 411-4

Web 2.0: "When the web began to harness the intelligence of its users." – Tim O'Reilly



Engelen, Derksen 2010

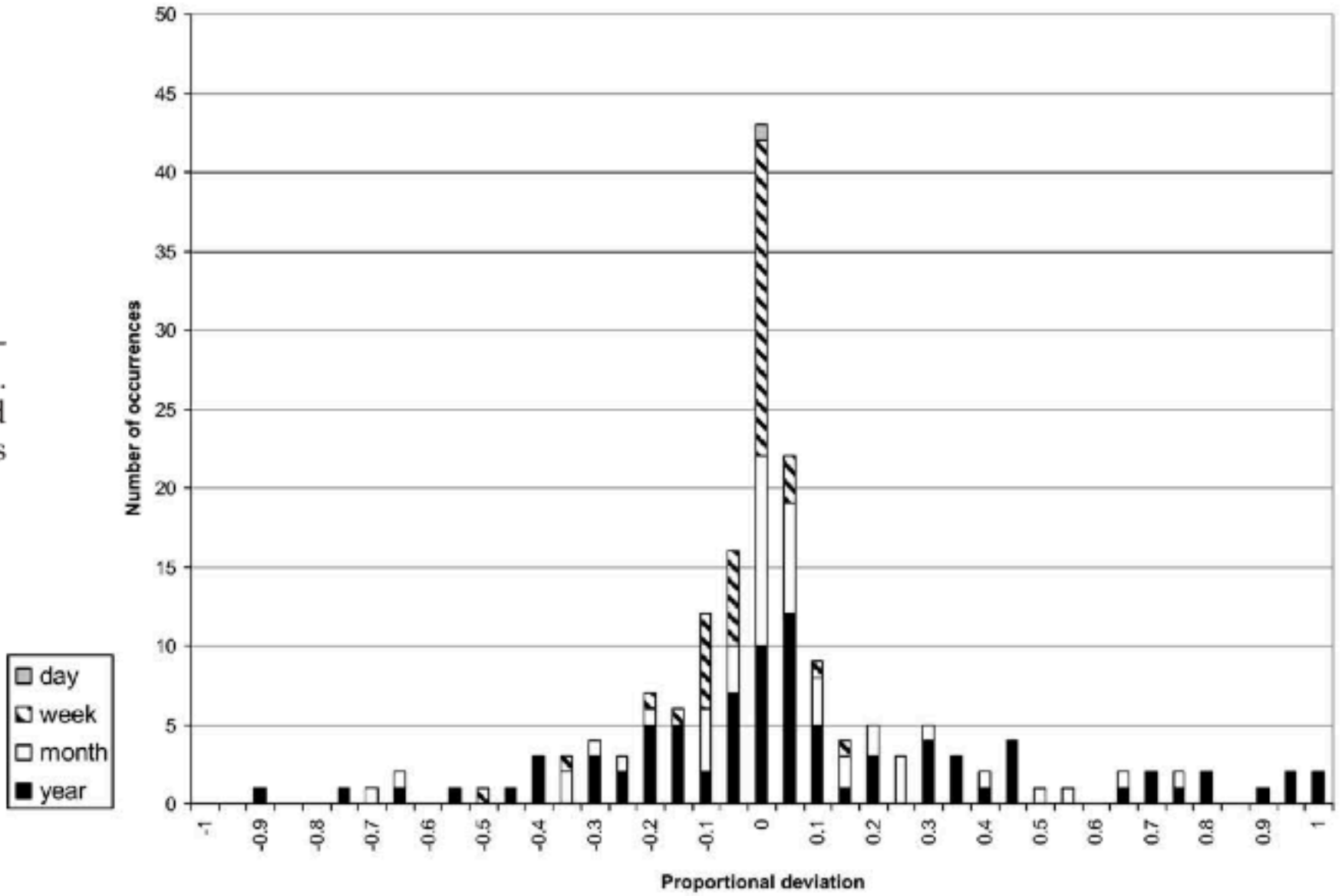
Quantified self



http://databetic.com

Challenges/open problems

- Data
- Replication/validation
- Privacy/sharing
- Translation



definite knowledge

language processing

Hripcsak G, Elhadad N, Chen YH, Zhou L, Morrison FP (2009) Using Empiric Semantic Correlation to Interpret Temporal Assertions in Clinical Texts. Journal of the American Medical Informatics Association 16(2):220-227.

Challenge: replication

Table 1: Reproducibility of research findings

Preclinical research generates many secondary publications, even when results cannot be reproduced.

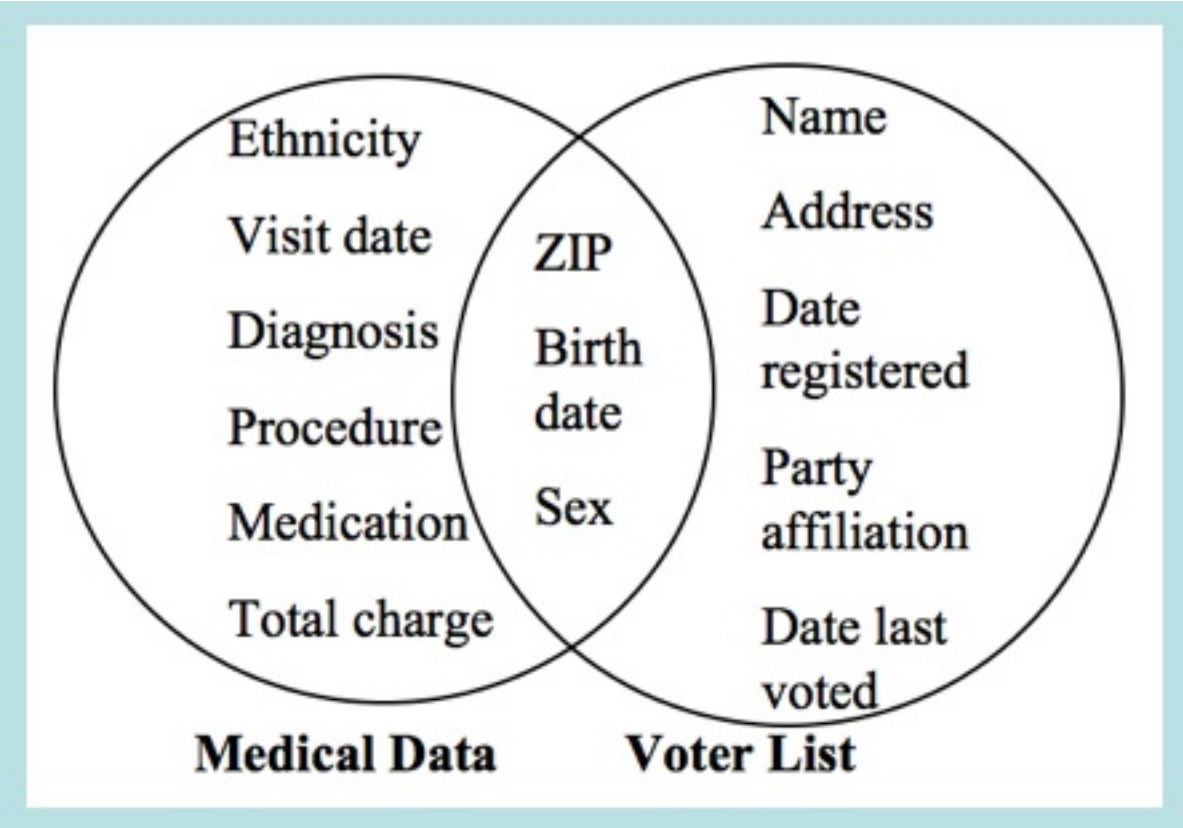
Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

*Source of citations: Google Scholar, May 2011.

Drug development: Raise standards for preclinical cancer research
C. Glenn Begley & Lee M. Ellis
Nature 483, 531–533 (29 March 2012) doi: 10.1038/483531a

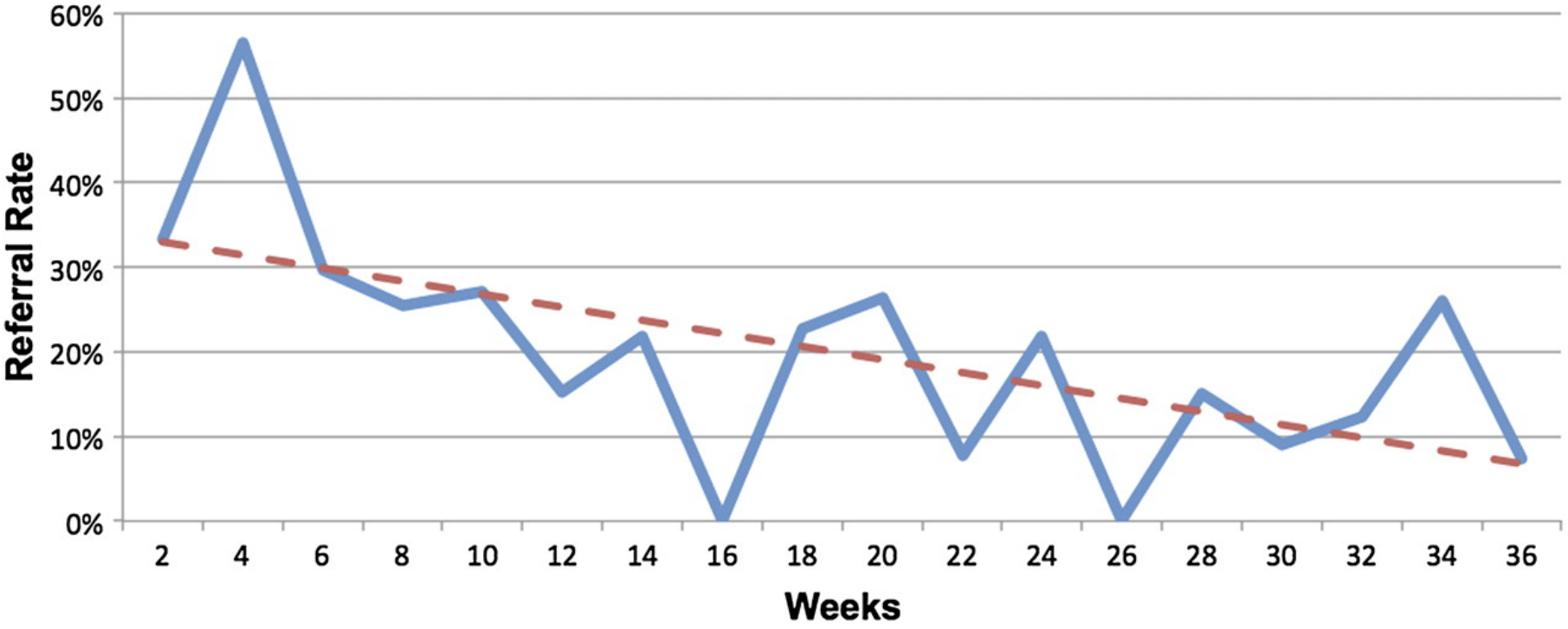
Challenge: Privacy



"87% of the U.S. Population are uniquely identified by {date of birth, gender, ZIP}."

http://latanyasweeney.org/work/identifiability.html

Challenge: Alert fatigue



Embi P J , and Leonard A C J Am Med Inform Assoc doi: 10.1136/amiajnl-2011-000743

Copyright © by the American Medical Informatics Association. All rights reserved.

For next week

- Be sure to read wicks et al. paper in syllabus!