Barriers to Expanded Data-Sharing and the Tremendous Good It Can Do:

6 Critical Ways *Re-identification "Science"* Has Failed to Support Sound Public Policies

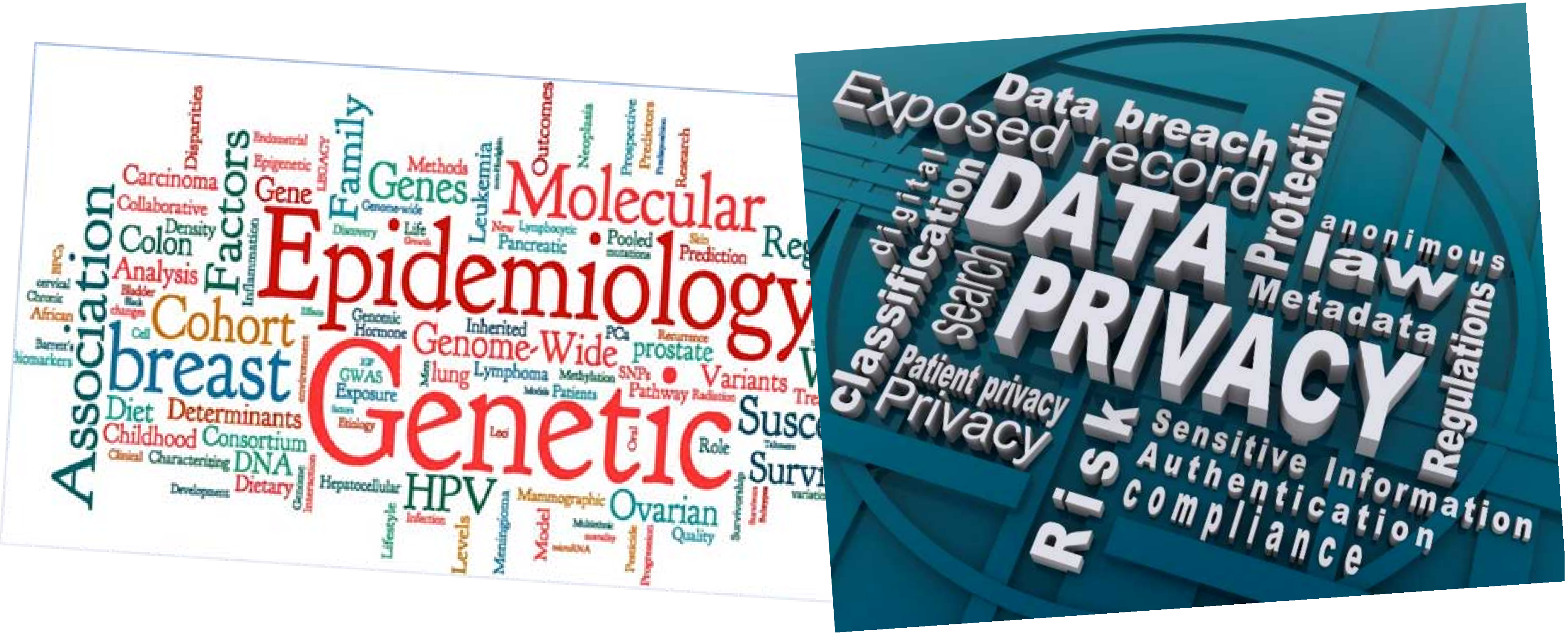
Daniel C. Barth-Jones, M.P.H., Ph.D.

Assistant Professor of Clinical Epidemiology,
Mailman School of Public Health
Columbia University

db2431@columbia.edu

@dbarthjones *on Twitter*

A Historic and Important Societal Debate is underway…



Public Policy Collision Course

The Research Value of De-identified Data

VARIATION IN HEALTH CARE SPENDING
Target Decision Making, Not Geography
Toward Quality Measures for Population Health and the Leading Health Indicators
INSTITUTE OF MEDICINE
PATIENT SAFETY
A NEW STANDARD FOR CARE
THE PRECISION MEDICINE INITIATIVE®
THE LEARNING HEALTH SYSTEM SERIES
ROUNDTABLE ON VALUE & SCIENCE-DRIVEN HEALTH CARE
CORE MEASUREMENT NEEDS FOR BETTER CARE, BETTER HEALTH AND LOWER COSTS
Counting What Counts
Workshop Summary
OBSERVATIONAL STUDIES IN A LEARNING HEALTH SYSTEM
Workshop Summary
Capturing Social and Behavioral Domains in Electronic Health Records
PHASE 1
FOR THE PUBLIC'S HEALTH
Investing in a Healthier Future
HOW FAR HAVE WE COME IN REDUCING HEALTH DISPARITIES?
Progress Since 2000
Workshop Summary
INSTITUTE OF MEDICINE
Sentinel Initiative
Transforming how we monitor the safety of FDA-regulated products
pcori®
Patient-Centered Outcomes Research Institute
ClinicalStudy DataRequest.com

Misconceptions about HIPAA De-identified Data:

"It doesn't work…" "easy, cheap, powerful re-identification" (Ohm, 2009 "*Broken Promises of Privacy*")

**Pre-HIPAA* Re-identification Risks {Zip5, Birth date, Gender} able to identify 87%?, 63%, 28%? of US Population (Sweeney, 2000, Golle, 2006, Sweeney, 2013)

- Reality: HIPAA compliant de-identification provides important privacy protections
  - Safe harbor re-identification risks have been estimated at 0.04% (4 in 10,000) (Sweeney, NCVHS Testimony, 2007)
- Reality: Under HIPAA de-identification requirements, re-identification is expensive and time-consuming to conduct, requires serious computer/mathematical skills, is rarely successful, and usually uncertain as to whether it has actually succeeded

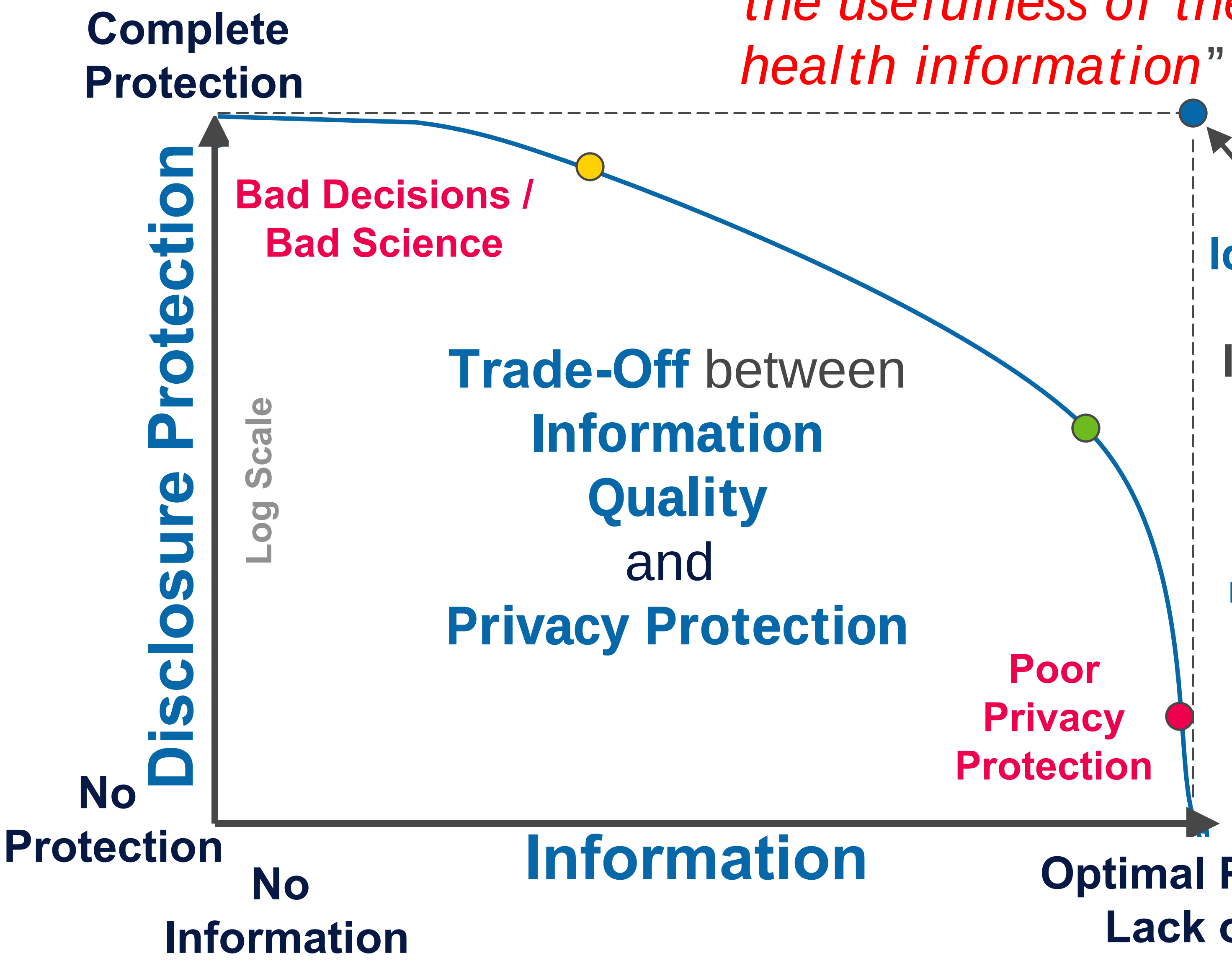
Misconceptions about HIPAA De-identified Data:

“It works perfectly and permanently...”

- Reality:
  - —Perfect de-identification is not possible
  - —De-identifying does not free data from all possible subsequent privacy concerns
  - —Data is never permanently “de-identified”... (There is no guarantee that de-identified data will remain de-identified regardless of what you do to it after it is de-identified.)

The Inconvenient Truth:

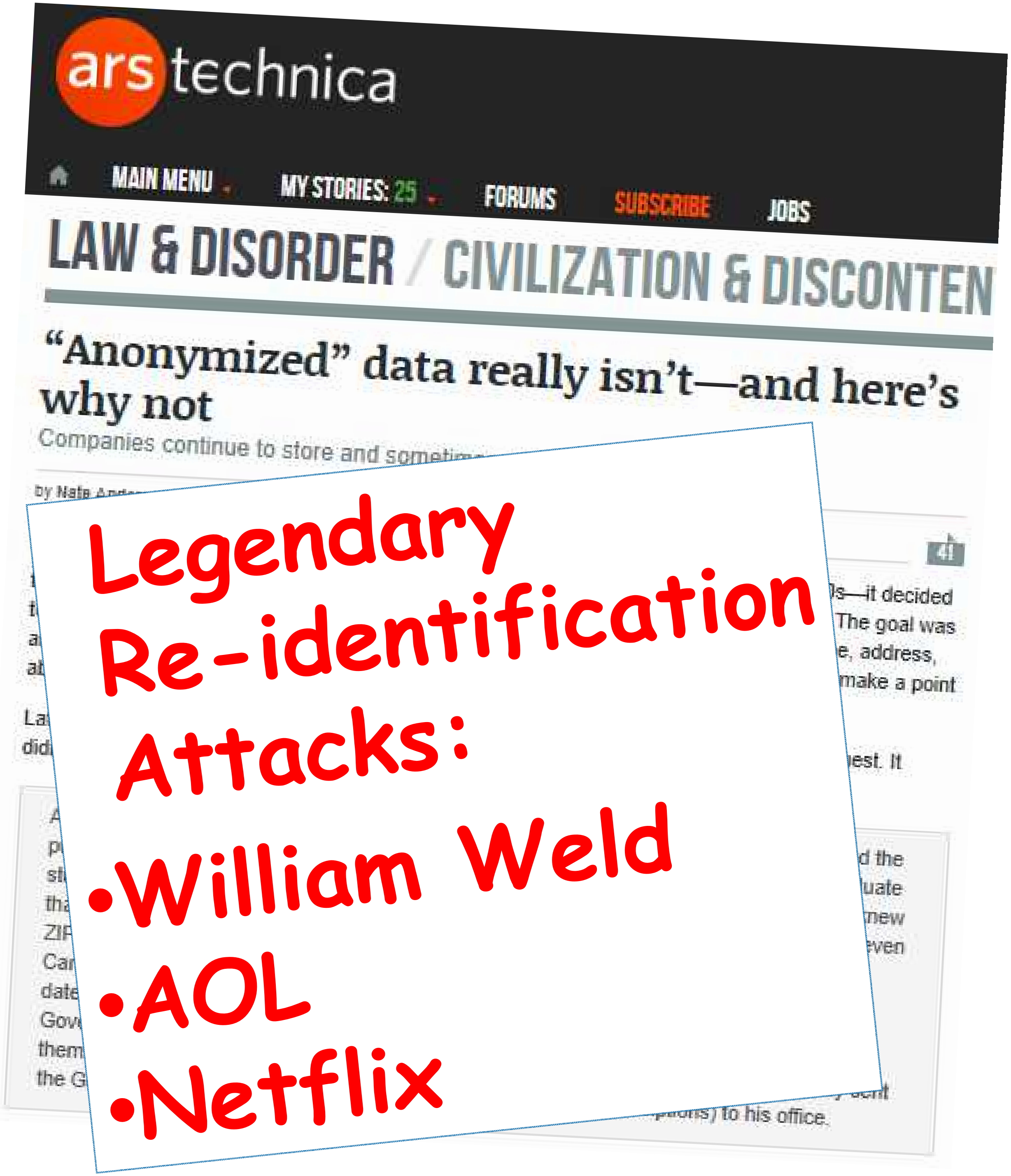
"De-identification leads to information loss which may limit the usefulness of the resulting health information" *(p.8, HHS De-ID Guidance Nov 26, 2012)*



Trade-Off between **Information Quality** and **Privacy Protection**

Ideal Situation (Perfect Information & Perfect Protection)

Unfortunately, not achievable due to mathematical constraints



"Anonymized" data really isn't—and here's why not

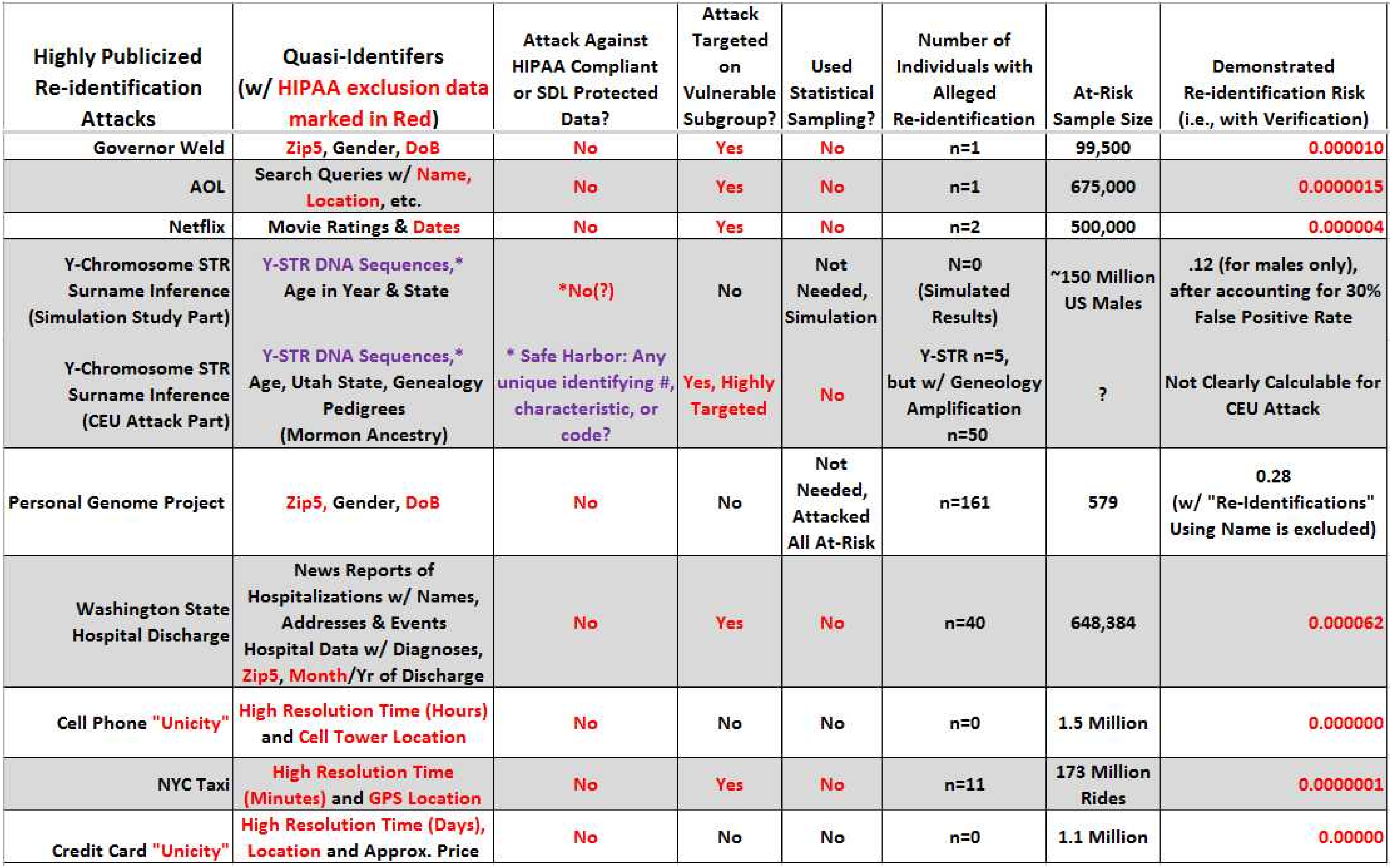
Companies continue to store and sometim

Legendary Re-identification Attacks:

- William Weld
- AOL
- Netflix

Unfortunately, de-identification public policy has often been driven by largely anecdotal and limited evidence, and re-identification demonstration attacks targeted to particularly vulnerable individuals, which fail to provide reliable evidence about real world re-identification risks

Re-identification Demonstration Attack Summary



Highly Publicized Re-identification Attacks	Quasi-Identifers (w/ HIPAA exclusion data marked in Red)	Attack Against HIPAA Compliant or SDL Protected Data?	Attack Targeted on Vulnerable Subgroup?	Used Statistical Sampling?	Number of Individuals with Alleged Re-identification	At-Risk Sample Size	Demonstrated Re-identification Risk (i.e., with Verification)
Governor Weld	Zip5, Gender, DoB	No	Yes	No	n=1	99,500	0.000010
AOL	Search Queries w/ Name, Location, etc.	No	Yes	No	n=1	675,000	0.0000015
Netflix	Movie Ratings & Dates	No	Yes	No	n=2	500,000	0.000004
Y-Chromosome STR Surname Inference (Simulation Study Part)	Y-STR DNA Sequences,* Age in Year & State	*No(?)	No	Not Needed, Simulation	N=0 (Simulated Results)	~150 Million US Males	.12 (for males only), after accounting for 30% False Positive Rate
Y-Chromosome STR Surname Inference (CEU Attack Part)	Y-STR DNA Sequences,* Age, Utah State, Genealogy Pedigrees (Mormon Ancestry)	* Safe Harbor: Any unique identifying #, characteristic, or code?	Yes, Highly Targeted	No	Y-STR n=5, but w/ Geneology Amplification n=50	?	Not Clearly Calculable for CEU Attack
Personal Genome Project	Zip5, Gender, DoB	No	No	Not Needed, Attacked All At-Risk	n=161	579	0.28 (w/ "Re-Identifications" Using Name is excluded)
Washington State Hospital Discharge	News Reports of Hospitalizations w/ Names, Addresses & Events Hospital Data w/ Diagnoses, Zip5, Month/Yr of Discharge	No	Yes	No	n=40	648,384	0.000062
Cell Phone "Unicity"	High Resolution Time (Hours) and Cell Tower Location	No	No	No	n=0	1.5 Million	0.000000
NYC Taxi	High Resolution Time (Minutes) and GPS Location	No	Yes	No	n=11	173 Million Rides	0.0000001
Credit Card "Unicity"	High Resolution Time (Days), Location and Approx. Price	No	No	No	n=0	1.1 Million	0.00000

Re-identification Science Policy Short-comings:

6 ways in which "Re-identification Science" has (thus far) typically failed to support sound public policies:

1. Attacking only trivially "straw man" de-identified data, where modern statistical disclosure control methods (like HIPAA) weren't used.
2. Targeting only especially vulnerable subpopulations and failing to use statistical random samples to provide policy-makers with representative re-identification risks for the entire population.
3. Making bad (often worst-case) assumptions and then failing to provide evidence to justify assumptions.

   Corollary: Not designing experiments to show the boundaries where de-identification finally succeeds.

Re-identification Science Policy Short-comings:

6 ways in which "Re-identification Science" has (thus far) typically failed to support sound public policies (Cont'd):

4. Failing to distinguish between sample uniqueness, population uniqueness and re-identifiability (ability to correctly link population unique observations to identities).

5. Failing to fully specify relevant threat models (using data intrusion scenarios that account for all of the motivations, process steps, and information required to successfully complete the re-identification attack for the members of the population).

6. Unrealistic emphasis on absolute "Privacy Guarantees" and *failure to recognize unavoidable trade-offs between data privacy and statistical accuracy/utility*.

Data Privacy Concerns are Far Too Important (and Complex) to be summed up with Catch Phrases or "Anecdata"

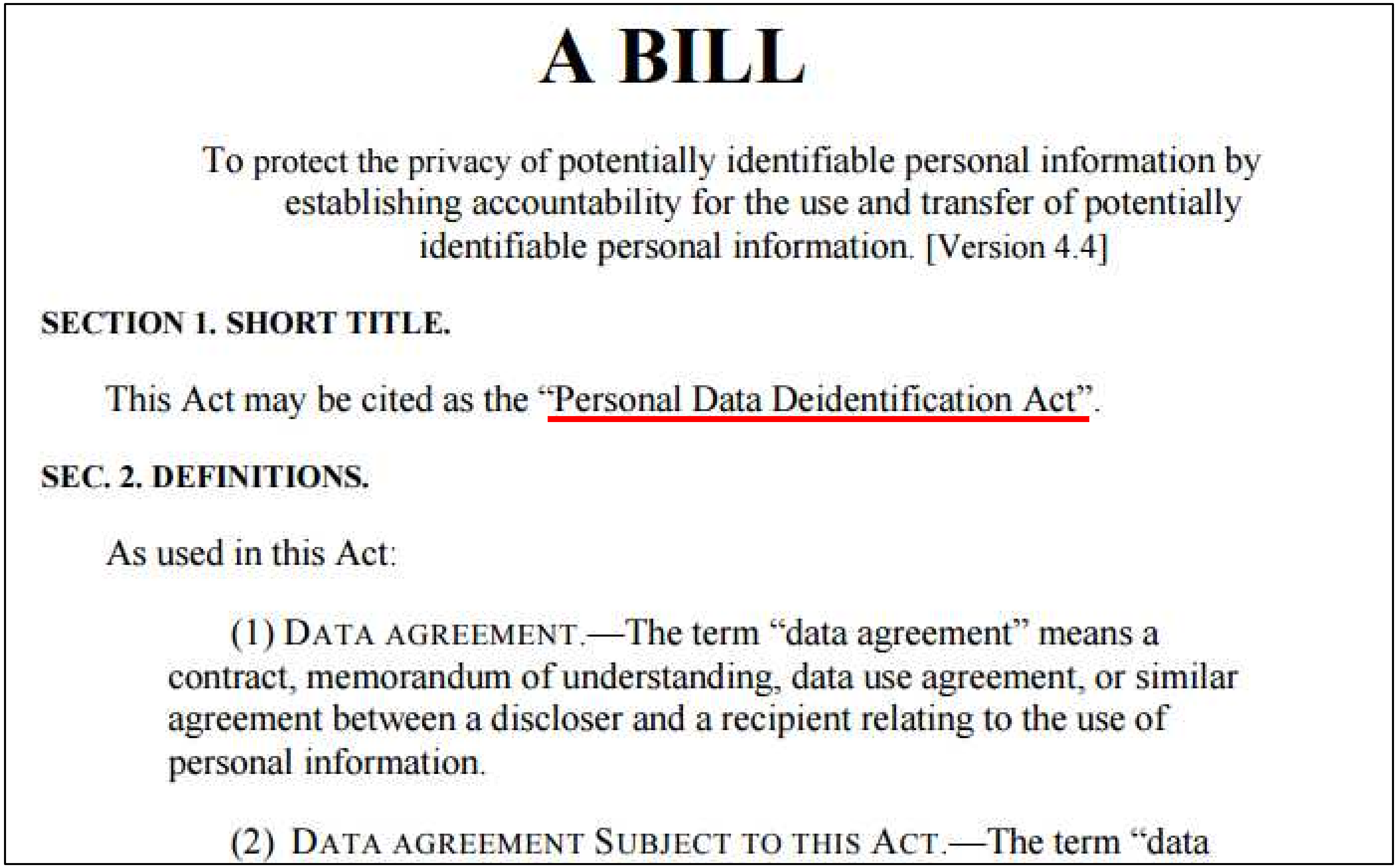
Eye-catching headlines and twitter-buzz announcing "*There's No Such Thing as Anonymous Data*" might draw the public's attention to broader and important concerns about data privacy in this era of "Big Data",

but such statements are essentially meaningless, even misleading, for further generalization without consideration of the specific de/re-identification contexts -- including the precise data details (e.g., number of variables, resolution of their coding schemas, special data properties, such as spatial/geographic detail, network properties, etc.) de-identification methods applied, and associated experimental design for re-identification attack demonstrations.

Good Public Policy demands reliable scientific evidence…

We also need…

Comprehensive Legislative Prohibitions Against Data Re-identification



A BILL

To protect the privacy of potentially identifiable personal information by establishing accountability for the use and transfer of potentially identifiable personal information. [Version 4.4]

SECTION 1. SHORT TITLE.

This Act may be cited as the "Personal Data Deidentification Act".

SEC. 2. DEFINITIONS.

As used in this Act:

(1) DATA AGREEMENT.—The term "data agreement" means a contract, memorandum of understanding, data use agreement, or similar agreement between a discloser and a recipient relating to the use of personal information.

(2) DATA AGREEMENT SUBJECT TO THIS ACT.—The term "data

Robert Gellman, 2010

https://fpf.org/wp-content/uploads/2010/07/The_Deidentification_Dilemma.pdf

Reserve Slides for Questions

Re-identification Science Can Better Inform Policy/Practice

1. Demonstrate re-identification risks on data where modern statistical disclosure control methods have actually been used.
2. Use proper statistical random samples and scientific study designs in order to provide *representative* risk estimates.
3. Design experiments to show the boundaries where de-identification finally succeeds and provide evidence to justify any data intruder knowledge assumptions.
4. Verify re-identifications and report false-positive rates for supposed re-identifications.
5. Investigate multiple realistic and relevant threats and fully specify these re-identification threat models.
6. Use modern probabilistic uncertainty analyses to examine impact of uncertainties in re-identification experiments.

Bill of Health

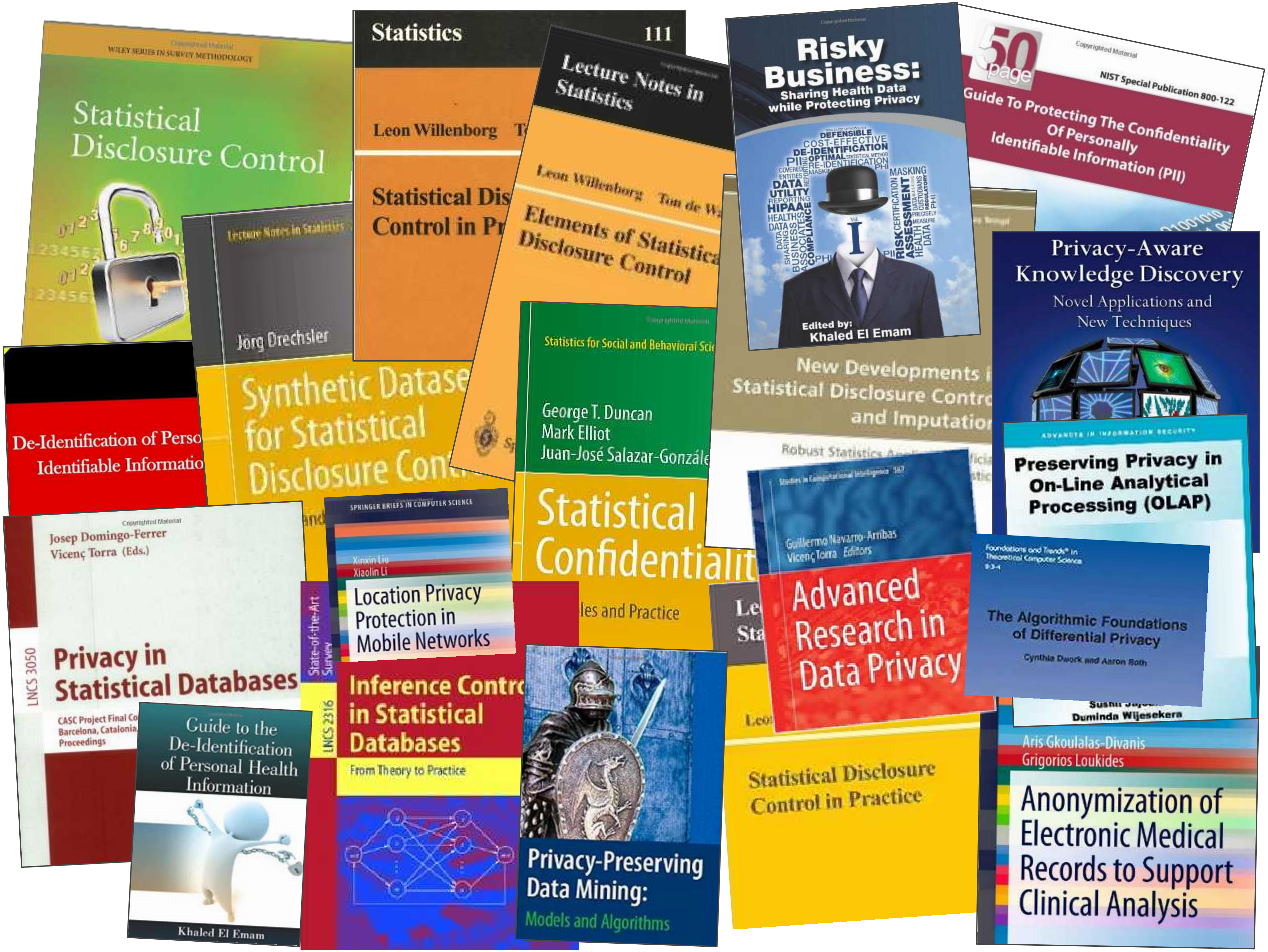
Examining the intersection of law and health care, biotech & bioethics
A blog by the Petrie-Flom Center and friends

Online Symposium on the Law, Ethics & Science of Re-identification Demonstrations

- http://blogs.law.harvard.edu/billofhealth/2013/05/29/public-policy-considerations-for-recent-re-identification-demonstration-attacks-on-genomic-data-sets-part-1-re-identification-symposium/

- https://blogs.law.harvard.edu/billofhealth/2013/10/01/press-and-reporting-considerations-for-recent-re-identification-demonstration-attacks-part-2-re-identification-symposium/

- http://blogs.law.harvard.edu/billofhealth/2013/10/02/ethical-concerns-conduct-and-public-policy-for-re-identification-and-de-identification-practice-part-3-re-identification-symposium/



WILEY SERIES IN SURVEY METHODOLOGY

Statistical Disclosure Control

Statistics 111

Leon Willenborg To

Statistical Dis Control in Pr

Lecture Notes in Statistics

Leon Willenborg Ton de W

Elements of Statistica Disclosure Control

Risky Business: Sharing Health Data while Protecting Privacy

Edited by: Khaled El Emam

50 page

NIST Special Publication 800-122

Guide To Protecting The Confidentiality Of Personally Identifiable Information (PII)

Lecture Notes in Statistics

Jörg Drechsler

Synthetic Datase for Statistical Disclosure Cont

De-Identification of Perso Identifiable Informatio

Statistics for Social and Behavioral Scie

George T. Duncan
Mark Elliot
Juan-José Salazar-Gonzále

Statistical Confidentialit

les and Practice

New Developments i Statistical Disclosure Contr and Imputation

Robust Statistics

Privacy-Aware Knowledge Discovery

Novel Applications and New Techniques

ADVANCES IN INFORMATION SECURITY

Preserving Privacy in On-Line Analytical Processing (OLAP)

Sushil Jajodia
Duminda Wijesekera

Josep Domingo-Ferrer
Vicenç Torra (Eds.)

LNCS 3050

Privacy in Statistical Databases

CASC Project Final Co
Barcelona, Catalonia
Proceedings

Guide to the De-Identification of Personal Health Information

Khaled El Emam

SPRINGER BRIEFS IN COMPUTER SCIENCE

Xinxin Liu
Xiaolin Li

Location Privacy Protection in Mobile Networks

State-of-the-Art Survey

LNCS 2316

Inference Contro in Statistical Databases

From Theory to Practice

Privacy-Preserving Data Mining:

Models and Algorithms

Studies in Computational Intelligence 567

Guillermo Navarro-Arribas
Vicenç Torra Editors

Advanced Research in Data Privacy

Le Sta

Leo

Statistical Disclosure Control in Practice

Foundations and Trends® in Theoretical Computer Science 9:3-4

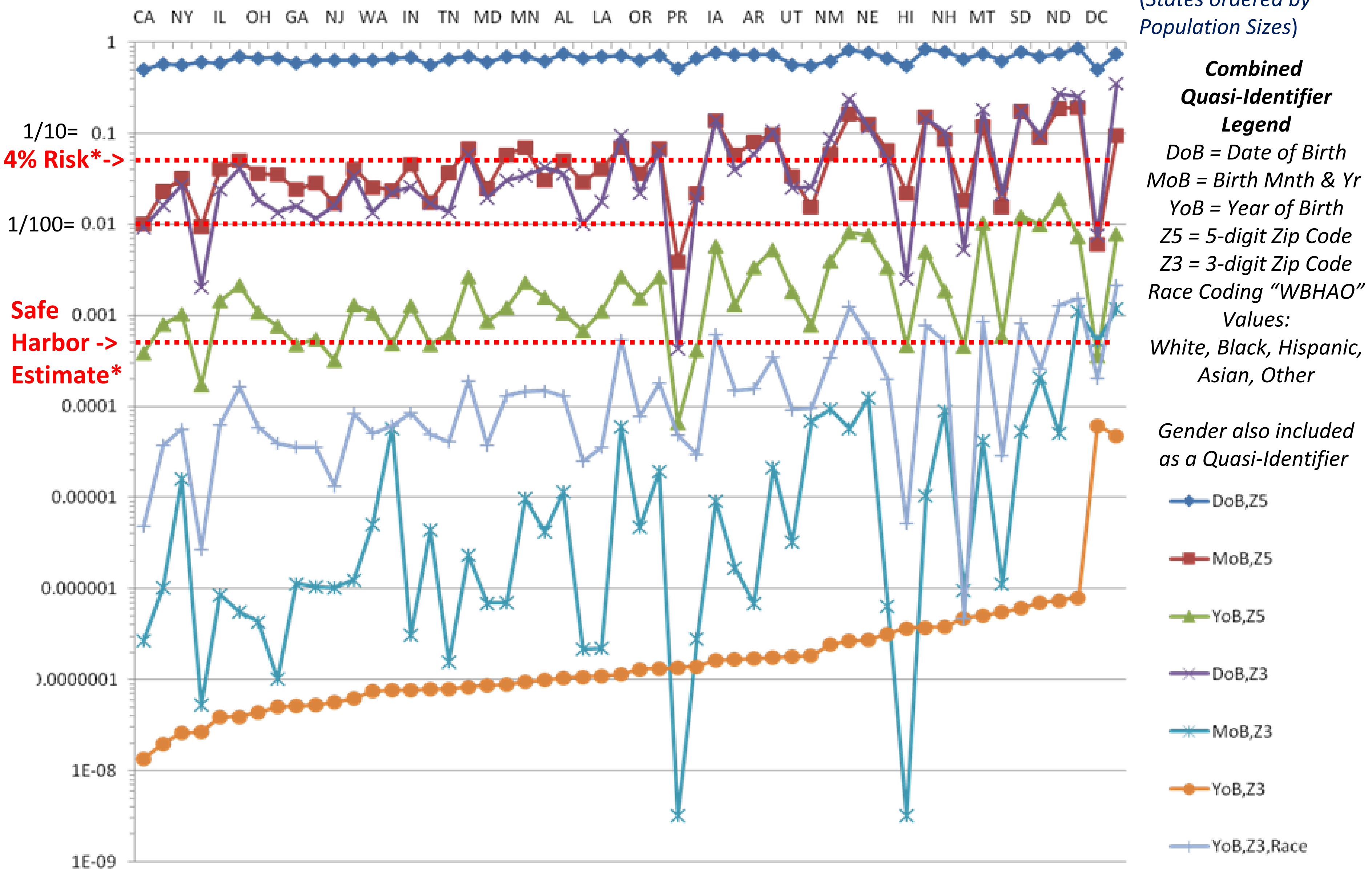
The Algorithmic Foundations of Differential Privacy

Cynthia Dwork and Aaron Roth

Aris Gkoulalas-Divanis
Grigorios Loukides

Anonymization of Electronic Medical Records to Support Clinical Analysis

State Specific Re-identification Risks: Population Uniqueness



(States ordered by Population Sizes)

Combined Quasi-Identifier Legend

DoB = Date of Birth
MoB = Birth Mnth & Yr
YoB = Year of Birth
Z5 = 5-digit Zip Code
Z3 = 3-digit Zip Code
Race Coding "WBHAO" Values:
White, Black, Hispanic, Asian, Other

Gender also included as a Quasi-Identifier

Data Source: 2010 U.S. Decennial Census

Graph © D-BJ 2013

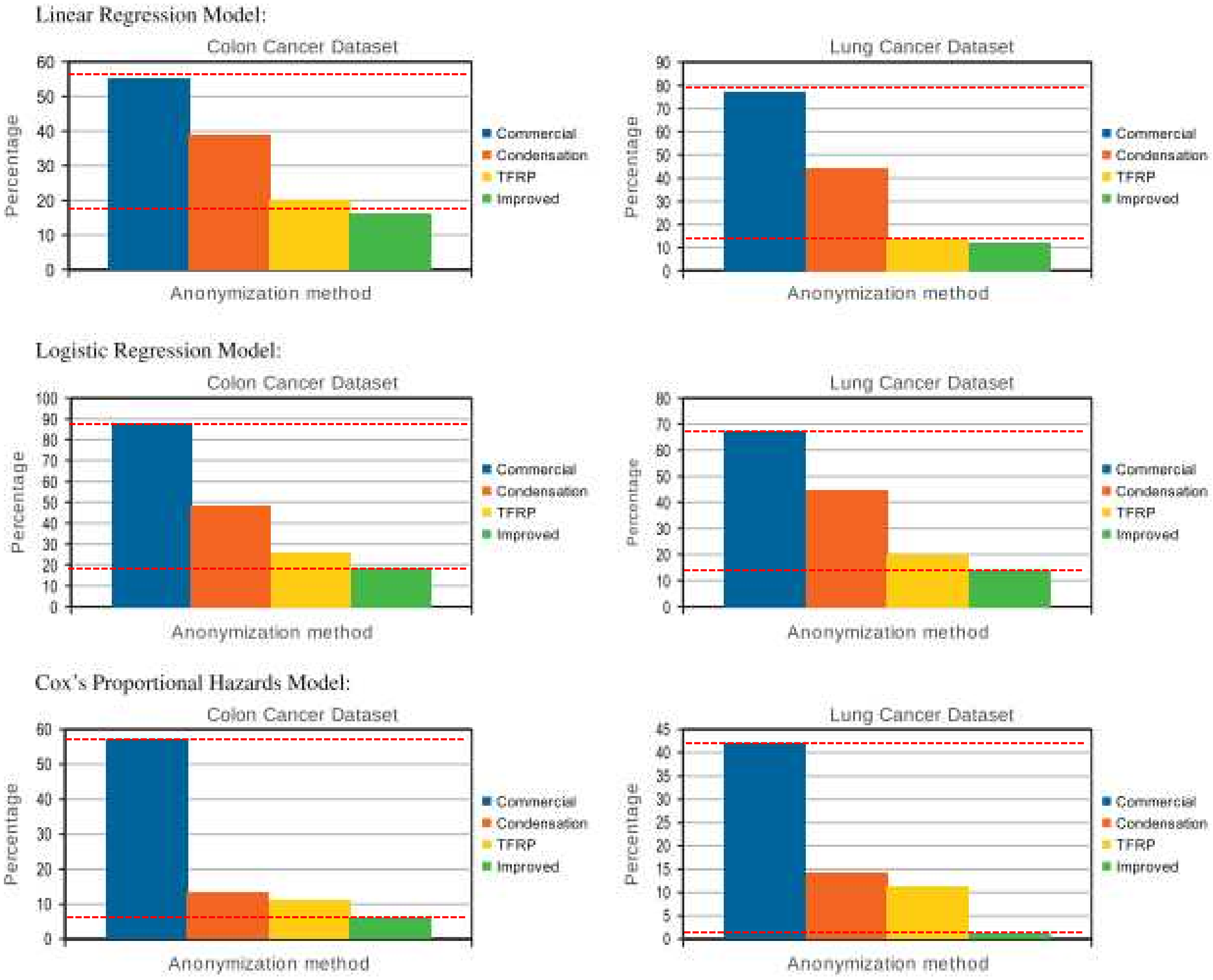
Balancing Disclosure Risk/Statistical Accuracy

- Balancing disclosure risks and statistical accuracy is essential because some popular de-identification methods (e.g. k-anonymity) can unnecessarily, and often undetectably, degrade the accuracy of de-identified data for multivariate statistical analyses or data mining (distorting variance-covariance matrixes, masking heterogeneous sub-groups which have been collapsed in generalization protections)
- This problem is well-understood by statisticians, but not as well recognized and integrated within public policy.
- Poorly conducted de-identification can lead to "bad science" and "bad decisions".

Reference: C. Aggarwal **http://www.vldb2005.org/program/paper/fri/p901-aggarwal.pdf**

Percent of Regression Coefficients which changed Significance:

T.S. Gal et al. / Journal of Biomedical Informatics xxx (2014) xxx–xxx


Linear Regression Model:

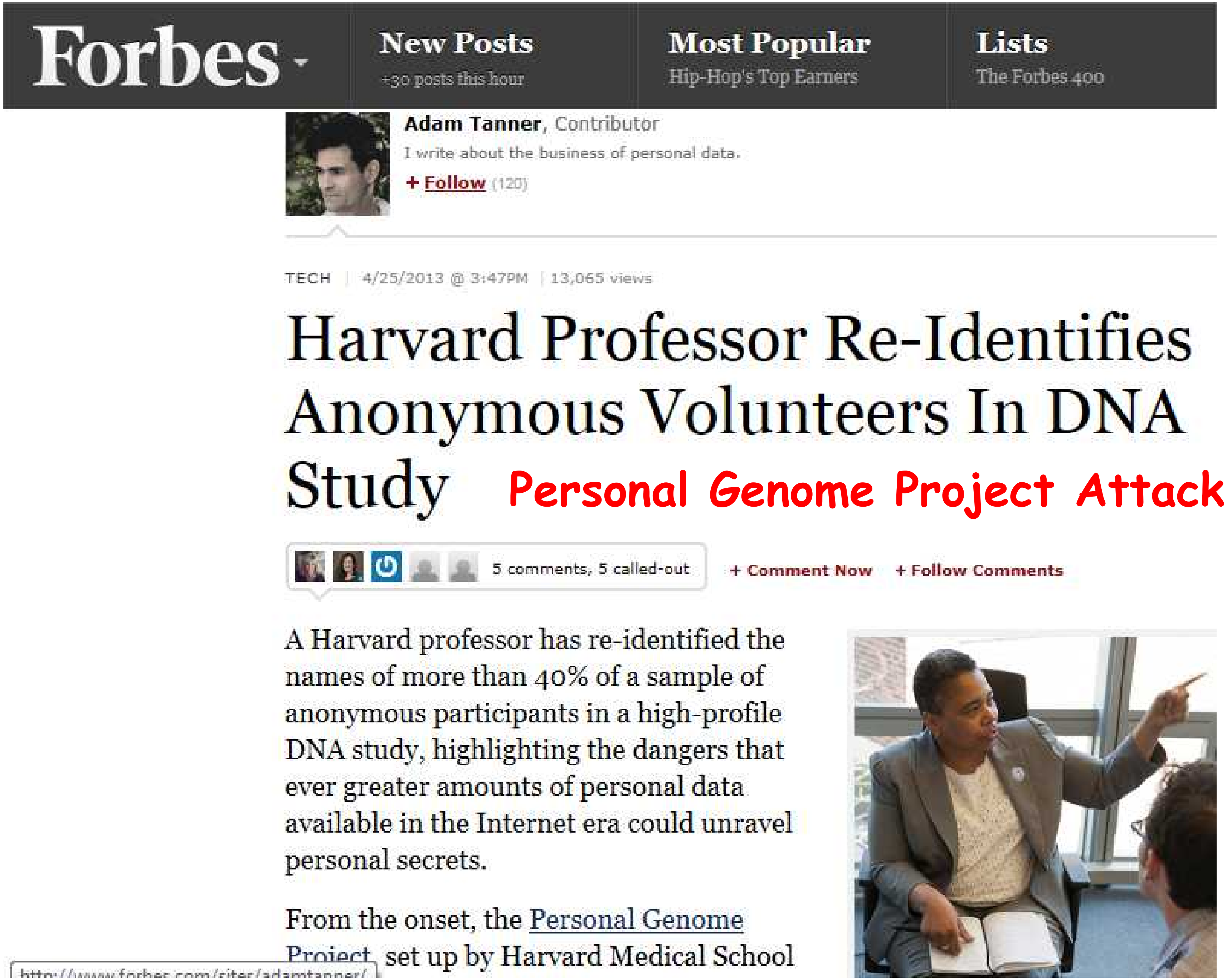
Logistic Regression Model:

Cox's Proportional Hazards Model:

Fig. 1. Coefficients changed significance.

If this is what we are going to do to our ability to conduct accurate research – then… we should all just go home.

- Although poorly conducted de-identification can distort our ability to learn what is true leading to “bad science/decisions”, this does not need to be an inevitable outcome.
- Well-conducted de-identification practice always carefully considers both the re-identification risk context and examines and controls the possible distortion to the statistical accuracy and utility of the de-identified data to assure de-identified data has been appropriately and usefully de-identified.
- But doing this requires a firm understanding/grounding in the extensive body of the statistical disclosure control/limitation literature.



Adam Tanner, Contributor

I write about the business of personal data.

TECH | 4/25/2013 @ 3:47PM | 13,065 views

Harvard Professor Re-Identifies Anonymous Volunteers In DNA Study

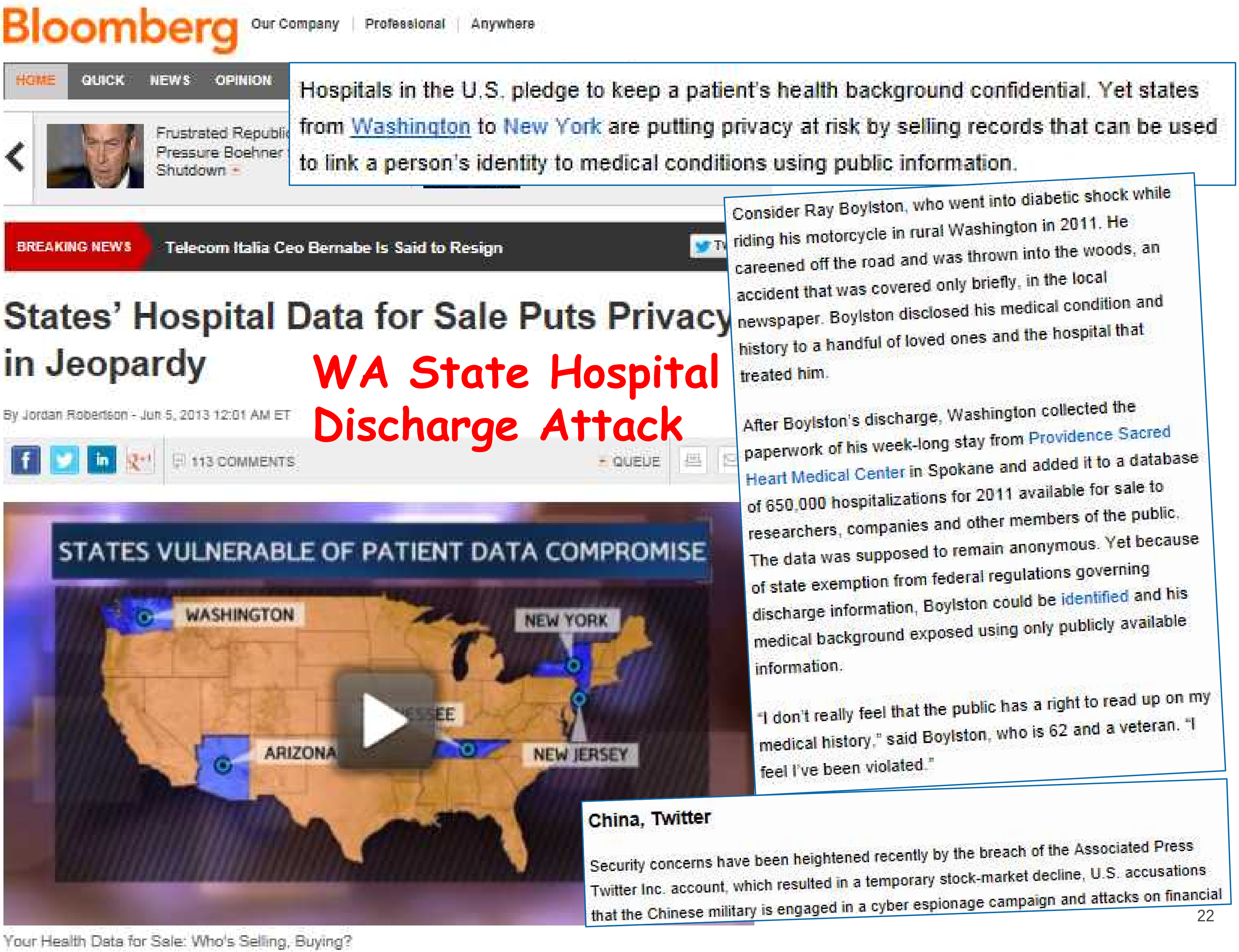
Personal Genome Project Attack

A Harvard professor has re-identified the names of more than 40% of a sample of anonymous participants in a high-profile DNA study, highlighting the dangers that ever greater amounts of personal data available in the Internet era could unravel personal secrets.

From the onset, the Personal Genome Project, set up by Harvard Medical School

http://www.forbes.com/sites/adamtanner/

WA State Hospital Discharge Attack



States' Hospital Data for Sale Puts Privacy in Jeopardy

By Jordan Robertson - Jun 5, 2013 12:01 AM ET

Your Health Data for Sale: Who's Selling, Buying?

Hospitals in the U.S. pledge to keep a patient's health background confidential. Yet states from Washington to New York are putting privacy at risk by selling records that can be used to link a person's identity to medical conditions using public information.

Consider Ray Boylston, who went into diabetic shock while riding his motorcycle in rural Washington in 2011. He careened off the road and was thrown into the woods, an accident that was covered only briefly, in the local newspaper. Boylston disclosed his medical condition and history to a handful of loved ones and the hospital that treated him.

After Boylston's discharge, Washington collected the paperwork of his week-long stay from Providence Sacred Heart Medical Center in Spokane and added it to a database of 650,000 hospitalizations for 2011 available for sale to researchers, companies and other members of the public. The data was supposed to remain anonymous. Yet because of state exemption from federal regulations governing discharge information, Boylston could be identified and his medical background exposed using only publicly available information.

"I don't really feel that the public has a right to read up on my medical history," said Boylston, who is 62 and a veteran. "I feel I've been violated."

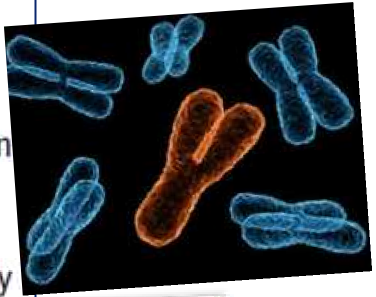
China, Twitter

Security concerns have been heightened recently by the breach of the Associated Press Twitter Inc. account, which resulted in a temporary stock-market decline, U.S. accusations that the Chinese military is engaged in a cyber espionage campaign and attacks on financial

Identifying Personal Genomes by Surname Inference

Science
AAAS

Melissa Gymrek,[1,2,3,4] Amy L. McGuire,[5] David Golan,[6] Eran Halperin,[7,8,9] Yaniv Erlich[1]*



Sharing sequencing data sets without identifiers has become a common practice in genomics. Here, we report that surnames can be recovered from personal genomes by profiling short tandem repeats on the Y chromosome (Y-STRs) and querying recreational genetic genealogy databases. We show that a combination of a surname with other types of metadata, such as age and state,
ca
ulate the identity of the target. A key feature of this technique is that it entirely
re
es. We quantitatively analyze the probabili
id
w

nature
International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audi

Archive > Volume 497 > Issue 7448 > News Feature > Article

NATURE | NEWS FEATURE

Privacy protections: The genome hacker

Yaniv Erlich shows how research participants can be identified from 'anonymou

Erika Check Hayden

08 May 2013

PDF | Rights & Permissions

Our analysis projects a success rate of ~12% (SD = 2%) in recovering surnames of U.S. Caucasian males (Fig. 1B and fig. S2). This rate can be accomplished with a conservative threshold that would return a wrong surname in 5% of cases and label 83% of cases as unknown. Higher success rates of up to 18% can be achieved at the price of increased probability to recover an incorrect surname. Because our input cohort is based

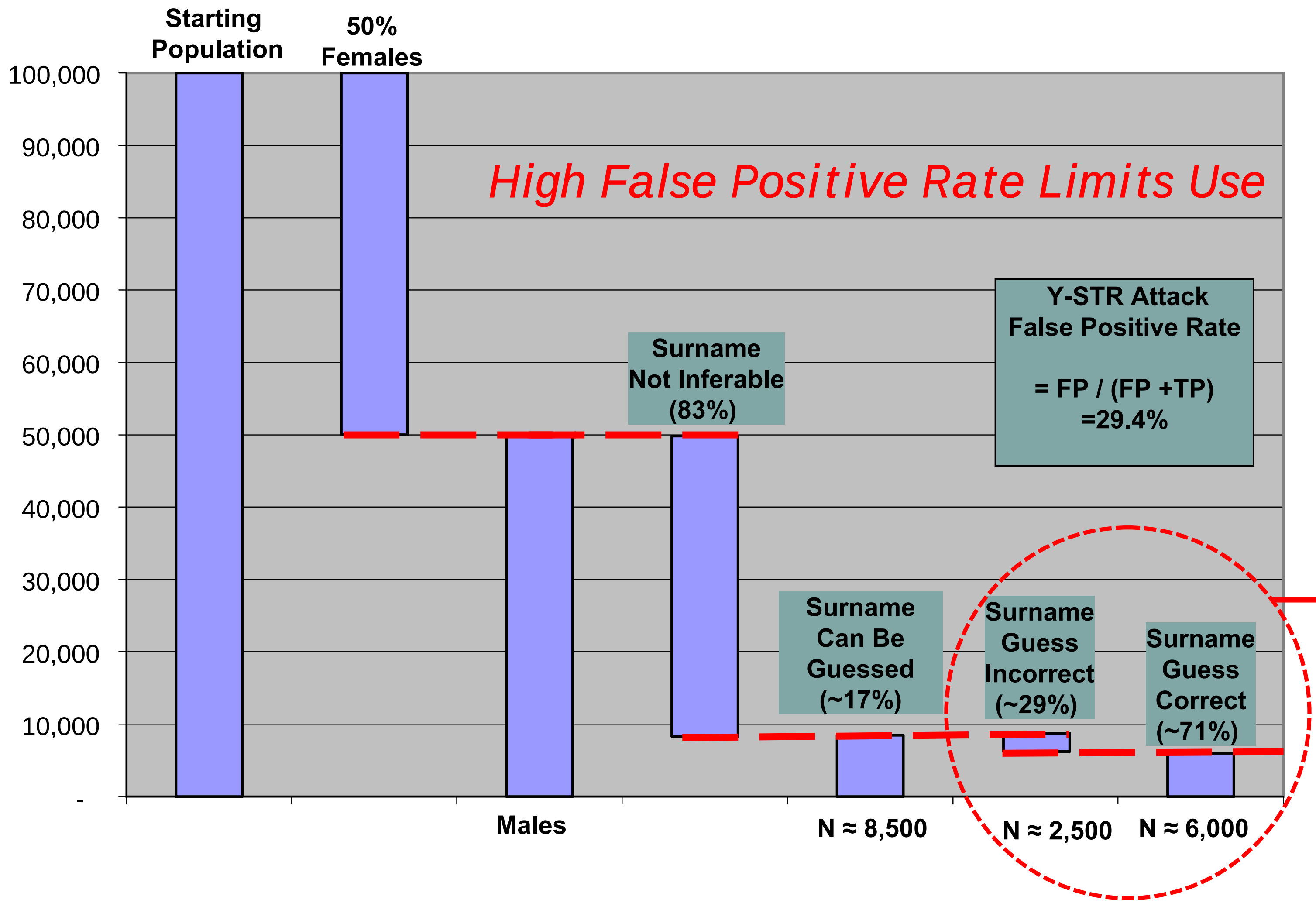
TG
TG
ACTACTACTACTACTACTAC
7 repeats

Question 1: Is Y-STR Attack Economically Viable?

Probably not -- unclear whether it eventually could be.

Question 2: Is "De-identification" pointless?

No, removing State, Grouping YoB would help importantly.



High False Positive Rate Limits Use

Y-STR Attack False Positive Rate

= FP / (FP +TP) =29.4%

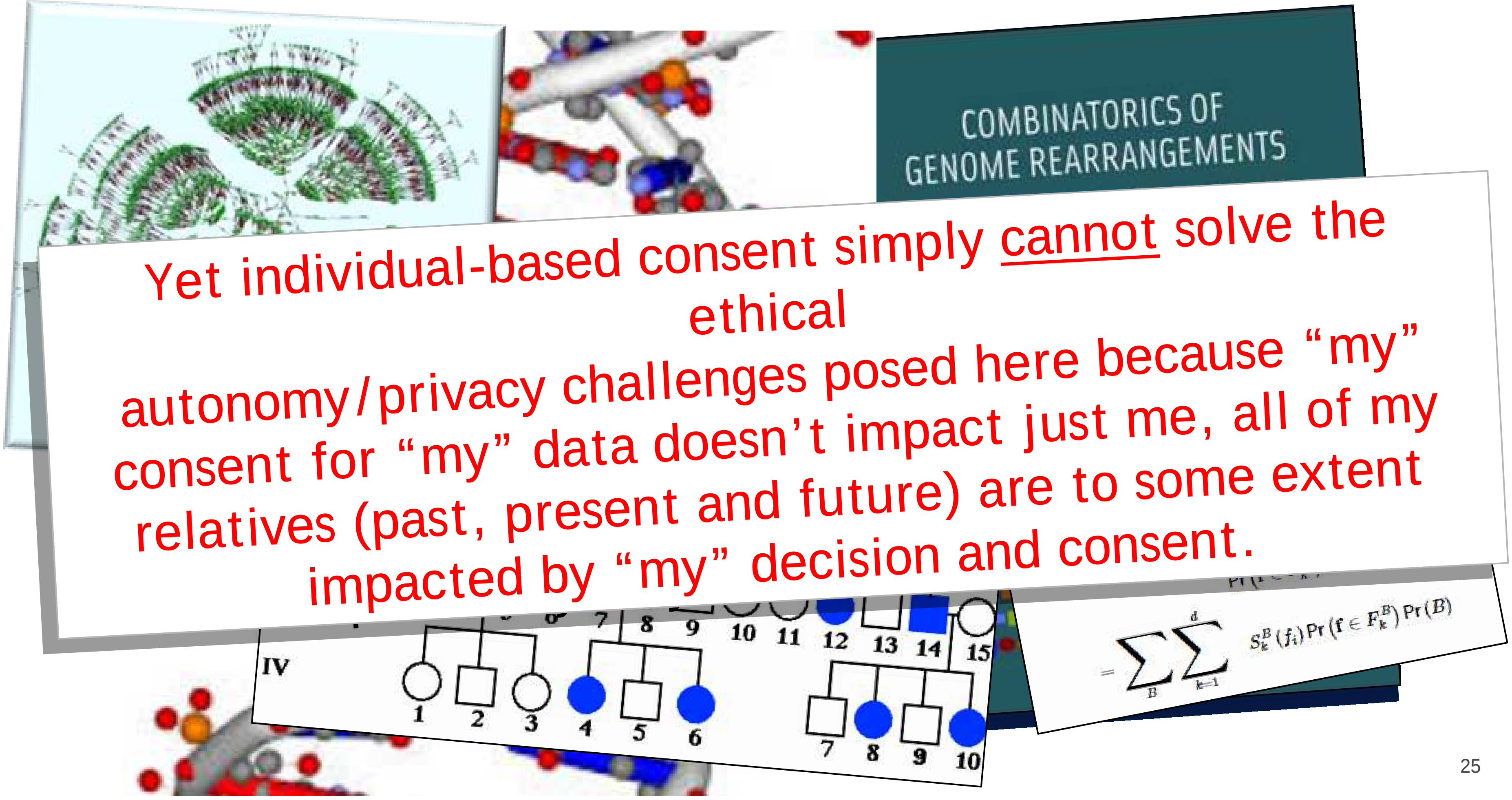
Re-ID isn't achieved by Surname Guess.

So what's the Threat Model?

Surname Guess Could Serve as a (Faulty) Quasi-identifier (e.g., w/ YoB & State)

But Will Produce Substantive Re-identification Errors

Given the inherent extremely large combinatorics of genomic data nested within inheritance networks which determine how genomic traits (and surnames) are shared with our ancestors/descendants, the degree to which such information could be meaningfully "de-identified" are non-trivial.



Yet individual-based consent simply cannot solve the ethical autonomy/privacy challenges posed here because "my" consent for "my" data doesn't impact just me, all of my relatives (past, present and future) are to some extent impacted by "my" decision and consent.

SCIENTIFIC REPORTS

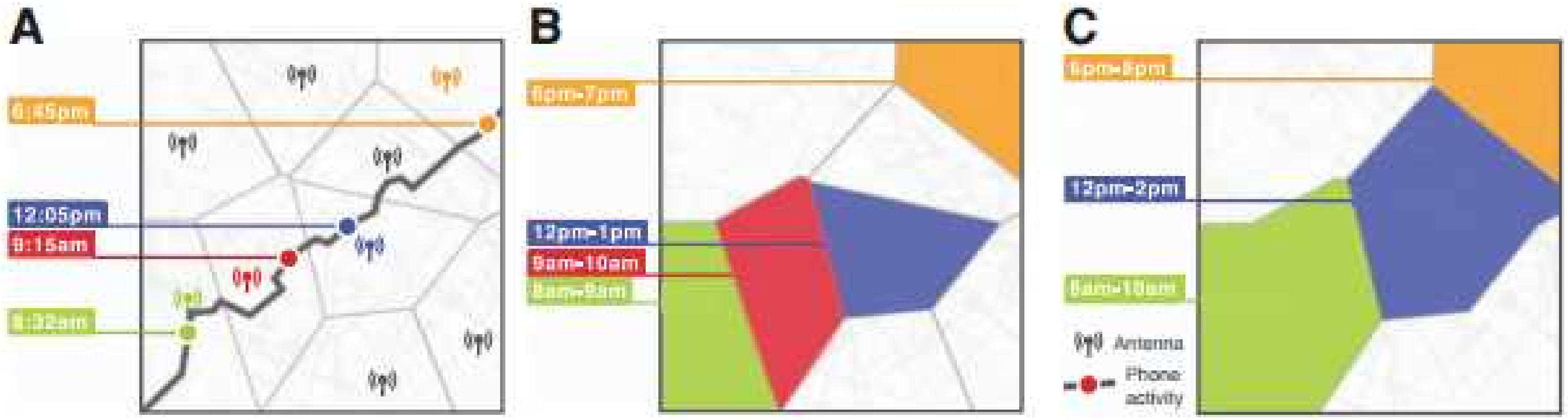
Published
25 March 2013

Unique in the Crowd: The privacy bounds of human mobility

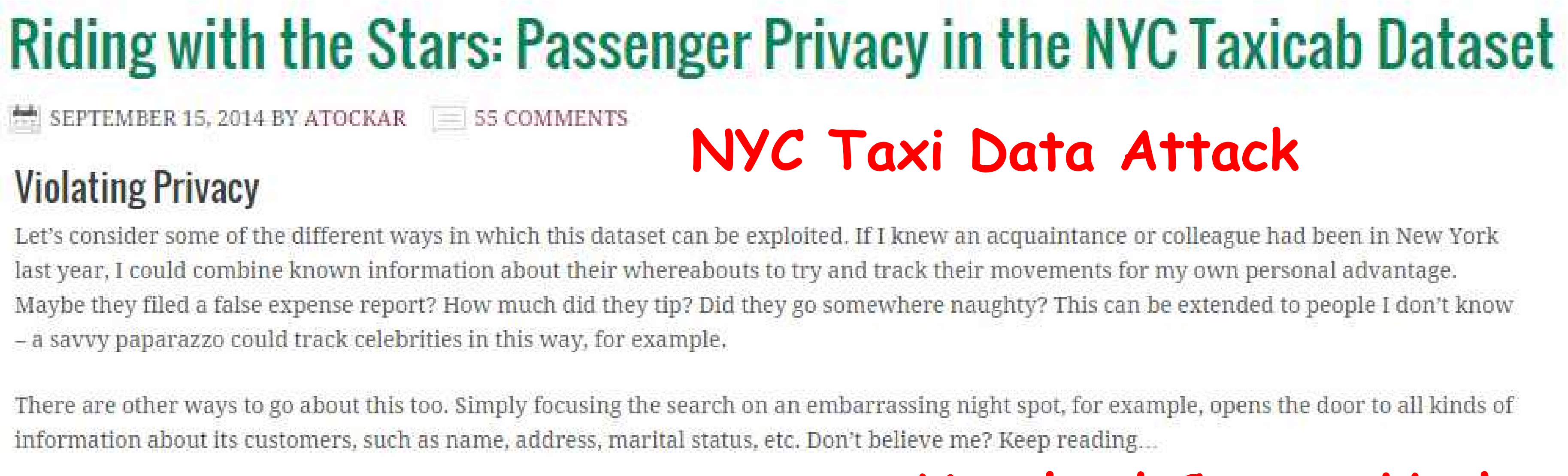
Yves-Alexandre de Montjoye[1,2], César A. Hidalgo[1,3,4], Michel Verleysen[2] & Vincent D. Blondel[2,5]

Cell Data Uniqueness

We study fifteen months of human mobility data for one and a half million individuals and find that human mobility traces are highly unique. In fact, in a dataset where the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier's antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals. We coarsen the data spatially and temporally to find a formula for the uniqueness of human mobility traces given their resolution and the available outside information. This formula shows that the uniqueness of mobility traces decays approximately as the 1/10 power of their resolution. Hence, even coarse datasets provide little anonymity. These findings represent fundamental constraints to an individual's privacy and have important implications for the design of frameworks and institutions dedicated to protect the privacy of individuals.



Sample Unique ≠ Re-identifiable



Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset

SEPTEMBER 15, 2014 BY ATOCKAR 55 COMMENTS

NYC Taxi Data Attack

Violating Privacy

Let's consider some of the different ways in which this dataset can be exploited. If I knew an acquaintance or colleague had been in New York last year, I could combine known information about their whereabouts to try and track their movements for my own personal advantage. Maybe they filed a false expense report? How much did they tip? Did they go somewhere naughty? This can be extended to people I don't know – a savvy paparazzo could track celebrities in this way, for example.

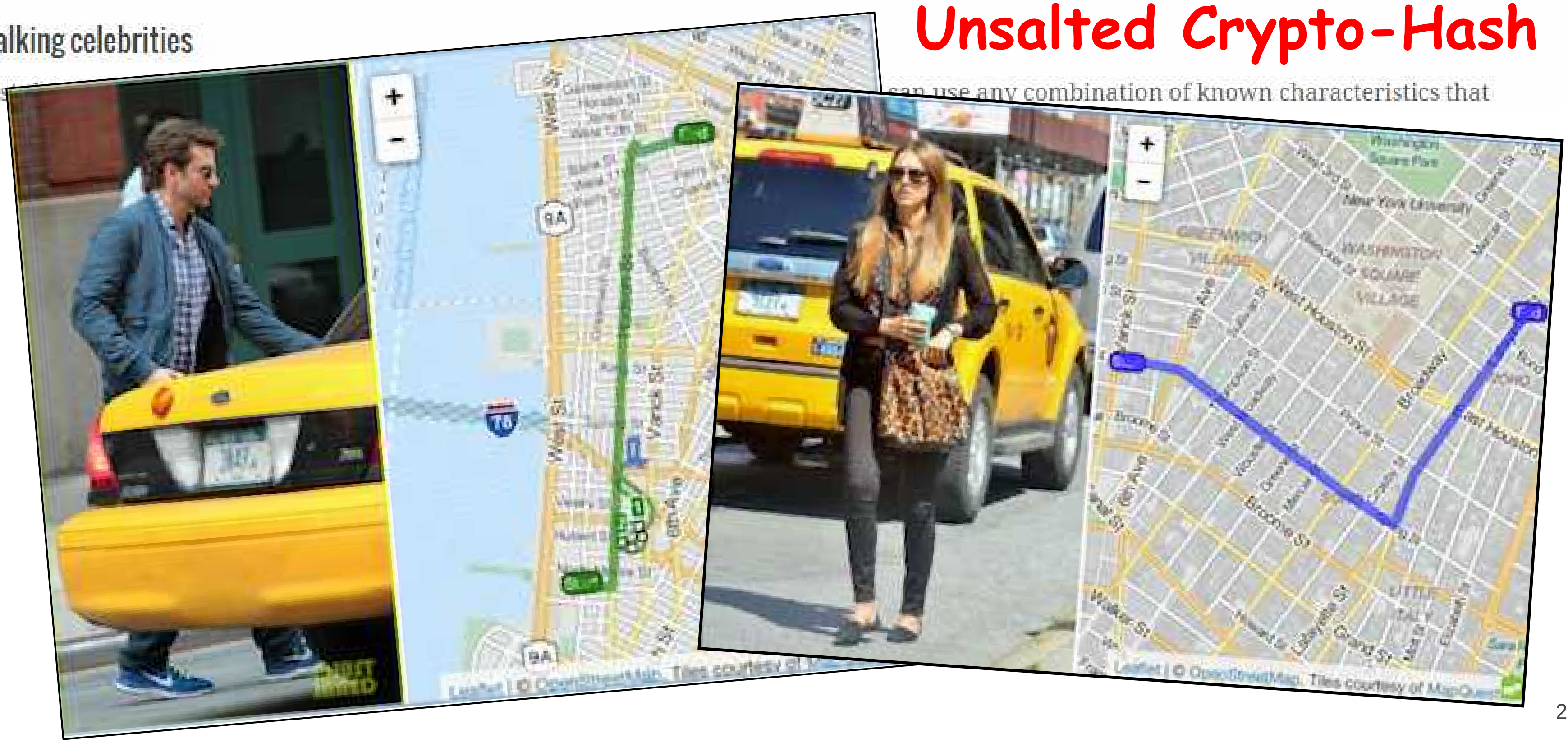
There are other ways to go about this too. Simply focusing the search on an embarrassing night spot, for example, opens the door to all kinds of information about its customers, such as name, address, marital status, etc. Don't believe me? Keep reading…

Unsalted Crypto-Hash

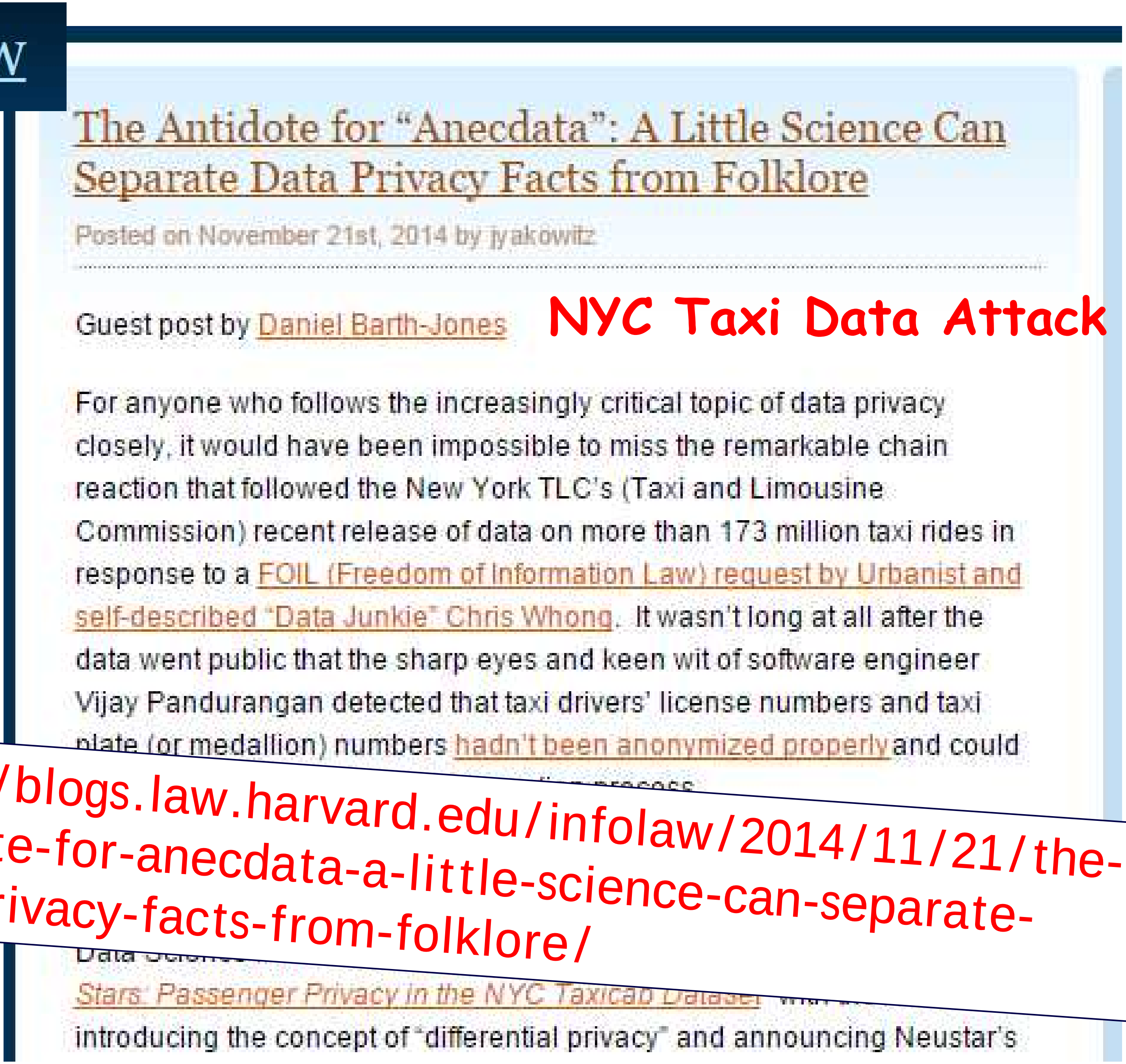
Stalking celebrities

Firs

can use any combination of known characteristics that



NYC Taxi Data Attack



The Antidote for "Anecdata": A Little Science Can Separate Data Privacy Facts from Folklore

Posted on November 21st, 2014 by jyakowitz

Guest post by Daniel Barth-Jones

For anyone who follows the increasingly critical topic of data privacy closely, it would have been impossible to miss the remarkable chain reaction that followed the New York TLC's (Taxi and Limousine Commission) recent release of data on more than 173 million taxi rides in response to a FOIL (Freedom of Information Law) request by Urbanist and self-described "Data Junkie" Chris Whong. It wasn't long at all after the data went public that the sharp eyes and keen wit of software engineer Vijay Pandurangan detected that taxi drivers' license numbers and taxi plate (or medallion) numbers hadn't been anonymized properly and could

...ion process.

Data Scienc...

Stars: Passenger Privacy in the NYC Taxicab Dataset

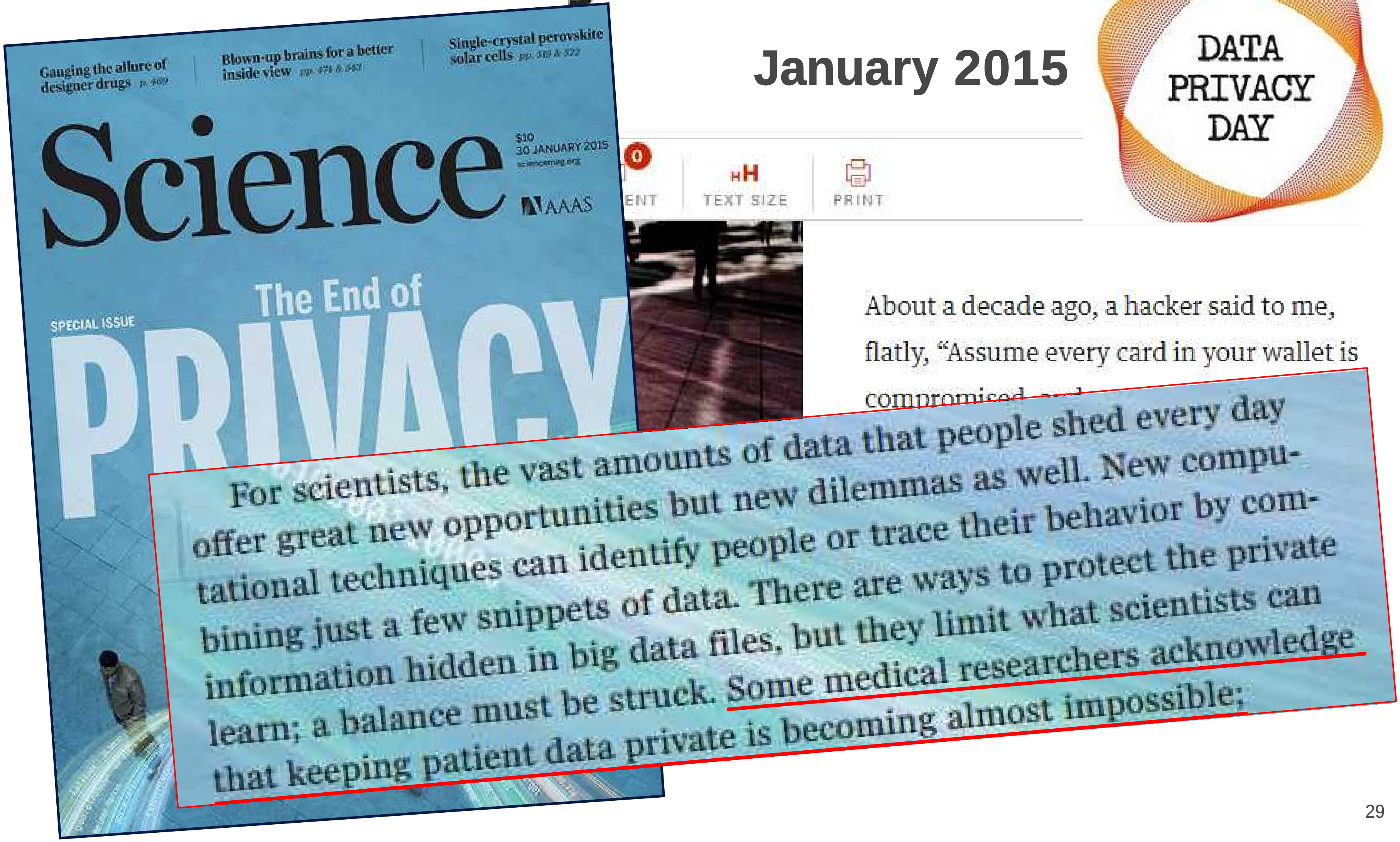
introducing the concept of "differential privacy" and announcing Neustar's

http://blogs.law.harvard.edu/infolaw/2014/11/21/the-antidote-for-anecdata-a-little-science-can-separate-data-privacy-facts-from-folklore/

Harvard
Business
Review

REGULATION

There’s No Such Thing as Anonymous Data



January 2015

Gauging the allure of designer drugs p. 469

Blown-up brains for a better inside view pp. 474 & 543

Single-crystal perovskite solar cells pp. 519 & 522

Science

$10
30 JANUARY 2015
sciencemag.org

AAAS

SPECIAL ISSUE

The End of PRIVACY

About a decade ago, a hacker said to me, flatly, “Assume every card in your wallet is compromised

For scientists, the vast amounts of data that people shed every day offer great new opportunities but new dilemmas as well. New computational techniques can identify people or trace their behavior by combining just a few snippets of data. There are ways to protect the private information hidden in big data files, but they limit what scientists can learn; a balance must be struck. Some medical researchers acknowledge that keeping patient data private is becoming almost impossible;

DATA PRIVACY DAY

Credit Card Data Uniqueness

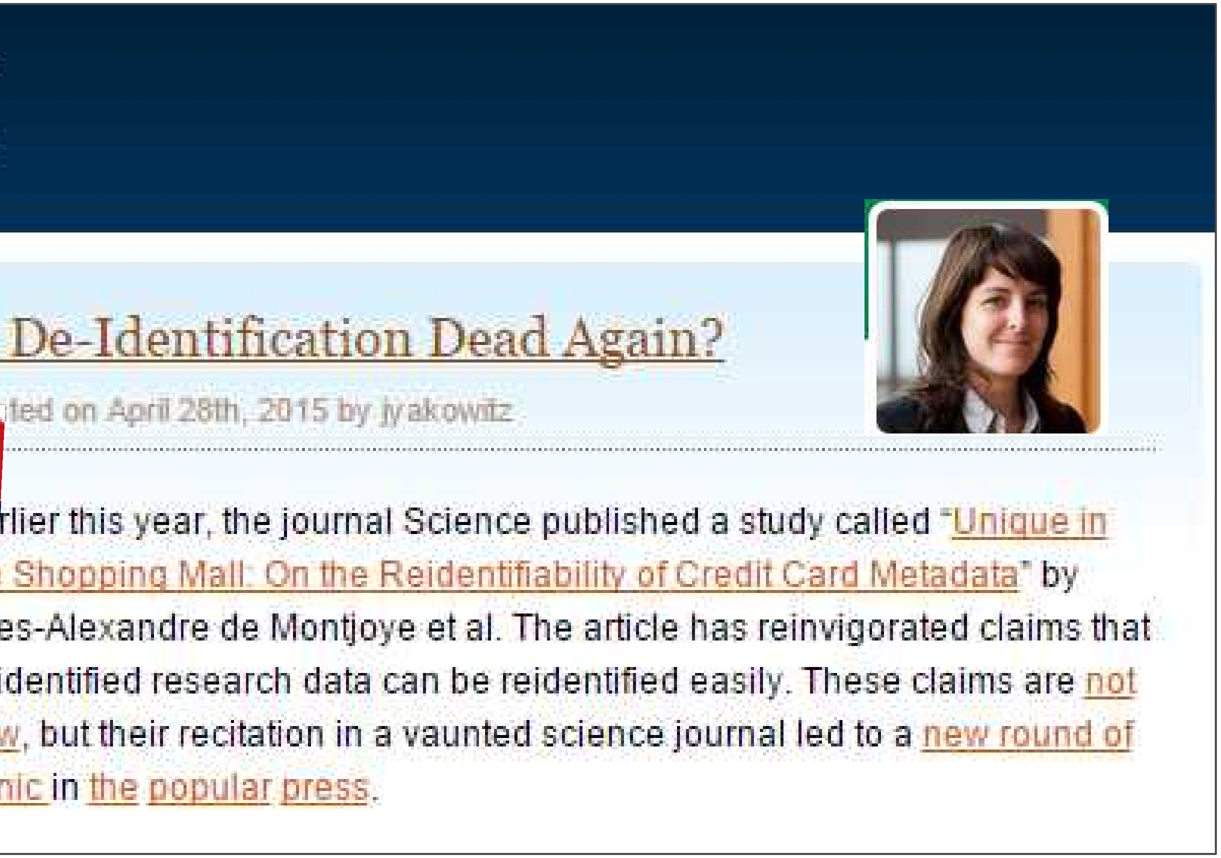
IDENTITY AND PRIVACY

Unique in the shopping mall: On the reidentifiability of credit card metadata

Yves-Alexandre de Montjoye,[1*] Laura Radaelli,[2] Vivek Kumar Singh,[1,3] Alex "Sandy" Pentland[1]

shop	user_id	time	price	price_bin
👟	7abc1a23	09/23	$97.30	$49 – $146
🍓	7abc1a23	09/23	$15.13	$5 – $16
🛒	3092fc10	09/23	$43.78	$16 – $49
🍞	7abc1a23	09/23	$4.33	$2 – $5

In fact, knowing just four random pieces of information was enough to reidentify 90 percent of the shoppers as unique individuals and to uncover their records, researchers calculated.



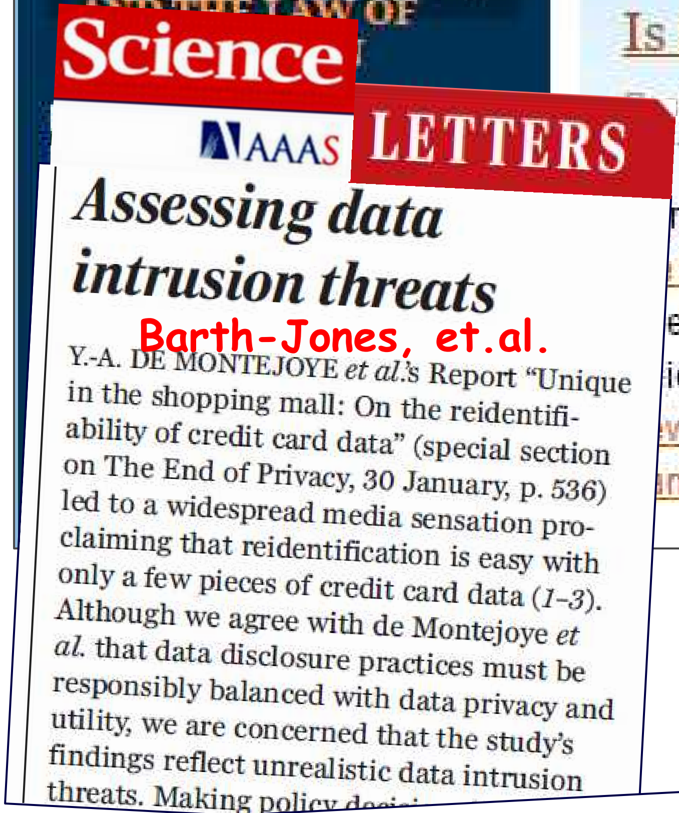
INFO/LAW

INFORMATION, LAW, AND THE LAW OF

Is De-Identification Dead Again?

Posted on April 28th, 2015 by jyakowitz

rlier this year, the journal Science published a study called "Unique in e Shopping Mall: On the Reidentifiability of Credit Card Metadata" by es-Alexandre de Montjoye et al. The article has reinvigorated claims that identified research data can be reidentified easily. These claims are not w, but their recitation in a vaunted science journal led to a new round of nic in the popular press.



Science

AAAS LETTERS

Assessing data intrusion threats

Barth-Jones, et.al.

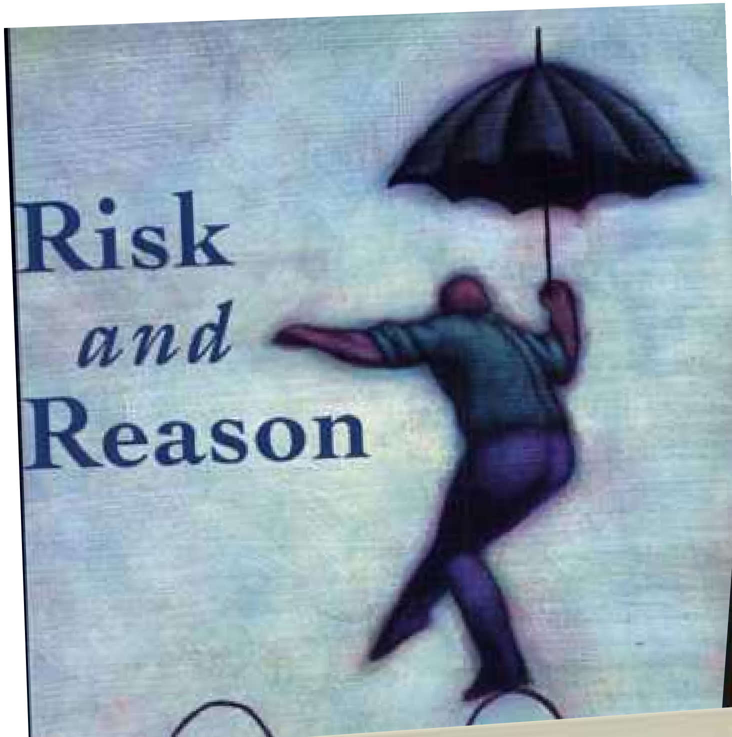
Y.-A. DE MONTEJOYE *et al.*'s Report "Unique in the shopping mall: On the reidentifiability of credit card data" (special section on The End of Privacy, 30 January, p. 536) led to a widespread media sensation proclaiming that reidentification is easy with only a few pieces of credit card data (*1–3*). Although we agree with de Montejoye *et al.* that data disclosure practices must be responsibly balanced with data privacy and utility, we are concerned that the study's findings reflect unrealistic data intrusion threats. Making policy decisi

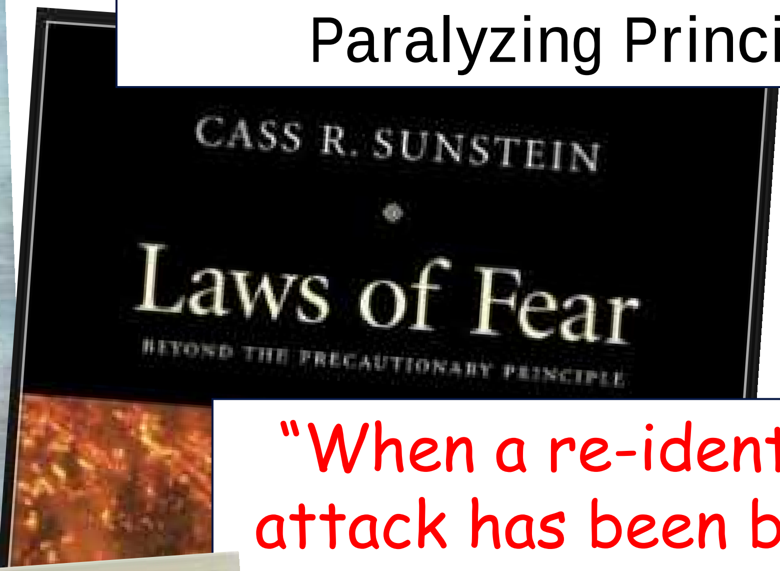
Sample Unique ≠ Re-identifiable

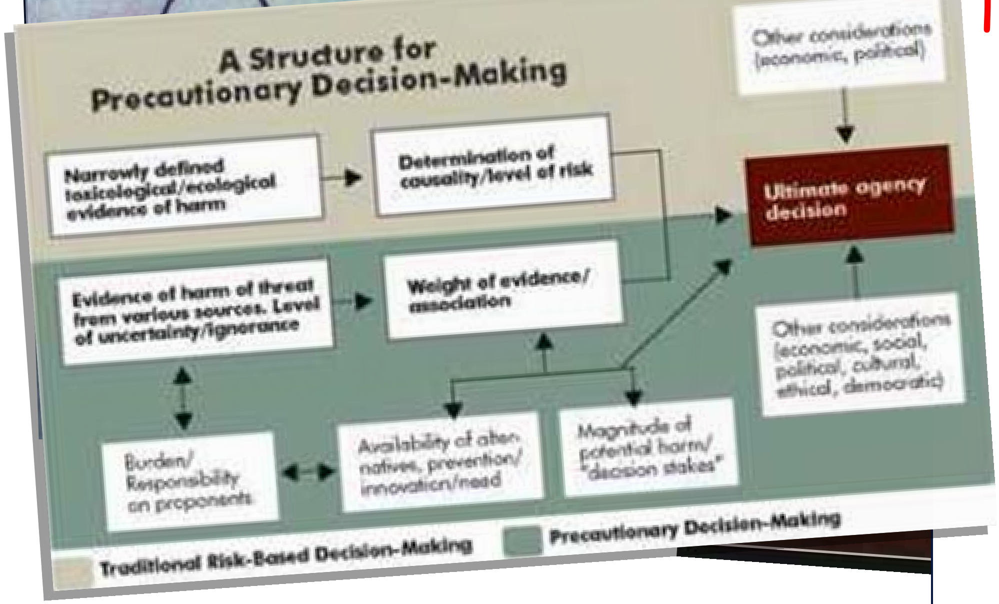
1.1 Million = small sample fraction

https://blogs.law.harvard.edu/infolaw/2015/04/28/is-de-identification-dead-again/

Precautionary Principle or Paralyzing Principle?






A Structure for Precautionary Decision-Making
Narrowly defined toxicological/ecological evidence of harm
Determination of causality/level of risk
Other considerations (economic, political)
Ultimate agency decision
Evidence of harm of threat from various sources. Level of uncertainty/ignorance
Weight of evidence/ association
Other considerations (economic, social, political, cultural, ethical, democratic)
Burden/ Responsibility on proponents
Availability of alternatives, prevention/ innovation/need
Magnitude of potential harm/ "decision stakes"
Traditional Risk-Based Decision-Making
Precautionary Decision-Making

"When a re-identification attack has been brought to life, our assessment of the probability of it actually being implemented in the real-world may subconsciously become 100%, which is highly distortive of the true risk/benefit calculus that we face." – DB-J