UNSW AUSTRALIA

Never Stand Still | Faculty of Arts & Social Sciences | Social Policy Research Centre

Lessons learned from record linkage to advance Indigenous mortality statistics in Australia: Unpacking sources of error and bias

Ching Choi, PhD (Social Policy Research Centre)

IAOS Conference 2016

SPRC Social Policy Research Centre

Acknowledgement

Disclaimer

The views expressed are those of the author and do not necessarily reflect those of the Australian Institute of Health and Welfare or the Australian Bureau of Statistics

Collaborating colleagues

- Dr Len Smith, Australian National University
- Dr Fadwa Al-Yaman, Senior Executive, Australian Institute of Health and Welfare
- Dr Tetteh Dugzaba, Senior Analyst, Australian Institute of Health and Welfare
- Professor Tony Barnes, Charles Darwin University, Australia

Map of Australia

Australian Indigenous population, 2011

SPRC
Social Policy Research Centre

	Indigenous ('000)	% distribution	Total population ('000)	Indigenous as % of total
New South Wales	209	31	7 261	2.9
Victoria	47	7	5 582	0.8
Queensland	189	28	4 518	4.2
Western Australia	88	13	2 391	3.7
South Australia	38	6	1 646	2.3
Tasmania	24	4	512	4.7
Northern Territory	69	10	232	29.7
Australia Capital Territory	6	1	371	1.6
Australia	670	100	22 520	3.0

ABS: 3238.0.55.001 August 2013

UNSW AUSTRALIA

National requirements

- There has always been a need to know the conditions of the Indigenous population at both the local and the national levels. The 1967 Referendum gave the federal government power to legislate for Indigenous people and to include Indigenous people in official statistics.
- In 2007, the then Prime Minister, Kevin Rudd, announced that he will report each year to the parliament the national gap in well-being between the Indigenous and the non-Indigenous people.
- The health indicators to be used for measuring gap were infant mortality and life expectancy. Other indicators include education and employment.
- At about the same time the State and Territory governments (COAG) asked the Productivity Commission to produce a bi-annual report on the effectiveness of government services in reducing various well-being gaps – using a very wide set of indicators.
- At that time and even before, it was known that death registration data (as well as population census data) for the Indigenous people were weak.
- There was a big push to improve Indigenous data to meet requests from the top.

SPRC
Social Policy Research Centre

Australian Indigenous Death data

- Death registration is the responsibility of State and Territory governments
- Full registration of deaths in all States and Territories, but Indigenous identification patchy
- Some state registrars introduced registration forms in the late 1990s that asked the standard question on Indigenous identity, eg Queensland in 1997.
- In 2000-2004 an estimated 57% of Indigenous deaths were identified nationally (using indirect demographic methods) and showed large variations (ABS 2004 3302.0 p.63):
  - Northern Territory 94%
  - Western Australia 72%
  - South Australia 66%
  - Queensland 53%
  - New South Wales 46%
  - Victoria 35%
- Death data for NT and some states are published, but no national data

Defining Indigenous status

- Official definition since 1978 has 3 parts:
  - Descent (blood)
  - Self identification
  - Recognition as Indigenous by Indigenous community (eg an Indigenous land council)

- Statistical collections
  - Standard question used by government in population census, surveys and administration:
    - Are you (is this person) of Aboriginal and Torres islander origin:
      - No
      - Yes, Aboriginal
      - Yes, Torres Strait Islander
        - » For persons of both Aboriginal and Torres Islander origin, mark both ‘yes’ boxes

Early Australian Bureau of Statistics (ABS) estimates

- ABS requires Indigenous life tables for Indigenous population estimates and projections. Mortality rates adjusted for under-identification are required.

- Indirect methods were used by the Australian Bureau of Statistics to estimate adjustment factors to correct for under-identification - 1991, 1996 and 2001 Censuses (Preston and Hill) 2006 Census (Bhat).

- Indirect methods (based on the Generalised Growth Method developed by W Brass in the early 1970s) use census and registration data jointly and contain some basic assumptions – such as accurate census counts, no net migration and constant under-identification rate by age. However, the Bhat method relaxes the ”no net migration” condition.

- Adjusted data used for life table construction, but not for routine release of mortality measures

- Result based on indirect methods showed large gaps in life expectancy of 18-19 years between Indigenous and non-Indigenous based on Preston and Hill method and 17 year based on Bhat method - much worse than US, Canada and New Zealand Indigenous estimates.

- Controversy over the assumptions behind the methods and the results

Move towards data linkage and direct methods

- Ken Hill research (2004-5) and Tony Barnes sensitivity tests (2008)
  - Different methods give very different results
    - The modified Hill method (that includes migration as an element of census data inaccuracy) show much higher life expectancy than the Bhat method
    - Results from both methods very sensitive to errors in the estimates of the size and the age structure of the population

- In 2004-5, Australian Bureau of Statistics and the Australian Institute of Health and Welfare began considering linking census data with other datasets to improve Indigenous death data – primarily for calculating life expectancy (ABS and AIHW) but also for developing other mortality measures (AIHW).

- The 2 government agencies had a joint statistical unit that published a bi-annual report of the health and wellbeing of Indigenous Australians, but they hold different data sets.

Source: K Hill et at 2007 and T Barnes et al 2008

ABS linking census to death registration

- First linkage 2006, repeated 2011, planned 2016.
- Census records linked to 11 months of death records after census date
- Australia does not have national ID cards – so no unique personal ID number system
- Linkage based on name, address, date of birth and other characteristics
- Sophisticated linkage methods - automatic assignment and clerical reviews
- Identification rates include census undercount (PES) adjustments

	2006	2011
Linkage rate	74% of death registration records	80% of death registration records
Identification rate	0.92	0.82
Life expectancy age 0	67.5 (M) 73.1 (F)	69.1 (M) 73.7 (F)
Gap between Indigenous and non-Indigenous	11.4 years (M) 9.6 years (F)	10.6 years (M) 9.5 years (F)

AIHW Linkage to Administrative Data

- Linkage 2001-2006; 2006 onwards being undertaken
- Deaths in hospitals (about 55% of all deaths), aged care facilities (about 35%)
- Death registration records linked sequentially to aged care facilities, hospitals, and peri-natal deaths records
- No unique ID number system
- Linkage based on name (age care facilities only), date of birth and date of death, address and other characteristics.
- Date of birth and date of death are powerful linkage combination
- Linkage rate : 86% of death registration records

	2001-2006
Life expectancy at age 0	66.6 (M) 72.7 (F)
Gap between Indigenous and non-Indigenous	11.5 years (M) 10.3 years (F)

SPRC
Social Policy Research Centre

Different data sources, similar results… But possible errors and bias

- Source data do not have full coverage
  - In both studies, a considerable proportion of death registration records not linked
  - Census data suffer from less than 100% enumeration and non-response on the Indigenous question on the census form.
  - Hospital, aged care data suffer from less than 100% coverage of all deaths

- Quality of identifying data for linkage?
  - Names, date of birth, date of death etc are not 100% accurate
  - Excellent matches, good matches, acceptable matches in linkage – where is the bottom line
  - In both studies, there was a need to clerically review considerable number of unlinked records - 17% of death records (ABS) and 15% of hospital records (AIHW)

- Data consistency in Indigenous identification - How consistent are the linked data sets
  - Which record to believe if Indigenous ID is different in a linked pair of records?

Consistency measures

- Assume two data sets, both aim to capture Indigenous identification. Records are linked to assess consistency in Indigenous response.

	Dataset 2		
Dataset 1	Indigenous	Non-Indigenous (including n/s)	Total
Indigenous	A	B	C
Non-Indigenous (including n/s)	D		
Total	G		

Overall consistency index = A/(B+D+A) (Indigenous in both datasets as % of Indigenous records in either or both datasets)
Consistent index for dataset 1 = A/C (Indigenous in both data sets as % of Indigenous records in dataset 1)
Consistent index for dataset 2 = A/G (Indigenous in both data sets as % of Indigenous records in dataset 2)

1 = totally consistent
0.5 = number of consistent linked records the same as the number of inconsistent linked records
Zero = totally inconsistent

SPRC
Social Policy Research Centre

2006 Census Versus Census Post Enumeration Survey (PES)

Index (1.0 = totally consistent; 0.0 = totally inconsistent)
Source: ABS 3302.0.55.003, 25 May 2009 and 15 Nov 2013

	Overall	2006 Census	2006 PES
AUSTRALIA	0.85	0.90	0.94
NSW	0.71	0.79	0.88
QLD	0.81	0.87	0.93
WA	0.87	0.93	0.95
NT	0.95	0.97	0.98

UNSW AUSTRALIA

2011 Census Versus PES

Index (1.0 = totally consistent; 0.0 = totally inconsistent)

Source: ABS 3302.0.55.003, 25 May 2009 and 22 Nov 2013

	Overall	2011 Census	2011 PES
AUSTRALIA	0.88	0.94	0.93
NSW	0.76	0.87	0.86
QLD	0.83	0.93	0.89
WA	0.87	0.93	0.93
NT	0.95	0.97	0.97
Major cities, inner regional	0.77	0.89	0.85
Outer regional and remote	0.92	0.96	0.95

2006 Census Versus Death Registration

Index (1.0 = totally consistent; 0.0 = totally inconsistent)

Source: ABS 3302.0.55.003, 25 May 2009 and 22 Nov 2013

	Overall	2006 Census	Deaths
AUSTRALIA	0.64	0.77	0.80
NSW	0.53	0.66	0.73
QLD	0.73	0.83	0.86
WA	0.66	0.85	0.74
NT	0.89	0.97	0.92

2011 Census Versus Death Registration

Index (1.0 = totally consistent; 0.0= totally inconsistent)

Source: ABS 3302.0.55.003, 25 May 2009 and 15 Nov 2013

	Overall	2011 Census	Deaths
AUSTRALIA	0.62	0.71	0.82
NSW	0.53	0.63	0.76
QLD	0.65	0.75	0.82
WA	0.78	0.86	0.89
NT	0.95	0.98	0.96
Major cities, inner regional	0.48	0.58	0.73
Outer regional and remote	0.80	0.87	0.91

SPRC
Social Policy Research Centre

2007-8 Hospital records Versus Hospital Data Quality Survey

Index (1.0 = totally consistent; 0.0 = totally inconsistent)

Source: AIHW Cat No HSE 85 and IHW 90, plus unpublished tables

	Overall	2007-8 Hospital	2007-8 survey
AUSTRALIA	0.92	0.99	0.93
Major cities	0.76	0.97	0.78
Inner regional	0.89	0.99	0.90
Outer regional	0.91	0.97	0.93
Remote & very remote	0.97	1.00	0.97

UNSW AUSTRALIA

SPRC
Social Policy Research Centre

2011-12 Hospital records Versus Hospital Data Quality Survey

Index (1.0 = totally consistent; 0.0 = totally inconsistent)

Source: AIHW Cat No HSE 85 and IHW 90, plus unpublished tables

	Overall	2011-2 hospital	2011-2 survey
AUSTRALIA	0.86	0.97	0.89
Major cities	0.71	0.94	0.75
Inner regional	0.77	0.92	0.81
Outer regional	0.94	0.99	0.95
Remote & very remote	0.96	0.98	0.98

UNSW AUSTRALIA

Census 2006 Versus Census 2011

Index (1.0 = totally consistent; 0.0 = totally inconsistent)

Source: Nicholas Biddle: “Indigenous Population”: Paper presented o “Symposium on New Ideas and Challenges for Demographic Research in Australia”, 23 October 2014 at ANU

	Overall	2006 Census	2011 Census
Australia	0.79	0.93	0.87

What do these consistency indices tell us?

SPRC
Social Policy Research Centre

1. Inconsistency is substantial
2. Inconsistency varies greatly between data sets – hospitals data more consistency than census and death data – death data the least consistent
3. Inconsistency varies greatly between States and Territories – NT more consistent than other States – NSW the least consistent
4. Inconsistency varies greatly between urban and rural areas – Outer region and Remote much more consistent than Major Urban areas

If consistent data are considered better than inconsistent data, then the data sets for NSW and major urban areas are of poor quality. (But these are the areas where the majority of Indigenous people live!)

UNSW AUSTRALIA

How robust are the consistency indices

- Depends on the effectiveness of the data linkage
  - Size of data set – death data are very small data sets – around 2,000 a year
  - Quality of identifying information
    - Spelling of names, name changes, poor date of birth etc.
    - Missing Indigenous information on records (not stated), eg census and death records
  - For population censuses – existence of large number of records that have little or no identifying information
    - 5% of 2011 census records or about 1.06 million records
    - of which 73% were “imputed” records – about 770,000 records
    - 17% under-count of Indigenous people at 2011 Census (6% non-Indigenous) ABS (2012) Cat No 2940.0

Possible biases in the use of linked records

- Assumptions need to be made if the indices are used for data adjustment
- Which data set(s) is superior and provides the standard?
  - The PES is probably superior than the Census – but even this can be challenged; and the PES is a small data set
  - Are census records better than death records in Indigenous identification?
  - Are quality survey records better than hospital records?
  - Is the 2011 Census better than the 2006 Census?

- What if it is not certain which data set is superior.
  - Inconsistent records are legitimate and real records and not necessarily “bad records”
  - Are there false positives?
  - Can both responses of an inconsistent record be right?

- Given the large inconsistencies in the linked data, biases can be substantial and have differential impacts on various groups.

Reference: Madden R et at 2012

Can data consistency be improved?

- Data collection methods can be improved - consistent definition, fuller coverage, better sampling in surveys, better collection instruments, fewer non-responses (to data collection and/or Indigenous question) , improved method of ascertaining Indigenous identification, better quality control in data processing etc

- But changes to Indigenous identification by respondents difficult to avoid – Indigenous identification is basically self-reported for most data collections, even if through an interviewer or an administrator.

- Data linking – privacy issues
  - Recent privacy debate re the 2016 Australian Census – 4 year retention of census names and address
  - Ethics approvals in the use of administrative data

A further issue – the population base

- Australian population censuses show high growth rates of Indigenous population
  - Not fully explainable through demography accounting
  - Illustrative of increasing propensity of people to identify
  - But some do drop out from census identification as well
- The size of the Indigenous population is revised after each census
- The instability of Indigenous death statistics and population estimates makes it problematic to measure Indigenous mortality in Australia
- What will the 2016 Census show?

4 basic causes of bias and error

- Quality of identifying data used for matching - name, age, sex, geography etc – affecting match success rates

- Basic assumptions made about the quality of Indigenous identification responses in different data sets
  - PES assumed to be better than Census
  - Census assumed to be better than death registration
  - Ever Indigenous - Hospital, aged care residential, death registration and peri-natal Indigenous Identification all assumed to be correct.

- Numeration/denominator bias possible in the AIHW method

- Consistency of the Indigenous population over time – This affects interpretation of mortality changes overtime if the Indigenous population at time 1 and time 2 are different because of movement of different people into and out of the Indigenous population.

References

Australian Bureau of Statistics (2005) Deaths 2004, (3302,0), Canberra

Australian Bureau of Statistics (2008) Discussion Paper: Assessment of methods for developing life tables for Aboriginal and Torres Strait Islander Australians, (Cat No 3302.0.55.002), Canberra

Australian Bureau of Statistics (2009 and 2013) Life tables for Aboriginal and Torres Strait Islander Australians (Cat No 3302.0.55.003), Canberra

Australian Bureau of Statistics (2012) 2011 Census of population and housing – details of undercount Australia, (Cat No 2940.0) June 2012, Canberra

Australian Bureau of Statistics (2013) Estimates of Australian Indigenous and Torres Islander population, June 2011 (3238.005.001), Canberra.

Australian Bureau of Statistics (2013) Information Paper: Death registration to census linkage project – Methodology and quality assessment, Australia, 2011-2012 (Cat No 3302.0.55.004), Canberra

Australian Institute of Health and Welfare (2010). Indigenous identification in hospital separations data: quality report. Cat. No. HSE 85. Canberra

Australian Institute of health and Welfare (2010) Comparing life Expectancy of Indigenous people in Australia, New Zealand, and the United states: Conceptual, Methodological and data issues. Canberra

Australian Institute of Health and Welfare (2012) An enhanced mortality database for estimating Indigenous life expectancy: A feasibility study. (Cat No IHW 75), Canberra

Australian Institute of Health and Welfare (2013). Indigenous identification in hospital separations data: quality report. Cat. No. IHW 90. Canberra

Barnes T et al (2008) A Comparative analysis of indirect methodologies for estimating Indigenous life expectancy, Charles Darwin University, Darwin

Hill K et al (2007) Excess Indigenous Mortality: Are Indigenous Australians more severely disadvantaged than other Indigenous populations. *International Journal of Epidemiology*, 10, 1093-1102.

Madden R et al Estimating Indigenous life expectancy: pitfalls with consequences, *Journal of Population Research*, (2012) 29; 269-281.