

Informatics for Translating Clinical Research Findings to Populations: Small and Big Data

Epidemiology Fall Seminar UAB School of Public Health

**James J. Cimino, MD
Director, Informatics Institute
University of Alabama at Birmingham
Birmingham, Alabama, USA
September 10, 2018**

Unpacking the Title of this Talk

- Informatics
- Translational science
- Big data

What is biomedical informatics?

“The field that concerns itself with the cognitive, information processing and communication tasks of medical practice, education and research, including the information science and the technology to support these tasks.”

- Greenes RA. Shortliffe EH. JAMA 1990 Feb 23; 263(8):1114-20.

The art and science of organizing knowledge of human health and disease, and making it useful for problem solving.

- Jim Cimino

Using computers to make people better.

- UAB Hospital Custodial Staff

What is translational science?

T0: basic biomedical research, including preclinical and animal studies

T1: translation to humans, including Phase 1 clinical trials

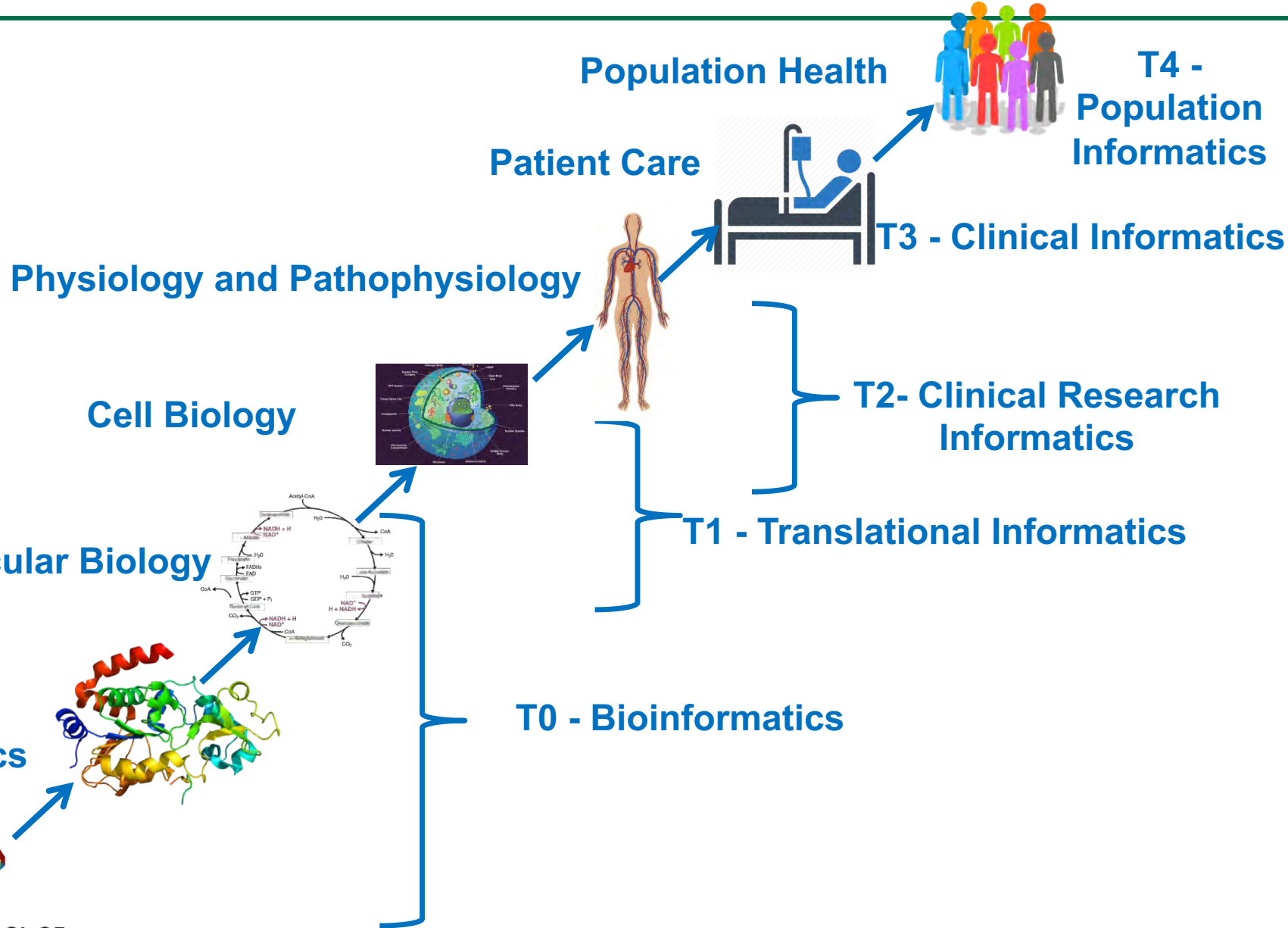
T2: translation to patients, including Phase 2 and 3 clinical trials

T3: translation to practice, including comparative effectiveness research, post-marketing studies, and clinical outcomes research

T4: translation to communities, including population level outcomes research and impacts of policy and change

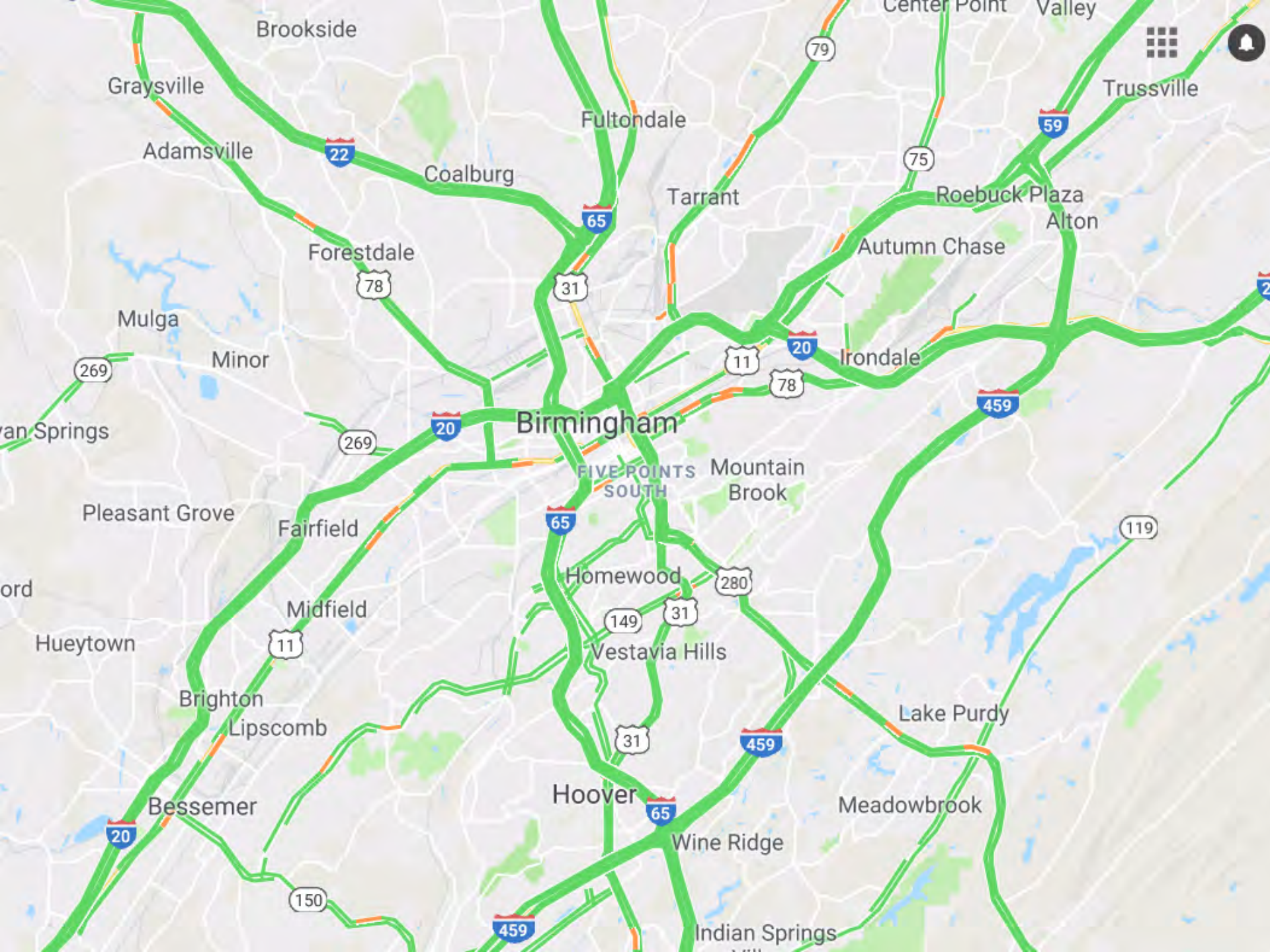
- UW-Madison Institute for Clinical and Translational Research

What is translational science?



What are big data?

- Volume – required
- Variety – optional



Brookside

Graysville

Adamsville

Coalburg

Fultondale

Tarrant

Roebuck Plaza

Trussville

Forestdale

Autumn Chase

Alton

Mulga

Minor

Irondale

an Springs

Birmingham

FIVE POINTS SOUTH

Mountain Brook

Pleasant Grove

Fairfield

Homewood

ord

Midfield

Vestavia Hills

Hueytown

Brighton

Lipscomb

Lake Purdy

Bessemer

Hoover

Meadowbrook

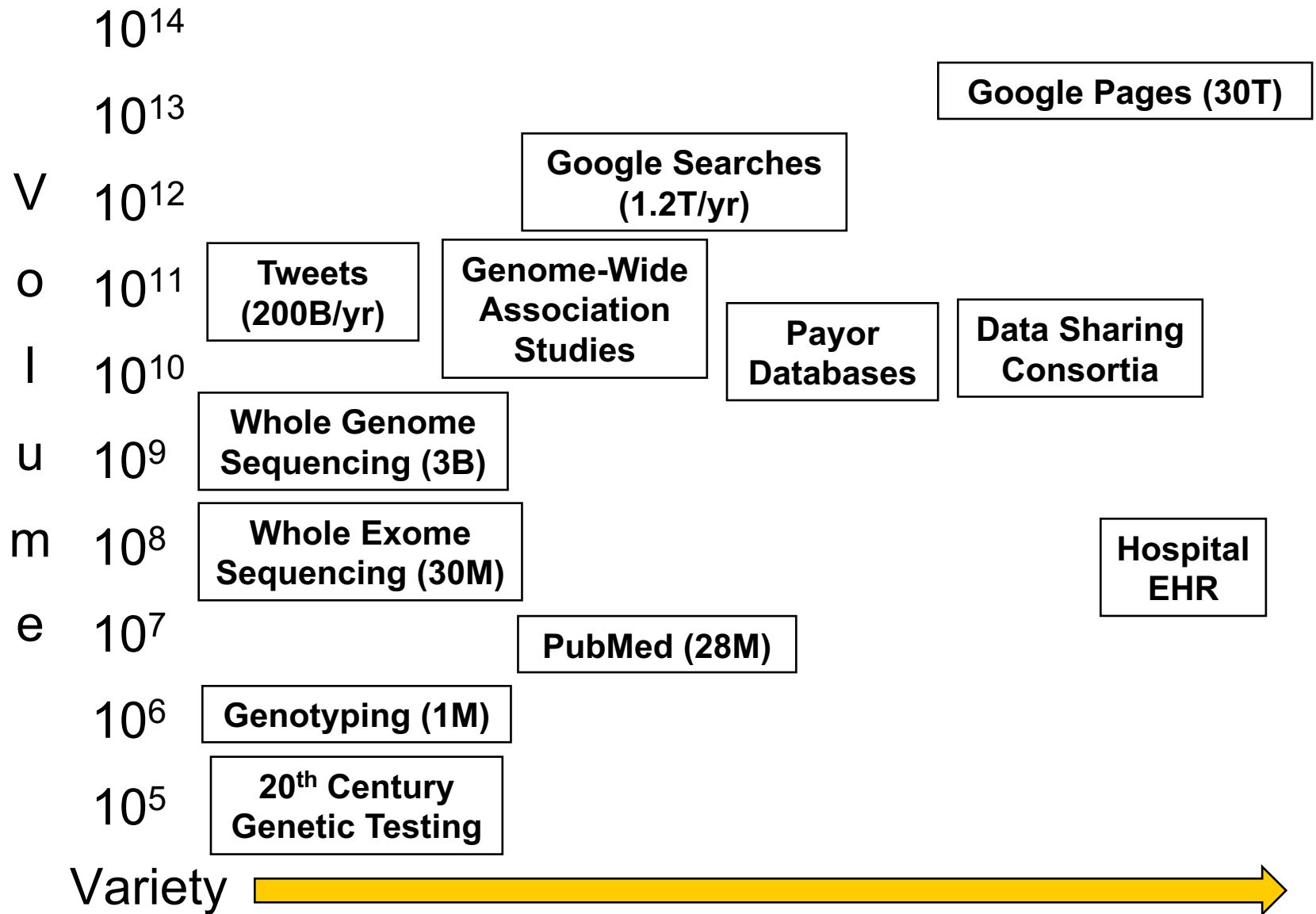
Wine Ridge

Indian Springs

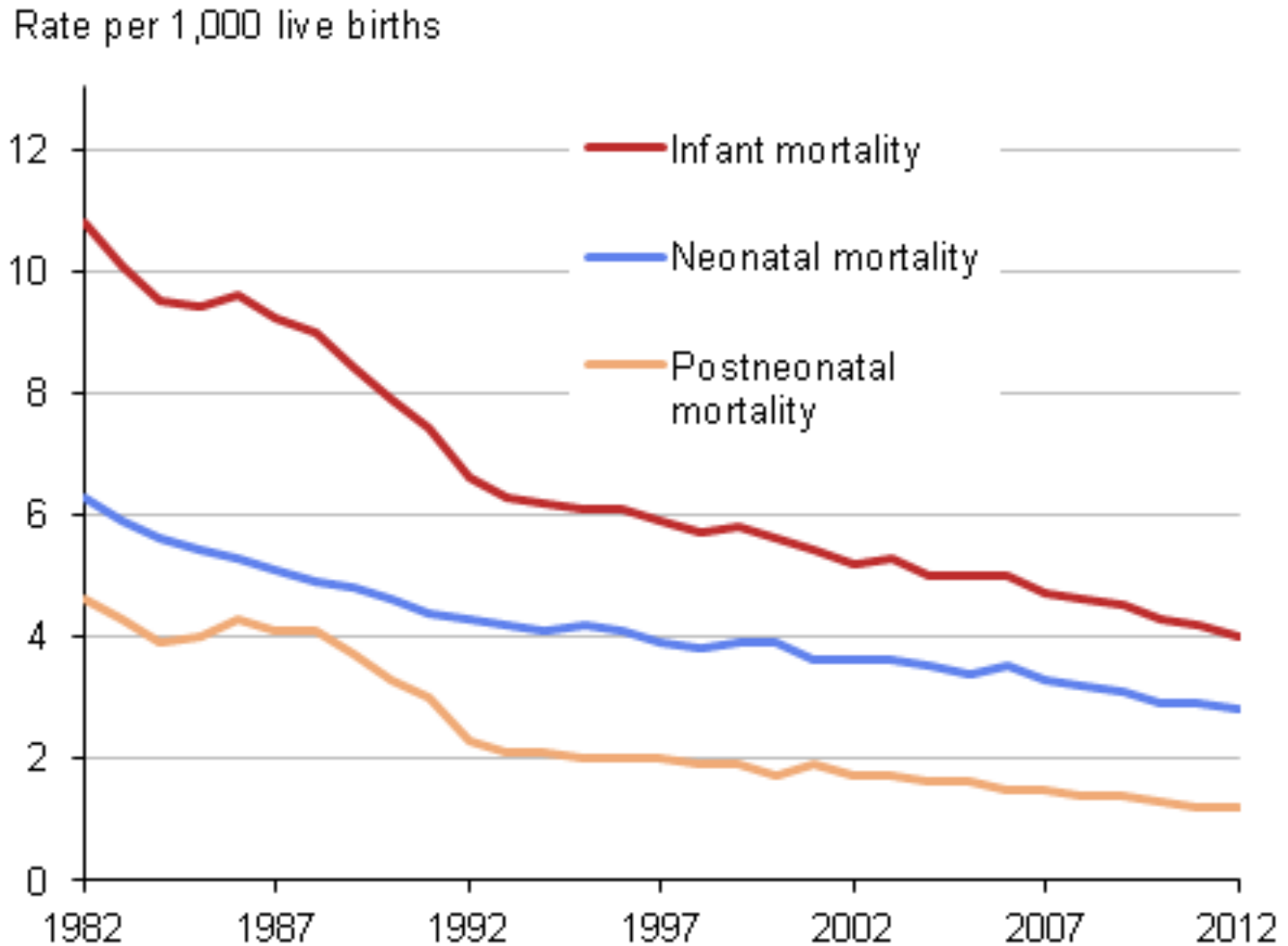
What are big data?

- Volume – required
- Variety – optional
- Velocity – maybe
- Value – enemy of the good
- Veracity – enemy of the good

What are big data in health?

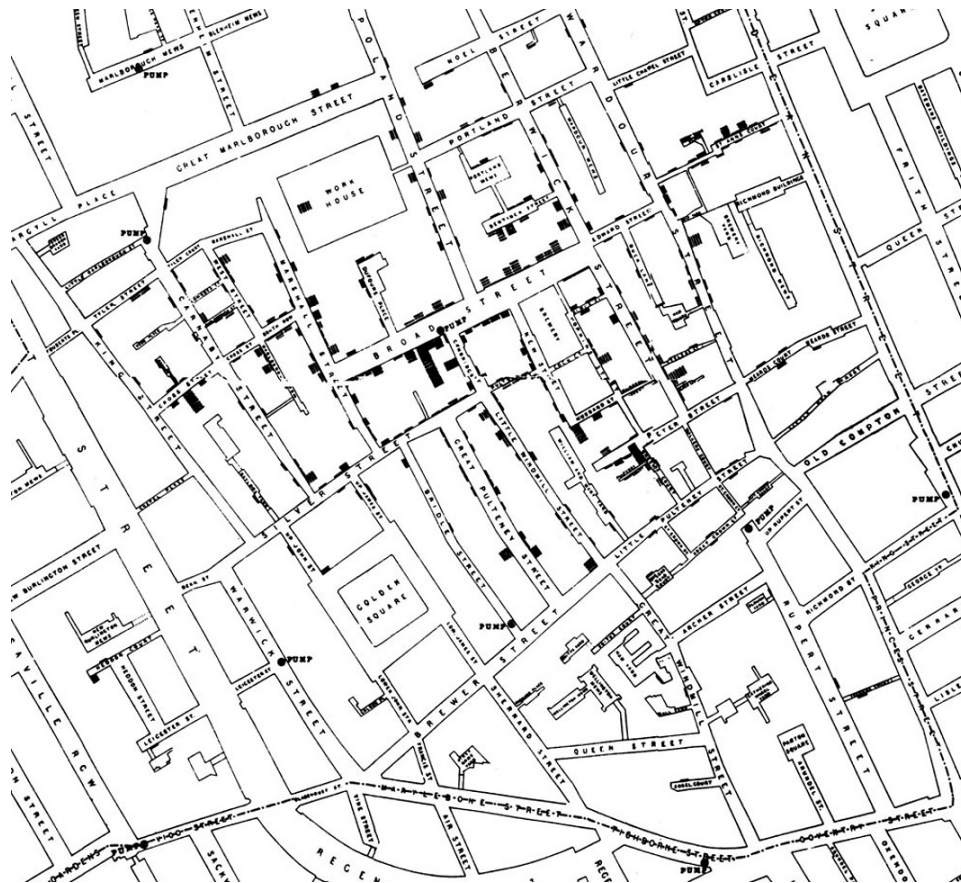


Big Data in Public Health – Temporal Trends



Big Data in Public Health – Mash-Ups

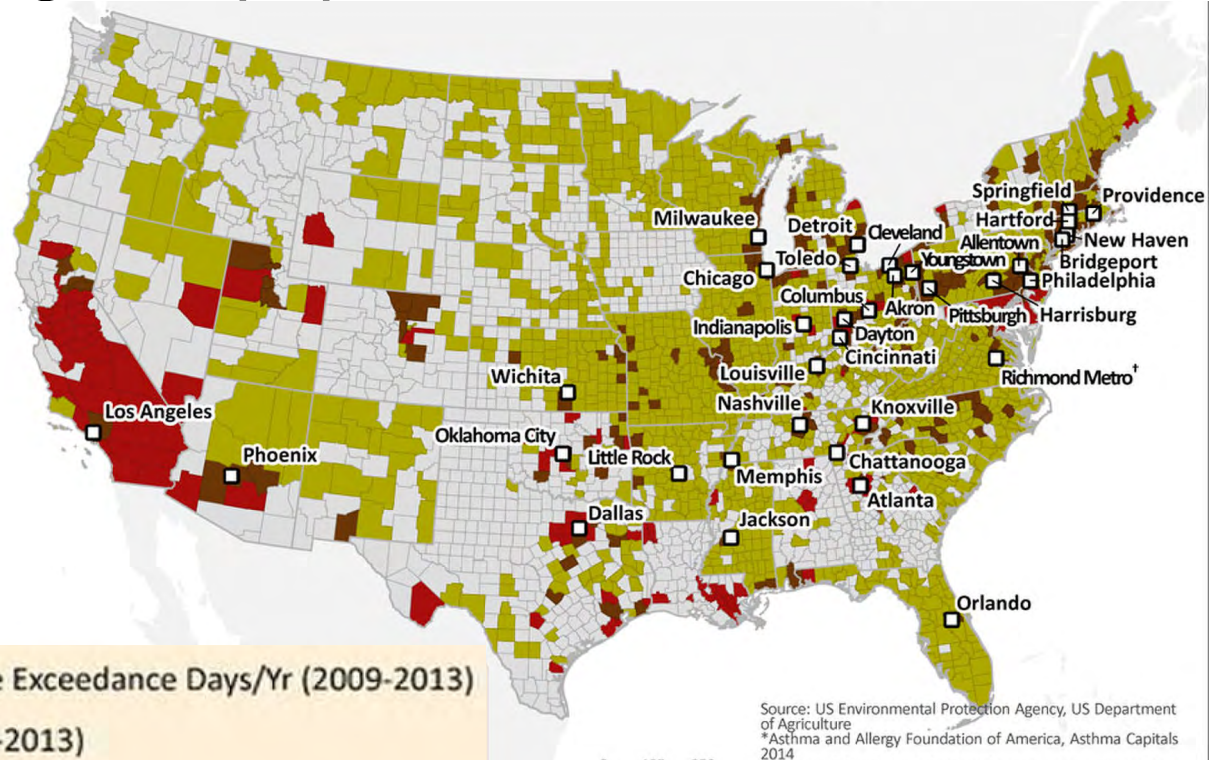
- Two or more data sets taken together
- Geocoding and population data



Cholera and the
Broad Street Pump
 $\sim 10^2$ data points

Big Data in Public Health – Mash-Ups

- Two or more data sets taken together
- Geocoding and population data



Source: US Environmental Protection Agency, USDA

Big Data in Public Health – Mash-Ups

- Two or more data sets taken together
- Geocoding and population data

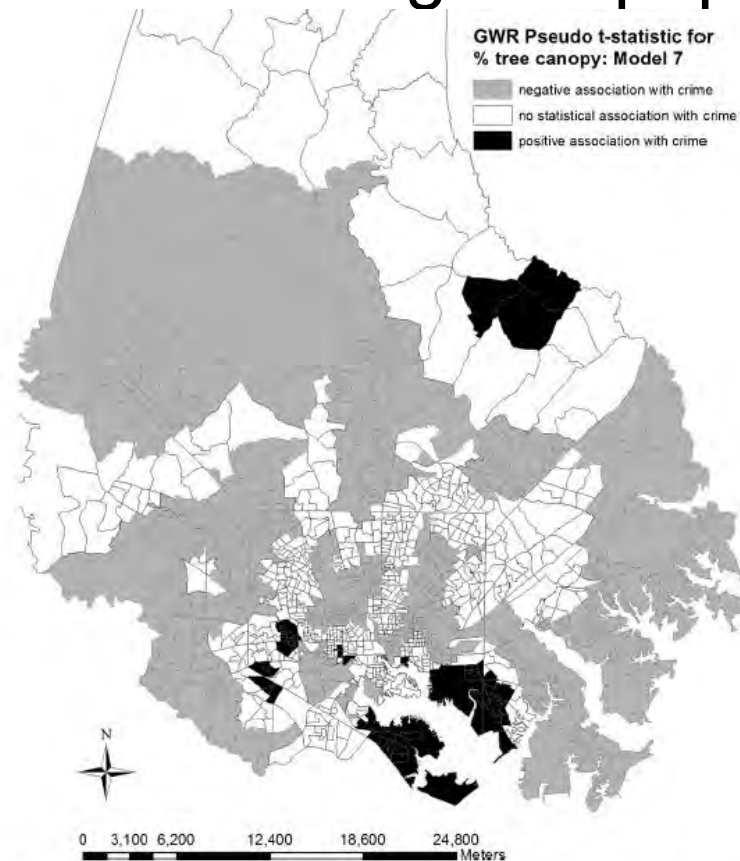


Fig. 1. Plot of pseudo t -statistic for percent tree canopy variable from GWR for model 7. The map shows Baltimore City and Baltimore County.



Troya A, Grove JM, O'Neil-Dunnea J. The relationship between tree canopy and crime rates across an urban–rural gradient in the greater Baltimore region. *Landscape and Urban Planning*. 2012; 106:262–270

Data Mining and Machine Learning (Wikipedia)

- Data mining: sorting through large data sets to identify patterns (typically “unsupervised”)
- Text mining: deriving high-quality information from text, typically through statistical pattern matching or natural language processing
- Machine learning: using statistical techniques to give computer systems the ability to “learn” (e.g., progressively improve performance on a specific task) with data, without explicit programming
- Deep learning: machine learning with neural nets

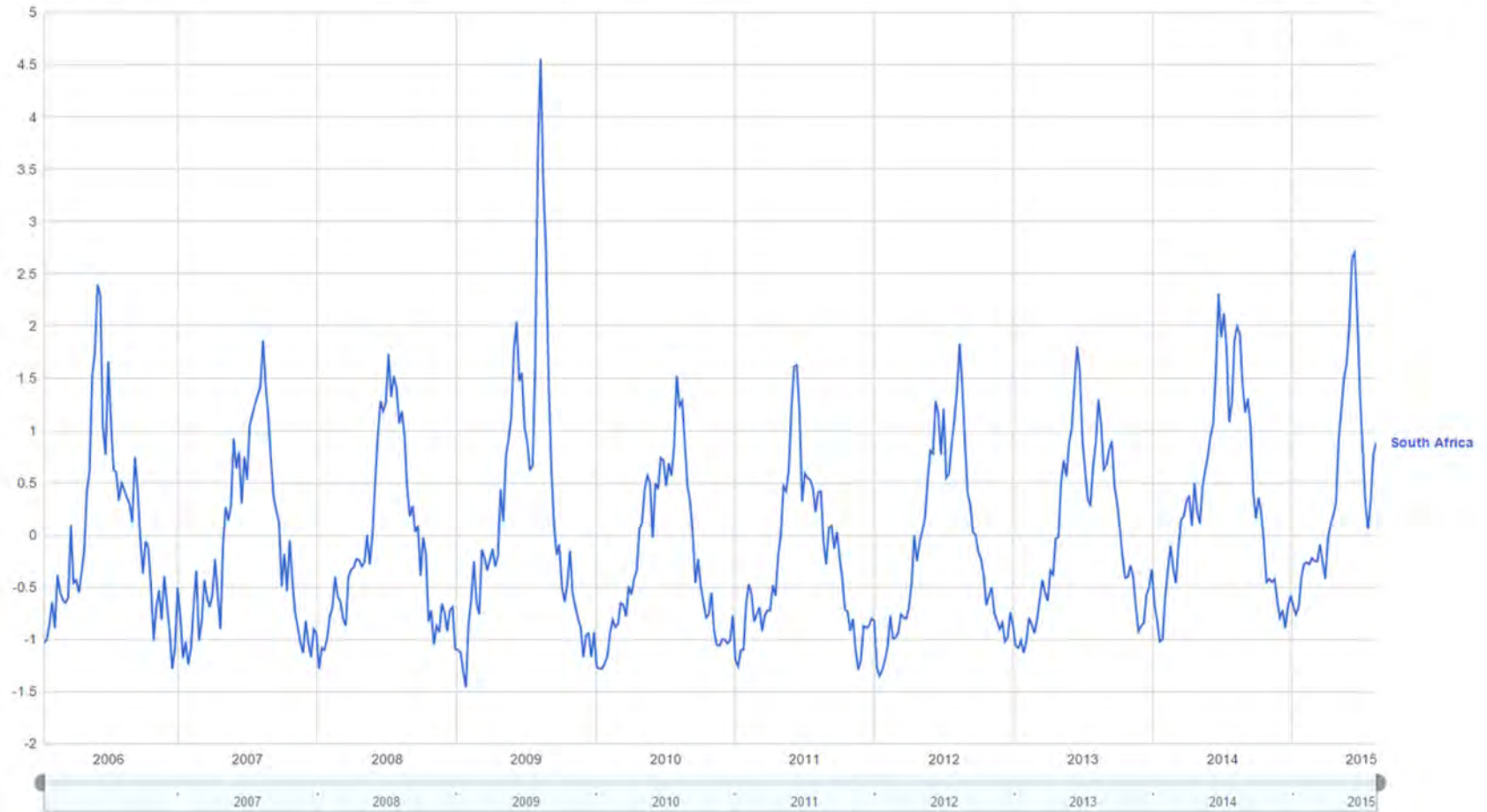
Data Mining and Machine Learning in Public Health

Herland M, Khoshgoftaar TM, Wald R. A review of data mining using big data in health Informatics. *Journal of Big Data*. 2014; 1:2

- Can message post data be used for dispersing clinically reliable information?
- Can search query data be used to accurately track epidemics throughout a population?
- Can Twitter post data be used to accurately track epidemics throughout a population?

Google Flu Trends

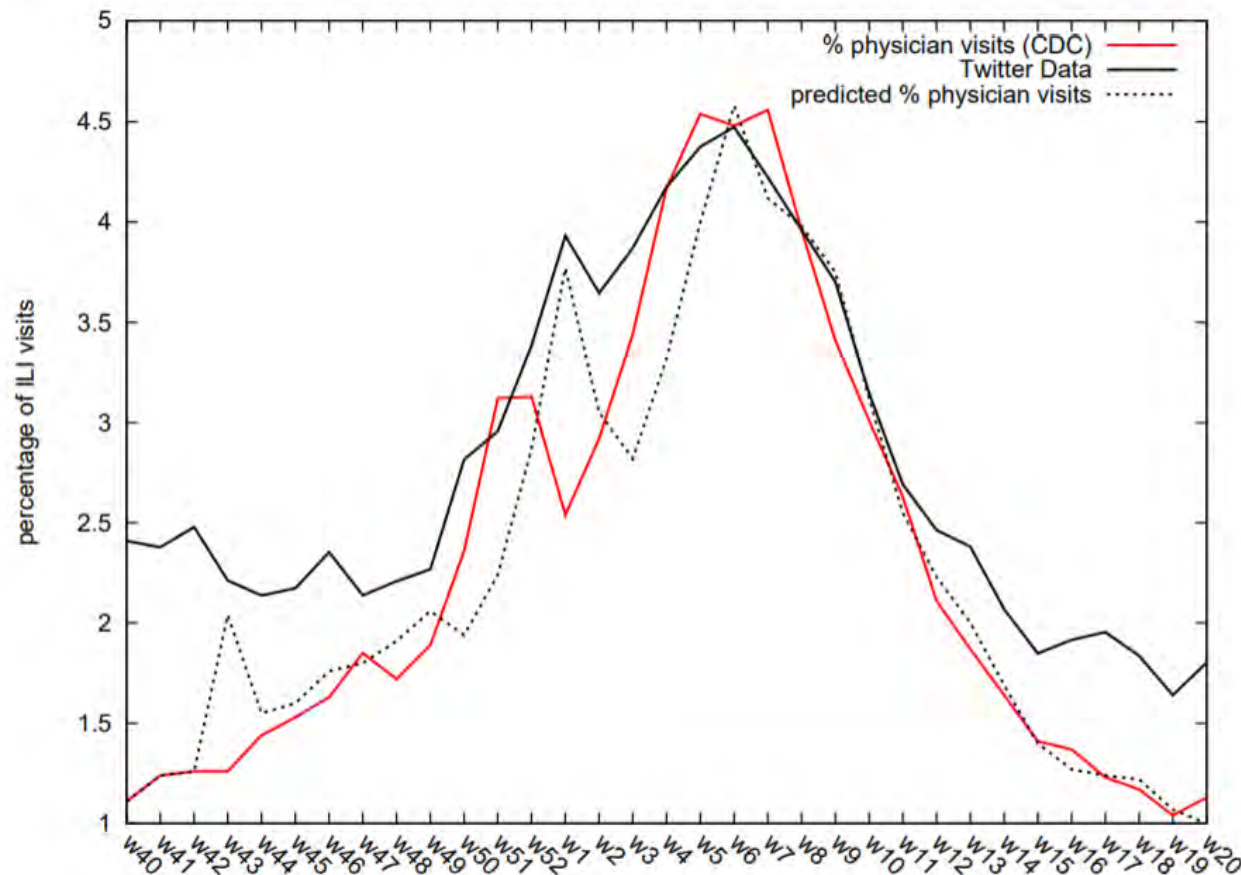
Flu search activity (standard deviation from baseline) ?



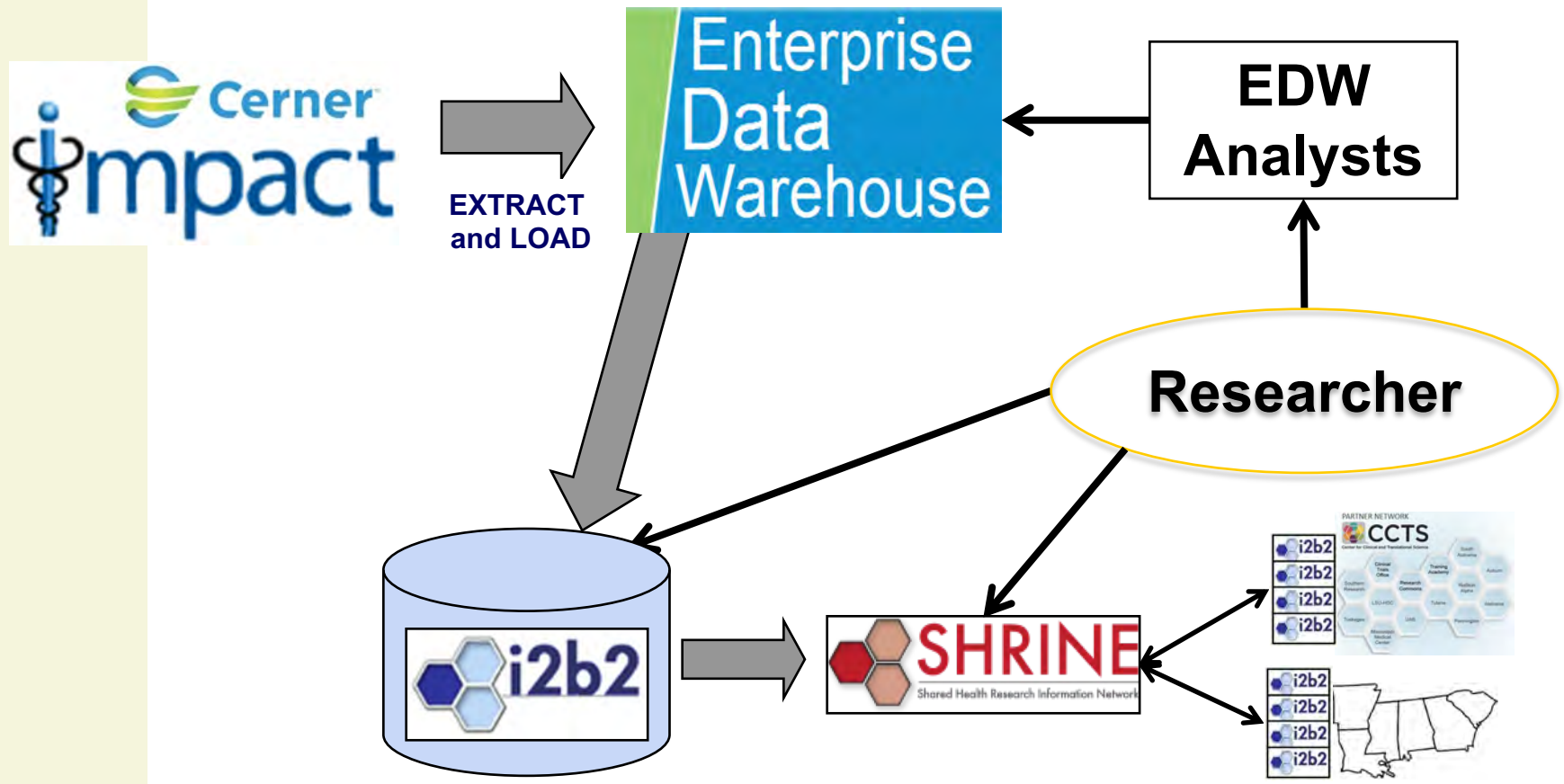
Data from Google Inc. Last updated: Aug 19, 2015

©2014 Google - [Help](#) - [Terms of Service](#) - [Privacy](#) - [Disclaimer](#) - [Discuss](#)

Achrekar H, Gandhe A, Lazarus R, Yu SH, Liu B: **Twitter improves seasonal influenza prediction.** In *International Conference on Health Informatics (HEALTHINF'12)*. 2012:61–70.



Electronic Health Record Data at UAB



What's in i2b2?

Row counts by Row Data Type			
Fact Family	Total Fact Counts	Distinct Patients	Earliest Date
Allergies	1,646,780	818,638	7/30/2003
BMI and Waist Measurements	3,782,881	562,852	5/12/2008
Clinical Diagnoses	4,805,169	383,643	5/19/1995
Lab Orders	141,875,875	529,950	6/25/2009
Home/Discharge Medications	11,197,426	533,216	4/28/1998
Tumor/Cancer Data	8,008,157	123,917	12/31/1965
Long-Term Problems	2,976,735	432,530	11/22/1965
Procedures	526,184	136,227	2/17/2014
DRGs	134,461	48,752	1/14/2014
Blood Pressure Measurements	48,425,976	574,164	8/5/2003
Inpatient Medications	36,100,284	341,748	5/18/2008
Height/Weight Measurements	17,570,643	593,767	8/28/2003
Lab Panels	9,677,955	453,094	10/26/2009
Encounter Insurance	11,369,918	695,140	12/1/2003
Immunizations	203,480	101,336	7/26/2008
Microbiology Tests	992,190	171,723	1/26/2008
Less Common Clinical Events	296	75	6/29/2011
Lab Powerplans	1,880,573	297,420	5/18/2008

Navigate Terms **Find**

Search by Names Search by Codes

Containing **proton pump**

Find Medications (Multum Ontology)

- proton pump inhibitors
 - dexlansoprazole - 3386
 - esomeprazole - 94153
 - lansoprazole - 7399
 - omeprazole - 65050
 - omeprazole-sodium bicarbonate - 1035
 - pantoprazole - 69107
 - rabeprazole - 1380

Query Tool

Query Name: Gastro--proton @10:10:46

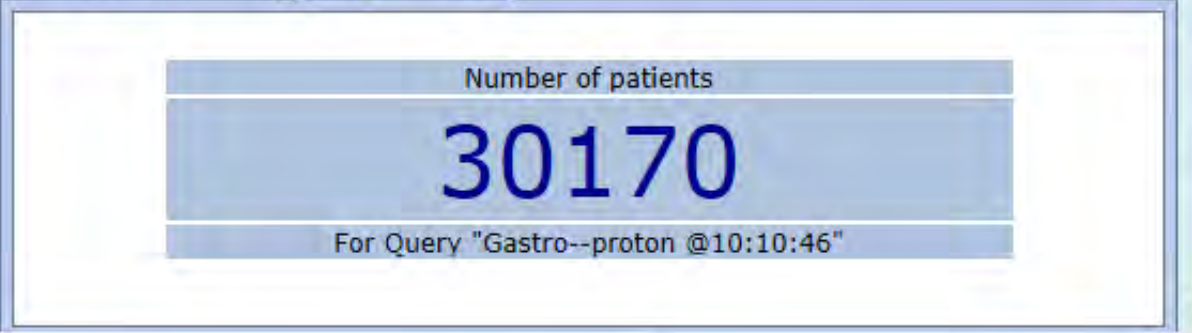
Temporal Constraint: Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
Gastro-esophageal reflux disease (ICD10:K21) - 37077			proton pump inhibitors - 181615					

one or more of these AND one or more of these AND drop a term on here

Run Query Clear Print Query 2 Groups New Group

Show Query Status Graph Results



UAB Central Authentication System

Enter your BlazerID and Password:

Log In

Having trouble logging in?

Visit [BlazerID Central](#) for help or to reset your password.

Contact AskIT at 205-996-5555 if you have any problems using this system or your BlazerID.

Security Notice

For security reasons, quit your web browser when finished accessing services that require authentication. Be wary of any program or web page that asks for your BlazerID and password. Legitimate UAB web pages prompting for your BlazerID and password should have addresses that begin with https:// (not http:) and contain "uab.edu". Also, your browser should visually indicate that you are accessing a secure page, usually a padlock icon.

This system is available only for authorized purposes by authorized users. Use for any other purpose may result in disciplinary action or criminal prosecution against the user.

	A	B	C	D	E	F	G	H	I
1	Patient Numk	Encounter Number	Concept Code	Start Date	End Date	Text/Desc	Numeric \	Units	
2	542822	1437307	ICD9:174.9	13-Feb-12	13-Feb-12	Breast cancer			
3	542822	3882648	SNMCT:4448450	6-Oct-12	6-Oct-12	Hypoglycemia			
4	542822	7214049	ICD9:174.9	6-Oct-12	6-Oct-12	Breast cancer			
5	542822	10843526	ICD9:441.7	26-Dec-14	26-Dec-14	Thoracoabdominal aneurysm, without mention of			
6	542822	2380775	ICD9:281.0	26-Dec-14	26-Dec-14	Pernicious anemia			
7	542822	2380775	ICD9:250.00	26-Dec-14	26-Dec-14	Type 2 diabetes mellitus without (mention of) com			
8	542822	5007014	ICD9:782.1	11-Jan-13	11-Jan-13	Rash.			
9	542822	5314589	VITALS:BMI	21-Sep-12	21-Sep-12	E	26.26	kg/m2	
10	542822	11428780	VITALS:BMI	27-Dec-14	27-Dec-14	E	26.34	kg/m2	
11	542822	10276801	UABLABS:315829	28-Dec-14	28-Dec-14	E	7	mEq/L	
12	542822	404417	UABLABS:316068	13-Feb-12	13-Feb-12	E	22	mg/dL	
13	542822	406870	UABLABS:317100	29-Jun-12	29-Jun-12	E	3.1	mg/dL	
14	542822	4318692	UABLABS:316068	26-Aug-12	26-Aug-12	E	48	mg/dL	
15	542822	3790992	UABLABS:315848	11-Jan-13	11-Jan-13	E	65	Units/L	
16	542822	5007014	UABLABS:317870	2-Sep-12	2-Sep-12	E	4.6	10^3/cmm	
17	542822	3790992	UABLABS:316050	2-Sep-12	2-Sep-12	E	0.8	mg/dL	
18	542822	7576461	UABLABS:315870	2-Sep-12	2-Sep-12	E	32	Units/L	
19	542822	404417	UABLABS:317100	8-Oct-12	8-Oct-12	E	2.9	mg/dL	
20	542822	5012603	UABLABS:317298	26-Dec-14	26-Dec-14	E	142	mEq/L-Historical	
21	542822	403278	UABLABS:317144	21-Sep-15	21-Sep-15	E	3.7	mEq/L-Historical	
22	542822	5012603	UABLABS:316588	2-Sep-12	2-Sep-12	E	24	mEq/L-Historical	
23	542822	2731270	UABLABS:316000	3-Sep-12	3-Sep-12	E	21	Units/L	
24	542822	3824200	UABLABS:752820	7-Oct-12	7-Oct-12	Performed		Unspecified	
25	542822	3824200	UABLABS:317144	22-Aug-12	22-Aug-12	E	4.1	mEq/L-Historical	
26	542822	3882648	UABLABS:316094	3-Sep-12	3-Sep-12	E	10.3	mg/dL	
27	542822	3882648	UABLABS:696694	8-Oct-12	8-Oct-12	E	3.06	ng/mL	

Reuse of EHR Data - Advantages

- From clinicians and instruments
- Data about the “phenome”
- Becoming universally available
- “Almost free”
- Post-market surveillance (phase 4 studies)
- Can be used to validate research studies
- Can be used to *replicate* research studies
- Understanding of the limitations is critical

Reuse of EHR Data - Caveats

1. EHRs may contain inaccurate (or incorrect) data
2. EHRs often do not tell a complete patient story
3. Many of the data have been transformed/coded for purposes other than research and clinical care
4. Data in clinical notes (text) may not be recoverable
5. EHRs may present multiple sources of data that affect data provenance
6. Data granularity in EHRs may not match the needs
7. Research protocols and clinical care are different
8. Beware of “controlled” terminologies

Annual Incidence of Shock (Normalized)

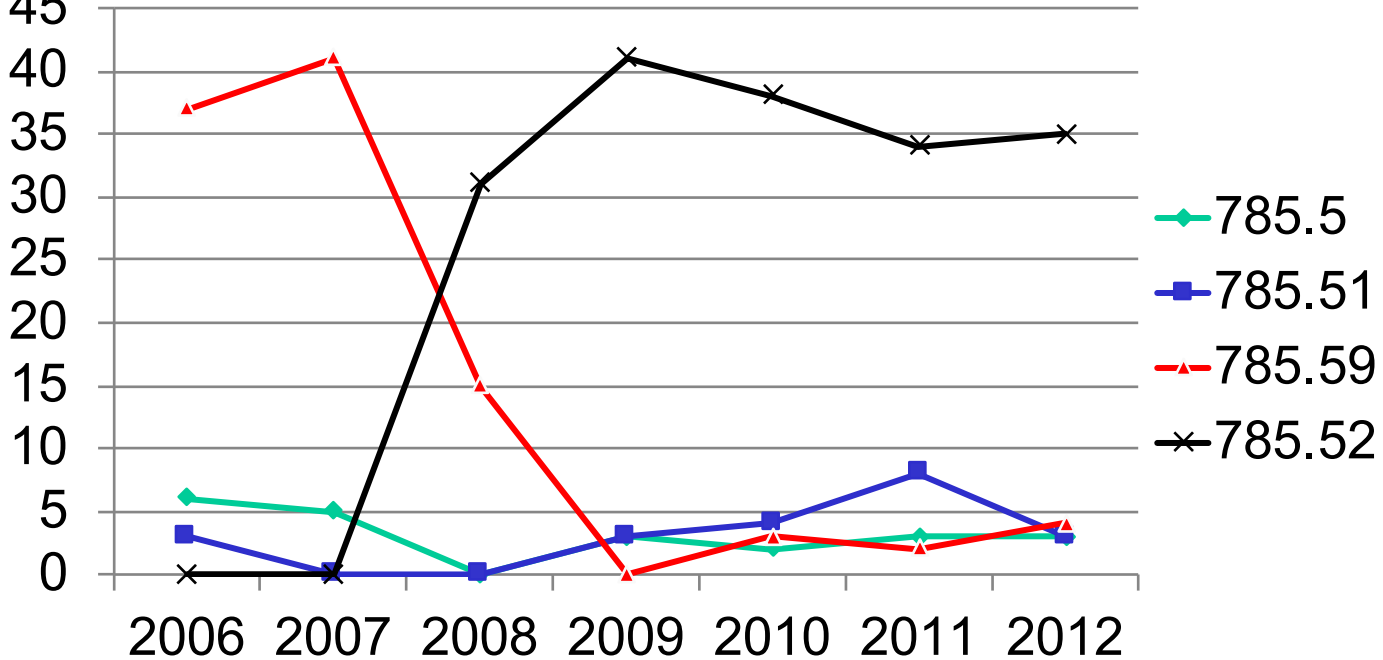
2007

Diagnosis	ICD9-CM Code	ICD9-CM Name
Cardiogenic Shock	785.51	Cardiogenic Shock

2008

Diagnosis	ICD9-CM Code	ICD9-CM Name
Cardiogenic Shock	785.51	Cardiogenic Shock

Septicemic Shock	45	Septic shock
Hypovolemic Shock	35	Shock without mention of infection, NEC



Caveats and Recommendations

Hersh WR, Weiner MG, Embi PJ, et al. **Caveats for the use of operational electronic health record data in comparative effectiveness research.** Med Care. 2013 Aug;51(8 Suppl 3):S30-7.

Hersh WR, Cimino J, Payne PR, et al. **Recommendations for the use of operational electronic health record data in comparative effectiveness research.** EGEMS (Wash DC). 2013 Oct 8;1(1):1018. doi: 10.13063/2327-9214.1018.

Alabama Genome Health Initiative

UAB MEDICINE

Knowledge that will change your world

[Patient Care](#) [Find A Provider](#) [Plan Your Visit](#) [Locations](#) [Contact](#)

[eMedicine](#) [myUABMEDICINE](#)



Telling Alabama's **GENOMIC STORY** ALABAMA GENOMIC HEALTH INITIATIVE



CLICK HERE TO SIGN UP NOW >

Online enrollment for the Alabama Genomic Health Initiative is now available.

The What, Why and How of AGHI

Your Questions Answered

What  AGHI?

▶ PLAYLIST

Or enroll by calling 855-462-6850.

Unlocking the potential in the human genome is part of the future of health care, and as the state's leader in genomic medicine, UAB Medicine is taking a major step forward with the Alabama Genomic Health Initiative (AGHI).

The program is aimed at preventing and treating disease, including certain types of cancer, heart problems, and genetic disorders. In collaboration with the

Southern All of Us Network (Y'all of Us)

[ABOUT](#)[HOW TO JOIN](#)[NEWS & EVENTS](#)[COMMUNITY](#)[LOG IN](#)[Español](#)

The future of health begins with you.



The *All of Us* Research Program has a simple mission. We want to speed up health research breakthroughs. To do this, we're asking one million people to share health information. In the future, researchers can use this to conduct thousands of health studies.



Our website uses cookies, as almost all websites do, to help provide you with the best experience we can. Cookies are small text files that are placed on your computer or mobile phone when you browse websites.

Cookies help us:

- Make our website work as you'd expect.
- Provide a message we believe is more relevant to you.

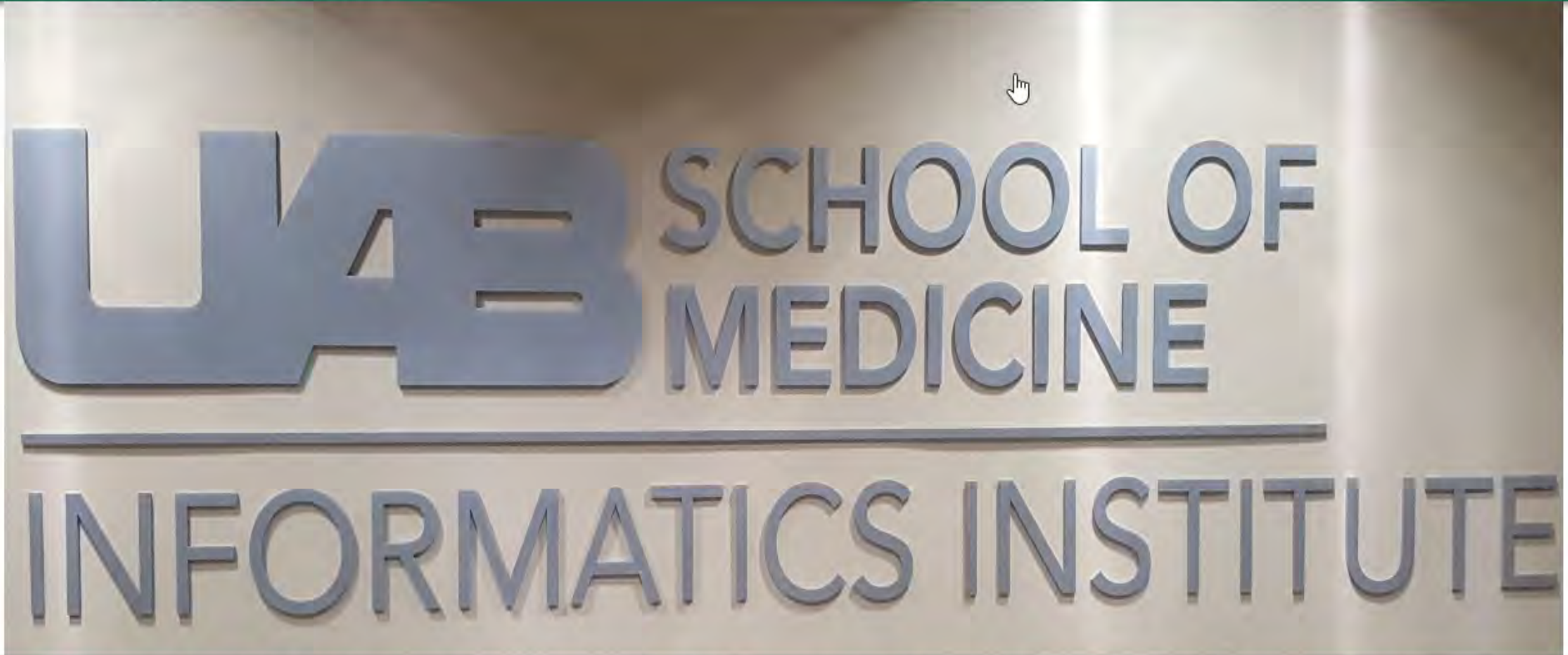
We do not use cookies to:

- Collect any personally identifiable information.
- Collect any sensitive information.
- Pass personally identifiable data to third parties.

You can learn more about all the cookies and the information we collect by reading our [Privacy Policy](#). If you don't want to use cookies you can either exit the website or [change your browser settings](#).

[Need Help?](#)

INFORMATICS INSTITUTE

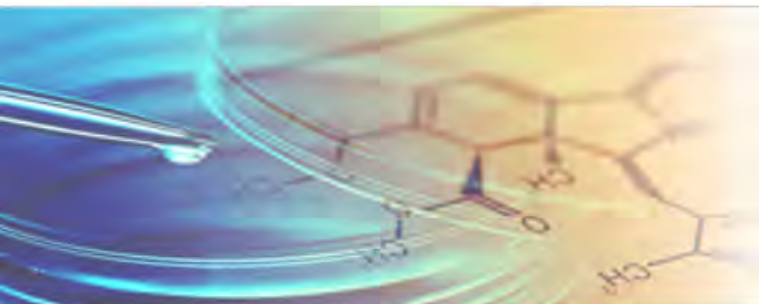


The Informatics Institute provides a home for focusing biomedical informatics activities in the UAB School of Medicine. The Institute comprises core and affiliated faculty and staff with expertise across the biomedical informatics spectrum, including bioinformatics, computational and systems biomedicine, translational informatics, clinical research informatics, and clinical informatics.

Our mission is fourfold:

Upcoming Events

SEP	Epidemiology Seminar Series: "Informatics for Translating Clinical Research Findings to Populations: Small and Big Data"
10	



RESEARCH COMMONS

Center for Clinical and Translational Science

Informatics

[Informatics Gateway](#)

[Bioinformatics](#)

[Clinical Research Informatics](#)

[i2b2](#)

[Access Data](#)

[Trainings](#)

[Projects](#)

[Team](#)

[BERD](#)

[Panels](#)

[Funding Opportunities](#)

[Grant Help](#)

[Home](#) | [Research Commons](#) | [Informatics](#)

Informatics

CCTS Informatics provides the resources and expertise (both bioinformatics and clinical informatics) to support biomedical collaboration and consultation across the translational research spectrum. We offer help with study design; access to summary, limited (de-identified), and fully identified data sets; innovative tools to support clinical, translational, and outcomes research; and data analytic services. Our vision is to build a vibrant community of collaborating informaticians not only across the CCTS Hub with its academic medical system, but also across the regional CCTS Partner Network and national CTSA Consortium.

Informatics Gateway

The Informatics Gateway is an enterprise-wide forum for collaborative project

RESEARCH

[Home](#) [About Us](#) [Academic Programs](#) [People](#) [Research](#) [Outreach](#) [News](#) [Give](#) [The College](#)

Research

Research Clusters

Research Labs

[Collaborative Computing Laboratory](#)

[Computer Forensics](#)

[iProgress](#)

[KDDM Lab](#)

[Shape Lab](#)

[SECRETLab](#)

[SPIES Group](#)

Undergraduate Research

[Home](#) [Research](#) [Research Labs](#) [KDDM Lab](#)

KDDM Lab



[KDDM Lab website](#)

The research conducted in the Knowledge Discovery and Data Mining Research (KDDM) Lab combines development of pattern matching algorithms, statistical techniques, distributed database techniques, and visualization methods.

The current research activities of KDDM focus on the following areas:

- multimedia data mining, in particular images and videos
- event/anomaly detection and analysis
- spatio-temporal data mining
- social network mining
- high performance data mining
- bioinformatics
- computer forensics

We have applied research to several domains, and collaborate closely with cyber-security specialists, colleagues in the departments of Physical Medicine & Rehabilitation, Biostatistics, and Government, and with industrial

Big Data Research and Analytics Lab

Directory Category: Core

The Big Data Research and Analytics Laboratory (BDRAL) is a core facility under the Center for Integrated Systems at the University of Alabama at Birmingham (UAB).

The BDRAL is a joint initiative of the Department of Electrical and Computer Engineering (School of Engineering), the Department of Neurology (School of Medicine) and UAB IT Research Computing. The lab has been set up to bring together engineers, physicians, computer scientists and statisticians to develop novel ways to manage, analyze, and visualize very large data sets. In addition to serving as a facility to process data generated (at UAB) and gathered (i.e. publicly available datasets, such as genomics data and medical imaging data) through research, the lab is also working on emerging technologies and approaches to big data analytics that would be applicable to a large number of fields, from engineering to medicine to business.

The BDRAL is the outcome of a partnership between two school-based administrative units at

- Public health research is the natural last step in translational science
- Big data provide volume and variety
- Mash-ups can provide new insights
- Informatics provides the conceptual abstraction
- UAB has EHR and genomic data
- UAB also has informatics expertise

