Managing Aboriginal and Torres Strait Islander Data for Public Health Research.

OJPHI

# Managing Aboriginal and Torres Strait Islander Data for Public Health Research.

D. van Gaans [1], S. Ahmed [1], K. D'Onise [1], S. M. Taylor [2], R. McDermott [3]

1. Centre for Research Excellence in the Prevention of Chronic Conditions in Rural and Remote Populations, University of South Australia, South Australia

2. Centre for Chronic Disease Prevention, Australian Institute of Tropical Health and Medicine, College of Public Health, Medical and Veterinary Sciences, James cook University, Queensland

3. College of Public Health, Medical and Veterinary Sciences, James Cook University, Queensland

## Abstract

Good quality data on Aboriginal and Torres Strait Islander peoples are needed to assess the effectiveness of programs and interventions, and to evaluate policies that are designed to improve the status of, and service delivery to, Aboriginal and Torres Strait Islander peoples. Due to the lack of longitudinal data it is difficult to gain knowledge on the specific causes or consequences of changes in indigenous outcomes. Variables such as name, date of birth and address for Aboriginal and Torres Strait Islanders may be subject to more variation and be less consistently reported than other Australians. Improving the collection and management of key identifying variables for Aboriginal and Torres Strait Islanders are key to providing more quality information on this population group.

**Key words:** Aboriginal, Research, Data management, Population Health, Informatics

## Introduction

The Council of Australian Governments (COAG) Closing the Gap commitments aim is to overcome key health disparities between Indigenous and non-Indigenous Australians [1]. Another fundamental aspect of achieving the COAG commitments is the capacity to measure health gaps between Indigenous and non-Indigenous people, and to monitor progress in closing those gaps [1]. Data or information collected and organised for analysis and interpretation, can shape, debate and guide policy decisions [2]. Data can be used to identify areas and populations with the greatest need and direct resources accordingly, monitor change and show where

1

outcomes are improving, stagnating or worsening. Most importantly, data can be used to evaluate policies and programs for their efficacy and cost effectiveness, thereby improving individual and collective wellbeing [2]. Accurate data about Aboriginal and Torres Strait Islander people are needed to guide policy formulation, program development and service delivery, as well as to monitor and evaluate the success of Government and community programs in reaching the 'Closing the Gap' targets [3].

Due to the lack of longitudinal information, we know little about the specific causes or consequences of changes in Indigenous outcomes [2]. Lee [4] states that data collected in the past have typically been inadequate to inform service planning. Biddle [2] highlights a consistent lack of information and data on the specific policies and programs that will lead to sustained improvement in Indigenous wellbeing or what some of the unintended consequences of particular programs might be.

Studies such as those by Li and McDermott [5] highlight the use and importance of linked data for Indigenous public health research. Li and McDermott [5] collected baseline data from 2787 adults in 19 rural Indigenous communities across three health districts in far north Queensland. They linked baseline data to hospital separation data using probabilistic linkage to quantify the risk of hospitalization for infections in Indigenous Australian adults with diabetes. This research showed that there was an extremely high background rate of community-acquired infection plus high prevalence of type 2 diabetes among Indigenous Australians leads to excess hospitalization for infections [5]. Population health research such as this are reliant on the ability to link individual patient records.

Evidence-based approaches to inform policy and program development continue to be of paramount importance [6]. The quality of Indigenous status data across key health data sets can be undermined by a range of issues including misclassification, structural limitations and high proportions of 'unknown' Indigenous status [1]. There is evidence that name, date of birth and address variables may be subject to more variation and be less consistently reported among Aboriginal and Torres Strait Islander Australians than among other Australians [3].

**Objectives:** To improve the collection and management of key identifying variables (name, address, and date of birth) for Aboriginal and Torres Strait Islanders for the creation of longitudinal quality information.

## Methods:

Clinical data was collected from 285 Aboriginal and Torres Strait Islanders from 1/1/2012 to 31/12/2015 as part of the project titled: "Primary Health Care Models and the Prevention and Management of Chronic Conditions in Rural and Remote High Risk Populations: A Collaboration between University of South Australia, the Aboriginal Health Council of South Australia, Queensland Aboriginal and Islander Health Council, The Royal Flying Doctor Service and James Cook University". Data was collected from Aboriginal and Torres Strait Islanders from the following Islands: Mabuiag Island, St. Pauls, Murray Island, Boigu Island, Stephen Island, Darnley Island, Yorke Island, Saibai Island, and Badu Island.

Data from the project was stored in the Public Health Research Data Management System at the University of South Australia (PHReDMS) [7]. Using the Australian Institute of Health and Welfare, Meteor Metadata Online Registry, National Health Data Dictionary (NHDD)

Version 16 as a guide for the project, participant's name, date of birth, and address were captured within the PHReDMS (table 1).

Table 1: Format of Person-Name, Person- Date of Birth, Address-purpose of address from Meteor Metadata Online Registry, National Health Data Dictionary (NHDD) Version 16 [8]

| Meteor Variable Name | Representation class | Data Type | Format | Maximum character length | Permissible values | |
|---|---|---|---|---|---|---|
| Person – name | Code | String | AAA | 3 | LGL | Legal name |
| | | | | | AKA | Also known as |
| | | | | | BTH | Name at birth |
| | | | | | MDN | Maiden name |
| | | | | | NEW | New born identification name |
| | | | | | PRV | Previous name |
| | | | | | PRF | Preferred name |
| | | | | | STG | Stage name |
| | | | | | TRB | Tribal name |
| | | | | | OTH | Non-specific name usage type |
| Person - Date of Birth | Date | Date/Time | MMYYYY | 6 | | |
| Address – purpose of address | Code | String | AA[A] | 3 | PR | Primary property address |
| | | | | | SEC | Secondary property address |
| | | | | | TEM | Temporary accommodation |
| | | | | | BUS | Business address |
| | | | | | OVS | Address when overseas |
| | | | | | DEL | Delivery address |
| | | | | | POS | Postal/correspondence address |
| | | | | | OTH | Other address |
| | | | | | UNK | Not stated/unknown |

## Results

### Name

Within the PHReDMS the participant's name has been managed the same way as described in the Meteor Metadata Online Registry, National Health Data Dictionary (NHDD) Version 16 [8]. The participant's name is stored using the following structure:

| | |
|---|---|
| NAME_Title | Prefix to the participants name. |
| NAME_FirstName | The participants first name. |
| NAME_MiddleName | The participants middle or second name. |
| NAME_LastName | The participants surname, family or last name. |
| NAME_NameType | The usage type of a person's family name that enables differentiation between each recorded name change. |

Name type included the following: Legal name, Also known as, Name at birth, Maiden name, New born identification name, Previous name, Stage name, Tribal name, and Non-specific name usage type. For example using the first authors name from this paper she could have 3 different name types associated with her name:

| | |
|---|---|
| Deborah van Gaans | Legal name |
| Deb van Gaans | Also known as |
| Deborah Duncombe-Wall | Maiden name |

Using this structure for a participant's name, the PHReDMS has been able to manage 54 participants with two name types and 10 participants with three name types. The use of name type can be seen in figure 1, which shows that all participants had a "legal name", 11 of those also had "Also known as" name type, which consisted mainly of nicknames, or shortened versions of their "legal name". Thirty one of the participants also had a "Non-specific Name Usage Type", which captured the variations in the spelling of their names.
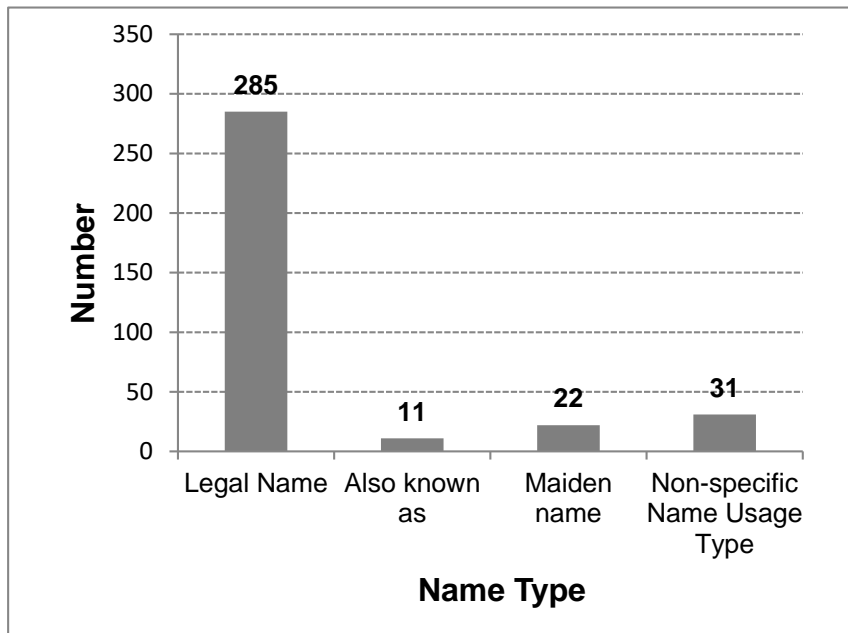


Figure 1: Number of Participants with Different Name Types (n= 285).

### Date of Birth
Within the PHReDMS the participant's date of birth was managed so that it captured the following:

| | |
|---|---|
| BD_BirthDate | Date of birth of the participant |
| BD_PreferredBirthDateFlag | Preferred date of birth |
| BD_Source | The origin of the date of birth |

BD_Source included the following: Primary health care centre, Pharmaceutical benefits scheme, Medicare, Hospital, Laboratory, Royal flying doctor, Database derived, Australian bureau of statistics, Births Deaths and Marriages, Unknown, and Hospital Admission. Handling the date of birth with this structure has allowed multiple birth dates to be recorded for each participant, with 3 participants having a second date of birth.

### Address
Within the PHReDMS the participant's address was managed so that it captured the following:

| | |
|---|---|
| ADD_Line1 | Property name |
| ADD_Line2 | Street number |
| ADD_Line3 | Street name |
| ADD_Suburb | The suburb and postcode |
| ADD_AddressType | The type of a person's address that enables differentiation between each address change. |

ADD_AddressType included the following: Mailing address, Postal address, and Residential address. Using this address structure the PHReDMS was able to capture the 5 participants that had two addresses.

## Discussion

The Australian Institute of Health and Welfare and Bureau of Statistics [9] identify that first and last names are critical linkage variables in both probabilistic and deterministic data linkage. Variations in the spelling of names can affect the quality of the linkage particularly for the spelling of traditional Aboriginal and Torres Strait Islander names which may have a different structure to European-type names, with its own set of nicknames, aliases and diminutions [3]. Lawrance et. al. [10] examined an Aboriginal birth cohort of mobile subjects belonging to diverse cultural and language groups in the Northern Territory and found that Aboriginal children had multiple names relating to kinship, clans and relationships with family groups, and that name changes often occurred following the death of another community member. Lawrance et al. [10] found that, out of a sample of 686 Aboriginal mother-child pairs living in the Top End of the Northern Territory, by the age of four years: 30% of children had changed their names at least once, 18% had changed address once, 2% had had three different name changes, and that 2% had had four different addresses. Using a data structure to be able to capture the name changes is critical if we are intending to build longitudinal datasets on Aboriginal and Torres Strait Islander populations.

There is evidence that some older Aboriginal and Torres Strait Islander people, particularly those living in remote communities, have had difficulties providing date of birth information. In the Northern Territory, for instance, date of birth was not recorded on death registration forms until 1994; only age at death was recorded in prior years [11]. A high

proportion of older Indigenous people in the Northern Territory do not know their exact age and have only an approximate year of birth [11]. Therefore date of birth information on some data sets may be incomplete (it may only include year of birth), or be inaccurate (it may only include approximate year of birth or be approximate date of birth calculated from an approximate age), these quality issues affect the quality of data linkage [3]. By capturing multiple dates of birth and their source it may be able to improve data linkage in the future by providing multiple fields to match the participant with.

According to Memmott et al. [12], there is strong evidence in remote Aboriginal communities of linked households or clustered households that are characterised by an extended family group dispersed across a number of places of residence. People may move between several residences and not regard themselves as having a single usual place of residence [9]. Mobility also affects the stability and completeness of reporting of the address variable, thereby limiting its use as a stable variable for linkage [9]. Memmott et. al. [12] found high levels of mobility among remote Aboriginal and Torres Strait Islander communities, which may result in different levels of reporting for the address variable. For instance, address may be recorded as: town, suburb, community region. Consequently, in some datasets the full address may be recorded, while in others, community name may be the only address information available [3]. Capturing multiple addresses and allowing for different levels of address quality appears to capture Aboriginal and Torres Strait Islander addresses in a culturally appropriate way.

## Conclusions

Variables such as name, date of birth and address variables for Aboriginal and Torres Strait Islanders may be subject to more variation and be less consistently reported among Aboriginal and Torres Strait Islander Australians than among other Australians. Improving the collection and management of key identifying variables for Aboriginal and Torres Strait Islanders are key to providing more quality information on this population group.

## Conflict of Interest

The Authors declare that there are no conflicts of interest.

## Human Subjects Protections

The data for this project forms part of the project titled : "Primary Health Care Models and the Prevention and Management of Chronic Conditions in rural and remote high risk populations: A Collaboration between University of South Australia, the Aboriginal Health Council of South Australia, Queensland Aboriginal and Islander Health Council, the Royal Flying Doctor Service and James Cook University". Ethics approval for this project was granted by the Far North Queensland Human Research Ethics Committee.

## Acknowledgements

# References

1. Australian Institute of Health and Welfare. Towards better Indigenous health data. Cat. no. IHW 93, 2013, Canberra: AIHW.

2. Biddle N. Data about and for Aboriginal and Torres Strait Islander Australians. Issues paper no. 10 prepared for the Closing the Gap Clearinghouse. Australian Government, Australian Institute of Health and Welfare, Australian Institute of Family Studies. July 2014

3. Australian Institute of Health and Welfare and Australian Bureau of Statistics. National best practice guidelines for data linkage activities relating to Aboriginal and Torres Strait Islander people. AIHW Cat. No. IHW 74, 2012, Canberra: AIHW.

4. Lee KSK. Better methods to collect self-reported alcohol and other drug use data from Aboriginal and Torres Strait Islander Australians.Drug and Alcohol Review (September 2014), 33, 466–472 DOI: 10.1111/dar.12159

5. Li M. and McDermott R. High Absolute Risk of Severe Infections Among Indigenous Adults in Rural Northern Australia is Amplified by Diabetes – A 7 Year Follow Up Study. Journal of Diabetes and Its Complications. 30 (2016) 1069-1073.

6. Australian Institute of Health and Welfare & Australian Bureau of Statistics. Recent developments in the collection of Aboriginal and Torres Strait Islander health and welfare statistics 2005. AIHW cat. no. IHW 15; ABS cat. no. 4704.0.55.001, 2006, Canberra: AIHW & ABS.

7. van Gaans D, D'Onise K, Cardone T, McDermott R. "The Development of the Public Health Research Data Management System." Electronic Journal of Health Informatics (2015) Vol. 9(1): e10

8. Australian Institute of Health and Welfare, Meteor Metadata Online Registry, National Health Data Dictionary (NHDD) Version 16. Accessed 25th October 2013, http://meteor.aihw.gov.au/content/index.phtml/itemId/274816

9. Australian Bureau of Statistics. Population Characteristics, Aboriginal and Torres Strait Islander Australians, 2006 ABS cat. no. 4713.0. 2010; Canberra: ABS.

10. Lawrance M, Sayers SM, Singh GR. Challenges and Strategies for cohort retention and data collection in an indigenous population: Australian Aboriginal birth cohort. BMC Medical Research Methodology 2014; 14:31 doi:10.1186/1471-2288-14-31

11. Condon J, Barnes A, Cunningham J, Smith L. Demographic Characteristics

and Trends of the Northern Territory Indigenous Population 1966 to 2001. Darwin: Cooperative Research Centre for Aboriginal Health, 2004

12. Memmott P, Long S, Bell M, Taylor J, Brown D. Between places: Indigenous mobility in remote and rural Australia. AHURI Positioning Paper No. 81, 2004; Melbourne: Australian Housing and Urban Research Institute, Queensland Research Centre, RMIT University.