Big data transforming health care

Louisa Jorm 17 May 2016
‘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

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

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.

Revascularisation: ‘unpacking’ the gap

Age, sex, year, MI type

Role of variation between hospitals?

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.

Revascularisation: ‘unpacking’ the gap

Comorbidity burden on admission

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

Once we adjust for comorbidities the gap is further reduced

Revascularisation: ‘unpacking’ the gap

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

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

Analysing EHR data

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.

**Natural language processing**

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

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

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.
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.

"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.

Empowering patients

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 obsessiona 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 half marathon. Coffee is my most frequent food. On average, I walk 11,720 steps a day, burning 3,089 calories, over 2.4 hours of activity, lasting for 5.6 hours and 9 minutes a night. This week, my sleep efficiency is 79% and my mood is 77% healthy. My BMI of 23.3 is 16 percentage points below the median for men my age, and my average daily Met score is 157, although I have exceeded what a Met score is.

I am, it seems, nothing more than a bundle of numbers and indicators, spewed out by LEDs and chattered 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 read from the chart, my heart pumping at the best of a heart rate monitor and forcing its variations to keep up with the national average. My head is able to imply from all the positive affirmation and grade matching, but it’s OK because my memes are being saved to my hard drive and my mood swings are extanting me “fragile” from strangers.

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

The QS movement, that I’ve temporarily joined, has three things to do. In San Francisco’s Spy Area in 2007, Two Wired magazine editors, Gary Wolf and Kevin

YOUR LIFE, UPLOADED
THE DIGITAL WAY TO BETTER MEMORY, HEALTH, AND PRODUCTIVITY

GORDON BELL AND JIM GEMMELL

FOREWORD BY BILL GATES

http://totalrecallbook.com/
TRACK YOURSELF!

A map of digital tools to help you quantify your life.

Take control of your health.
Explore HealthVault

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 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

1,000,000+

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.
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

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

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

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

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

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

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

**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

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

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

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

**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

**Women's Health and Pregnancy**
- Infertility, Endometriosis, Menopause, PCOS,
- Postpartum Depression
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
<table>
<thead>
<tr>
<th>Name</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chris Durall (Chris_Durall)</td>
<td></td>
</tr>
<tr>
<td>Beau Gunderson (beau)</td>
<td></td>
</tr>
<tr>
<td>Gabriel Dean (gsdcomo)</td>
<td></td>
</tr>
<tr>
<td>Cliff (Clifford_G_Andrew_PGP84)</td>
<td></td>
</tr>
<tr>
<td>Stephen M Albers (salbers12)</td>
<td></td>
</tr>
<tr>
<td>Jaak R (vbcoach)</td>
<td></td>
</tr>
<tr>
<td>James Larry Vick (jameslvick)</td>
<td></td>
</tr>
<tr>
<td>Jason Yang (jasonhyang)</td>
<td></td>
</tr>
<tr>
<td>Juhan Sonin (jsonin)</td>
<td></td>
</tr>
<tr>
<td>Dario X. Musso (Dario)</td>
<td></td>
</tr>
<tr>
<td>Dan Bergman (Dyfflinviking)</td>
<td></td>
</tr>
<tr>
<td>Jen Wagner (JKW)</td>
<td></td>
</tr>
<tr>
<td>Crystal Butler (SuperLuckyCat)</td>
<td></td>
</tr>
<tr>
<td>Alex Hoekstra (ahoekstra)</td>
<td></td>
</tr>
<tr>
<td>Anna (annallull)</td>
<td></td>
</tr>
</tbody>
</table>
Genetic studies of body mass index yield new insights for obesity biology


Affiliations | Contributions | Corresponding authors

Received 20 November 2013 | Accepted 23 December 2014 | Published online 11 February 2015
Big data transforming health care: ‘P4 Medicine’

- Predictive
- Preventive
- Personalized
- Participatory

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