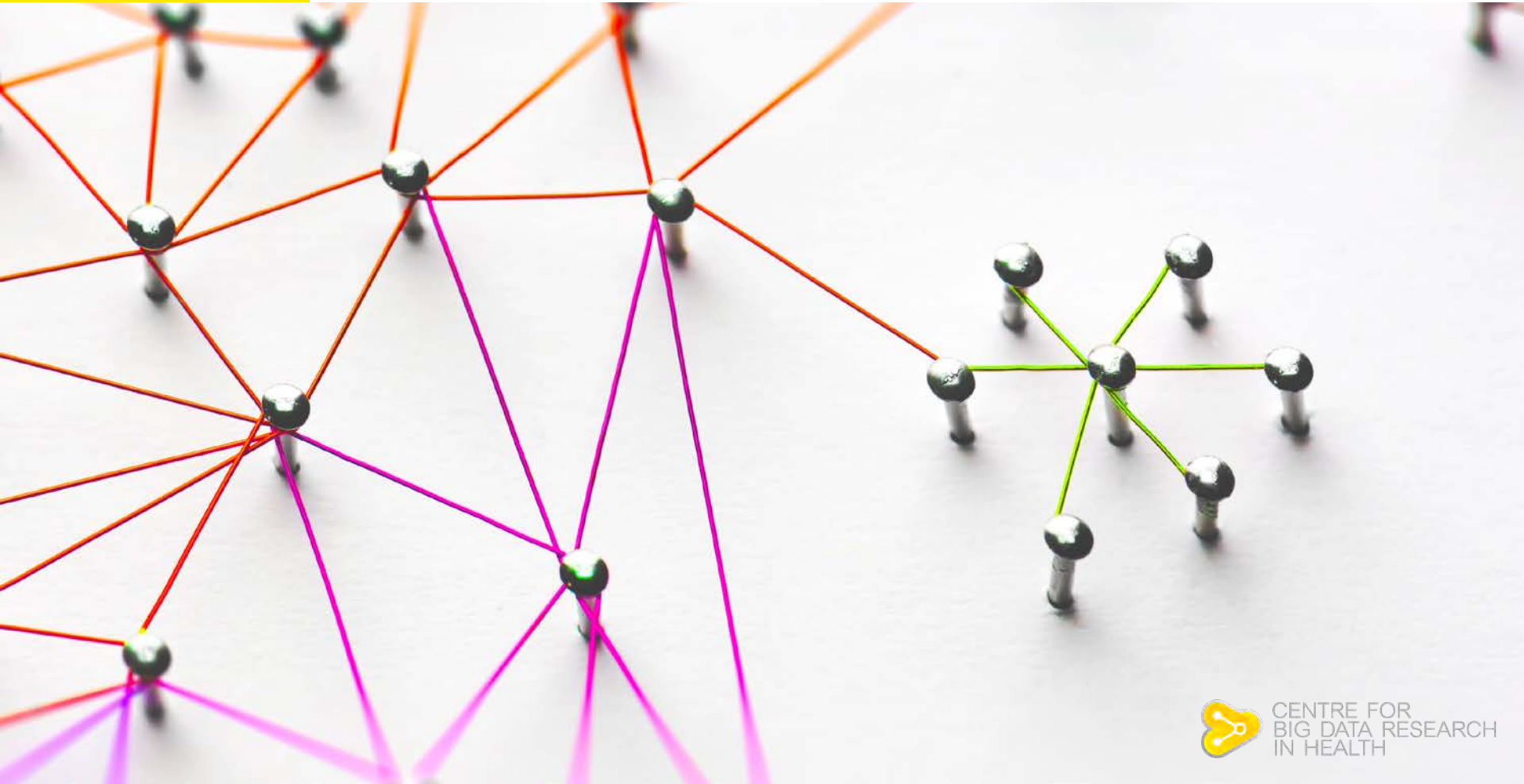
Big data in health and medicine: issues and challenges

Louisa Jorm

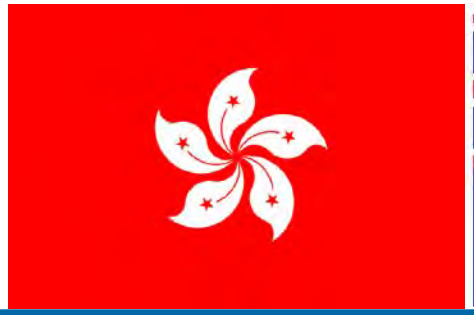
Health Research Symposium, Hong Kong, 16 June 2017



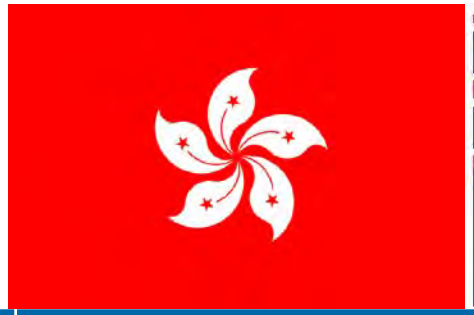
CENTRE FOR BIG DATA RESEARCH IN HEALTH

Outline

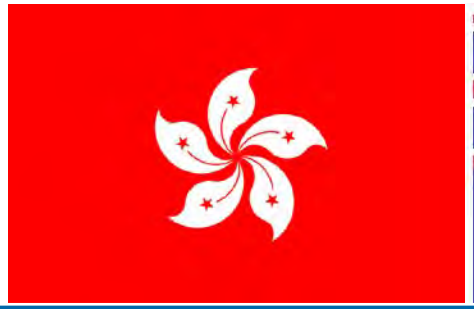
- What are big data?
- How are they being used in health and medicine?
- Issues and challenges:
  1. Upscaling technology
  2. New analytic paradigms
  3. Workforce shortages
  4. “Open data” vs. privacy protection



	Hong Kong SAR	Australia
Population (‘000)	7,242	23,470
Total fertility rate	1.2	1.8
% population aged 65+	19.2	15.0
Life expectancy at birth M/F	81.2/86.9	80.4/84.5
Total health expenditure,% of GDP	5.2	9.0
Human development index	0.90	0.93
Per capita GDP (US$)	38,074	44,820



	Hong Kong SAR	Australia
Population (‘000)	7,242	23,470
Total fertility rate	1.2	1.8
% population aged 65+	19.2	15.0
Life expectancy at birth M/F	81.2/86.9	80.4/84.5
Total health expenditure,% of GDP	5.2	9.0
Human development index	0.90	0.93
Per capita GDP (US$)	38,074	44,820



	Hong Kong SAR	Australia
Public hospitals	42	736
Hospital beds	25,000	60,300
Doctors per 10,000 population	18.3	35
Hospital beds per 10,000 population	34.5	25.7
Health ranking (Legatum prosperity index*)	7	8

*Combines indices of physical and mental health, health infrastructure, preventive care

https://lif.blob.core.windows.net/lif/docs/default-source/publications/2016-legatum-prosperity-index-pdf.pdf?sfvrsn=2

Outline

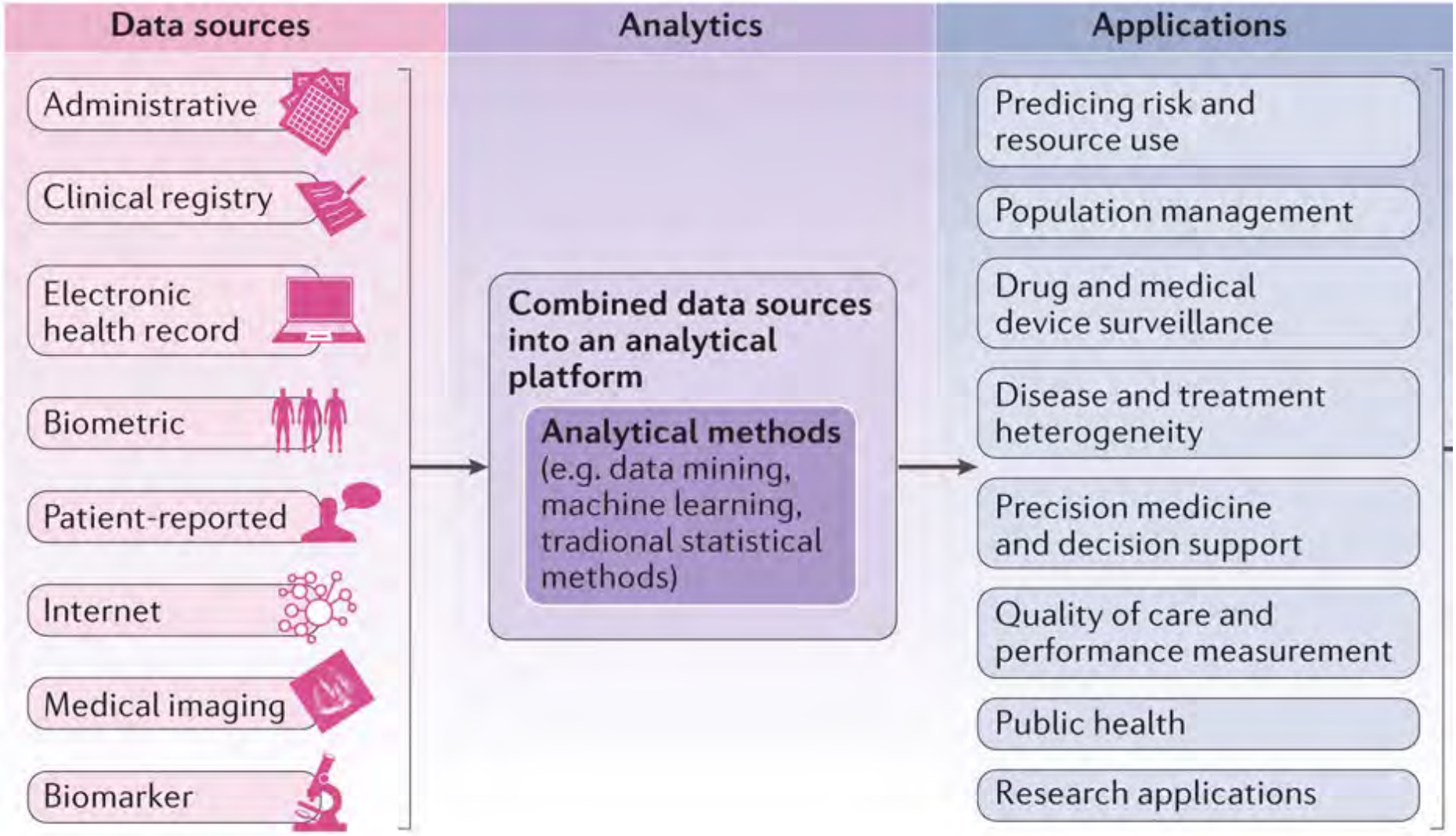
- What are big data?
- How are they being used in health and medicine?
- Issues and challenges:
  1. Upscaling technology
  2. New analytic paradigms
  3. Workforce shortages
  4. “Open data” vs. privacy protection

What are big data?

- Volume: Large scale of data (terabytes or petabytes)
- Variety: Variable format of data (structured, semi structured and unstructured)
- Velocity: Speed at which data are produced, processed, and analysed
- Value: Worth of information to stakeholders and decision makers

Outline

- What are big data?
- How are they being used in health and medicine?
- Issues and challenges:
  1. Upscaling technology
  2. New analytic paradigms
  3. Workforce shortages
  4. “Open data” vs. privacy protection



Rumsfeld, J. S. *et al.* (2016) Big data analytics to improve cardiovascular care: promise and challenges. *Nat. Rev. Cardiol.* doi:10.1038/nrcardio.2016.42

Real-Time Risk Prediction on the Wards: A Feasibility Study

Critical Care Medicine. 44(8):1468–1473, AUG 2016

Michael A. Kang; Matthew M. Churpek; and 4 more

Objective:
Failure to detect clinical deterioration in the hospital is common and associated with poor patient outcomes and increased healthcare costs. Our objective was to evaluate the feasibility and accuracy of real-time risk stratification using the electronic Cardiac Arrest Risk Triage score, an electronic health record-based early warning score.

Design:
We conducted a prospective black-box validation study. Data were transmitted via HL7 feed in real time to an integration engine and database server wherein the scores were calculated and stored without visualization for clinical providers. The high-risk threshold was set a priori. Timing and sensitivity of electronic Cardiac Arrest Risk Triage score activation were compared with standard-of-care Rapid Response Team activation for patients who experienced a ward cardiac arrest or ICU transfer.

Setting:
Three general care wards at an academic medical center.

Patients:
A total of 3,889 adult inpatients.

Conclusions:
Electronic Cardiac Arrest Risk Triage score identified significantly more cardiac arrests and ICU transfers than standard Rapid Response Team activation and did so many hours in advance.

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Single Reading with Computer-Aided Detection for Screening Mammography

Fiona J. Gilbert, F.R.C.R., Susan M. Astley, Ph.D., Maureen G.C. Gillan, Ph.D., Olorunsola F. Agbaje, Ph.D., Matthew G. Wallis, F.R.C.R., Jonathan James, F.R.C.R., Caroline R.M. Boggis, F.R.C.R., and Stephen W. Duffy, M.Sc., for the CADET II Group*

BACKGROUND

The sensitivity of screening mammography for the detection of small breast cancers is higher when the mammogram is read by two readers rather than by a single reader. We conducted a trial to determine whether the performance of a single reader using a computer-aided detection system would match the performance achieved by two readers.

RESULTS

The proportion of cancers detected was 199 of 227 (87.7%) for double reading and 198 of 227 (87.2%) for single reading with computer-aided detection ($P=0.89$). The overall recall rates were 3.4% for double reading and 3.9% for single reading with computer-aided detection; the difference between the rates was small but significant ($P<0.001$). The estimated sensitivity, specificity, and positive predictive value for single reading with computer-aided detection were 87.2%, 96.9%, and 18.0%, respectively. The corresponding values for double reading were 87.7%, 97.4%, and 21.1%. There were no significant differences between the pathological attributes of tumors detected by single reading with computer-aided detection alone and those of tumors detected by double reading alone.

CONCLUSIONS

Single reading with computer-aided detection could be an alternative to double reading and could improve the rate of detection of cancer from screening mammograms read by a single reader. (ClinicalTrials.gov number, NCT00450359.)

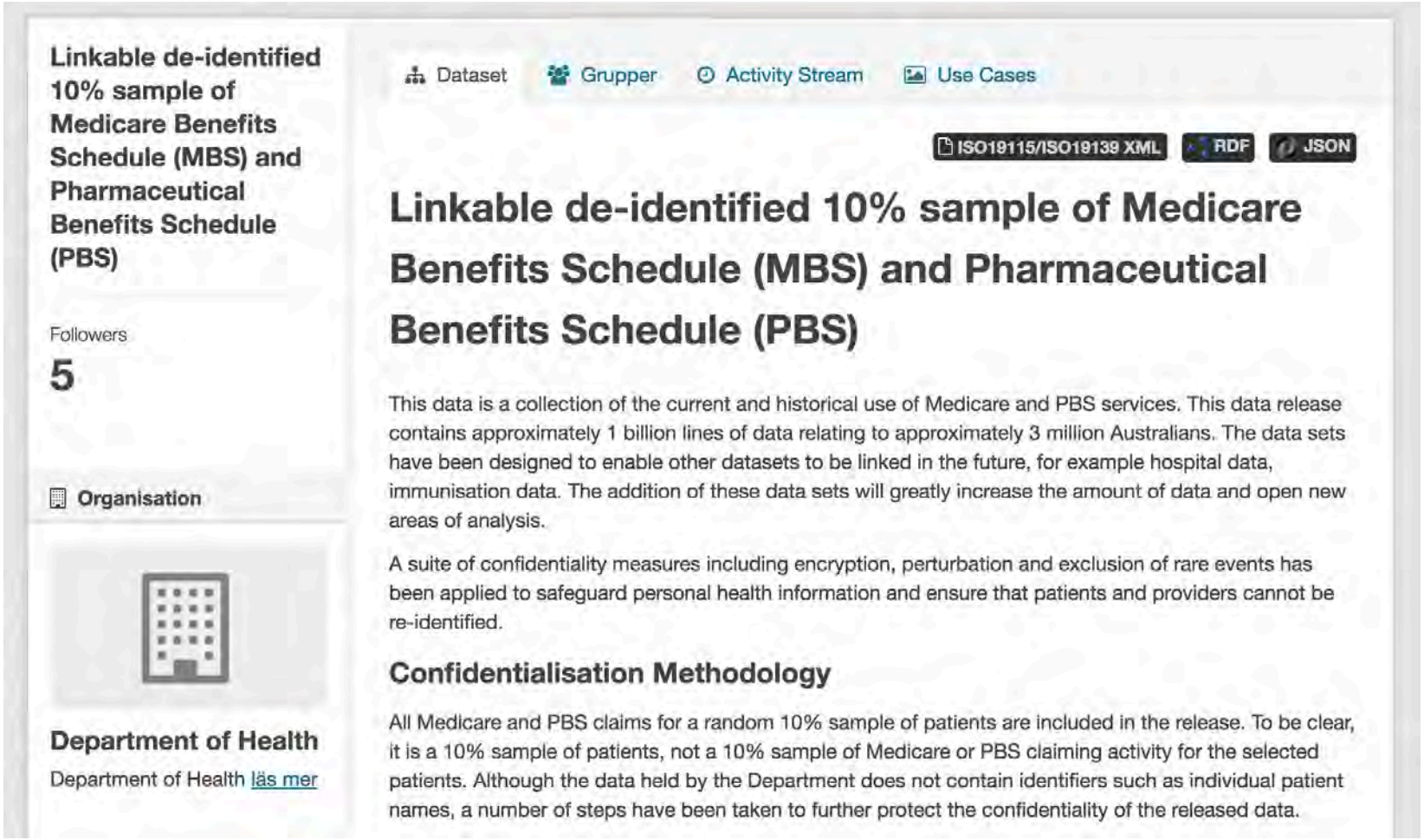
Interrupted Time Series Analysis of the Effect of Rescheduling Alprazolam in Australia: Taking Control of Prescription Drug Use

Alprazolam is significantly more toxic, has no add
therapeutic benefit, and is increasingly misused cor
with other benzodiazepines.[1-3] Due to concerns ab
increasing use of alprazolam, in February 2014, the
lian Therapeutic Goods Administration selectively re
uled alprazolam from Schedule 4 (Prescription Only
cine) of the Poisons Standard to Schedule 8 (Con
Drug), equivalent to Schedule II in the United State

Discussion | In Australia, selectively rescheduling alprazolam led to a reduction in overall use and adverse events and increased switching to less toxic benzodiazepines. More

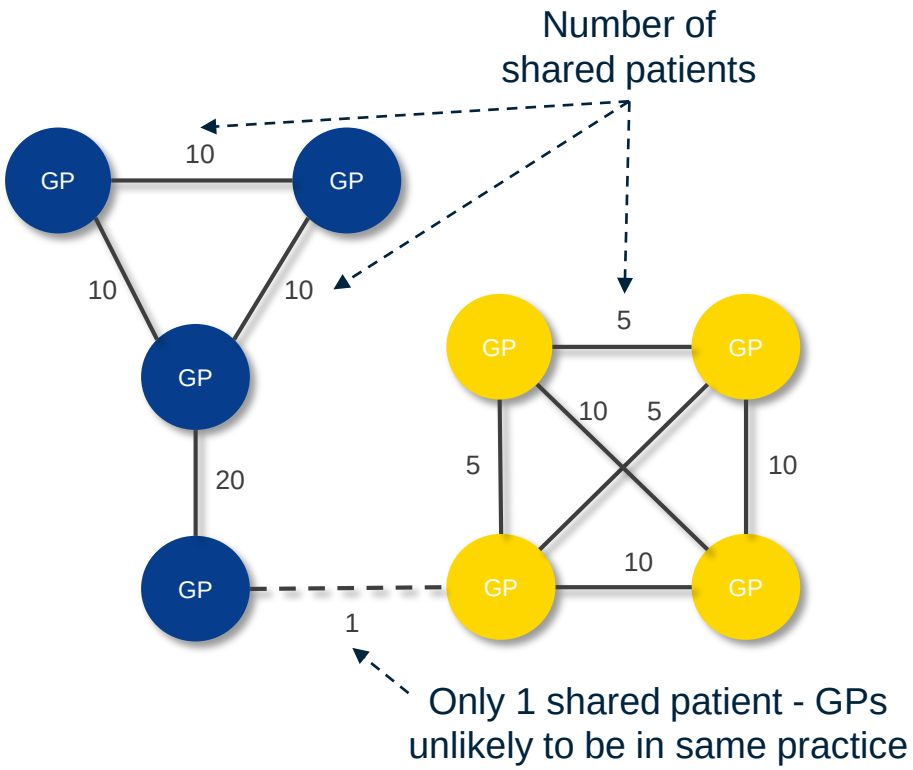
Figure. Monthly Time Series of Alprazolam Prescriptions per 100 000 Population, Dispensings, and Calls to Poisons Information Centre

Schaffer AL, Buckley NL, Cairns R, Pearson S.
JAMA Intern Med. Published online July 05, 2016.
doi:10.1001/jamainternmed.2016.2992

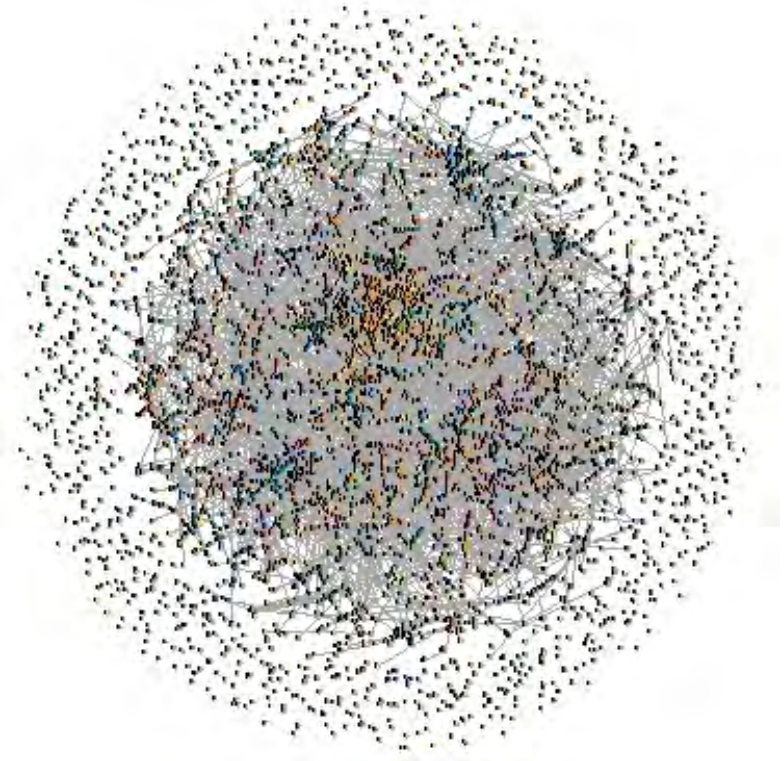

Linkable de-identified 10% sample of Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Schedule (PBS)
Followers
5
Organisation
Department of Health
Department of Health läs mer
Dataset
Grupper
Activity Stream
Use Cases
ISO19115/ISO19139 XML
RDF
JSON
Linkable de-identified 10% sample of Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Schedule (PBS)
This data is a collection of the current and historical use of Medicare and PBS services. This data release contains approximately 1 billion lines of data relating to approximately 3 million Australians. The data sets have been designed to enable other datasets to be linked in the future, for example hospital data, immunisation data. The addition of these data sets will greatly increase the amount of data and open new areas of analysis.
A suite of confidentiality measures including encryption, perturbation and exclusion of rare events has been applied to safeguard personal health information and ensure that patients and providers cannot be re-identified.
Confidentialisation Methodology
All Medicare and PBS claims for a random 10% sample of patients are included in the release. To be clear, it is a 10% sample of patients, not a 10% sample of Medicare or PBS claiming activity for the selected patients. Although the data held by the Department does not contain identifiers such as individual patient names, a number of steps have been taken to further protect the confidentiality of the released data.

Creating networks of health practitioners

- Connect GPs based on patterns of shared patients
  - GPs who share patients are more likely to work in the same practice



- e.g. National Medicare data for 10% of all Australians
  - 2,923 GP communities created from 25,338 GPs



We can explore the impact of different practice patterns on patient care

www.healthmap.org/en/

7 alerts for all diseases, current location, in the past week

Outbreaks in Current Location

Zika outbreak

4 Respiratory Alerts

Avian Influenza H5N1 (1), Avian Influenza H5N6 (1), Avian Influenza H7N9 (2)

1 Animal Alerts

Avian Influenza (1)

1 Skin/Rash Alerts

Hand, Foot and Mouth Disease (1)

Potentially Preventable Hospitalisations

Potentially prevented by timely and effective provision of primary and preventative care

Vaccine preventable

- Influenza and pneumonia
- Other vaccine-preventable conditions

Acute

- Dehydration & gastroenteritis
- Pyelonephritis
- Perforated/bleeding ulcer
- Cellulitis
- Pelvic inflammatory disease
- Ear, nose & throat infections
- Dental conditions
- Appendicitis with generalised peritonitis
- Convulsions & epilepsy
- Gangrene

Chronic

- Asthma
- Congestive heart failure
- Diabetes complications
- Chronic obstructive pulmonary disease
- Angina
- Iron deficiency anaemia
- Hypertension
- Nutritional deficiencies
- Rheumatic heart disease

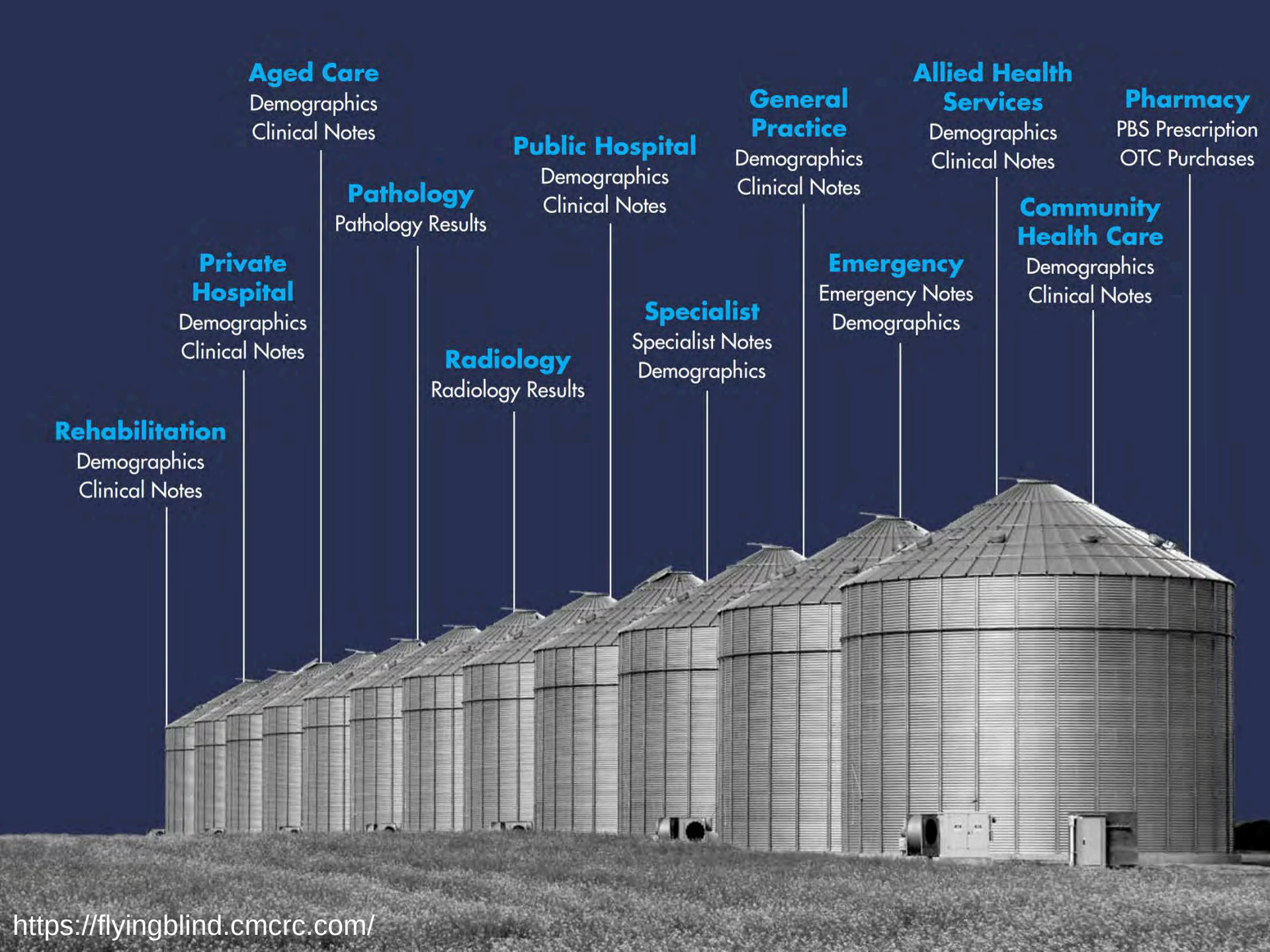
Impact Of Socioeconomic Status On Hospital Use In New York City

by John Billings, Lisa Zeitel, Joanne Lukomnik, Timothy S. Carey, Arthur E. Blank, and Laurie Newman

Abstract: This DataWatch examines the potential impact of socioeconomic differences on rates of hospitalization, based on patterns of hospital use in New York City in 1988. The research suggests that lack of timely and effective outpatient care may lead to higher hospitalization rates in low-income areas. For certain conditions identified as ambulatory care sensitive, hospitalization rates were higher in low-income areas than they were in higher-income areas where appropriate outpatient care was more readily available. Further study is needed to determine the relative impact of various economic, structural, and cultural factors that affect access to care.

Cite this article as:
J Billings, L Zeitel, J Lukomnik, T S Carey, A E Blank and L Newman
Impact of socioeconomic status on hospital use in New York City
Health Affairs, 12, no.1 (1993):162-173

doi: 10.1377/hlthaff.12.1.162



https://flyingblind.cmcrc.com/

The APHID Study

45 & Up Study

- Prospective cohort of 267,091 people aged over 45 in NSW.
- Study entry 2006-2008
- Questionnaire
  - Demographics
  - Health status
  - Risk factors

NSW Admitted Patient Data Collection

- Census of all hospital separations in NSW public and private hospitals and day procedure centres.
- Linked data, 2000-2013
- N=1,761,178 records

MBS

- Claims for subsidised medical and diagnostic services in Australia
- Linked data, 2004-2011
- N=45,754,339 records

PBS

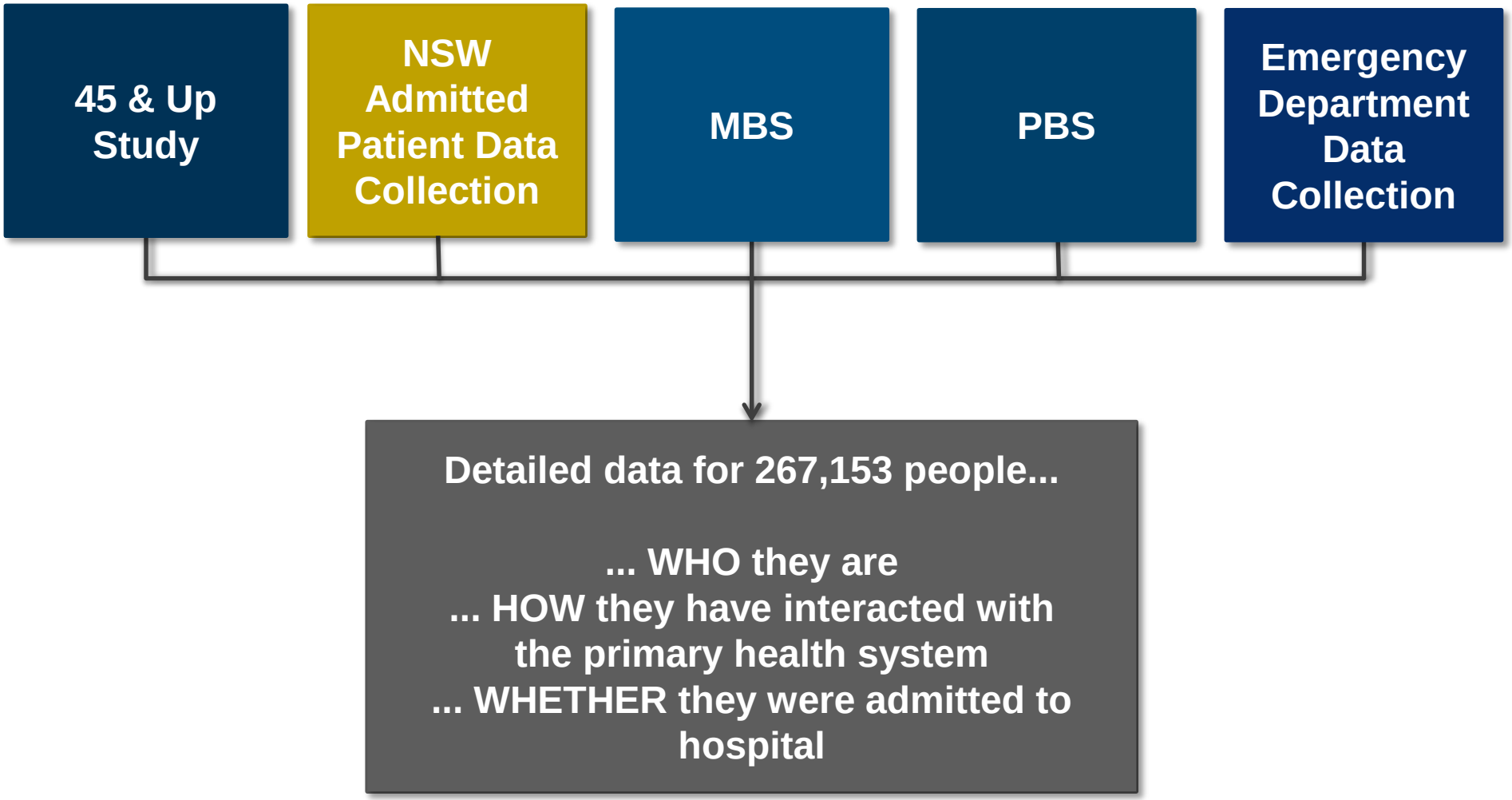
- Claims for subsidised pharmaceuticals In Australia
- Linked data, 2004-2011
- N= 34,978,006 records

NSW Emergency Department Data Collection

- Presentations to 80 EDs (75% 0f NSW presentations)
- Linked data, 2006-2013
- N= 586,131 records

+ Fact of death to 2013

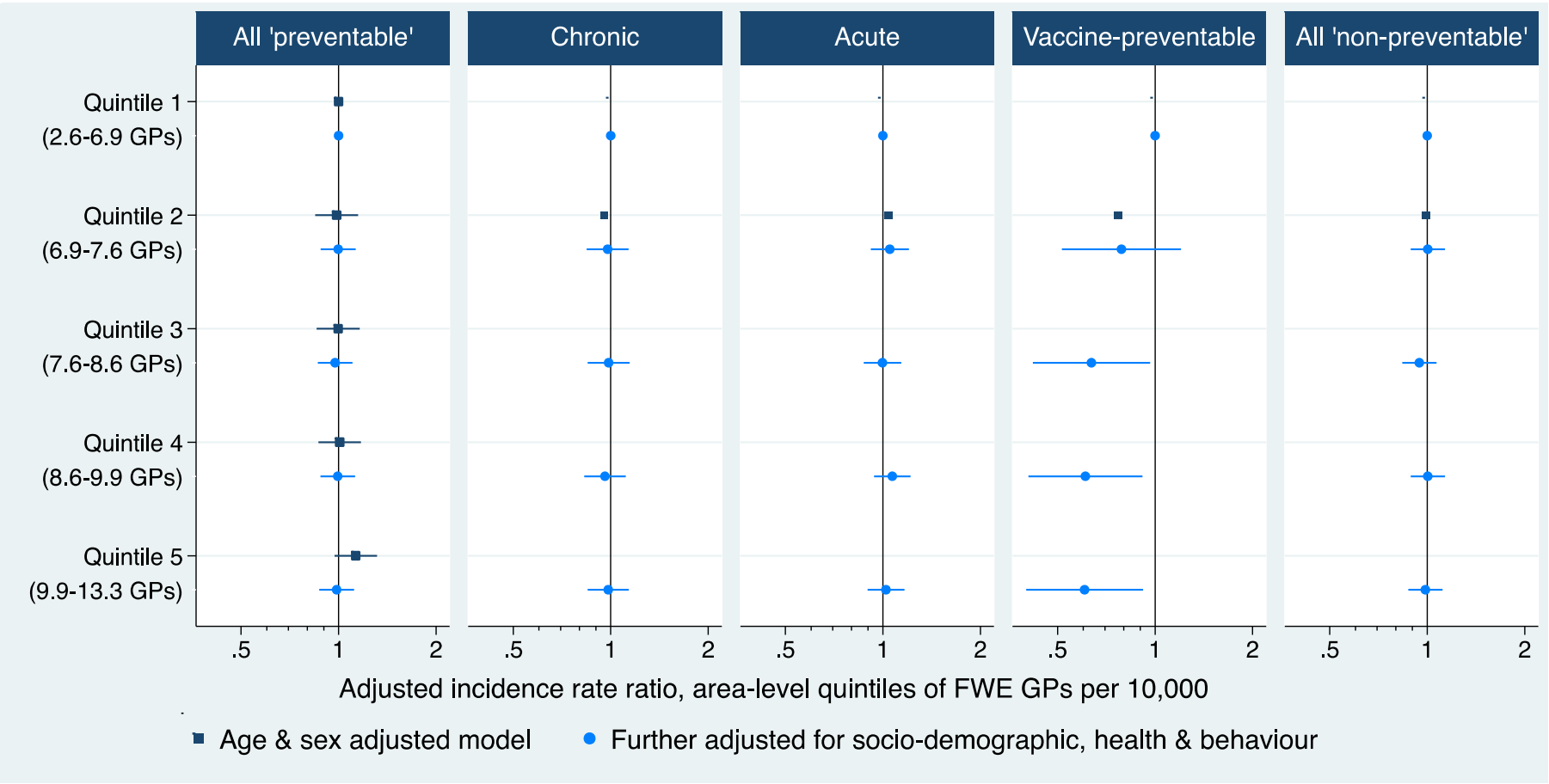
Data linkage



What factors explain geographic variation in admission?

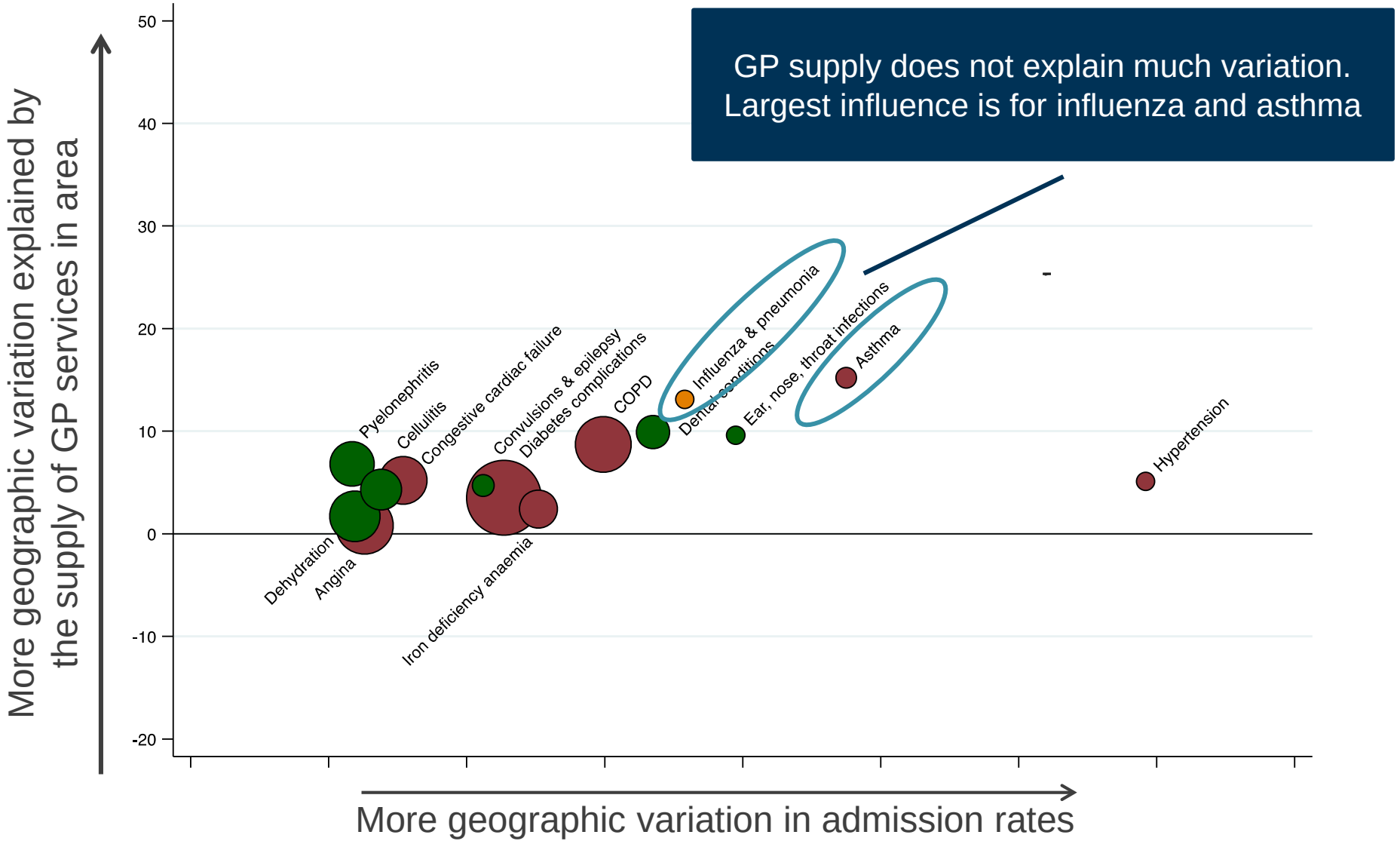
Primary care supply	Socio-demographic factors	Factors amenable to behaviour change & disease management
▪ Number of full time workload equivalent (FWE) GPs, per 10,000 residents in area	▪ Education ▪ Language spoken at home ▪ Marital status ▪ Aboriginal status ▪ Income ▪ Employment ▪ Health insurance status ▪ Number of people can depend on	▪ Healthy behaviours ▪ Body Mass Index ▪ Self-reported health ▪ Number of co-morbidities ▪ Functional status ▪ Psychological distress

Area-level supply of full time workload equivalent GPs and PPH admissions



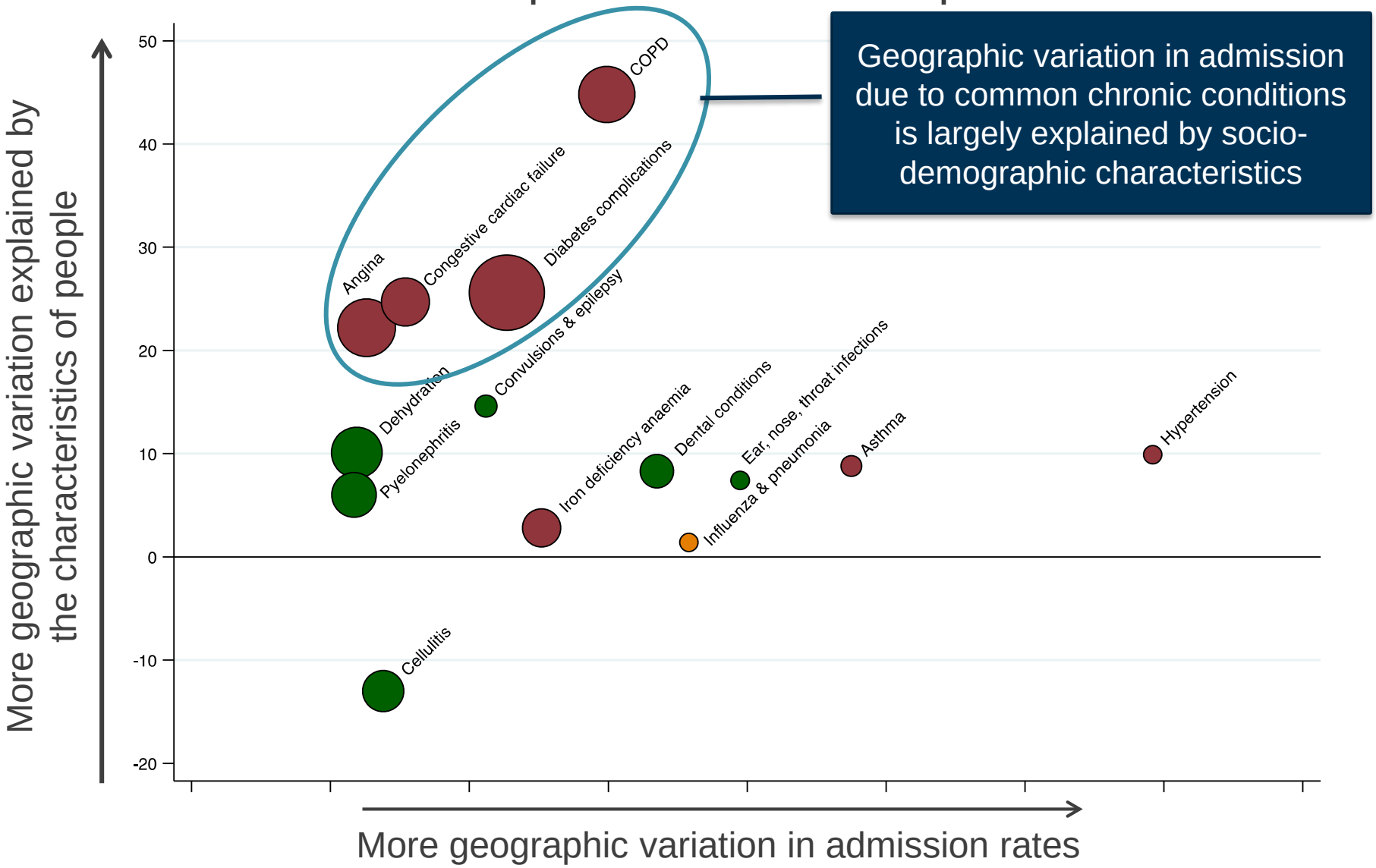
Falster MO, Jorm LR, Douglas KA, et al. Sociodemographic and health characteristics, rather than primary care supply, are major drivers of geographic variation in preventable hospitalisations in Australia. Medical Care 2015; 53:436-445

For which conditions does GP supply explain most variation?



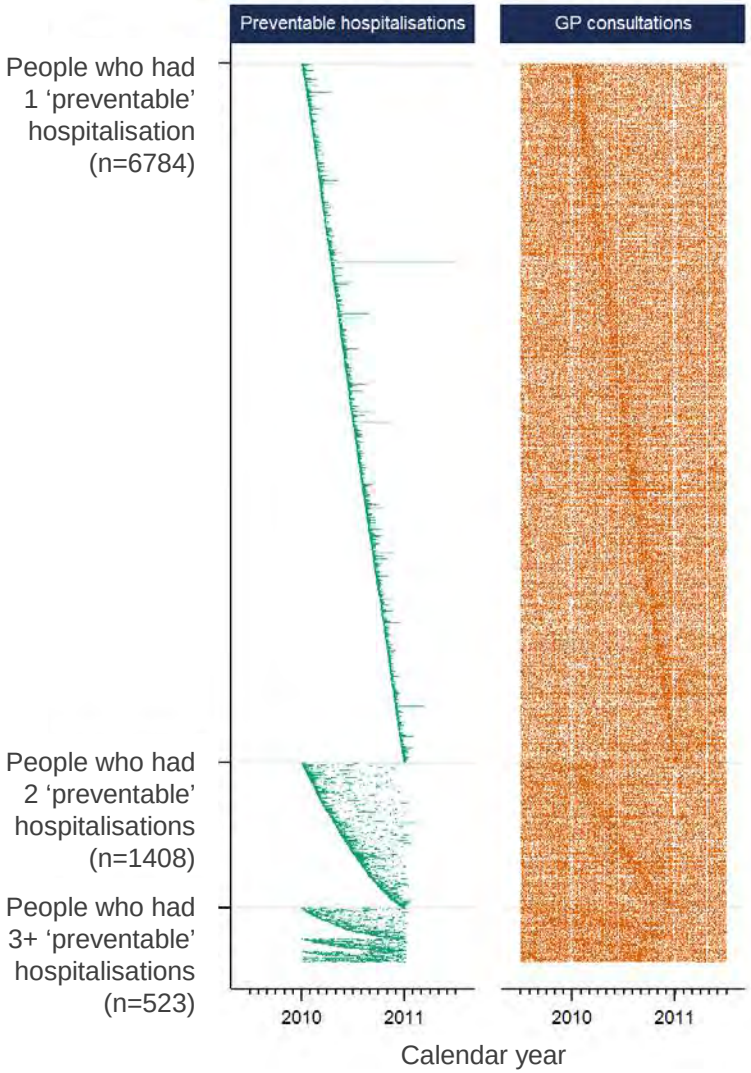
Falster MO, Jorm LR, Douglas KA, et al. Sociodemographic and health characteristics, rather than primary care supply, are major drivers of geographic variation in preventable hospitalisations in Australia. Medical Care 2015; 53:436-445

For which conditions do personal factors explain most variation?



Falster MO, Jorm LR, Douglas KA, et al. Sociodemographic and health characteristics, rather than primary care supply, are major drivers of geographic variation in preventable hospitalisations in Australia. Medical Care 2015; 53:436-445

Visualising unit record data on preventable hospitalisations, GP consultations… and other health events



Each row is a person

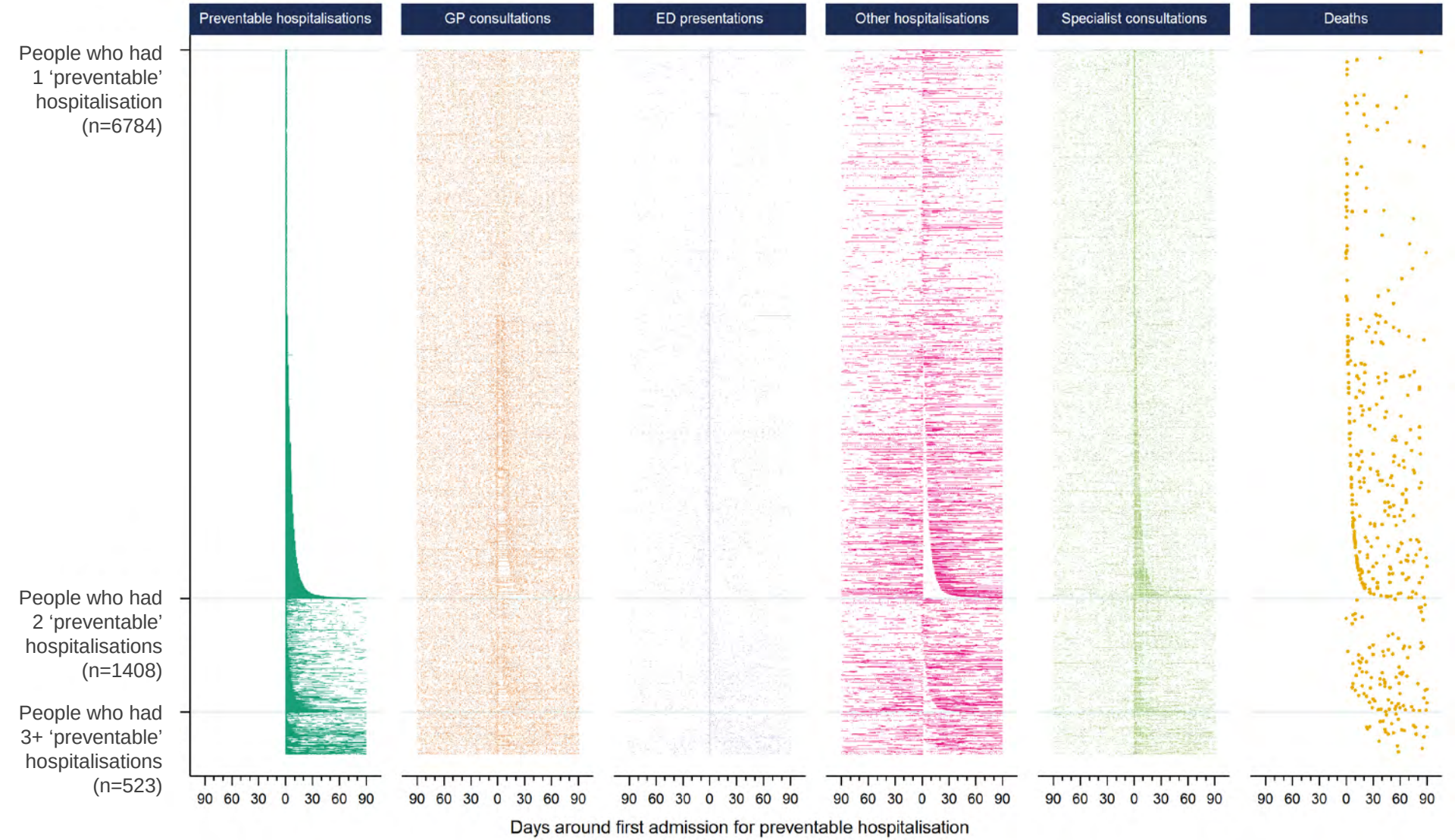
Each line/dot is a health event over time

Patterns become clear in the position and density of health events in the plot

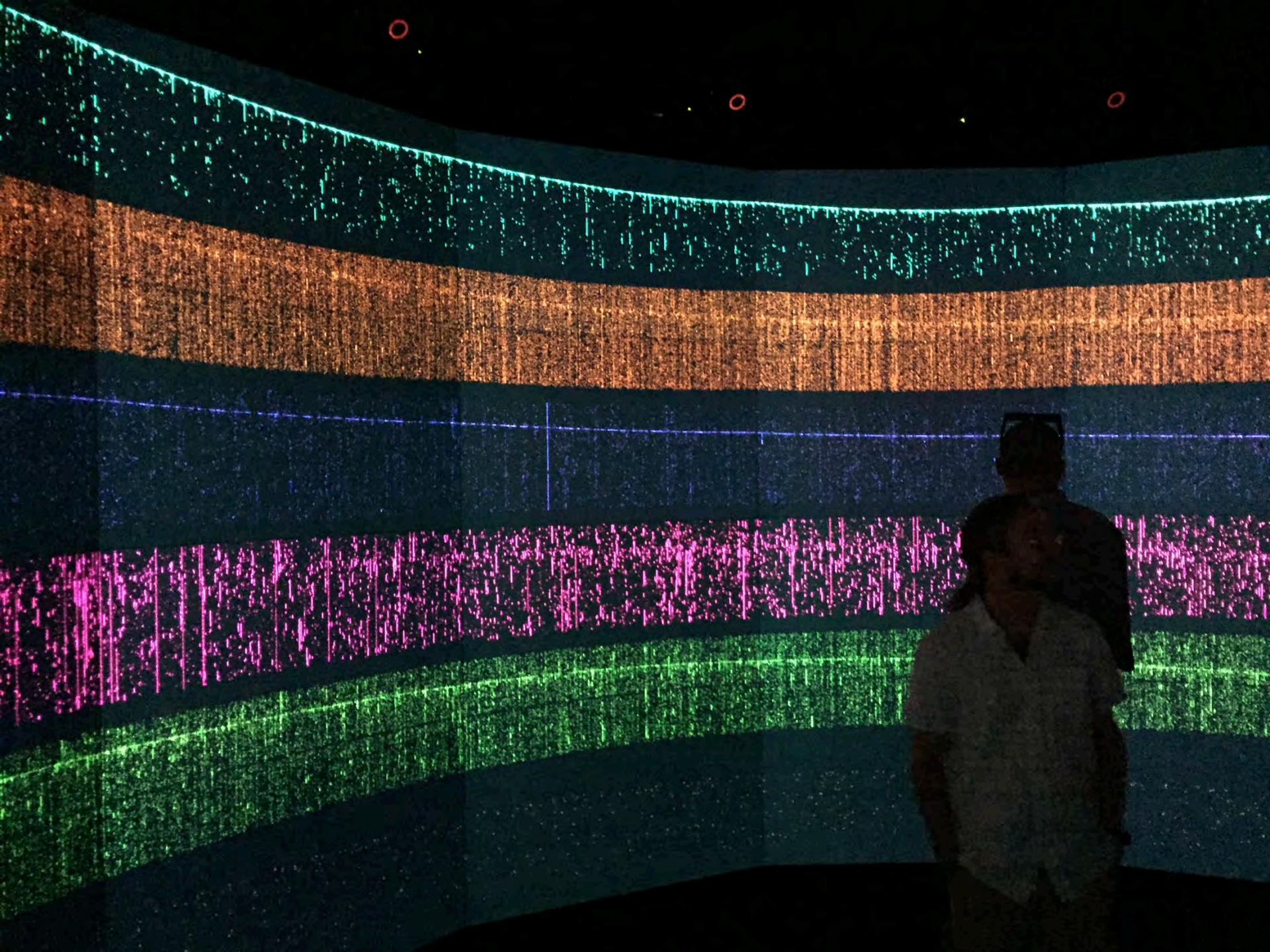
e.g. Patients visiting GPs at the time of hospitalisation, not at Christmas

Falster MO, Jorm LR, Leyland AH. Visualising linked health data to explore health events around preventable hospitalisations in NSW Australia. BMJ Open 2016; 6:e012031

Zooming in on the time around preventable hospitalisation



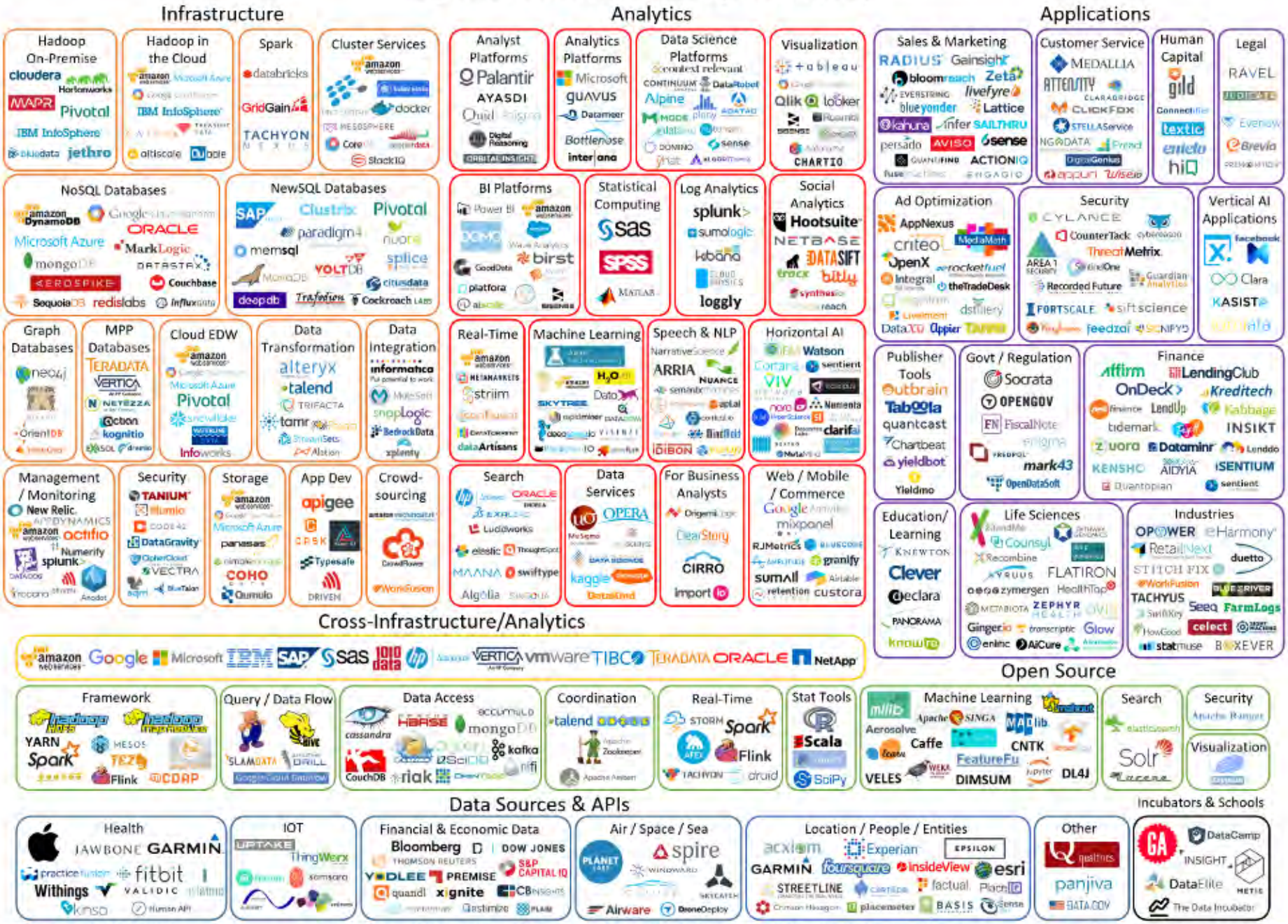
Falster MO, Jorm LR, Leyland AH. Visualising linked health data to explore health events around preventable hospitalisations in NSW Australia. BMJ Open 2016; 6:e012031



Outline

- What are big data?
- How are they being used in health and medicine?
- Issues and challenges:
  1. Upscaling technology
  2. New analytic paradigms
  3. Workforce shortages
  4. “Open data” vs. privacy protection

Big Data Landscape 2016 (Version 2.0)



Last Updated 2/12/2016

FIRSTMARK

Secure Unified Research Environment

- SURE hosts virtual project workspaces that are accessed remotely over encrypted internet connections
- Members of a research team have shared access to project workspaces
- Each project workspace has its own security perimeter so data from different projects cannot be combined
- All data are stored on dedicated servers housed in a highly secure data centre
- Movement of data in and out of SURE is carefully controlled via specially designed software called the Curated Gateway

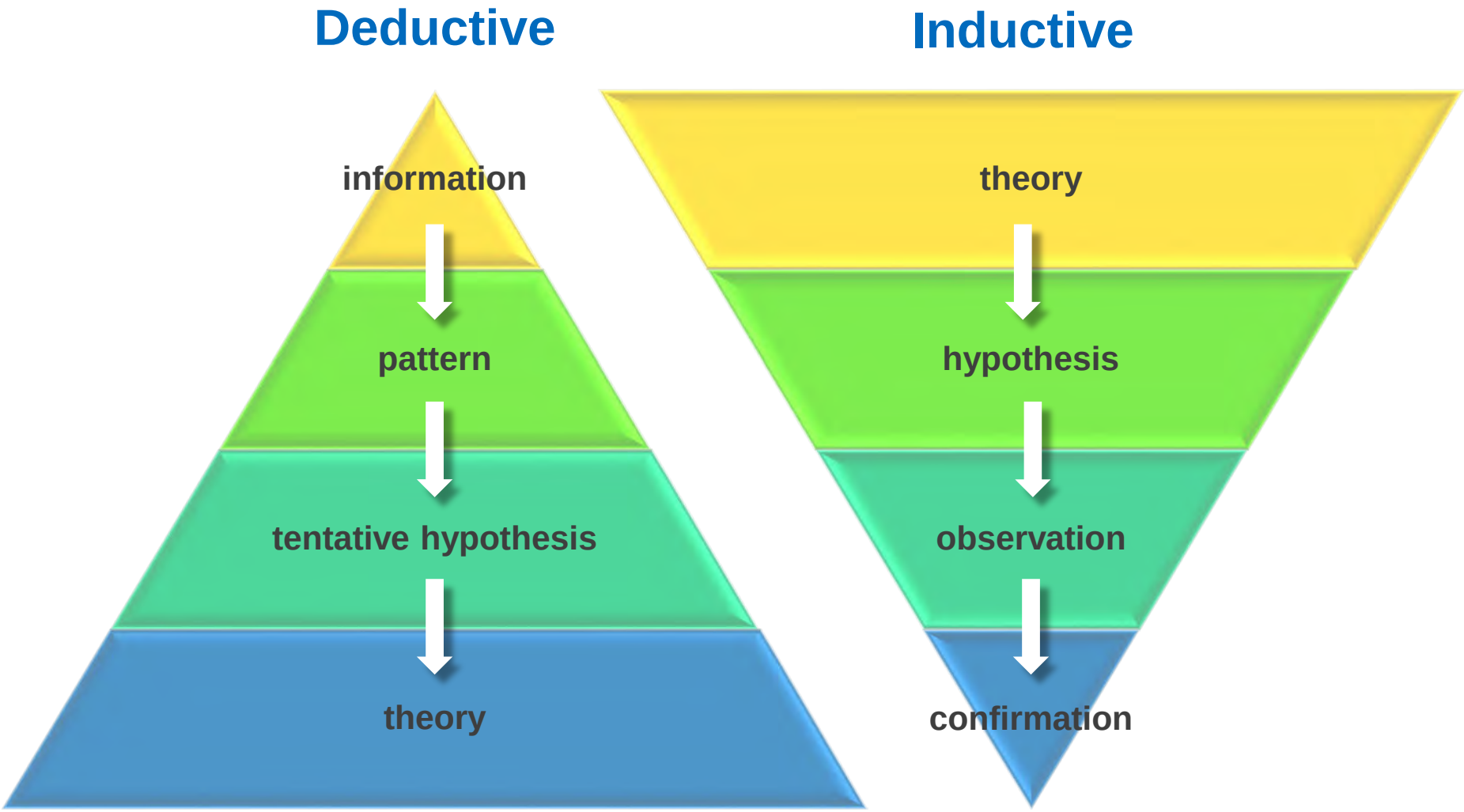
SURE
Secure Unified Research Environment

saxinstitute

Outline

- What are big data?
- How are they being used in health and medicine?
- Issues and challenges:
  1. Upscaling technology
  2. New analytic paradigms
  3. Workforce shortages
  4. “Open data” vs. privacy protection

Deductive vs. inductive reasoning



Statistics vs. machine learning

	Statistics	Machine learning
Focus	Formal statistical inference (why)	Prediction (what)
Problem space	Low dimensional problems	High dimensional problems
Models	Models to explain and predict	Network/graphs to train and test
Assumptions	Explicit a priori assumptions	None (learn from the data)
Distribution	Defined *a priori*	Unknown *a priori*
Fit	Fit to the distribution	Best fit to learning models (generalisation)
Language	Estimation Data point Regression Classification Covariate	Learning Example/Instance Supervised Learning Unsupervised Learning Feature

Describing the Relationship between Cat Bites and H[illegible] Depression Using Data from an Electronic Health

Abst[illegible]

Data mining [illegible]ingly applied to the electronic health record and have led to the discovery of numerous clinical [illegible]ing studies have suggested a potential association between cat bites and human depression. To [illegible]iation in more detail we first used administrative diagnosis codes to identify patients with eithe[illegible] from a population of 1.3 million patients. We then conducted a manual chart review in the elect[illegible]nts with a code for a bite to accurately determine which were from cats or dogs. Overall there were [illegible]108 with dog bites, and approximately 117,000 patients with depression. Depression was found in [illegible]tes and 28.7% of those with dog bites. Furthermore, 85.5% of those with both cat bites and de[illegible]pared to 64.5% of those with dog bites and depression. The probability of a woman being diag[illegible]me point in her life if she presented to our health system with a cat bite was 47.0%, compared to [illegible]a similar bite. The high proportion of depression in patients who had cat bites, especially amo[illegible]reening for depression could be appropriate in patients who present to a clinical provider with a [illegible]o causative link is known to explain this association, there is growing evidence to suggest that the[illegible]d human mental illness, such as depression, warrants further investigation.

depression. The probability of a woman being diagnosed with depression at some point in her life if she presented to our health system with a cat bite was 47.0%, compared to 24.2% of men presenting with a similar bite. The high proportion of depression in patients who had cat bites, especially among women, suggests that screening for depression could be appropriate in patients who present to a clinical provider with a cat bite. Additionally, while no causative link is known to


SHOULD I FOLLOW THE DATA
OR MY INSTINCTS?
memegenerator.net

TECHNOLOGY NEWS 29 April 2016

Revealed: Google AI has access to huge haul of NHS patient data

A data-sharing agreement obtained by **New Scientist** shows that Google DeepMind's collaboration with the NHS goes far beyond what it has publicly announced

Gathering information
Oli Scarff/AFP/Getty Images

By **Hal Hodson**

It's no secret that Google has broad ambitions in healthcare. But a document obtained by New Scientist reveals that the tech giant's collaboration with the UK's National Health Service goes far beyond what has been publicly announced.

The agreement gives DeepMind access to a wide range of healthcare data on the 1.6 million patients who pass through three London hospitals run by the Royal Free NHS Trust – Barnet, Chase Farm and the Royal Free – each year. The agreement also includes access to patient data from the last five years.

https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data

Outline

- What are big data?
- How are they being used in health and medicine?
- Issues and challenges:
  1. Upscaling technology
  2. New analytic paradigms
  3. Workforce shortages
  4. “Open data” vs. privacy protection

At e-commerce site operator Etsy Inc., a biostatistics Ph.D. who spent years mining medical records for early signs of breast cancer now writes statistical models to figure out the terms people use when they search Etsy for a new fashion they saw on the street.

Another 28-year-old at Yelp, with a Ph.D. in applied mathematics, turned his dissertation research on genome mapping into a product used by the company's advertising team. The same genome-mapping algorithm is now used to measure the effect on consumers when multiple small changes are made to online ads.

http://www.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-

Harvard
Business
Review

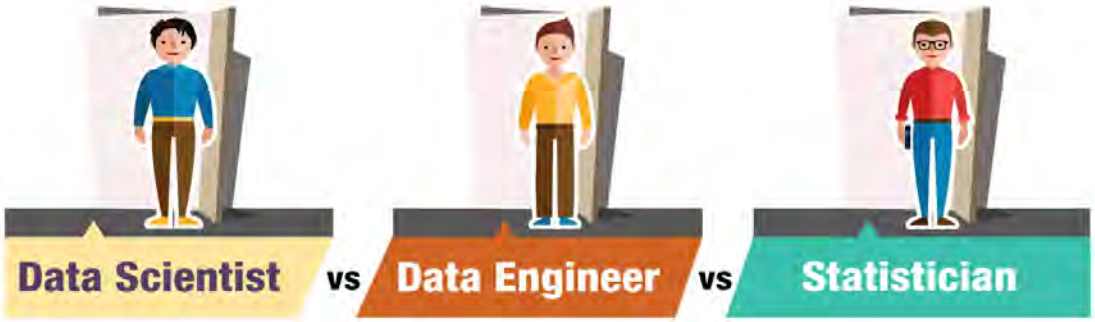
DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

COMPARISON
— Jobs in Data Science —

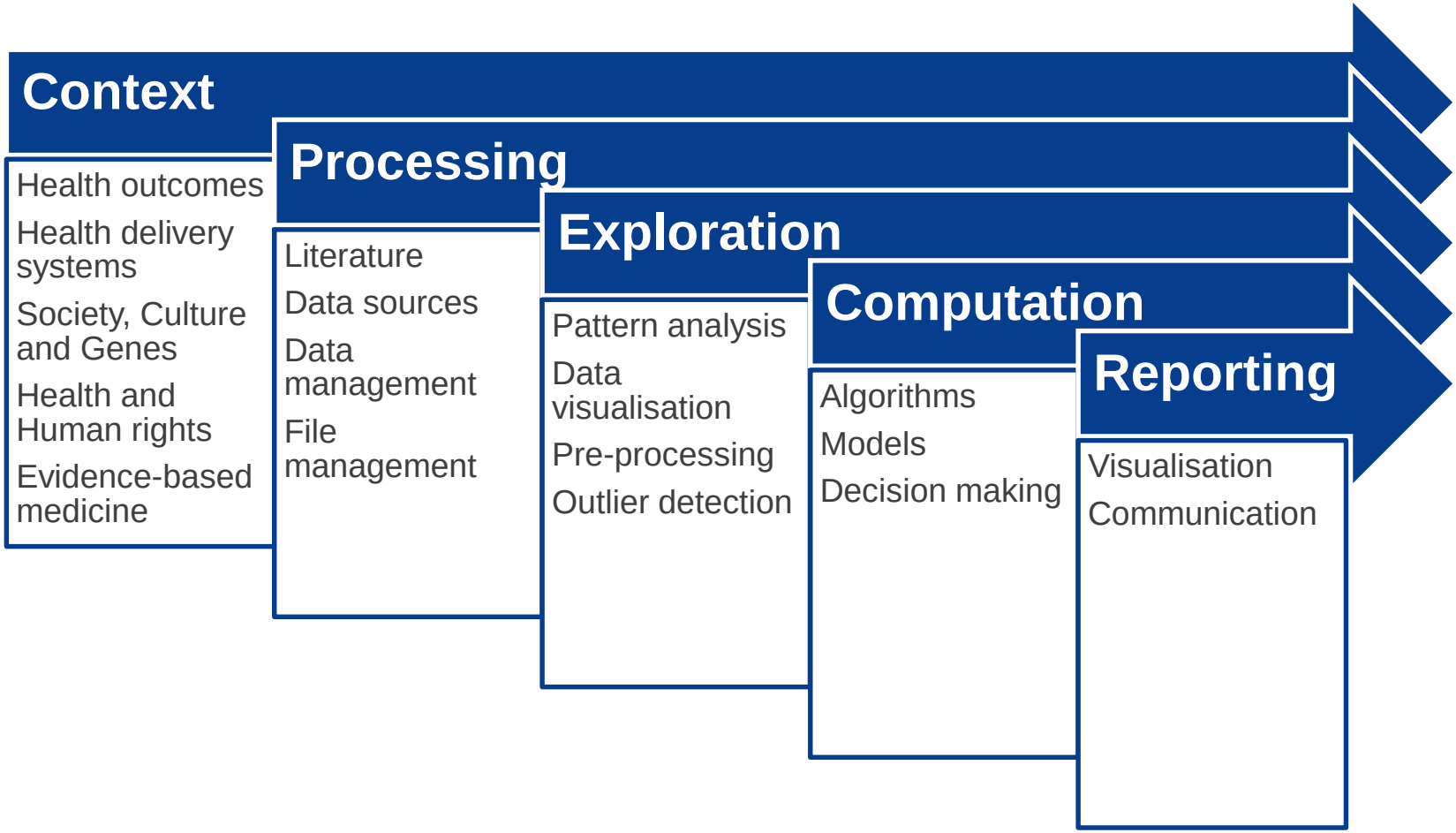


	Data Scientist	Data Engineer	Statistician
	These people use their analytical and technical capabilities to extract meaning insights from data.	These people ensure uninterrupted flow of data between servers and applications. They are responsible for data architecture.	These people understand statistics theoretically and apply them to real life problems.
Responsibilities	Develop and plan required analytic projects in response to business needs. Contribute to data mining architectures, modeling standards, reporting, and data analysis methodologies. Collaborate with stakeholders to integrate data mining results with existing systems. Monitor data mining system performance and implement efficiency improvements.	Design, construct, install, test and maintain highly scalable data management systems Improve data foundational procedures, guidelines and standards Integrate new data management technologies and software engineering tools into existing structures Create custom software components (e.g. specialized UDFs) and analytics applications	Apply statistical theories and methods to solve practical problems of various industries Determine methods for finding or collecting data Design surveys or experiments or opinion polls to collect data Analyze, interpret & undertake data analysis Report conclusions from their analyses
Skills	Programming, Mathematics, Business Understanding, Statistics, Data Visualization, Machine Learning, Attention to detail	Database design, Production coding, Data collection, data warehousing, Data transformation, Work diligently with data	Technical and Analytics Skills, Mathematics, Operational Research, Writing skills,Ability to Analyze, Model and interpret data, Flair of explaining difficult concepts in simple manner
Salary (US)	 Mean $122,000	 Mean $115,000	 Mean $92,000

Tools

https://www.analyticsvidhya.com/wp-content/uploads/2015/10/infographic.jpg

UNSW MSc in Health Data Science (commencing 2018)



UNSW MSc in Health Data Science: Overview

Who?

– Broad local and international target student base

What?

Grad Cert	Grad Dip	MSc	
Context of Health Data Science	Health Data Analytics: Machine Learning and Data Mining	Dissertation	Capstone
Statistical Foundations for Health Data Science	Health Data Analytics: Statistical Modelling I		Elective
Principles of Programming	Health Data Analytics: Statistical Modelling II		Elective
Management and Curation of Health Data	Visualisation and Communication of Health Data		Elective

When?

– Semester 1 2018

Outline

- What are big data?
- How are they being used in health and medicine?
- Issues and challenges:
  1. Upscaling technology
  2. New analytic paradigms
  3. Workforce shortages
  4. “Open data” vs. privacy protection

Benefits of data sharing

- Accelerates the pace of discovery
- Promotes open inquiry
- Supports diversity in analysis and interpretation
- Allows results to be replicated or alternative hypotheses to be tested
- Avoids unnecessary duplication of data collection

Barriers to data sharing

Political

- Lack of trust

Legal

- Ownership, IP
- Preserving privacy

Ethical

- Lack of proportionality
- Lack of reciprocity

Technical

- Data not preserved or not found
- Lack of metadata and standards

Motivational

- No incentives
- Opportunity cost
- Potential criticism/embarrassment

Economic

- Lack of resources

van Panhuis WG, Paul P, Emerson C, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health* 2014, 14:1144

http://www.booktopia.com.au/http_coversbooktopiacomau/big/9781743529805/1000-piece-jigsaw-puzzle.jpg

1 General privacy legislation currently in place in Australia

Jurisdiction	Legislation	Regulator
Federal	*Privacy Act 1988* (Cwlth)	Federal Privacy Commissioner
Australian Capital Territory	*Privacy Act 1988* (Cwlth)	Federal Privacy Commissioner
New South Wales	*Privacy and Personal Information Protection Act 1998*	NSW Privacy Commissioner
Northern Territory	*Information Act 2002*	NT Information Commissioner
Queensland	*Information Privacy Act 2009*	QLD Information Commissioner
South Australia	*Cabinet Administrative Instruction 1/89 2009*	Privacy Committee of South Australia
Tasmania	*Personal Information Protection Act 2004*	Ombudsman Tasmania
Victoria	*Information Privacy Act 2000*	Victorian Privacy Commissioner
Western Australia	No laws	Not applicable

2 Health privacy legislation currently in place in Australia

Jurisdiction	Health privacy legislation	Regulator
Federal	*Privacy Act 1988* (Cwlth)	Federal Privacy Commissioner
Australian Capital Territory	*Health Records (Privacy and Access) Act 1997*	Community and Health Services Complaints Commissioner
New South Wales	*Health Records and Information Privacy Act 2002*	Public sector: internal review Private sector: Privacy NSW
Northern Territory	None currently in place	Not applicable
Queensland	*Information Privacy Act 2009*	Health Quality and Complaints Commission
South Australia	*Code of Fair Information Practice*	Not applicable
Tasmania	None currently in place	Not applicable
Victoria	*Health Records Act 2001*	Health Services Commissioner
Western Australia	None currently in place	Not applicable

Cwlth = Commonwealth. ◆

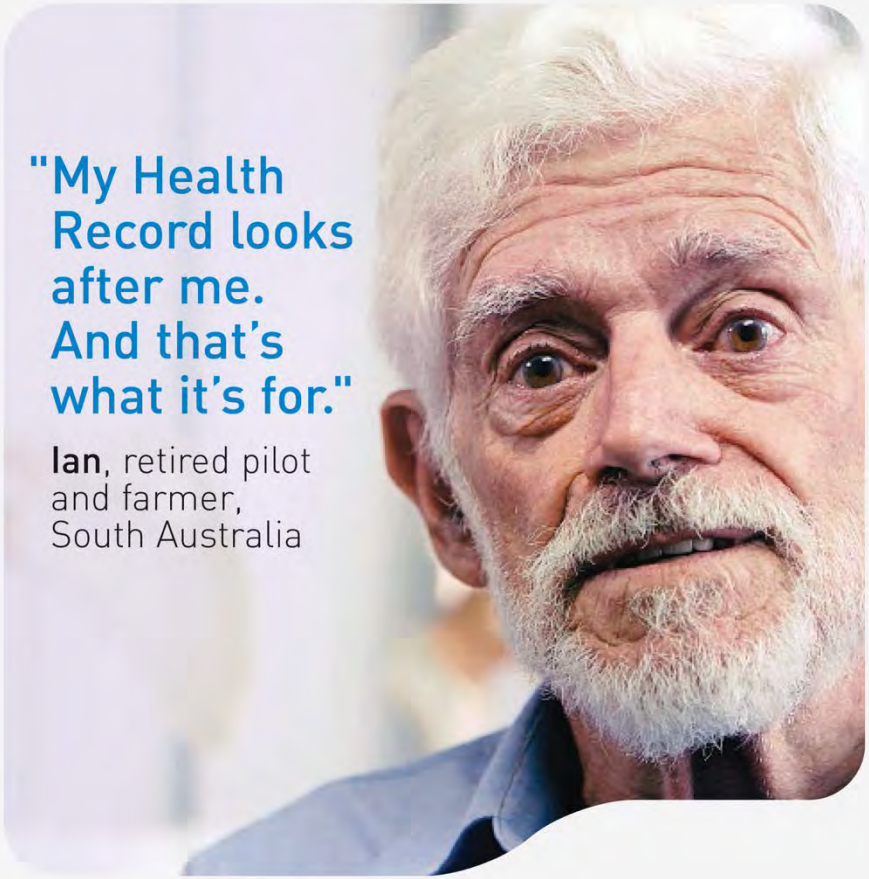
Australian Government
Department of the Prime Minister and Cabinet

PUBLIC SECTOR DATA MANAGEMENT

July 2015

Principles for accessing and using publicly funded data for health research

The National Health and Medical Research Council, th
Department of Health; the Australian Institute of Health a
Bureau of Statistics, the Australian Government Depart
the Australian Electoral Commission, the Austra
and Torres Strait Islander Studies, Universities Aus



Data Availability and Use

Productivity Commission Inquiry Report

No. 82, 31 March 2017

CENTRE FOR BIG DATA RESEARCH IN HEALTH

UNSW SYDNEY

My Health Record 'dumb and useless': Australian Privacy Foundation

Forget last week's Census debacle. Far more has been spent on an e-health system with little clinical value and fewer than 17 percent of Australians on board.

By Stilgherrian for The Full Tilt | August 19, 2016 -- 04:59 GMT (14:59 AEST) | Topic: Security

" My Health Record (MyHR) is not yet a f*** up because hardly anybody's using it, [but] it's a f*** up in terms of how much money the government has spent, and how little they've got for that expenditure," Dr Bernard Robertson-Dunn, who chairs the health committee of the Australian Privacy Foundation (APF), said.

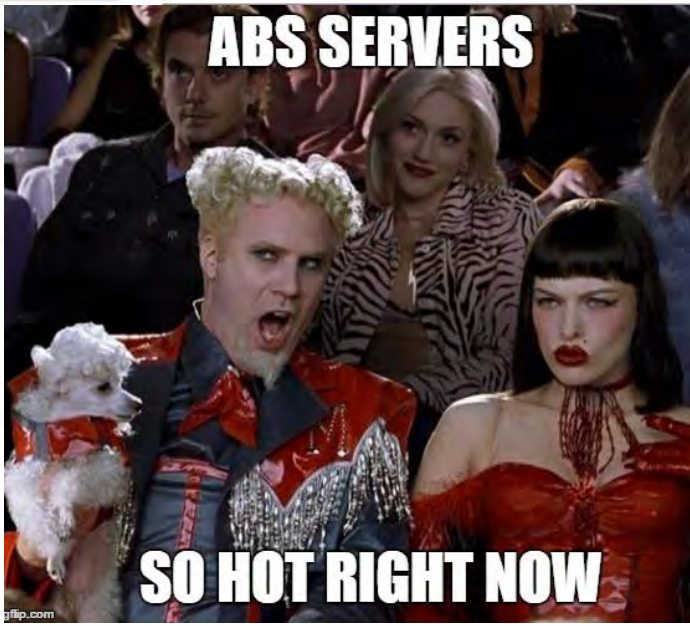
"It's cost AU$2 billion so far, it's costing over AU$400 million every year, but the government has never told us how it has improved health care or reduced health costs. All it is doing is putting patient data at risk."

Blame all around as the Bureau of Statistics deflects criticism of Census 2016

Peter Martin

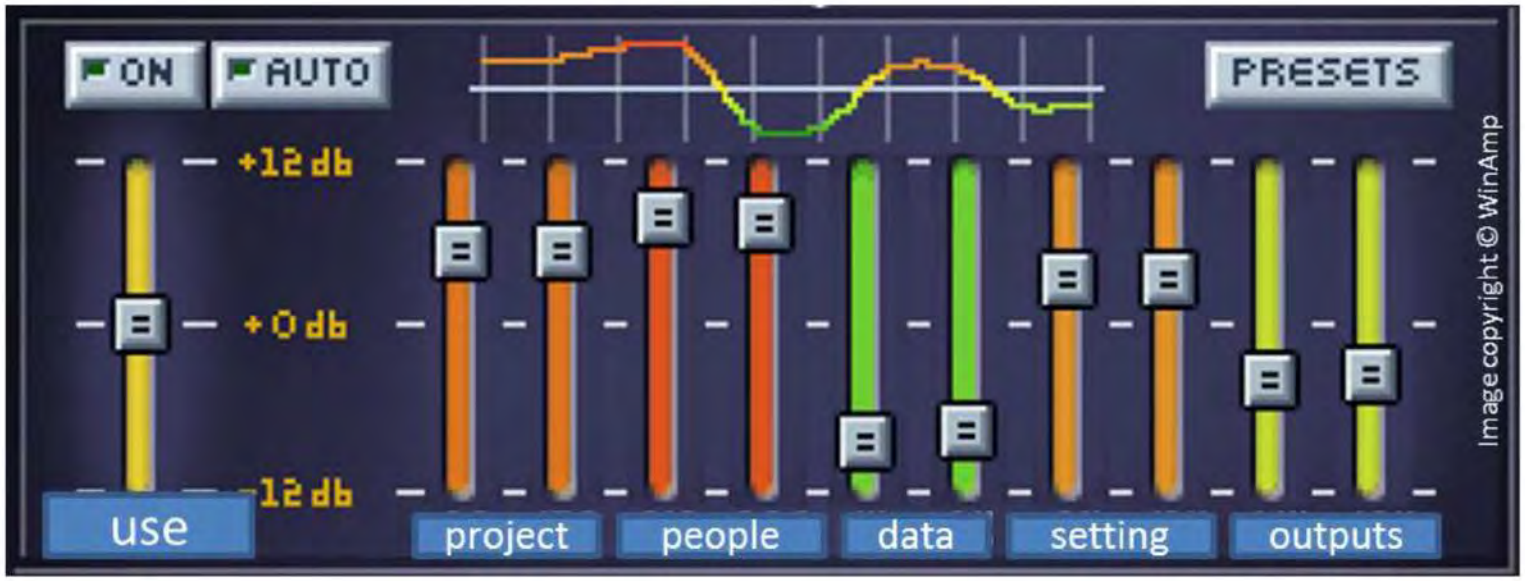
The Australian Bureau of Statistics has blamed the media for the failure of its census hotline and blamed an overseas denial-of-service attack for the failure of its census website.

In a strongly worded submission to a Senate inquiry, the Bureau also attempts to deflect blame for the overwhelming of its website on to its contractor, IBM.



‘Five safes’ framework

1. Safe Projects — Is this use of the data appropriate?
2. Safe People — Can the researchers be trusted to use it appropriately?
3. Safe Data — Is there a disclosure risk in the data itself?
4. Safe Settings — Does the access facility limit unauthorised use?
5. Safe Outputs — Are the statistical results non-disclosive?

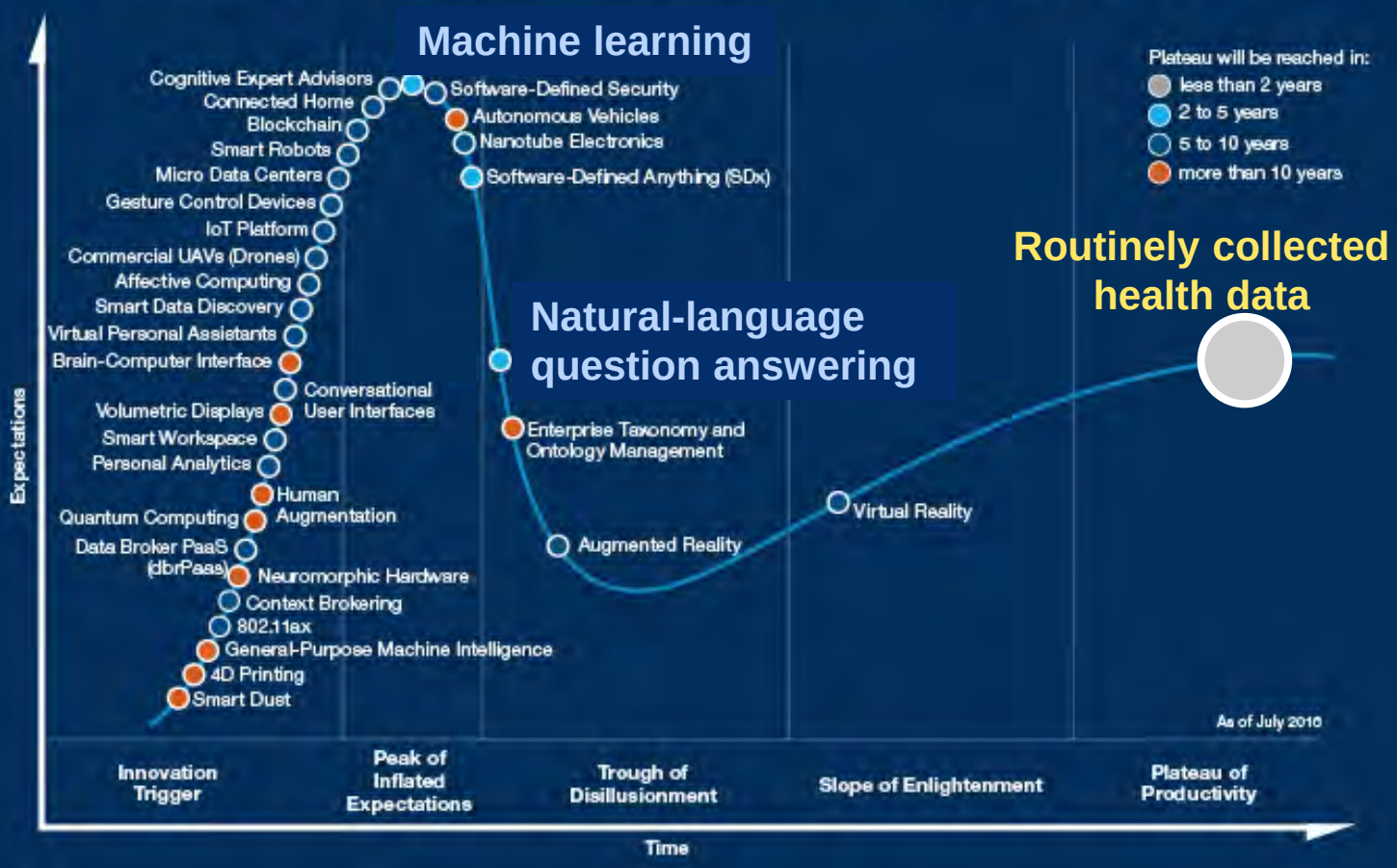


Desai T, et al. *Five Safes: designing data access for research*. Economics working paper series 1601. Bristol: University of the West of England, 2016.

Outline

- What are big data?
- How are they being used in health and medicine?
- Issues and challenges:
  1. Upscaling technology
  2. New analytic paradigms
  3. Workforce shortages
  4. “Open data” vs. privacy protection

Gartner Hype Cycle for Emerging Technologies, 2016



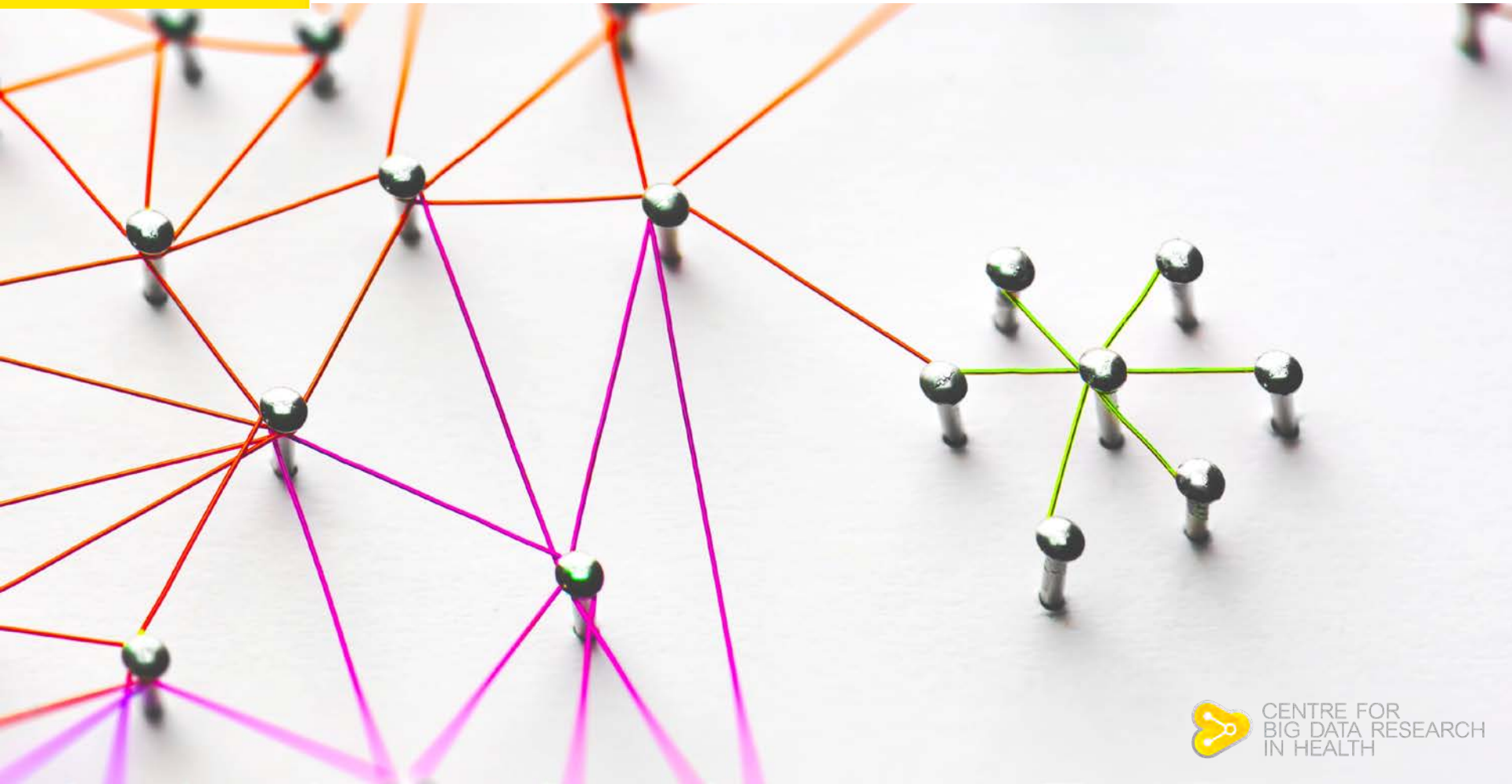
gartner.com/SmarterWithGartner

Gartner.

UNSW SYDNEY | Australia's Global University

Thank you!

l.jorm@unsw.edu.au



CENTRE FOR BIG DATA RESEARCH IN HEALTH