UNSW AUSTRALIA

CENTRE FOR BIG DATA RESEARCH IN HEALTH

Big data transforming health care

Louisa Jorm 17 May 2016

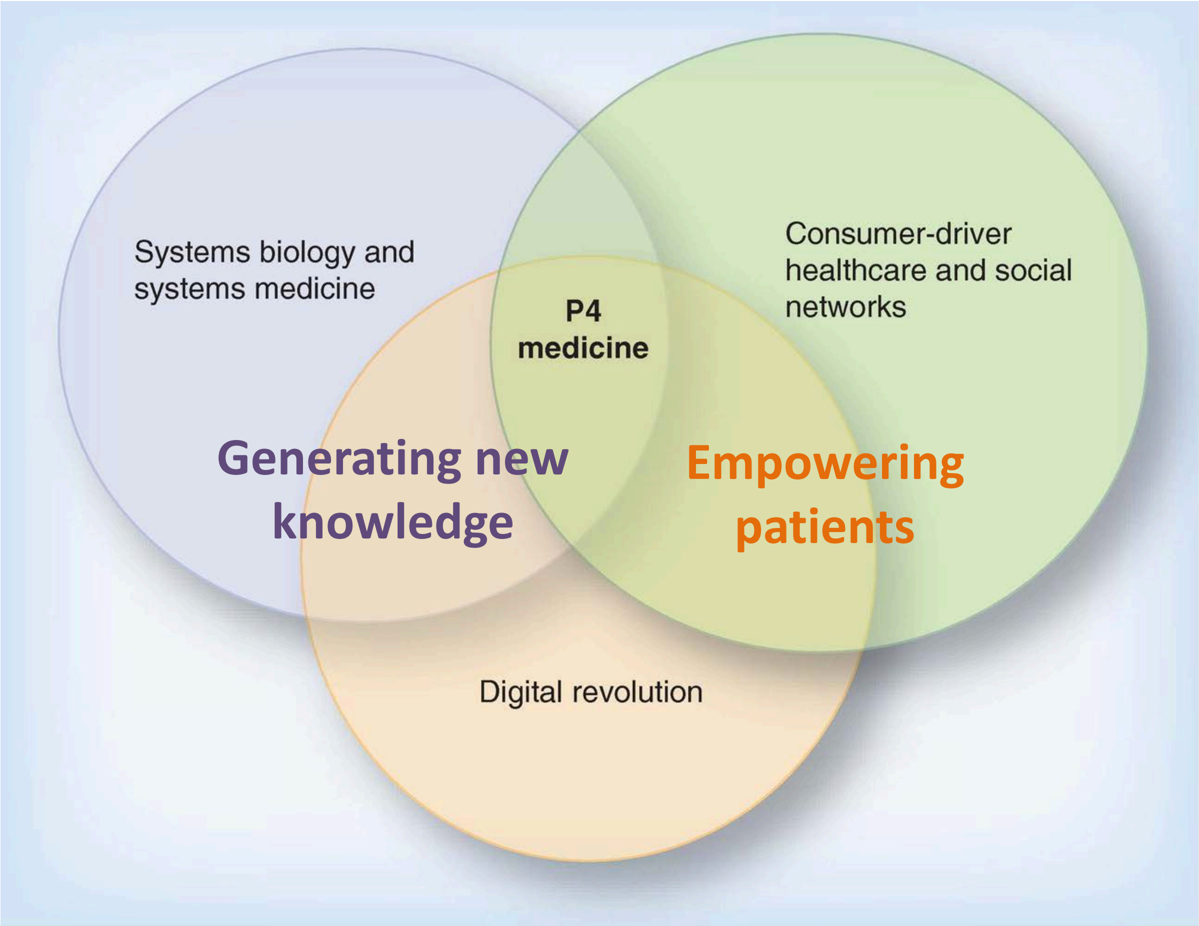
Never Stand Still | Medicine | Centre for Big Data Research in Health

‘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
- Veracity
  - Quality, relevance, predictive value and meaning of data
- Value
  - Worth of information to stakeholders and decision makers

Big data transforming health care: ‘P4 Medicine’

- Predictive
- Preventive
- Personalized
- Participatory

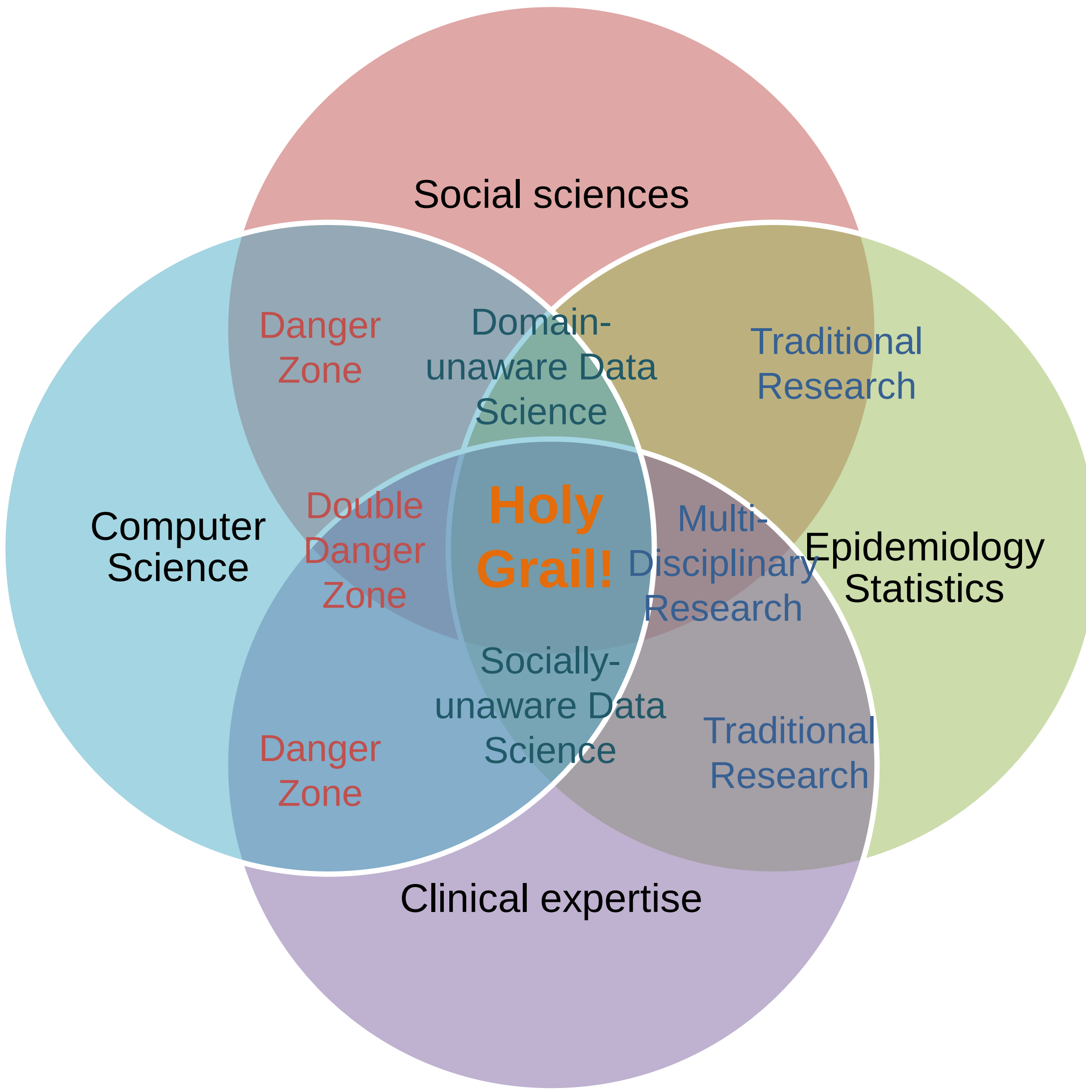


Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. *Future Medicine* 2013;10(6):565-576.

Generating new knowledge

- RCTs and quasi-experimental studies have been the foundation of evidence-based medicine
  - Cost, logistics and ethics preclude using these methods to answer many (most?) clinical questions
- ‘Big data’ offer the potential to create a new observational evidence base
  - Administrative data
  - Electronic health records
  - ‘New’ data sources
- Traditional research methods will not suffice!

The rise of 'data science'



Adapted from: http://www.datascienceassn.org/content/fourth-bubble-data-science-venn-diagram-social-sciences

Using administrative data:
Indigenous Health Outcomes Patient Evaluation (IHOPE)

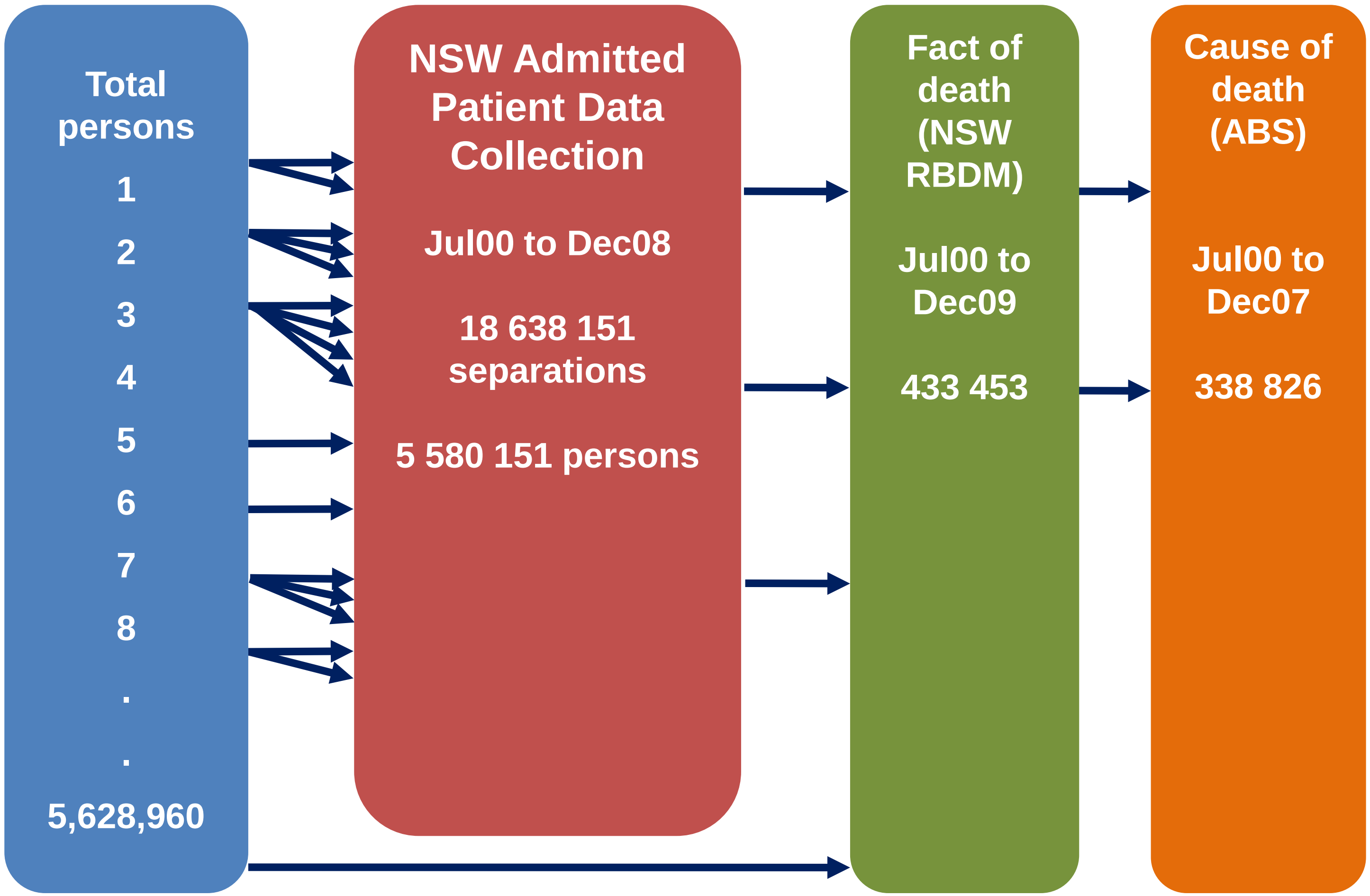
Where are the gaps?

In health outcomes

In treatment and access

What is the contribution of:
Area, SES, Remoteness?
Hospitals, Health system?

IHOPE data


Total persons
1
2
3
4
5
6
7
8
.
.
5,628,960
NSW Admitted Patient Data Collection
Jul00 to Dec08
18 638 151 separations
5 580 151 persons
Fact of death (NSW RBDM)
Jul00 to Dec09
433 453
Cause of death (ABS)
Jul00 to Dec07
338 826

Research focus

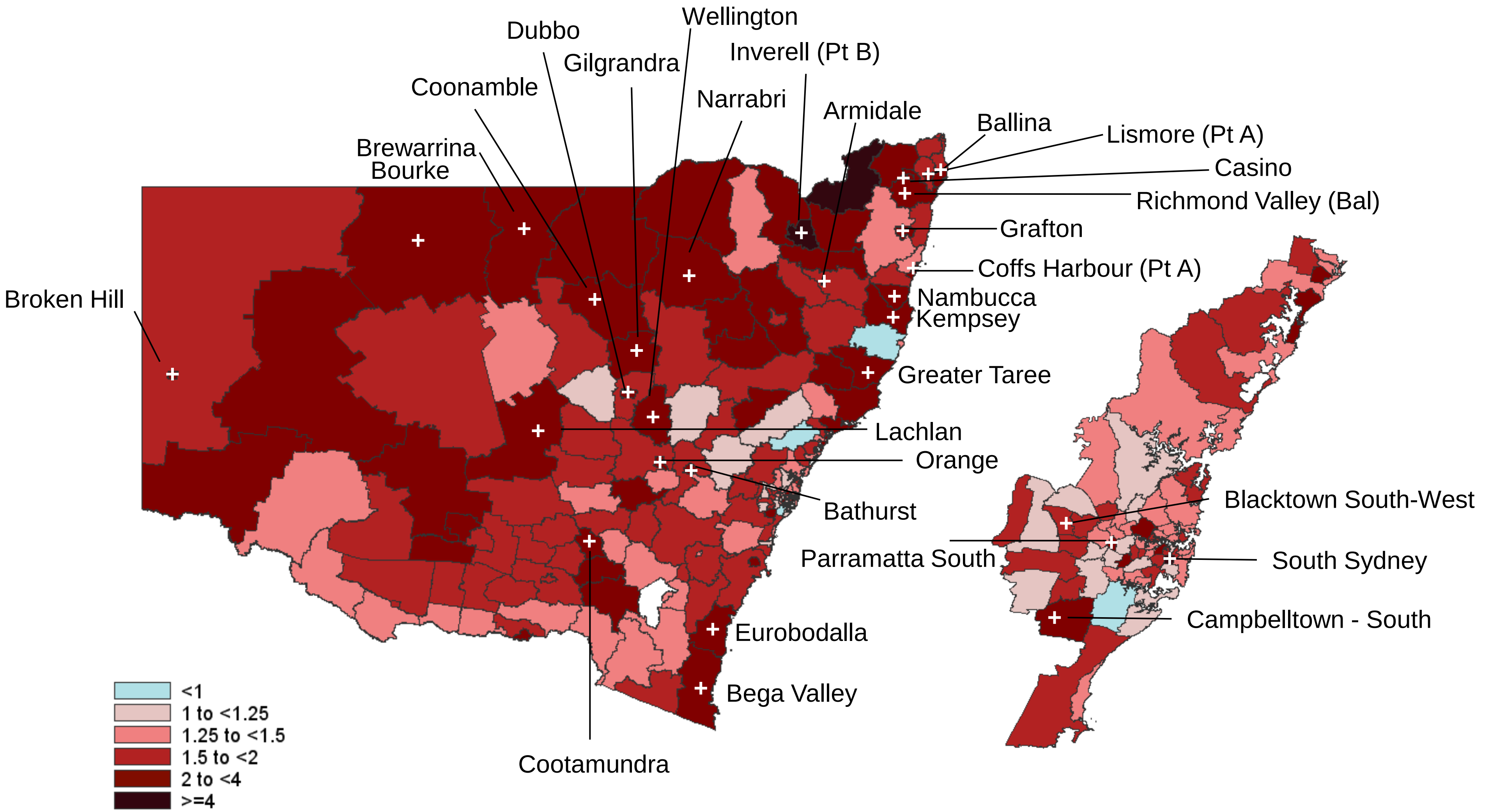
- Acute myocardial infarction
- Road traffic injuries
- Unintentional injuries in children
- Cataract procedures
- Otitis media procedures in children
- Potentially preventable hospitalisations
- Breast conserving surgery
- ……

Multilevel modelling

- Models data that are clustered
  - e.g. live in same neighbourhood, go to the same hospital
  - more similar than those in other areas or hospitals because of shared exposure (often unmeasured)
  - can impact on standard errors and parameter estimates if not taken into account

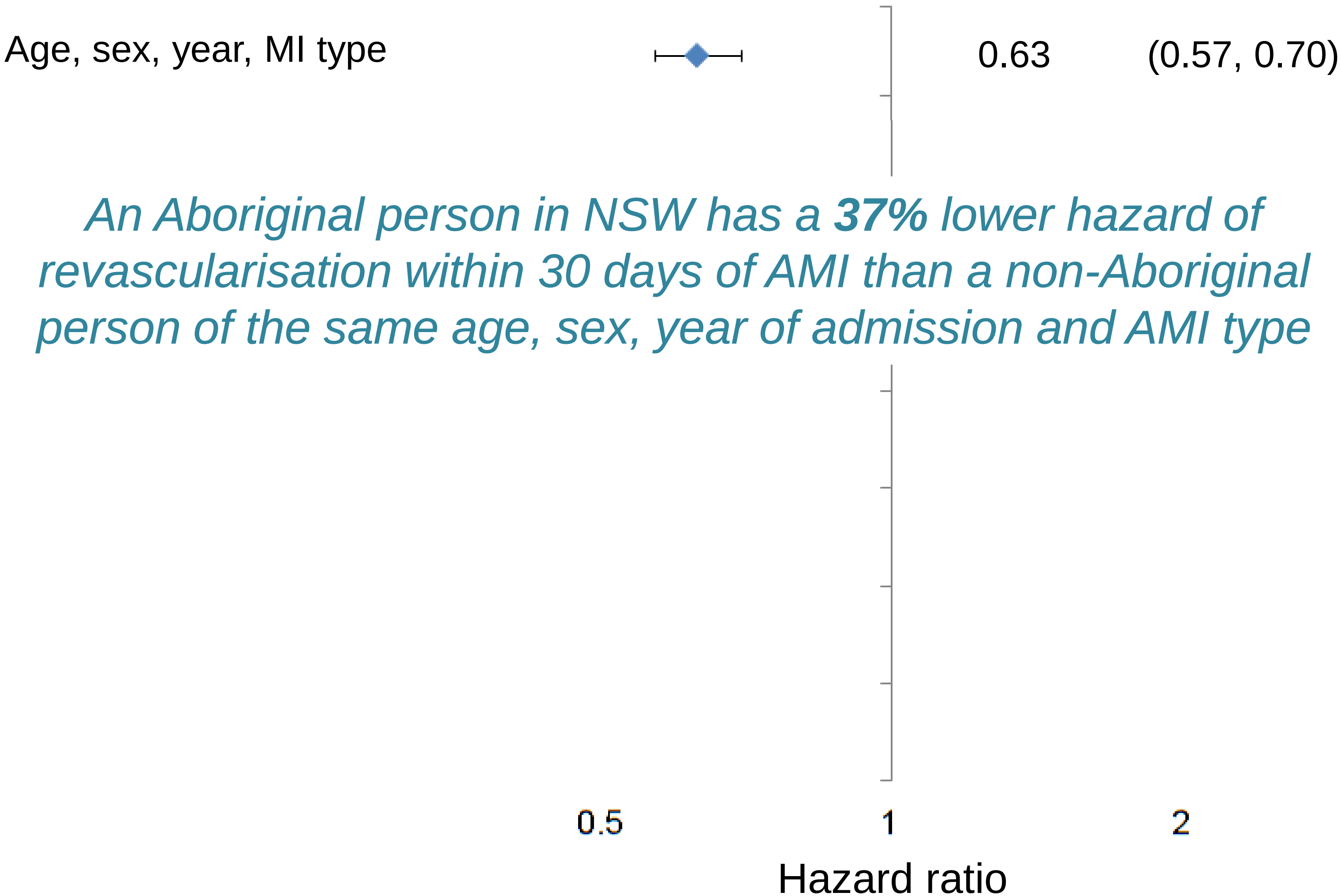
- Particular issue for Aboriginal health research
  - geographic distribution of Aboriginal people in NSW
  - ~40% of Aboriginal people live in major cities compared with ~70% of non-Aboriginal people

AMI: ‘High incidence, high disparity’ areas



Randall DA, Jorm LR, Lujic S, et al. Exploring disparities in acute myocardial infarction events between Aboriginal and non-Aboriginal Australians: roles of age, gender, geography and area-level disadvantage. *Health and Place* 2014; 28: 58-66.

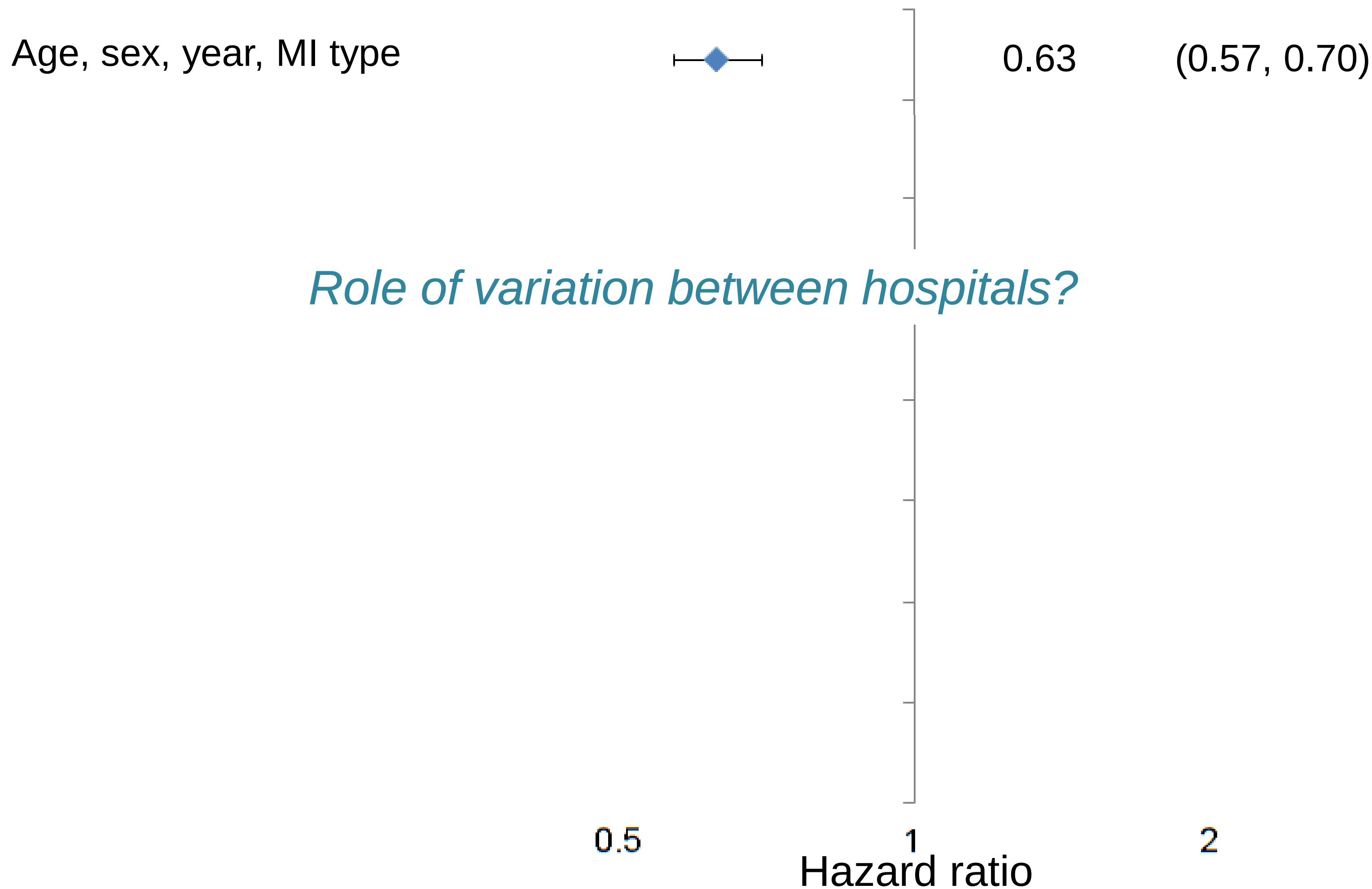
AMI: disparity in revascularisation rates



*An Aboriginal person in NSW has a **37%** lower hazard of revascularisation within 30 days of AMI than a non-Aboriginal person of the same age, sex, year of admission and AMI type*

Randall DA, Jorm LR, Lujic S, et al. Disparities in revascularization rates after acute myocardial infarction between Aboriginal and non-Aboriginal people in Australia. *Circulation*, 2013 Feb 19;127(7):811-9.

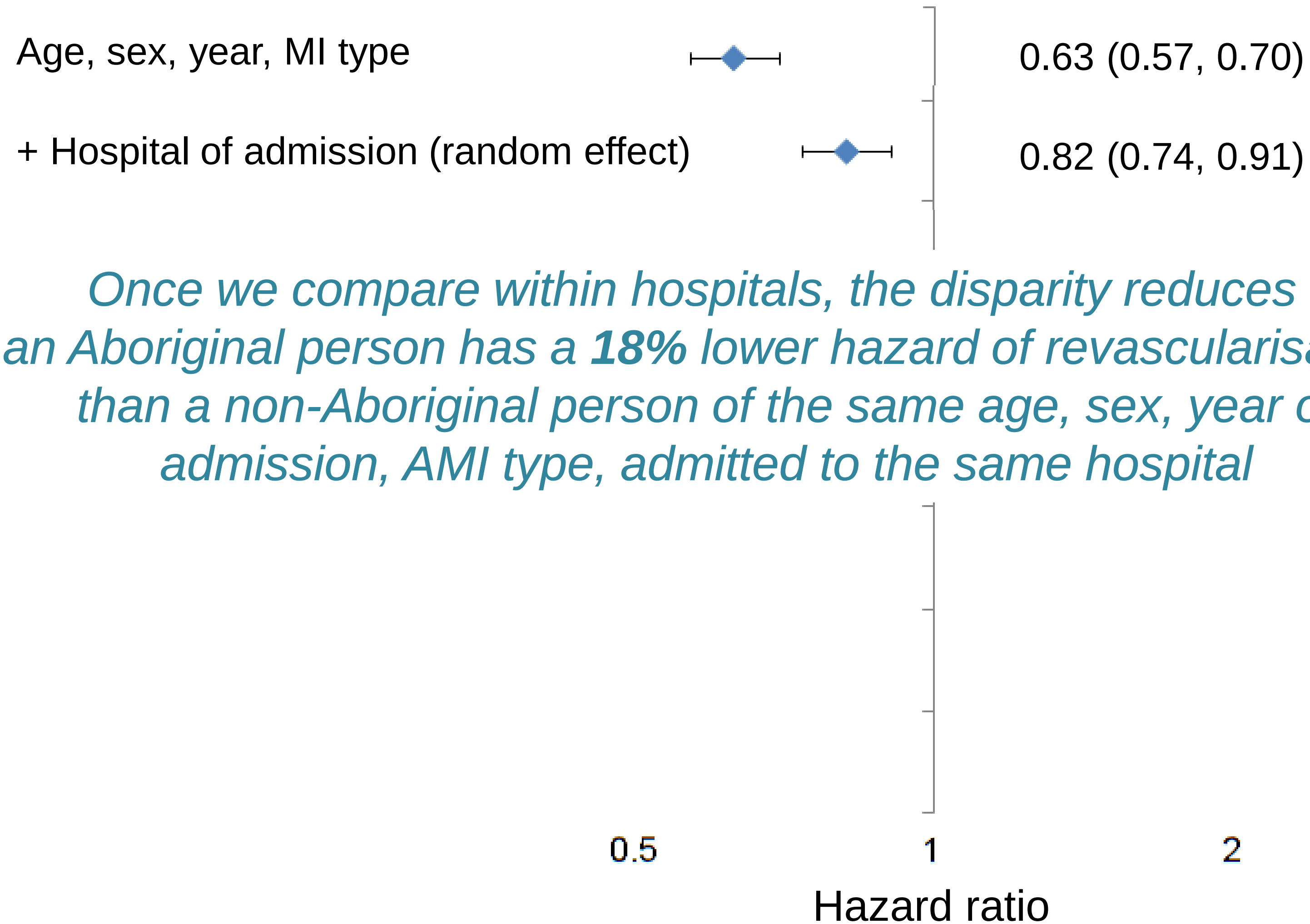
Revascularisation: ‘unpacking’ the gap



Role of variation between hospitals?

Randall DA, Jorm LR, Lujic S, et al. Disparities in revascularization rates after acute myocardial infarction between Aboriginal and non-Aboriginal people in Australia. *Circulation*, 2013 Feb 19;127(7):811-9.

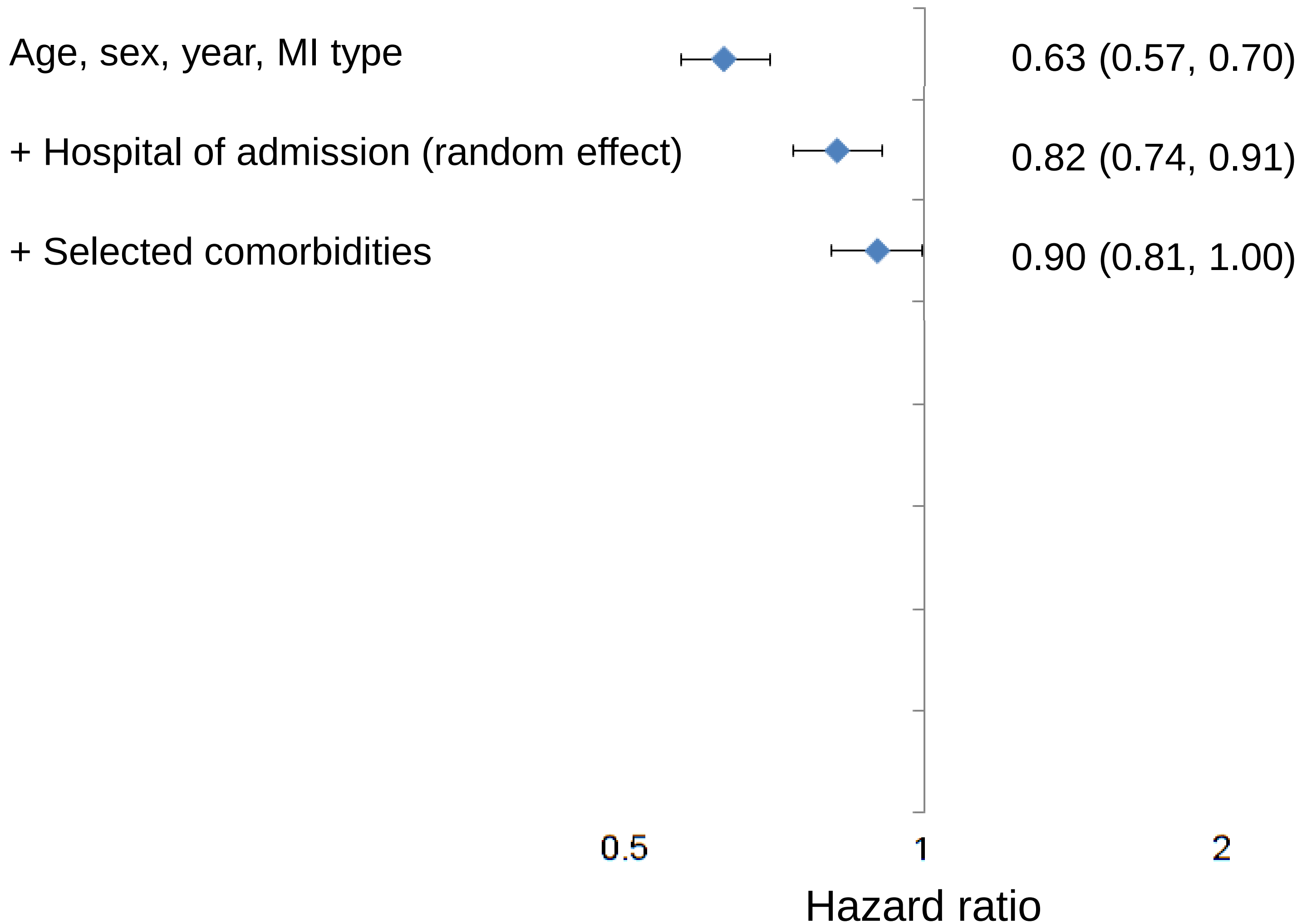
Revascularisation: ‘unpacking’ the gap



*Once we compare within hospitals, the disparity reduces - an Aboriginal person has a **18%** lower hazard of revascularisation than a non-Aboriginal person of the same age, sex, year of admission, AMI type, admitted to the same hospital*

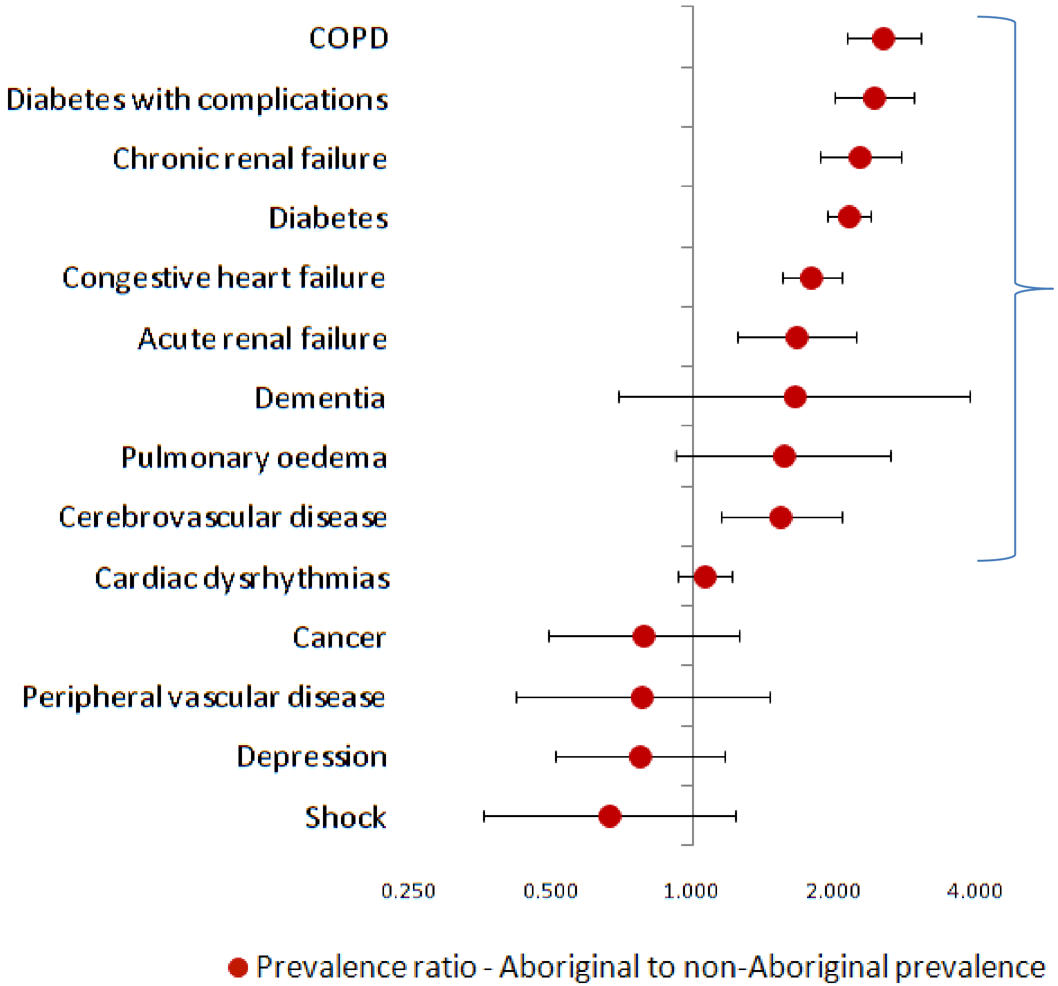
Randall DA, Jorm LR, Lujic S, et al. Disparities in revascularization rates after acute myocardial infarction between Aboriginal and non-Aboriginal people in Australia. *Circulation*, 2013 Feb 19;127(7):811-9.

Revascularisation: ‘unpacking’ the gap



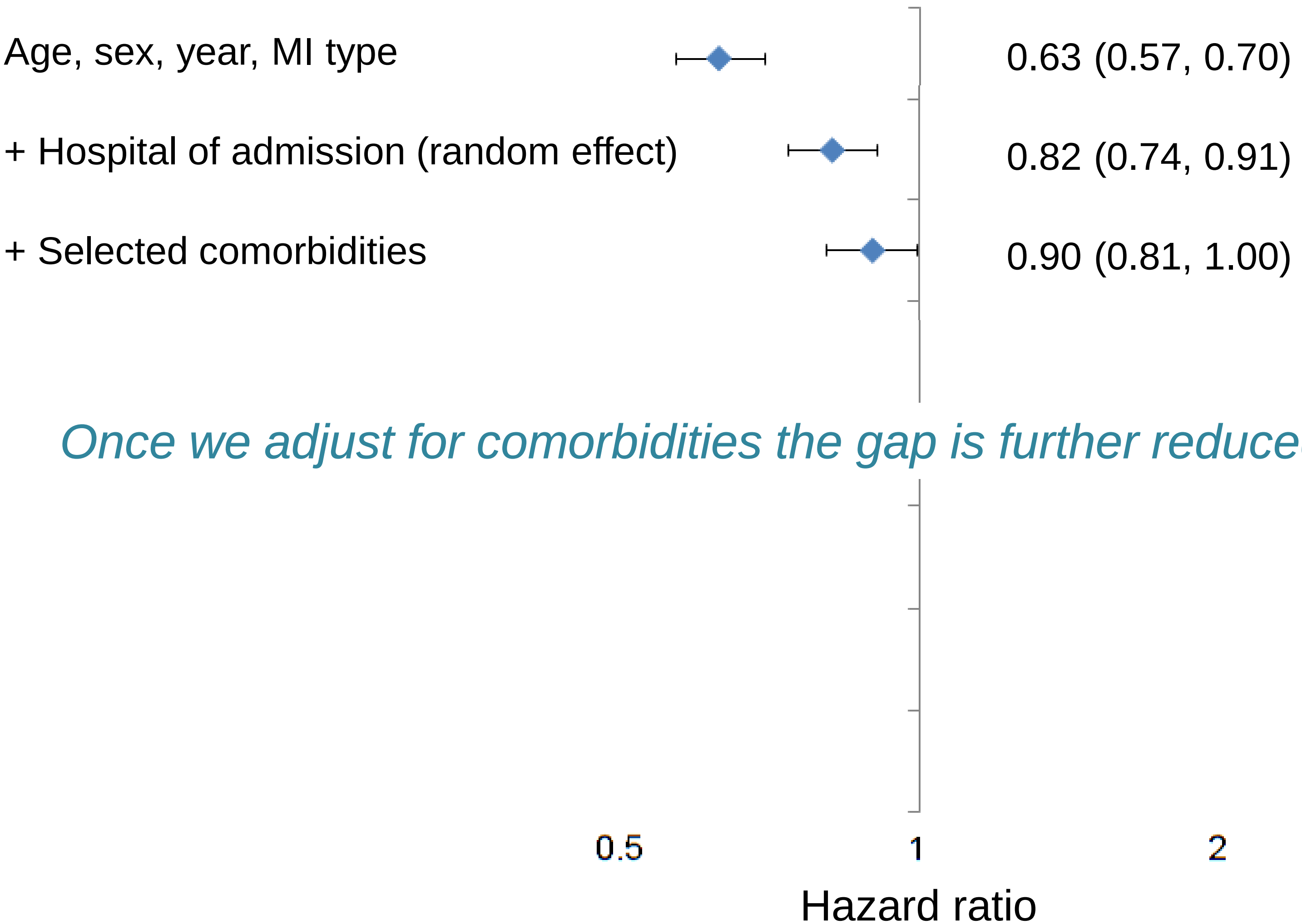
Randall DA, Jorm LR, Lujic S, et al. Disparities in revascularization rates after acute myocardial infarction between Aboriginal and non-Aboriginal people in Australia. *Circulation*, 2013 Feb 19;127(7):811-9.

Comorbidity burden on admission



Aboriginal people have higher rates of these conditions recorded in hospital data than non-Aboriginal people

Revascularisation: ‘unpacking’ the gap

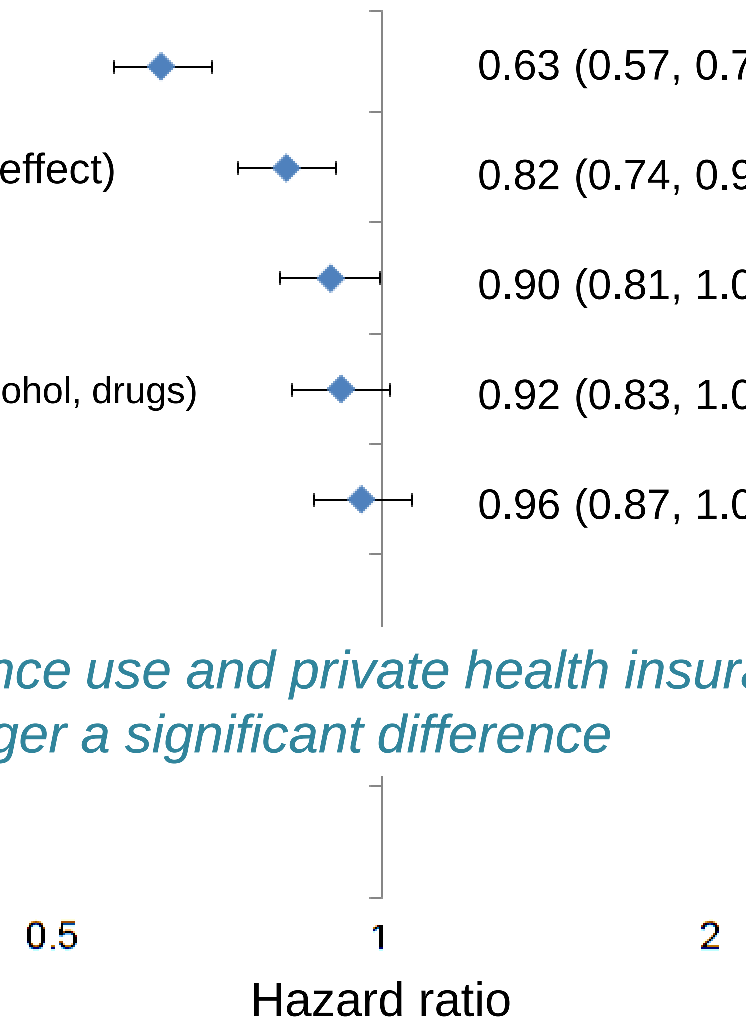


Once we adjust for comorbidities the gap is further reduced

Randall DA, Jorm LR, Lujic S, et al. Disparities in revascularization rates after acute myocardial infarction between Aboriginal and non-Aboriginal people in Australia. *Circulation*, 2013 Feb 19;127(7):811-9.

Revascularisation: ‘unpacking’ the gap

Age, sex, year, MI type	0.63 (0.57, 0.70)
+ Hospital of admission (random effect)	0.82 (0.74, 0.91)
+ Selected comorbidities	0.90 (0.81, 1.00)
+ Substance use (incl. smoking, alcohol, drugs)	0.92 (0.83, 1.02)
+ Private health insurance	0.96 (0.87, 1.07)



After adjusting for substance use and private health insurance, there is no longer a significant difference

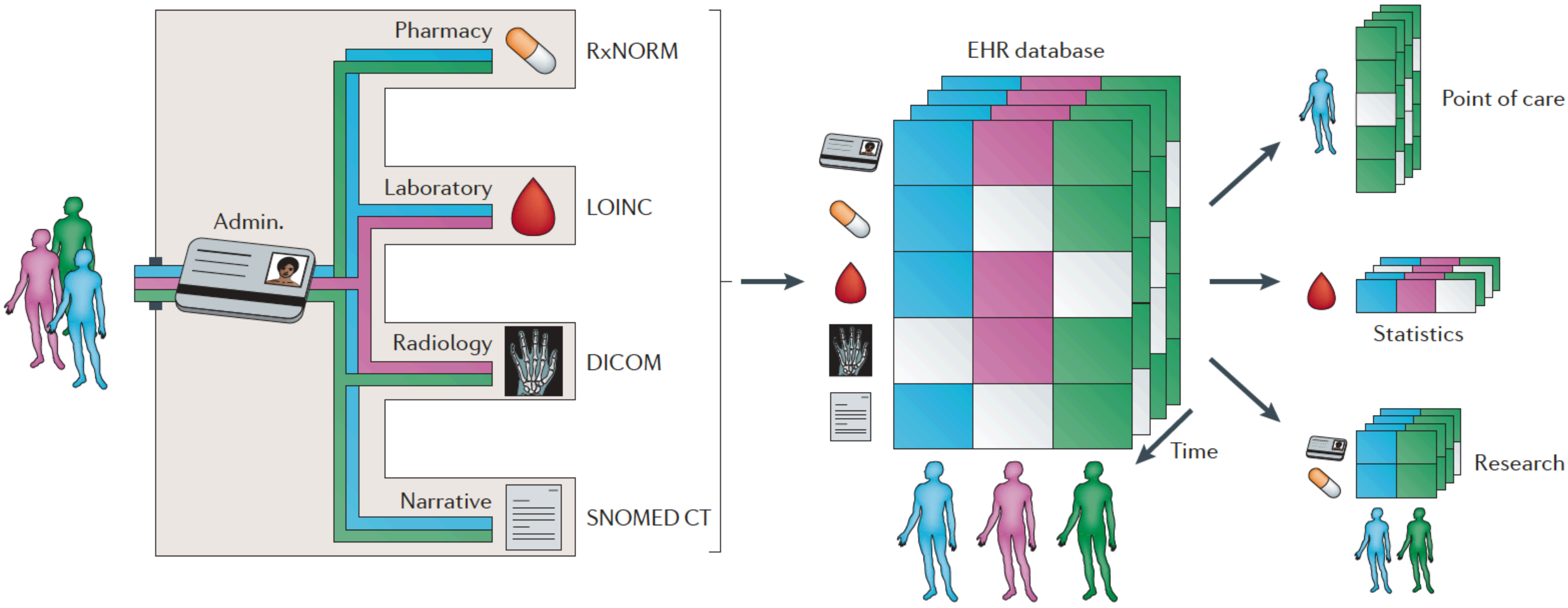
Randall DA, Jorm LR, Lujic S, et al. Disparities in revascularization rates after acute myocardial infarction between Aboriginal and non-Aboriginal people in Australia. *Circulation*, 2013 Feb 19;127(7):811-9.

IHOPE AMI: Summary

Where are the gaps?

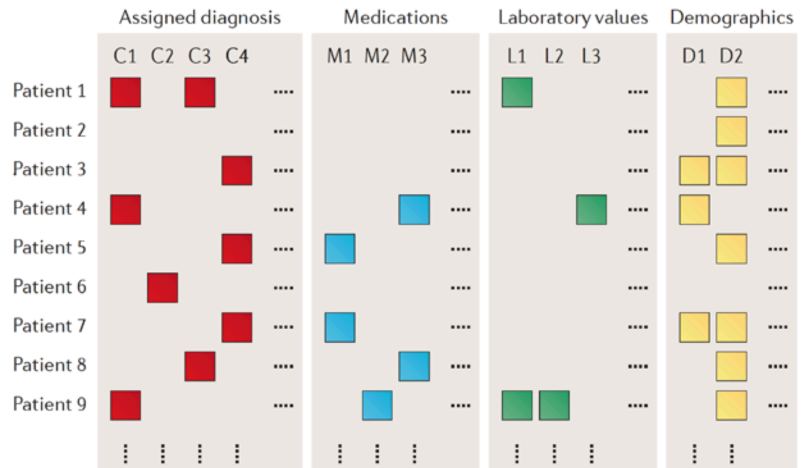
Age at first heart attack	Aboriginal people on average 12 years younger at first AMI Greater disparity in young and women	Importance of prevention and management of early heart disease symptoms
Cardiac procedures	Aboriginal people admitted with AMI less likely to get revascularisation	Related to hospital of admission and higher rate of comorbidities such as diabetes and renal failure
Mortality after AMI	No difference in 30-day mortality after AMI, but Aboriginal people more likely to die within 1 year	Importance of follow-up care and managing multimorbidity. Possible link to lower procedure rates?

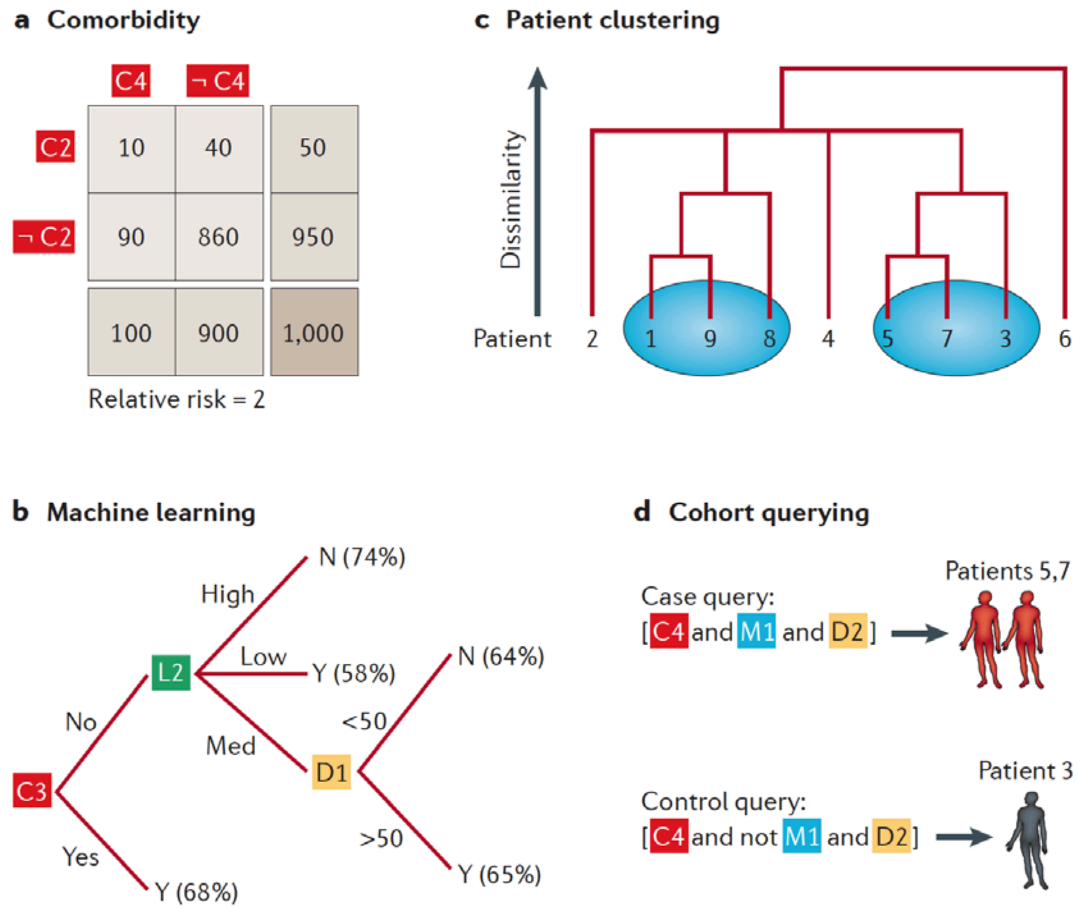
Using EHRs



Jensen PB1, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012 May 2;13(6):395-405.

Analysing EHR data

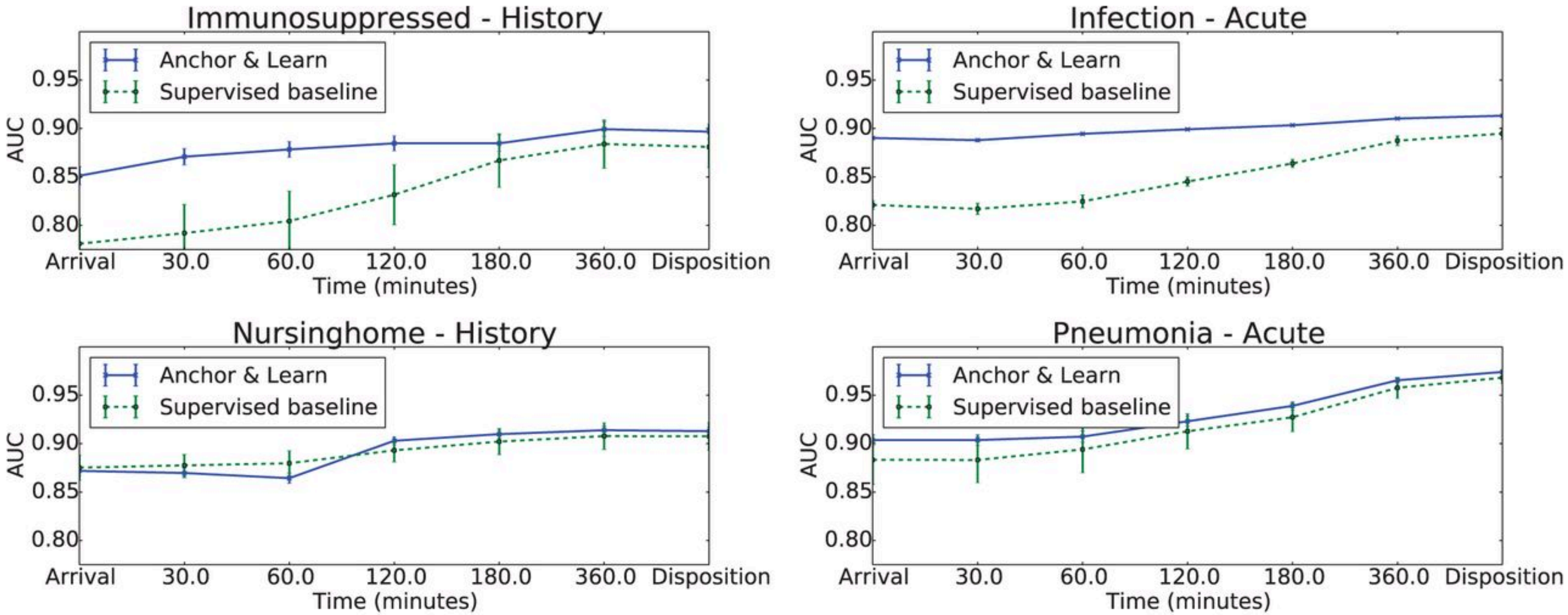




Jensen PB1, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012 May 2;13(6):395-405.

Machine learning

- Data-driven approaches that discover statistical patterns in multivariate data sets
- Starting point is a data set of training examples
- Supervised training methods derives a model from a set of ‘labelled’ examples
  - e.g. naive Bayes, artificial neural networks, support vector machines, random forests
- Unsupervised methods take an unlabelled data set and find groups sharing similar features
  - e.g. self-organizing maps and clustering algorithms
- Data from EHR systems are challenging
  - have many dimensions but are sparse
    - many features describe patients but most of them are typically absent for any given patient
  - heterogeneous, encompassing quantitative data, categorical data and text
  - subject to random errors and systematic biases



We developed a phenotype library that uses both structured and unstructured data from the EMR to represent patients for real-time clinical decision support.... Learning with anchors presents a method of efficiently learning statistically driven phenotypes with minimal manual intervention

Halpern Y, Horng S, Choi Y, Sontag D. Electronic medical record phenotyping using the anchor and learn framework. *JAMIA* 22 April 2016. DOI http://dx.doi.org/10.1093/jamia/ocw011

Natural language processing

Boundary detection	Splits text into individual sentences
Tokenization	Splits text into individual words (with rules for handling e.g. dates)
Normalization	Normalizes e.g. case, inflection or spelling variants
Part-of-speech tagging	Assigns part-of-speech tags to each word (e.g. NN for noun)
Shallow parsing	Identifies syntactic units, most importantly noun phrases (NPs)
Entity recognition	Maps NPs to controlled vocabularies, accounting for negating (e.g. 'no', 'never')

Disease or disorder UMLS ID: C0028754 Status: family history Negated: no	Disease or disorder UMLS ID: C0010054 Status: family history Negated: yes	Anatomy UMLS ID:C0205042

Jensen PB1, *Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care.* Nat Rev Genet. 2012 May 2;13(6):395-405.

From: **Automated Identification of Postoperative Complications Within an Electronic Medical Record Using Natural Language Processing**

JAMA. 2011;306(8):848-855. doi:10.1001/jama.2011.1204

Table 3. Comparison of a Natural Language Processing–Based Approach to the Agency for Healthcare Research and Quality Patient Safety Indicators in Identifying Postoperative Complications

Occurrence	Event Rate	Test Characteristic	Natural Language Processing	Patient Safety Indicator	*P* Value
Acute renal failure	39/1924	Sensitivity	0.82 (0.67-0.91)	0.38 (0.25-0.54)	<.001
		Specificity	0.94 (0.93-0.95)	1.00 (1.00-1.00)	<.001
Pulmonary embolism/ deep vein thrombosis	46/2327	Sensitivity	0.59 (0.44-0.72)	0.46 (0.32-0.60)	.30
		Specificity	0.91 (0.90-0.92)	0.98 (0.98-0.99)	<.001
Sepsis	61/866	Sensitivity	0.89 (0.78-0.94)	0.34 (0.24-0.47)	<.001
		Specificity	0.94 (0.93-0.96)	0.99 (0.98-0.99)	<.001
Pneumonia	222/1405	Sensitivity	0.64 (0.58-0.70)	0.05 (0.03-0.09)	<.001
		Specificity	0.95 (0.94-0.96)	0.99 (0.99-1.00)	<.001
Myocardial infarction	35/1822	Sensitivity	0.91 (0.78-0.97)	0.89 (0.74-0.96)	.67
		Specificity	0.95 (0.94-0.96)	0.99 (0.98-0.99)	<.001

Among patients undergoing inpatient surgical procedures at VA medical centers, natural language processing analysis of EMRs to identify postoperative complications had higher sensitivity and lower specificity compared with patient safety indicators based on discharge coding

Copyright © 2016 American Medical Association. All rights reserved.

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

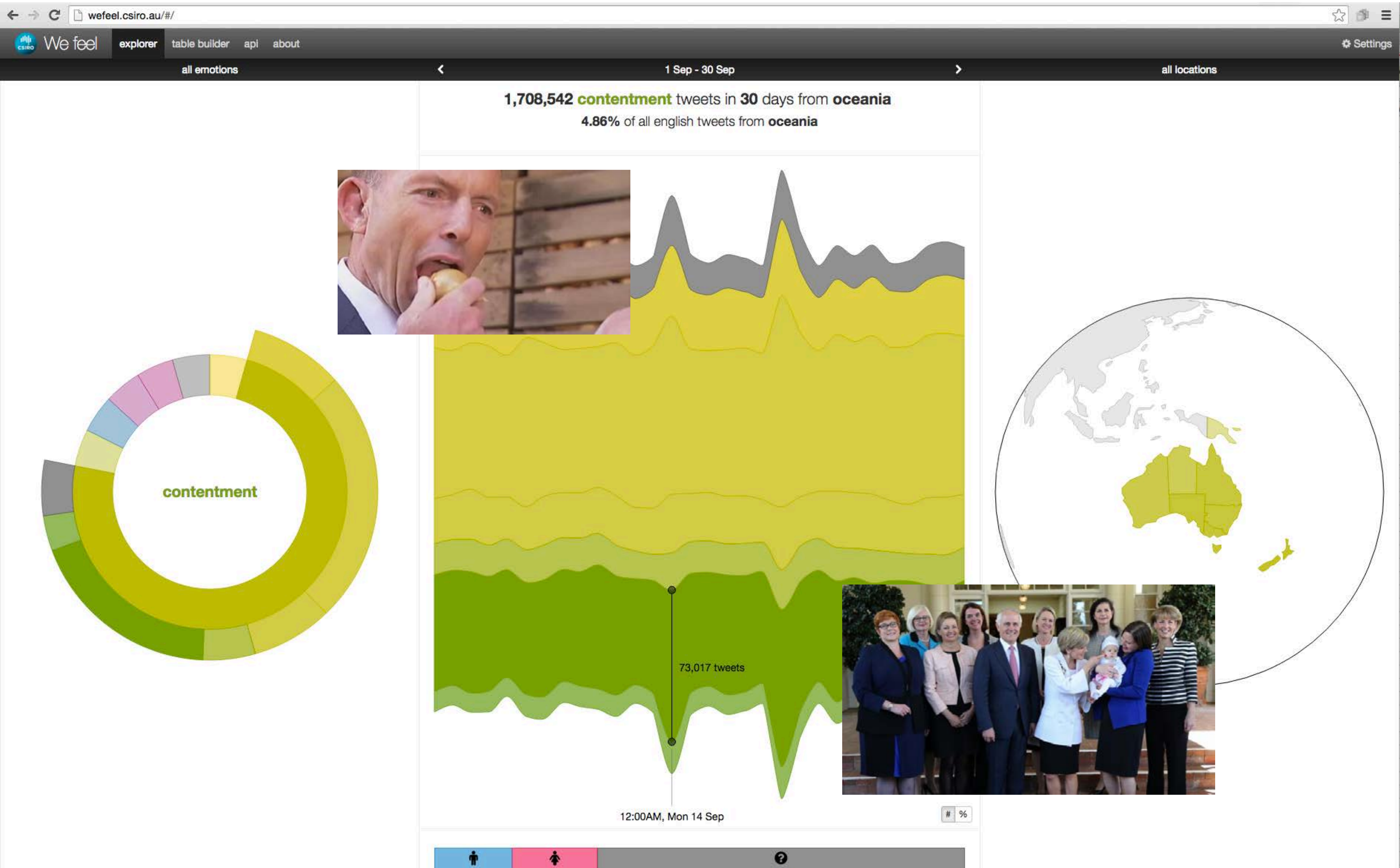
By **Hal Hodson**

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.

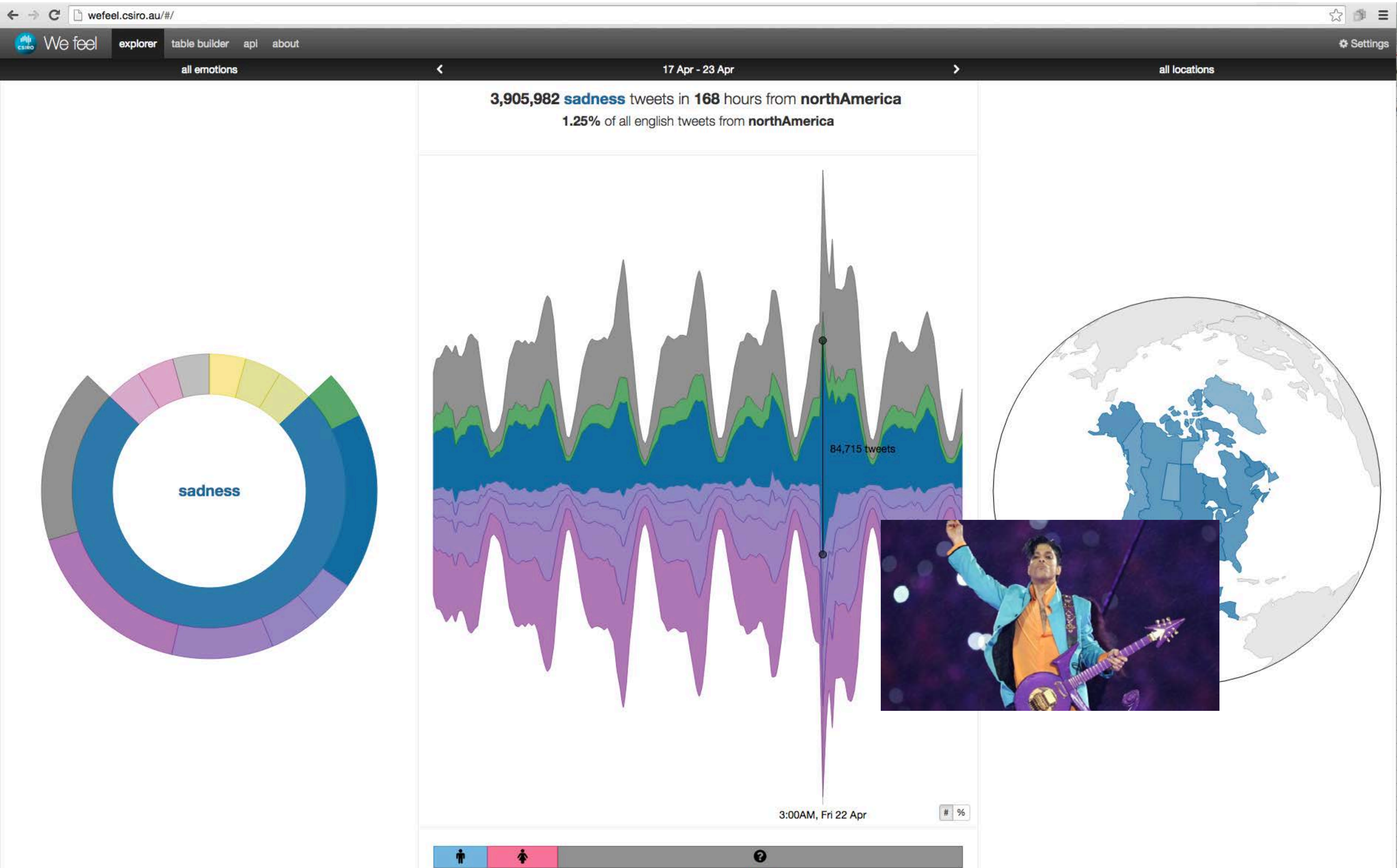
"The agreement gives DeepMind access to a wide range of healthcare data on the 1.6 million patients who pass through three London hospitals run by the Royal Free NHS Trust – Barnet, Chase Farm and the Royal Free – each year. The agreement also includes access to patient data from the last five years."

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

Using ‘new’ data sources: Twitter

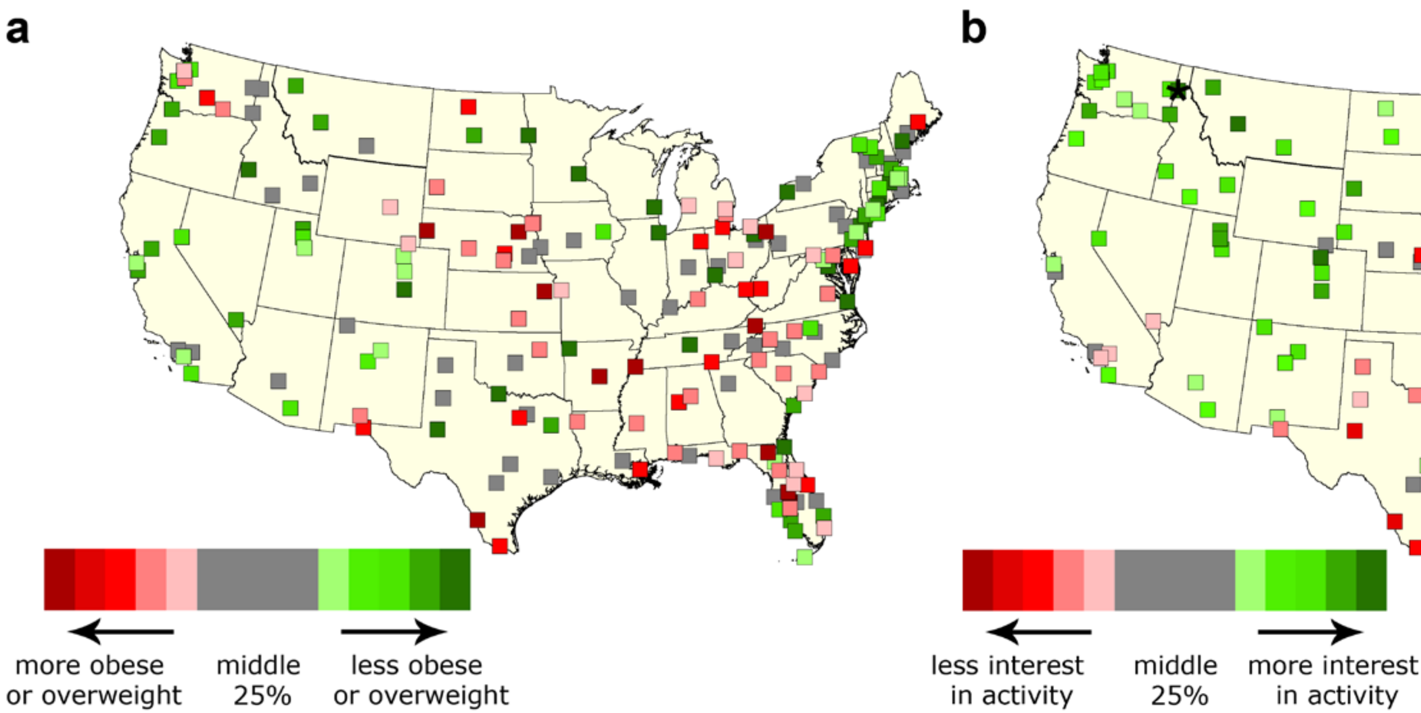


Using ‘new’ data sources: Twitter



Using 'new' data sources: facebook

Figure 3. Prevalence of activity-related interests and obesity in the USA.



Chunara R, Bouton L, Ayers JW, Brownstein JS (2013) Assessing the Online Social Environment for Surveillance of Obesity Prevalence. *PLoS ONE* 8(4): e61373. doi:10.1371/journal.pone.0061373

Empowering patients

QUANTIFIED SELF

YOU ARE JUST A NUMBER

Can you make yourself healthier and happier by logging every snore, step and mood swing? As a Californian trend for obsessional data-tracking makes its way over here, **Tim Chester** covers his body in gadgets to find out if self-knowledge is power. Photograph by Paul Stuart

Today I have climbed the equivalent of a tall giraffe. Coffee is my most frequent food. On average, I walk 11,726 steps a day, burning 3,089 calories, over 2.4 hours of activity. I sleep for 6 hours and 9 minutes a night. This week, my sleep efficiency is 72% and my food is 77% healthy. My BMI of 23.5 is 14 percentage points below the median for men my age, and my average daily Met score is 1.71, although I have no idea what a Met score is.

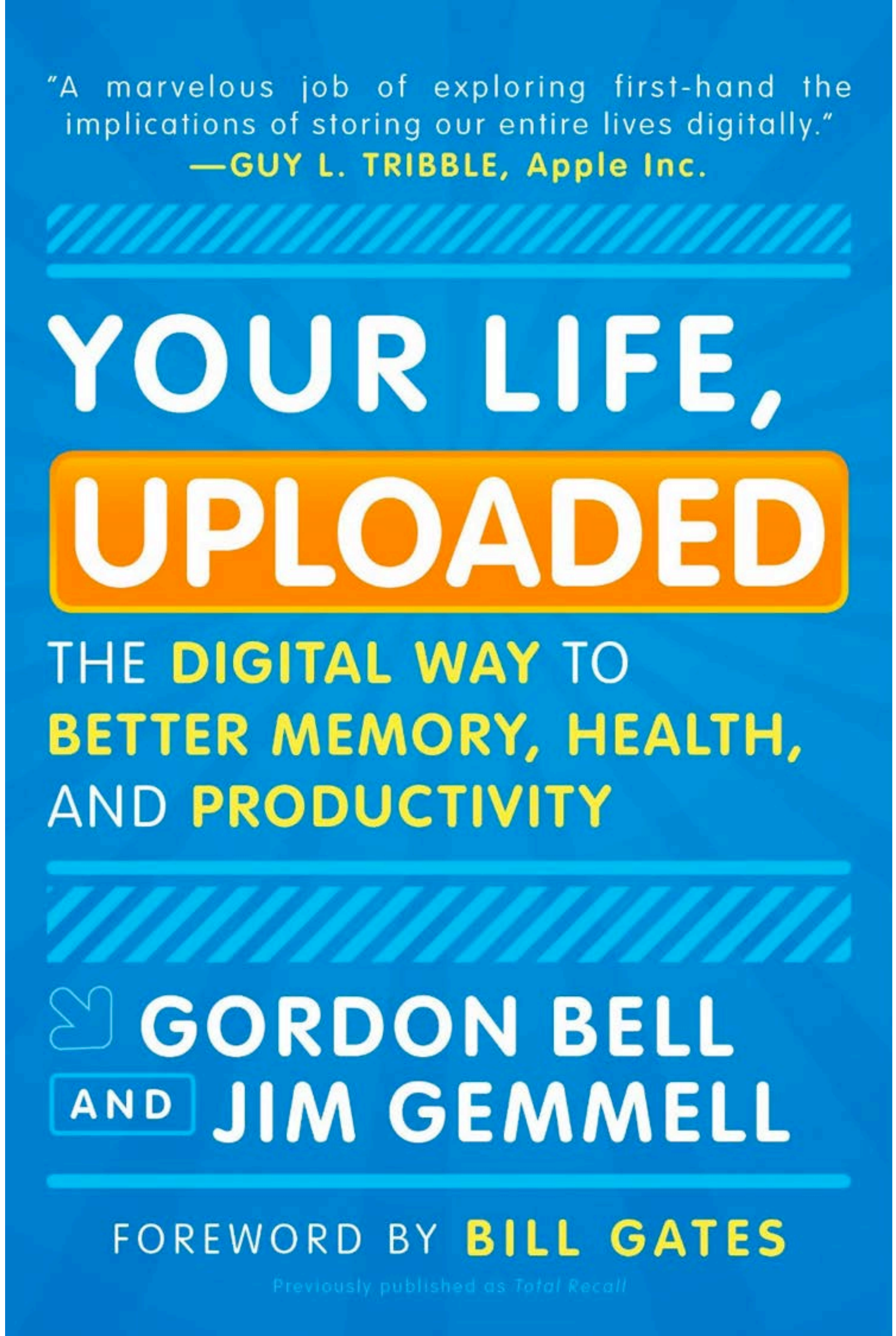
I am, it seems, nothing more than a bundle of numbers and milestones, spurred on by LEDs and chided by pop-up messages. A wireless accessory for the iPhone; perhaps its most sophisticated yet.

My arms are covered in bands, my pockets augmented with accelerometers, my eyes numb from all the charts, my heart pumping to the beat of a heart-rate monitor and forcing its ventricles to keep up with the national average. My head is about to implode from all the positive affirmation and gentle nudging, but it's OK because my memories are being saved to my hard drive and my mood swings are earning me *hugs* from strangers.

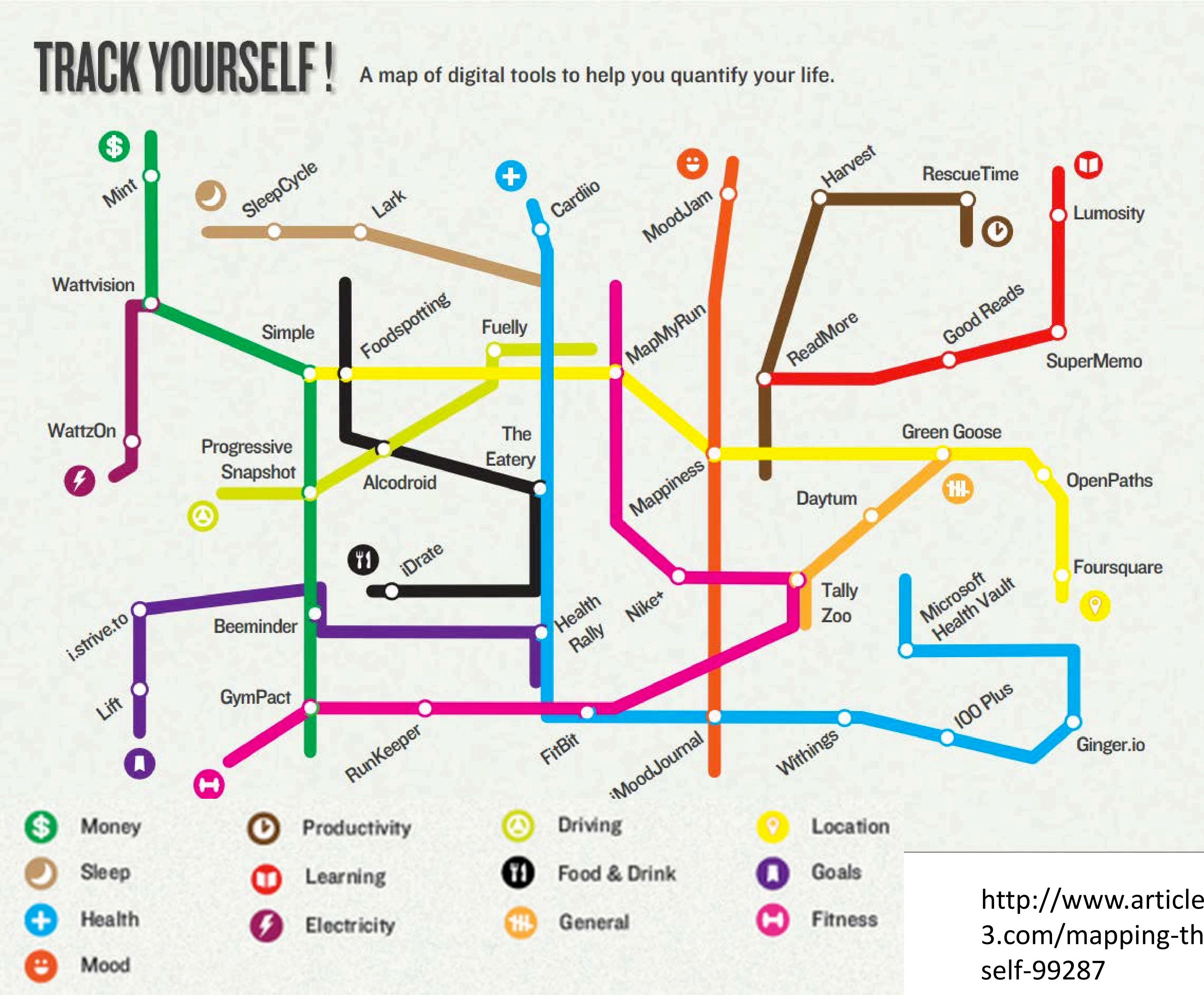
I am producing, analysing and socially sharing personal data. I am becoming fitter, happier, and more productive. I am staying motivated by earning badges. I have become a Quantified Self (QS).

The QS movement that I've temporarily joined began, as these things tend to do, in San Francisco's Bay Area in 2007. Two Wired magazine editors, Gary Wolf and Kevin

http://timchester.com/wp-content/uploads/2013/08/QS-spread.jpg



http://totalrecallbook.com/



http://www.article-3.com/mapping-the-quantified-self-99287

HealthVault

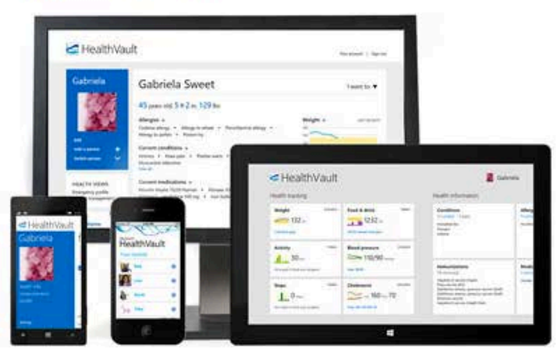
Explore HealthVault   Discover apps & devices   Sign up or sign in

What is HealthVault?

Microsoft HealthVault is a trusted place for people to gather, store, use, and share health information online. Learn more

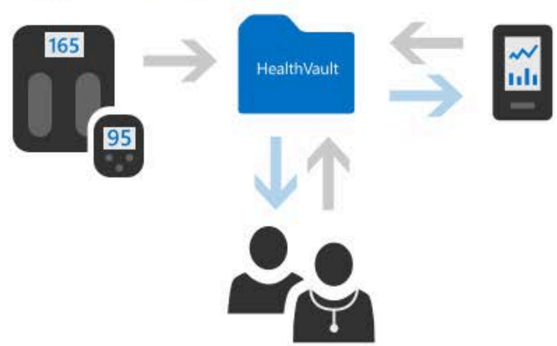
- Organize your family's health information.
- Be better prepared for doctor visits and unexpected emergencies.
- Create a more complete picture of your health, with you at the center.
- Achieve your fitness goals.

Connect anywhere



Connect from the web, Windows, Windows Phone, iPhone, and more.

Connect your health data



Learn how HealthVault works with connected apps and devices and helps you share information with people you trust.

How it works

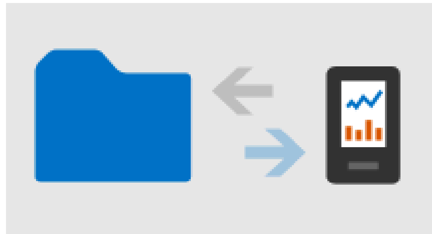
HealthVault lets you gather, store, use, and share health information for you and your family, putting you in control of your health information.

Your health data

Microsoft HealthVault gives you one place to access all of your health information online.

There are many ways to add information and connect with data from your healthcare providers.

If you're managing the health of a child, parent, or other family member, you can add records for them too.



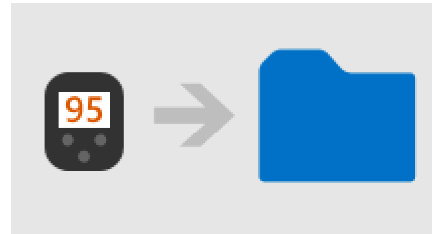
Health apps

HealthVault-connected apps are websites, computer software, and mobile apps that can help you get more out of—and put more into—HealthVault.

You can choose apps to help you stay motivated, analyze trends, and receive education and recommendations to keep you at your best.

That's the great thing about HealthVault: you only need to gather your information once and then you can use it in all kinds of ways.

Browse the App Directory

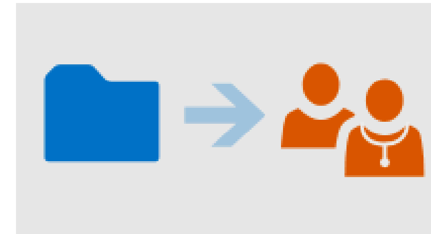


Personal health devices

A growing list of devices such as pedometers, blood pressure monitors, blood glucose monitors, and even weight scales work with HealthVault.

The best part is that you don't have to enter anything by hand, just upload your data directly to HealthVault from compatible devices.

Browse the Device Directory



Sharing

You can share any part of your health record with anyone you choose, whenever you like, to make sure everyone's in the loop.

Find out what your DNA says about you and your family.

- Learn what percent of your DNA is from populations around the world
- Contact your DNA relatives across continents or across the street
- Build your family tree and enhance your experience with relatives

order now **$149**

Watch Greta and Stacy's story.

Find out how these two women discovered their connection as sisters.

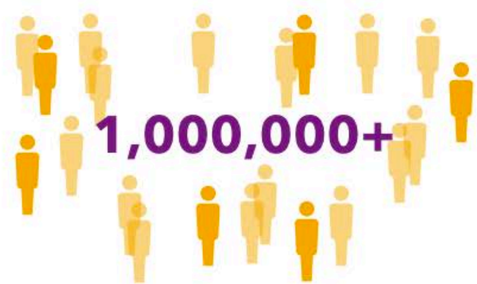
See all stories →

Bring your ancestry to life through your DNA.

Discover your ancestral origins and trace your lineage with a personalized analysis of your DNA.

- Ancestry composition
- DNA relatives
- Neanderthal percentage
- Family tree tool
- Maternal and paternal lineages

order now

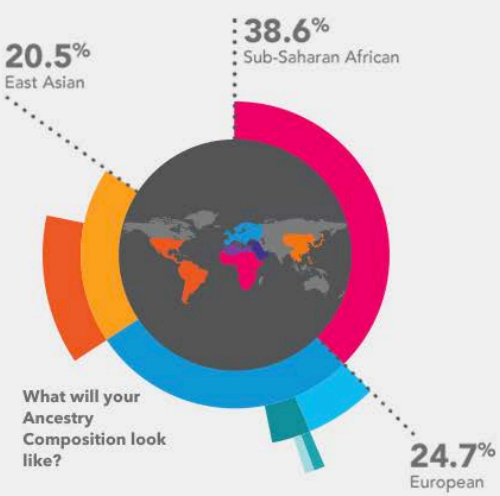


Bigger database. More matches.

With over 1,000,000 people worldwide in our database, you could discover hundreds of matches, more data and more discoveries.

23 pairs of chromosomes. One unique you.

Find out what percent of your DNA comes from populations around the world, ranging from East Asia, Sub-Saharan Africa, Europe, and more. Break European ancestry down into distinct regions such as the British Isles, Scandinavia and Italy. People with mixed ancestry, African Americans, Latinos, and Native Americans will also get a detailed breakdown.



patientslikeme®

Already a member? **Sign in.**

conditions, symptoms, treatments...

Live better, together!™

Making healthcare better for everyone through sharing, support, and research

Join now

(it's free!)

Learn from others

Compare treatments, symptoms and experiences with people like you and take control of your health

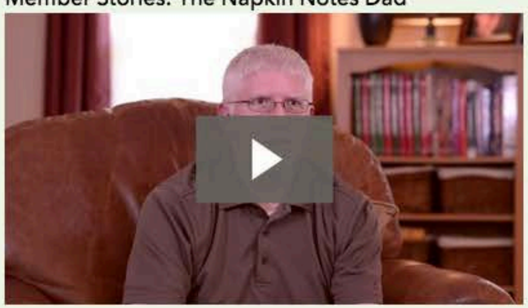
Connect with people like you

Share your experience, give and get support to improve your life and the lives of others

Track your health

Chart your health over time and contribute to research that can advance medicine for all

Member Stories: The Napkin Notes Dad



"Our brother Stephen was living with ALS and we thought, 'there has to be a better way.' There is. By sharing our experiences, we can all contribute new data that can accelerate research and help create better treatments. Our experiences can actually change medicine… for good."

Jamie & Ben Heywood

Co-founders, PatientsLikeMe

News

PatientsLikeMe and AstraZeneca Announce Global Research Collaboration

Our five-year agreement is a major step forward to make patient-centric evidence a cornerstone of scientific discovery and development.

PatientsLikeMe Appoints Ed Godber as First Executive Vice President of Life Sciences Ventures

Ed is responsible for managing PatientsLikeMe's life sciences business and bringing the patient agenda to the forefront of our partners' development and delivery operations.

See all

325,000 members | 2,400+ conditions | 60+ published research studies | 27 million data points about disease

Conditions at PatientsLikeMe

Cancer
Breast, Lung, Liver, Testicular, Prostate, Pancreatic, CLL (Chronic Lymphocytic Leukemia), Non-Hodgkin's Lymphoma, Thyroid

Developmental and Chromosomal
Tay-Sachs, Autism Spectrum, Down Syndrome

Digestive and Intestinal
Crohn's Disease, IBS, Ulcerative Colitis

Endocrine
Diabetes: Type I, Type II, Hypothyroidism, Hyperthyroidism

Eye, Ear, Nose and Throat
Hearing Loss, Glaucoma, Macular Degeneration

Heart, Blood and Circulatory
Coronary Artery Disease, Hypertension, Iron Deficiency Anemia, Raynaud's Syndrome, Congestive Heart Failure, Cardiomyopathy, Aplastic Anemia

Immune, Inflammatory and Infections
Rheumatoid Arthritis, Lupus, HIV, Lyme Disease, AIDS

Kidneys and Urinary
Polycystic Kidney Disease, Chronic Kidney Disease, Interstitial Cystitis

Liver, Pancreas and Gallbladder
Hepatitis C, Pancreatitis, Polycystic Liver Disease

Lungs and Respiratory
Pulmonary Fibrosis, Asthma, COPD, Cystic Fibrosis, Emphysema, Pulmonary Hypertension

Men's Health
Infertility, Erectile Dysfunction, Benign Prostatic Hypertrophy

Mental Health and Behavior
Depression, Bipolar I, Bipolar II, Social Anxiety, ADHD/ADD, Dysthymia, Generalized Anxiety Disorder, Panic Disorder, Eating Disorder, OCD, Phobia, PTSD, Schizophrenia, Drug Addiction, Tobacco Addiction, Alcohol Addiction

Metabolism and Nutrition
Hypercholesterolemia, Hemochromatosis, Obesity

Muscle, Bone and Joint
Fibromyalgia, OA, Osteoporosis, TMJ, Muscular Dystrophy

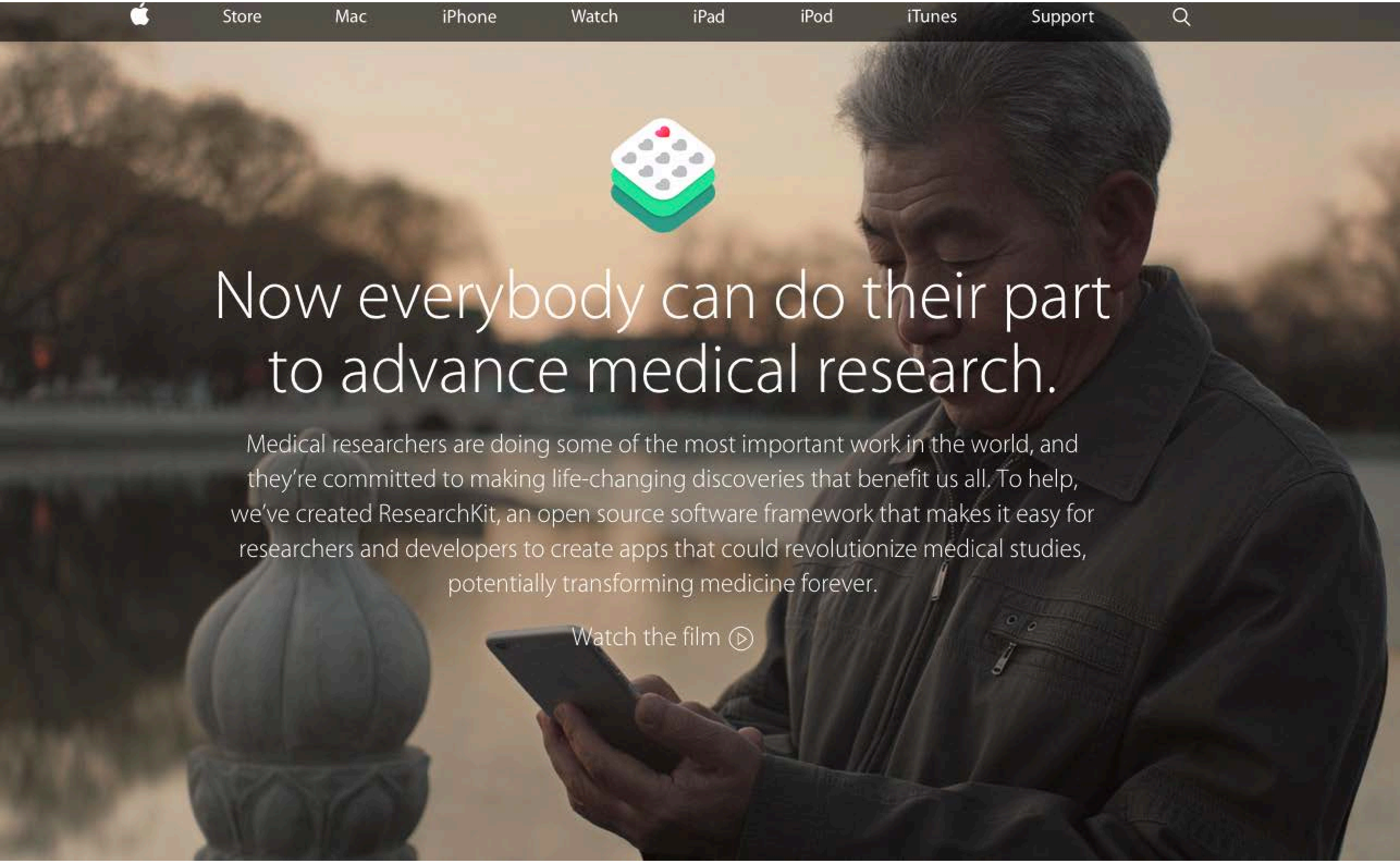
Neurological and Brain
ALS, MS, PD, Epilepsy, RLS, CFS, MSA, NMO, PLS, PSP, CBD, Stroke, Migraine

Skin, Hair and Nails
Psoriasis, Eczema, Rosacea

Transplants
Heart Transplant, Kidney Transplant, Liver Transplant, Lung Transplant, Pancreas Transplant

Women's Health and Pregnancy
Infertility, Endometriosis, Menopause, PCOS, Postpartum Depression

Last updated: March 16, 2016


Store
Mac
iPhone
Watch
iPad
iPod
iTunes
Support
Now everybody can do their part to advance medical research.
Medical researchers are doing some of the most important work in the world, and they're committed to making life-changing discoveries that benefit us all. To help, we've created ResearchKit, an open source software framework that makes it easy for researchers and developers to create apps that could revolutionize medical studies, potentially transforming medicine forever.
Watch the film

Open Humans

Awesome studies. Your data. Advancing research.

Become a member | Log in

Researchers, connect your studies

Open Humans members have valuable data – and they can share it with you!

Learn about connecting

Why become a member?

Open Humans helps you connect to exciting and innovative research studies – and gives you access to the data they produce from you.

- find **exciting studies** to join
- access to your **raw data**
- **share** your data with others
- **connect** with other members
- let others see your **contribution to science**!

Open source yourself.

Open Humans members can choose to publicly share their data.
Make your data a public resource! Which data you share is always up to you.

Learn about public sharing

Members 1 to 50 of 1,902

Sort by username

Name	Studies
Chris Durall (Chris_Durall)	
Beau Gunderson (beau)	
Gabriel Dean (gsdcomo)	
Cliff (Clifford_G_Andrew_PGP84)	
Stephen M Albers (salbers12)	
Jaak R (vbcoach)	
James Larry Vick (jameslvick)	
Jason Yang (jasonhyang)	
Juhan Sonin (jsonin)	
Dario X. Musso (Dario)	
Dan Bergman (Dyfflinviking)	
Jen Wagner (JKW)	
Crystal Butler (SuperLuckyCat)	
Alex Hoekstra (ahoekstra)	
Anna (annallull)	


SHARE
NHS
SCOTLAND
NHS RESEARCH SCOTLAND
CONNECT WITH US
SEARCH THIS SITE
MEET THE TEAM
RESEARCHERS
PARTNERS
CONTACT US
SHARE
STUDIES
WHY I SHARED
REGISTER
WHY
WHAT'S NEW
QUESTIONS?
SPARE BLOOD
GET INVOLVED
SHARE
SAVE LIVES
Sign up now.
It takes less than a minute to submit your details
REGISTER
TELL YOUR FRIENDS
Like
268
Tweet
SHARE YOUR STORY
WATCH THE VIDEO
101741
101741
People registered so far
SHARE
ANY QUESTIONS?

Genetic studies of body mass index yield new insights for obesity biology

Adam E. Locke, Bratati Kahali, Sonja I. Berndt, Anne E. Justice, Tune H. Pers, Felix R. Day, Corey Powell, Sailaja Vedantam, Martin L. Buchkovich, Jian Yang, Damien C. Croteau-Chonka, Tonu Esko, Tove Fall, Teresa Ferreira, Stefan Gustafsson, Zoltán Kutalik, Jian'an Luan, Reedik Mägi, Joshua C. Randall, Thomas W. Winkler, Andrew R. Wood, Tsegaselassie Workalemahu, Jessica D. Faul, Jennifer A. Smith, Jing Hua Zhao + *et al.*

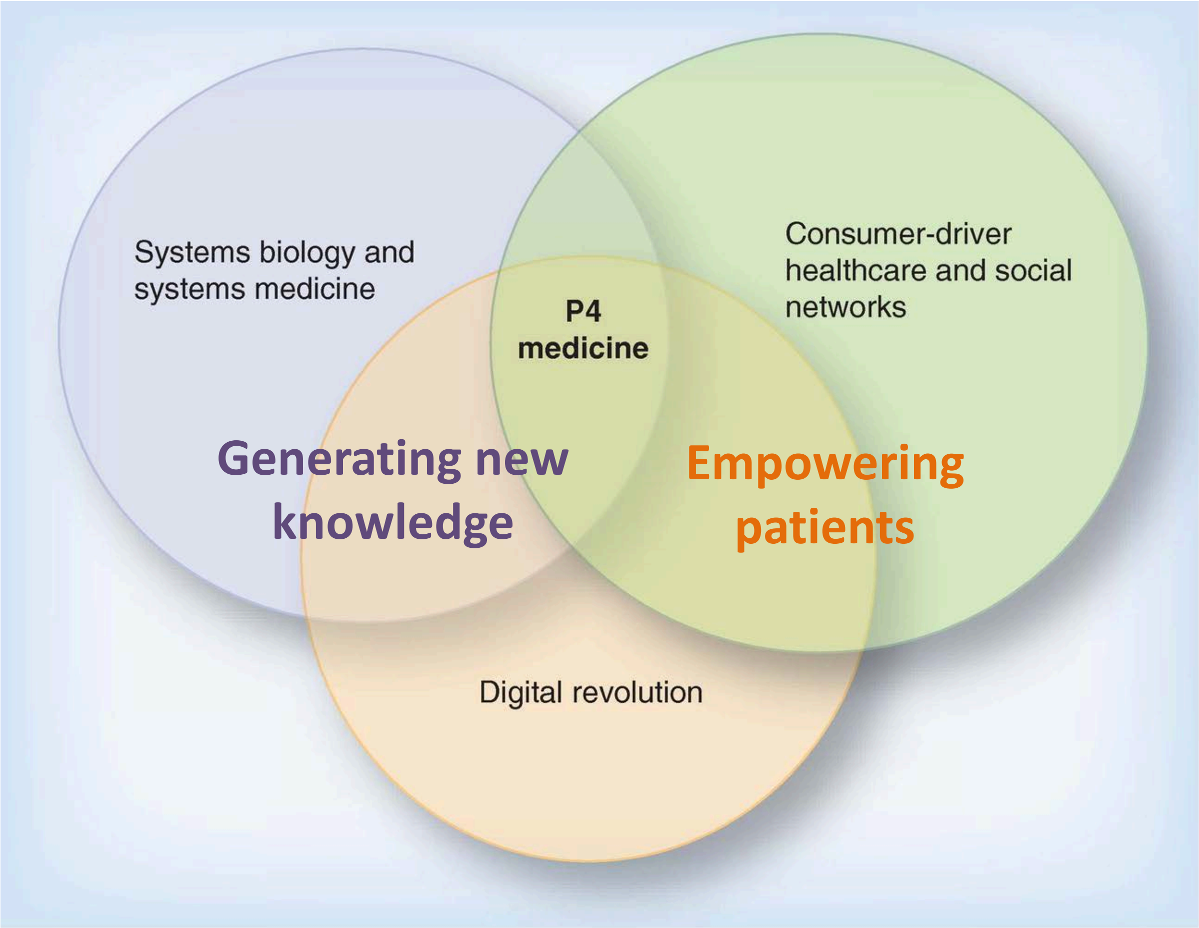
Affiliations | Contributions | Corresponding authors

Nature **518**, 197–206 (12 February 2015) | doi:10.1038/nature14177

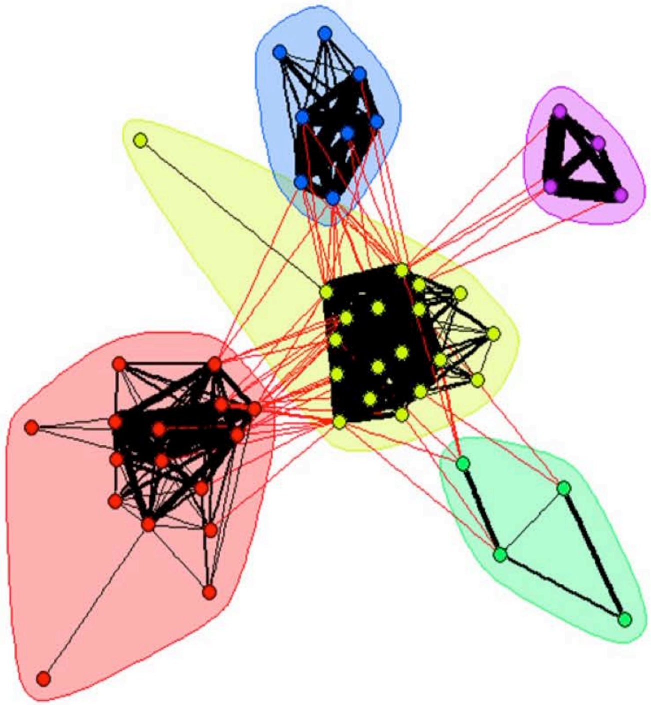
Received 20 November 2013 | Accepted 23 December 2014 | Published online 11 February 2015

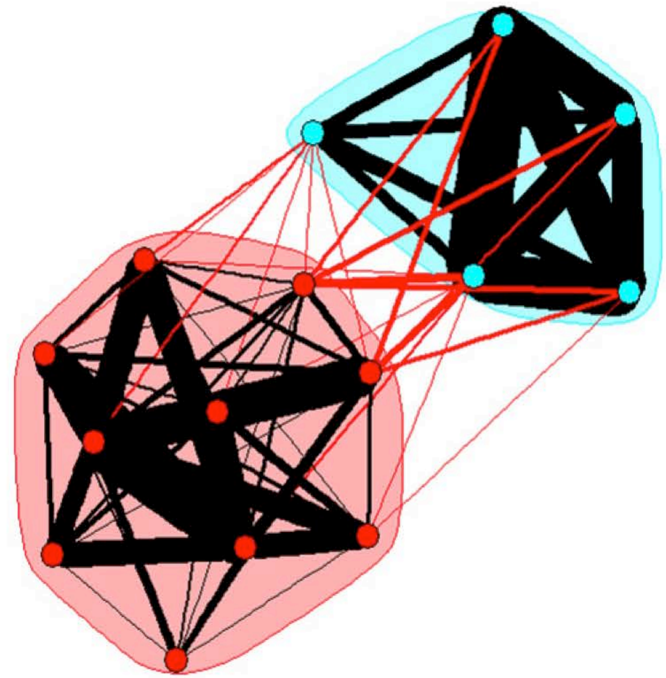
Big data transforming health care: ‘P4 Medicine’

- Predictive
- Preventive
- Personalized
- Participatory



Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. *Future Medicine* 2013;10(6):565-576.





Network characteristics

Number of providers: 51

Number of PPCs: 5

Provider-level characteristics

Mean number of shared patients : 41.3

Mean adjusted degree: 40.8

Mean betweenness centrality: 40

PPC-level characteristics

Mean number of providers: 10.2

Mean number of patients: 151.2

Mean adjusted degree: 6.5

Network characteristics

Number of providers: 15

Number of PPCs: 2

Provider-level characteristics

Mean number of shared patients: 85.7

Mean adjusted degree: 15

Mean betweenness centrality: 7.9

PPC-level characteristics

Mean number of providers: 7.5

Mean number of patients: 343

Mean adjusted degree: 2.9