

UNSW AUSTRALIA CENTRE FOR BIG DATA RESEARCH IN HEALTH

UNSW AUSTRALIA Routinely collected data – a strategic resource for research and policy
Louisa Jorm 22 June 2016

Never Stand Still Medicine Centre for Big Data Research in Health

cbrdh.med.unsw.edu.au

Data live utilitarian lives. From the moment they are conceived, as measurements of some thing or system or person, they are conscripted to the cause of being useful. They are fed into algorithms, clustered and merged, mapped and reduced. They are graphed and charted, plotted and visualized... Always, though, the measure of the life of data is in its utility.

Data that are collected but not used are condemned to a quiet life in a database.

Jer Thorp
Data Artist
(<http://blog.blprnt.com/blog/blprnt-on-data-and-performance>)

Summary

- Value of routinely collected data
- Routinely collected data in health
 - sources
 - benefits and limitations
- Examples of use
 - acute myocardial infarction
 - serious road transport injury
- Priorities and opportunities

Big value in big data

"... making data more widely available in shareable formats... has the potential to unlock large amounts of economic value, by improving the efficiency and effectiveness of existing processes; making possible new products, services, and markets; and creating value for individual consumers and citizens"

Big data: The next frontier for innovation, competition, and productivity

Big value in big data

... we estimate the aggregate direct and indirect value of government data in Australia at up to AUD 25 billion per annum

... assuming a doubling of accessibility and use, we estimate the return on Australia's investment in government and research data at around AUD 34 billion per annum

Figure 11 Estimates of the value of open data (AUD billions pa)

| Category | Value (AUD billions pa) |
|-------------------------------|-------------------------|
| Government Data (PSD) | 25 |
| Publicly-Funded Research Data | 9 |
| Private Data (Research) | 30 |

Big value in big data

Figure 9 Value of open data for Australia (AUD billions pa)

| Sector | Value (AUD billions pa) |
|-------------------|-------------------------|
| Education | 18 |
| Transport | 14 |
| Consumer Products | 10 |
| Electricity | 6.7 |
| Oil and Gas | 4.8 |
| Health Care | 5.9 |
| Consumer Finance | 4.2 |

Australian policy climate supports data use

Principles for accessing and using publicly funded data for health research

Supported by
The National Health and Medical Research Council, the Australian Government Department of Health, the Australian Institute of Health and Welfare, the Australian Bureau of Statistics, the Australian Government Department of Human Services, the Australian Electoral Commission, the Australian Institute of Aboriginal and Torres Strait Islander Studies, Universities Australia, and the Consumers Health Forum of Australia

Select Committee on Health

Sixth interim report

Big health data: Australia's big potential

PUBLIC SECTOR DATA MANAGEMENT

data.gov.au

Search data: Big health data

8.2k 4.3k 2.4k 21

Sources of routinely collected data in health

- By-product of operating health services
 - hospital inpatient stays, MBS and PBS claims, private health insurance claims
- Meeting regulatory requirements
 - births, deaths, health practitioner registration
- Monitoring health
 - cancer notifications, communicable disease notifications, perinatal data
- Produced through operations of "non-health" sectors
 - child development, education, housing, environment

Benefits of routinely collected data

- Population reach
 - can be used to study rare outcomes and population subgroups
- Longitudinal (when linked!)
 - supports studies across the lifecourse, enables long-term follow-up
- Avoids nonresponse and bias
 - surveys are increasingly non-viable
- Cost-effective
 - studies over many decades can be undertaken time- and cost-efficiently
- "Real world"
 - often the only way to evaluate outcomes of services or interventions where there is no evidence from randomised trials

Limitations of routinely collected data

- Event-based
 - difficult to define denominators or appropriate comparison groups
- Data quality
 - rely on correct information being present and entered correctly
 - 'rare' values in large datasets may be more likely to represent keystroke or coding errors than valid entries
- Limited data items
 - often limited information on confounders and risk factors e.g. smoking, BMI
- Lack of metadata
 - detailed metadata and other documentation may not be readily, or publicly, available.

Summary

- Value of routinely collected data
- Routinely collected data in health
- Examples of use
 - disparities in acute myocardial infarction
 - disparities in serious road transport injury
- Priorities and opportunities

Indigenous Health Outcomes Patient Evaluation (IHOPE)

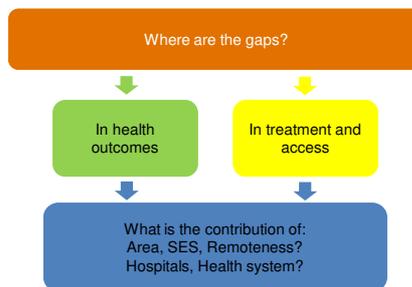
| | |
|------------------|---------------------------------|
| Deborah Randall | Holger Möller |
| Sandra Eades | Aiden O'Loughlin |
| Alastair Leyland | Rebecca Ivers |
| Sanja Lujic | Tim Harold |
| Timothy Churches | Tracie Reinten |
| Mary Haines | |
| Michael Falster | IHOPE Community Reference Group |
| Kathleen Falster | NHMRC Project Grant #573113 |

Aboriginal health gap

- Life expectancy 11.5 years less for males and 9.7 years less for females
- Estimated that 59% of total burden of disease for Aboriginal Australians could be avoided if have same rate of disease burden as the general population (Vos et al, data for 2003)



Aims of IHOPE



IHOPE data



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



Characteristics of people admitted to hospital with AMI

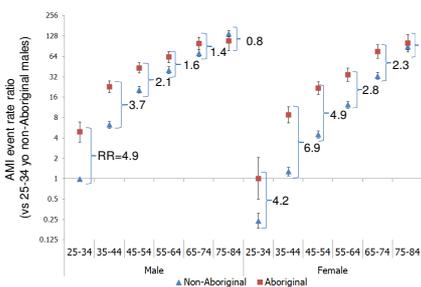
| | Aboriginal | Non-Aboriginal |
|---|------------|----------------|
| Average age | 54 yo | 66 yo |
| Current smokers | 51% | 27% |
| Private health insurance | 16% | 45% |
| Live in most disadvantaged areas | 48% | 26% |
| First admitted to: | | |
| - major city hospital | 33% | 67% |
| - hospital with specialist cardiac facilities | 27% | 44% |

Findings: AMI incidence rates

- The age-standardised incidence of AMI in NSW was:
 - 464 per 100,000 for Aboriginal people
 - 234 per 100,000 for non-Aboriginal people
- An Aboriginal person had 2.1 (2.0-2.2) times the risk of an AMI as a non-Aboriginal person of the same age, sex and year of event, from the same area of residence

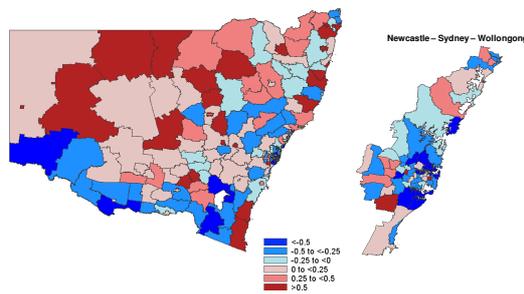
Findings: AMI incidence rates

The disparity is greatest in younger age groups and for females



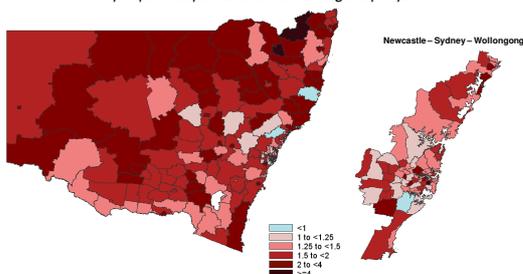
Where are the higher rates for Aboriginal people?

Rates of AMI for Aboriginal people vary by Statistical Local Area, with higher rates generally in regional and rural areas

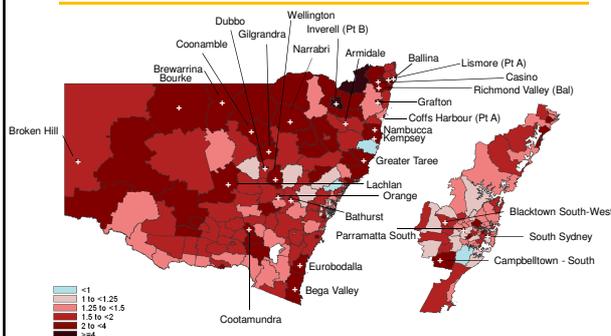


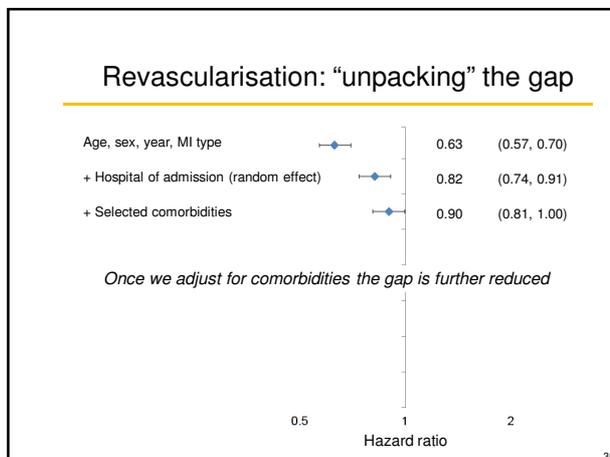
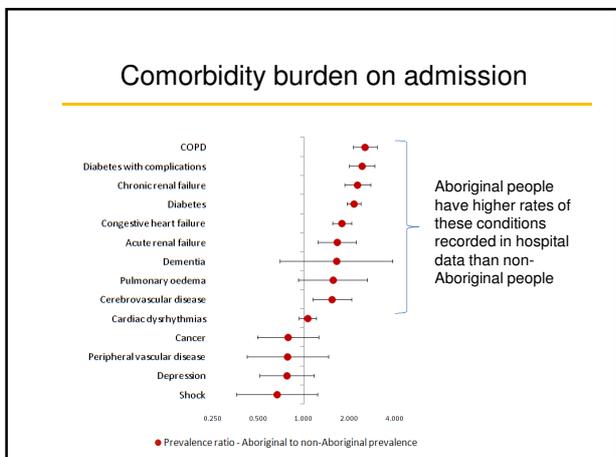
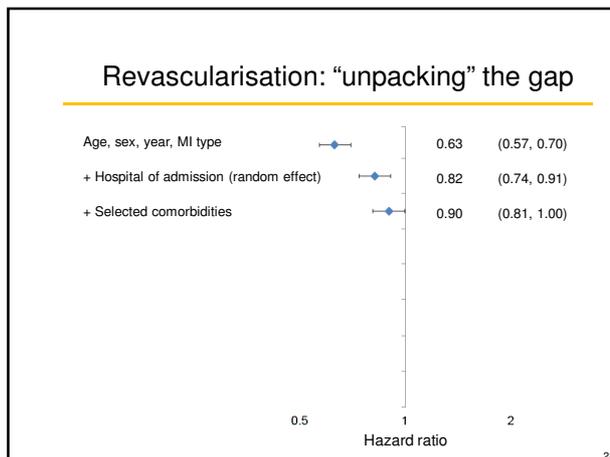
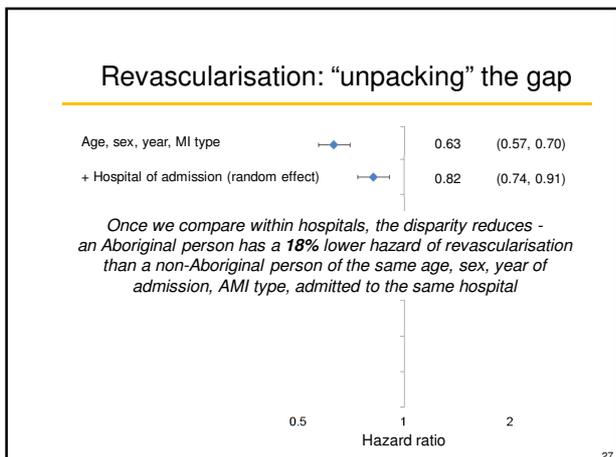
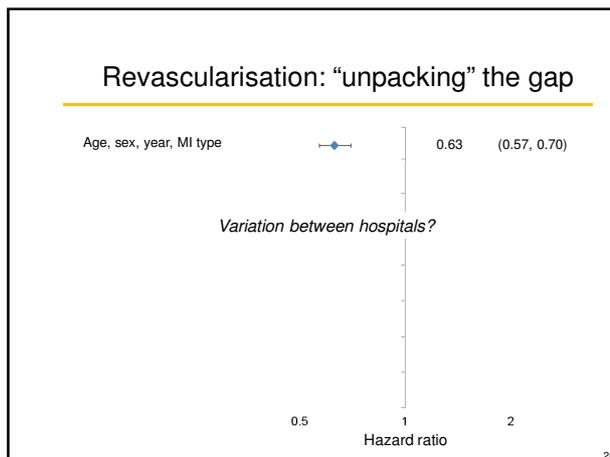
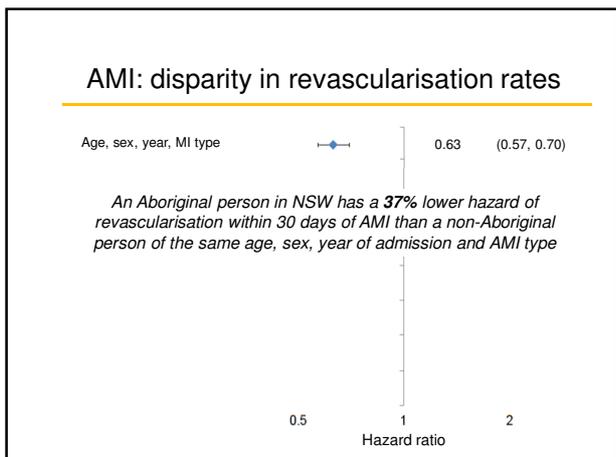
Where is the higher disparity for Aboriginal people?

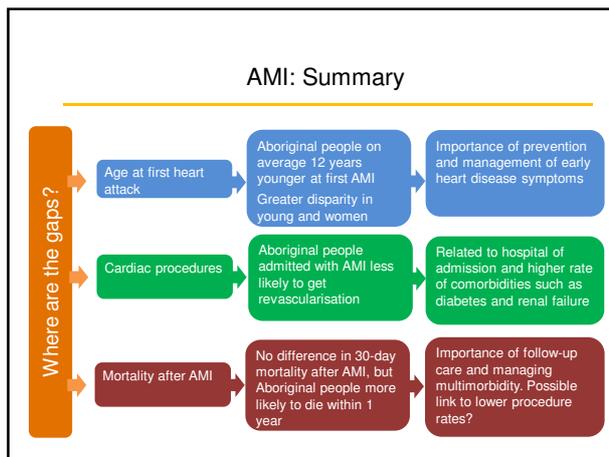
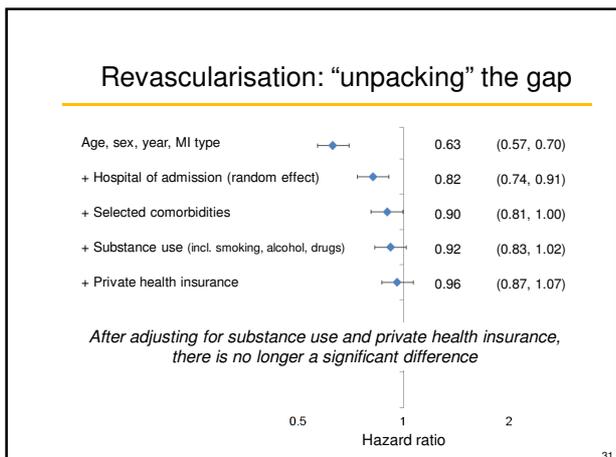
Almost all areas in NSW have a higher incidence of AMI for Aboriginal people compared with non-Aboriginal people



“High incidence, high disparity” areas



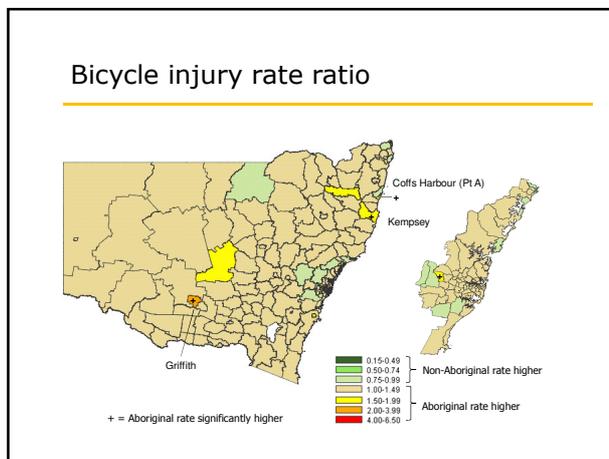
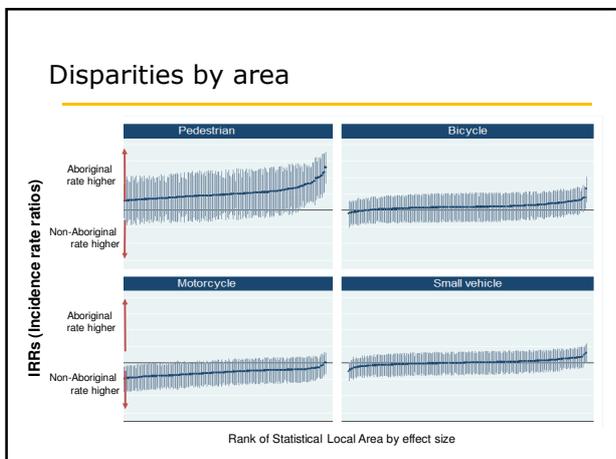




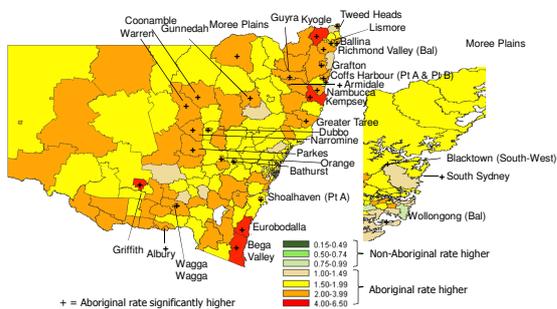
Is there a disparity in injury rates?

| | IRR adjusted for age and sex | IRR adjusted for age, sex and area |
|-----------------------------|------------------------------|------------------------------------|
| All road transport injuries | 1.18 (1.09-1.28) | 1.00 (0.96-1.04) ↓ |
| Small vehicle | 1.14 (1.03-1.27) | 1.01 (0.94-1.08) ↓ |
| Pedestrian | 1.76 (1.55-1.99) | 1.96 (1.75-2.19) ↑ |
| Bicycle | 1.24 (1.12-1.37) | 1.18 (1.08-1.29) ↓ |
| Motorcycle | 0.98 (0.82-1.17) | 0.64 (0.59-0.70) ↓ |

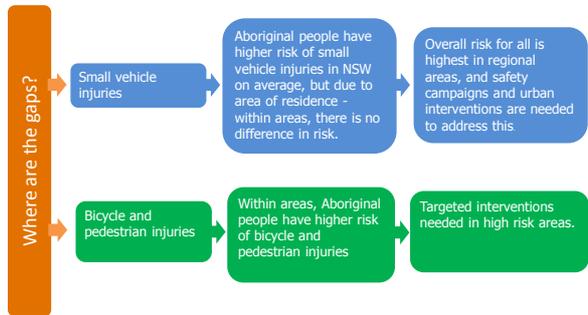
Legend: 0.5 1 2. Non-Aboriginal rate higher (left of 1), Aboriginal rate higher (right of 1). Arrows indicate direction of change.



Pedestrian injury rate ratio



Summary - Serious road traffic injuries



Conclusions

- Whole-of-population linked routinely collected data and multilevel modelling methods have unique power to explore health disparities
 - "unpack" contributions of personal, geographic and service factors
 - identify targets for intervention
- It is essential that geography is taken into account in studies of health disparities
 - especially where there are significant urban-rural differences in the distribution of disadvantaged populations and health services
- The simplest of data linkages hugely increases the value of routinely collected data!

Summary

- Value of routinely collected data
- Routinely collected data in health
- Examples of use
 - disparities in acute myocardial infarction
 - disparities in serious road transport injury
- **Priorities and opportunities**

Priorities and opportunities

- Address current data limitations
- Link across sectors
- Address barriers to data access and use
 - especially state-national divides
- Involve consumers and patients
- Pre-empt workforce needs

Priorities and opportunities

- **Address current data limitations**
- Link across sectors
- Address barriers to data access and use
 - especially state-national divides
- Involve consumers and patients
- Pre-empt workforce needs

Limitations of routinely collected data

- **FIX:** Create a population "spine" (e.g. using Medicare data)
- **FIX:** Use the data!
Feedback to data providers
Optimise use of technology
- **FIX:** More interaction between "IT people" and data users
Optimise use of technology
- **FIX:** "Surface" existing metadata
Optimise use of technology



Priorities and opportunities

- Address current data limitations
- **Link across sectors**
- Address barriers to data access and use
 - especially state-national divides
- Involve consumers and patients
- Pre-empt workforce needs



- (slide showing unpublished data)



Priorities and opportunities

- Address current data limitations
- Link across sectors
- **Address barriers to data access and use**
 - especially state-national divides
- Involve consumers and patients
- Pre-empt workforce needs



- (slide showing unpublished data)



Priorities and opportunities

- Address current data limitations
- Link across sectors
- Address barriers to data access and use
 - especially state-national divides
- **Involve consumers and patients**
- Pre-empt workforce needs



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, Zoltan 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

Bring your ancestry to life through your DNA.
 Discover your ancestral origins and trace your lineage with a personalized analysis of your DNA.

1,000,000+
 With over 1,000,000 samples 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.

20.5% East Asian
 38.6% Sub-Saharan African
 24.7% European

Priorities and opportunities

- Address current data limitations
- Link across sectors
- Address barriers to data access and use
 - especially state-national divides
- Involve consumers and patients
- Pre-empt workforce needs

Table 2. Workforce capabilities for research using routinely collected health data

| Skills in | Knowledge of |
|---------------------------------|---------------------------------------|
| Applying biostatistical methods | Bioinformatics |
| Conference presentations | Biostatistics |
| Data management | Communications |
| Data manipulation | Computer programming |
| Data security | Computer science |
| Database design | Data governance |
| Grantmanship | Data provenance and interpretation |
| Literature review and synthesis | Data security and privacy |
| Managing contracts | Data standards |
| Project management | Data structures |
| Research design | Epidemiology |
| Specifying research questions | Ethics |
| Visualisation design | Health and clinical domains |
| Working with databases | Health system structure and operation |
| Working with policy partners | Machine learning |
| Working with the media | Meta-analysis |
| Writing blogs and commentaries | Metadata standards |
| Writing ethics applications | Research methodology |
| Writing for policy | Research translation |
| Writing grant applications | Social media |
| Writing scientific papers | Unstructured data (e.g. images, text) |
| Writing tenders | Visualisation |

Jorm L. Routinely collected data as a strategic resource for research: priorities for methods and workforce. *Public Health Res Pract.* 2015;25(4):e2541540.

