Government of **Western Australia**
Department of **Health**

Review of Cultural and Linguistic Diversity (CaLD) Data Collection Practices in the WA Health System

Epidemiology Branch
Public and Aboriginal Health Division
Western Australian Department of Health

February 2018

Suggested citation

Forrest S (2018). Review of Cultural and Linguistic Diversity (CaLD) Data Collection Practices in the WA Health System 2018. Department of Health, Western Australia.

For enquiries contact:

Epidemiology Branch
Public and Aboriginal Health Division
Western Australian Department of Health

189 Royal Street
East Perth WA 6004
Email: epi@health.wa.gov.au

Acknowledgements

The author would like to acknowledge the assistance of staff within both the Epidemiology Branch and Cultural Diversity Unit of the WA Department of Health, and public health training supervisor, Dr Suzanne McEvoy.

Contents

Glossary of terms

Culturally and linguistically diverse (CaLD): Groups and individuals who differ according to religion, race, language or ethnicity, except those whose ancestry is Anglo Saxon, Anglo Celtic, Aboriginal or Torres Strait Islander.

CaLD-related data or CaLD data: data variables or parameters that measure those attributes of persons that relate to their cultural or language background.

WA health system: health services publicly funded and administered by the Western Australian Government. Consists of the WA Department of Health, five Health Service Providers and Health Support Services.

WA Department of Health: the 'System Manager' of the WA health system, responsible for the overall management, performance and strategic direction of the health system as a whole.

List of key abbreviations

ABS	Australian Bureau of Statistics
AIHW	Australian Institute of Health and Welfare
CaLD	Culturally and linguistically diverse
Census	Australian Bureau of Statistics' Census of Population and Housing
EDDC	Emergency Department Data Collection
HCare	Health Care and Related Information Systems
HMDC	Hospital Morbidity Data Collection
HMDS	Hospital Morbidity Data System
LASS	Language Services System
METeOR	Metadata Online Registry
MHIS	Mental Health Information System
NESB	Non-English Speaking Background
PAS	Patient Administration System
PSOLIS	Psychiatric Online Information System
The Standards	Australian Bureau of Statistics' Standards for Statistics on Cultural and Language Diversity
TOPAS	The Open Patient Administration System
webPAS	Web-based Patient Administration System

Executive Summary

The WA health system is committed to equal opportunity and respect for diversity. The initial step in being able to achieve an equitable health service for all Western Australians is identifying disadvantaged groups. Culturally and linguistically diverse (CaLD) individuals or groups may have poorer health outcomes or access issues across multiple levels of the health service. It is therefore essential health services are collecting appropriate data to assist in identifying, measuring and addressing the level of disadvantage in CaLD populations. This report represents an initial review into CaLD data collection practices across key WA health services.

The project aimed to review current CaLD data collection practices and make preliminary recommendations for improvement. Specific objectives narrowed the scope of the project to focus particularly on the data variables for measuring CaLD, with a further smaller secondary focus on data quality issues. The review concentrated on public health services funded and administered by the WA Government.

Three key steps were undertaken to address the project's objectives:

- literature review of CaLD data variables and data quality issues
- desktop and key stakeholder review of large core data sets held by the WA Department of Health
- stakeholder consultation of health services seeking feedback on current CaLD data collection practices and priorities for improvement.

Measuring 'ethnicity' and related concepts is a complex issue worldwide. In Australia, it is generally accepted that a surrogate combination of variables provides optimal CaLD data. In 1999, the Australian Bureau of Statistics (ABS) developed Standards for Statistics on Cultural and Language Diversity which included a Minimum Core Set plus a number of optional variables recommended for collection across service settings. The Minimum Core Set includes the CaLD variables: Country of Birth of Person, Main Language Other Than English Spoken at Home and Proficiency in Spoken English. However, the literature demonstrated that most health jurisdictions have not implemented this Minimum Core Set, but rather are often solely collecting Country of Birth of Person. Some health settings are also collecting Interpreter Service Required and Preferred Language. Minimal discussion of data quality issues existed in the reviewed literature.

The seven large core data sets held by the WA Department of Health (relating to health events such as emergency department presentations, hospital admissions, births and deaths) are collecting: Country of Birth of Person most commonly, Interpreter Service Required by some and Preferred Language by a few. None are collecting the ABS Minimum Core Set. Several issues were identified with data quality, including lack of mandatory collection, poor data dictionary guidance, suboptimal software functionality, non-standardised response options and poor staff training. In addition, data collected on primary software packages such as the widely used Patient Administration System (PAS) is not always transmitted through to core data set collections.

A survey of stakeholders received 72 responses from a range of services and staff roles. Primary data collection tools reported reflected the predominance of PAS as either the user interface or for its function to auto-populate to other software packages. Most services had used their CaLD data for service reporting or informing individual care in the last five years, and few had completed a quality assurance process. Data quality issues identified were similar to those identified in the core data set review. There was strong support to improve CaLD data collection, with 'optimisation of collected variables' the highest ranking measure for improvement. A strong theme was the desire to increase the capacity to share data, including via the PAS auto-populating function.

In line with other health jurisdictions, most WA health system services are not collecting the ABS Minimum Core Set, but are instead collecting Country of Birth of Person and sometimes Interpreter Service Required and Preferred Language. It is likely this reflects the health system's requirement to focus on service needs, rather than statistical measures of the CaLD population's health. Proficiency in English is not likely to be an adequate substitute for Interpreter Service Required, and Preferred Language is important for informing the use of interpreters. However, failing to include the Minimum Core Set variables beyond Country of Birth of Person severely limits the capacity to calculate population-level trends because numerator data is in a non-compatible format in comparison to denominator CaLD data obtained from the ABS Census. Currently CaLD data collection in the health setting is an unresolved issue nationally with no stand out or agreed combination of variables. A number of data quality and data collation issues were also evident across WA health system services.

It is recommended the WA health system develops its own 'minimum core set' for interim use across most health services. It is proposed this include the variables Country of Birth of Person, Interpreter Service Required, Preferred Language, Indigenous Status and Main Language Other Than English Spoken at Home. Main Language Other Than English Spoken at Home would act as a filter question to Interpreter Service Required which then prompts the Preferred Language question. It is also recommended the WA health system works with other health jurisdictions and the ABS over coming years to develop a standardised set for collection across health services nationally. Several further recommendations relating to data quality and data sharing are made, along with the recommendation to establish a working group to further pursue the issue.

This project makes a number of recommendations to address the issue of CaLD data collection in relation to both variables collected, and the quality of data collection.

1 Introduction

The WA health system is committed to equal opportunity and diversity, evident in the *Western Australian Language Services Policy 2014* and the *WA Charter of Multiculturalism 2004*, and supported by the *Equal Opportunity Act 1984*. The WA health system's *Strategic Intent 2015-2020* has a vision to deliver a safe, high quality, sustainable health system for all Western Australians. It is comprised of strategic priorities focused on a continuum of care through integrated service delivery from prevention and health promotion, early intervention, primary care through to diagnosis, treatment, rehabilitation and palliation.

Identifying disadvantaged groups is the first step in being able to achieve an equitable health service for all Western Australians. CaLD individuals or groups may have poorer health outcomes or access issues across multiple levels of the health service. Quality data collection is therefore crucial in being able to adequately identify and measure the extent of challenges faced by CaLD populations.

Amidst the increasing cultural and linguistic diversity of the Australian population, there is increasing interest in identifying CaLD service users both in terms of developing tailored programs and ensuring equitable service delivery. The Epidemiology Branch of the WA Department of Health has received several documented CaLD data requests annually over recent years. The Assistant Director General of the WA Department of Health requested a review into the WA health system's current data collection practices in relation to cultural and linguistically diversity. This report represents an initial investigation into current practices and makes a number of recommendations for improvement.

2 Background

2.1 Defining cultural and linguistic diversity (CaLD)

It is important to define the term cultural and linguistic diversity before considering related health and data collection implications. Cultural and linguistic diversity (CaLD) refers to the range of different cultures and language groups represented in a population. In popular usage, CaLD communities are those whose members identify as having non-mainstream cultural or linguistic affiliations by virtue of their place of birth, ancestry or ethnic origin, religion, preferred language or language spoken at home (Victoria University, 2009). When it comes to information about Aboriginal people, there may be different experiences, perspectives, expectations, requirements, preferences and needs. It is recommended that the needs of Aboriginal Australians be considered separately, rather than under the framework of cultural and linguistic diversity (Department of Human Services Victoria, 2006; Cultural and Indigenous Research Centre Australia, 2017). Aboriginal people have a unique place in Australia as the original inhabitants of the land and hold distinctive rights as Australia's First People (Australian Human Rights Commission, 2014). Aboriginal cultures and communities may differ considerably and often have specific geographic characteristics that are varied across metropolitan, regional and remote areas (Neckowaya, Brownlea and Castellana, 2007). Providing health and health care information, and planning and delivering services for Aboriginal people needs to be approached with consideration and in the context of individual, family, community, social, cultural and historical factors (Australian Institute of Family Studies, 2011).'Non-English speaking background' (NESB) is no longer considered an appropriate measure of culturally-related disadvantage predominantly because it is an oversimplified indicator of disadvantage which may result in inappropriate service provision, and neglects the positive aspects of cultural and linguistic diversity (ABS, 1999).

This report will use the definition assumed by the Western Australian Office of Multicultural Interests: "Culturally and linguistically diverse (CaLD) includes groups and individuals who differ according to religion, race, language or ethnicity, except those whose ancestry is Anglo Saxon, Anglo Celtic, Aboriginal or Torres Strait Islander" (Office of Multicultural Interests, 2010).

In the Australian context, CaLD relates particularly to migrants from developing countries and/or countries where the predominant language(s) is not English, and the children of these migrants.

2.2 Cultural and linguistic diversity in Western Australia

The results of the 2016 Australian Census indicate Western Australia is a culturally diverse state, consistent with a long history of migration to the nation. Nearly half (49%) of Australians were born overseas or have a parent who was born overseas. WA was the state with the largest proportion of residents born outside of Australia (32%), with an increase from 31 per cent reported in the 2011 Census (ABS, 2017a). In 2016, 16 per cent of the total WA population was born in non-main English speaking countries (Supreme Court of Western Australia, 2017). The 2016 Census also reveals that those born overseas were more likely to live in a capital city (83%), a much higher percentage than for people born in Australia (ABS, 2017b).

In 2016, there were nearly 300 separately identified languages spoken in Western Australian homes. Within the Western Australian population, 18 per cent spoke a language other than English at home. After English, the next most common languages spoken in WA homes were Mandarin, Italian, Vietnamese and Cantonese (ABS, 2017d). Results available at a national level demonstrate 10.5 per cent of those born overseas could not speak English well or at all (ABS, 2017d).

The latest Census data highlights that Western Australia is a religiously diverse state, with Christianity remaining the most commonly reported religious affiliation (49.8% of the total population). Buddhism was the second most common religion (2.1% of the total population) closely followed by Islam (2.0%) (ABS, 2017d).

The current profile of the CaLD population reflects the history of migration to Australia subsequent to European settlement. Since World War II, two permanent migration streams have emerged: migration of skilled and family migrants, and resettlement of people in humanitarian need. Generally the former is much larger than the latter. In 2015, there were 189 770 places for skilled and family migrants and 17 555 for humanitarian entrants (Phillips & Simon-Davies, 2017). As a proportion of the total population, Australia's overseas-born population has grown in the last 50 years – increasing from approximately one in five Australians (18%) in 1966 to one in four (26%) in 2016. Of the overseas-born population in 2016, migrants came from over 190 different countries and nearly one in five (18%) had arrived since the start of 2012 (ABS, 2017c). Predominant countries of birth have changed over time; European migration featured throughout the 20th century, and there has been a marked increase in migration from China, India, Vietnam and the Philippines since 1975 (ABS, 2017c). At the time of the 2016 Census, England and New Zealand were still the next most common countries of birth after Australia nationally, however of the people born overseas the proportion born in China or India increased since 2011 (from 6% to 8.3%, and 5.6% to 7.4%, respectively) (ABS, 2017c). Given the changing source countries over time, older migrants in Australia are most likely to have been born in European countries, while younger people are relatively more likely to have been born in New Zealand or countries throughout Asia (ABS, 2017c).

In 2015–2016, Australia granted 17 555 Humanitarian visas to refugees and others in refugee-like situations (Department of Home Affairs, 2017). The top five countries of origin for those arriving via the offshore resettlement component of the programme were Iraq, Syria, Burma, Afghanistan, and the Democratic Republic of Congo (Department of Home Affairs, 2017). This reflects a focus in recent years to resettle predominantly refugees from the Middle East and South West Asia.

2.3 Relevance of cultural and linguistic diversity to health and wellbeing

Ethnic, religious and linguistic background create a range of influences that have an ongoing impact on physical and mental health status and needs throughout the life course. These influences are particularly significant during the early settlement period in a new country and especially significant for some high needs groups like refugees. Impacts can extend beyond the first generation to second-generation migrants.

There has been a relative dearth of academic research on the health status and challenges of CaLD groups in recent years. A 2010 systematic review of three major Australian health journals found just 2.2 per cent of total articles were primarily based on multicultural issues, and some communities and health issues were essentially invisible or unrepresented in research (Garrett, Dickson, Whelan, & Whyte, 2010). Government census or survey CaLD data is more frequently summarised at a national level and not readily available for WA. The following sections summarise the current understanding of health in CaLD populations.

2.3.1 Mortality and morbidity

In general, overseas-born residents have better health than Australia-born persons for measures related to mortality, such as life expectancy and mortality rates (AIHW, 2004, Anikeeva et. al., 2010). These inequalities are likely explained by the 'healthy migrant effect', which ensures that, for the most part, only those migrants in good health migrate to Australia. However, this effect diminishes over time with increasing duration of residence in Australia (Anikeeva et al., 2010).

Evidence demonstrates generally greater variation in morbidity outcomes than mortality, with certain diseases and health risk factors more prevalent among some migrant groups (Vang, Sigouin, Flenon, & Gagnon, 2015). Reported patterns are listed below.

- Migrants originating from regions of the globe where particular diseases are more widespread have ongoing increased rates of such diseases. These include infectious diseases such as tuberculosis, and some diseases of genetic or biological origin (e.g. haemoglobinopathies or regional differences in malignancy patterns respectively (Gushulak et al., 2011)).
- Higher cervical cancer incidence and mortality exists in some migrant groups in Australia and overseas (Aminisani, Armstrong, Egger, & Canfell, 2012; Vang et al., 2015).
- Higher rates of diabetes and hypertension and lower rates of mental illness and asthma is present in the migrant population overall, and migrants most often present later in an illness (ABS, 2015; Anikeeva et al., 2010; Gushulak et al., 2011).
- There is strong foreign-born health advantage in adulthood but less so in the perinatal period, childhood/adolescence, and later life (Vang et al., 2015). Children are particularly at risk from suboptimal health due to the impact of resettlement stresses on parents' ability to care for their children (Davidson et al., 2004).
- Migrants show frequent poor perinatal use or outcomes. For example, mothers born in Lebanon, New Zealand, Fiji, Iraq, Pakistan, Korea, China, Indonesia, Vietnam and the Philippines are less likely to have their first antenatal visit before 20 weeks' gestation and mothers born in Italy, Fiji, the Philippines and New Zealand are more likely to have premature babies (NSW Department of Health, 2010). Immigrant women had worse maternal health than Canadian-born women in a systematic review, with mental health among immigrant mothers especially poor (Vang et al., 2015).
- The physical and psycho-social welfare of older people from CALD backgrounds is affected by a diverse range of factors. For example, cultural and language barriers, social supports, migration circumstances (time since arrival, education, trauma experiences), as well as current geographical location (Rao, Warburton, & Bartlett, 2006). Bilingual older migrants can lose ability to speak English as a second language, especially for people suffering from dementia.

- There is evidence of variable hospitalisation rates e.g. higher rates of admissions for diabetes or its complications (people born in Lebanon and the Philippines), for coronary heart disease (people born in Lebanon, Fiji, Sri Lanka and Iraq) and cardiac revascularisation procedures (people born in Fiji, Lebanon, Sri Lanka, Greece, Indonesia, India, Italy and Iraq) (NSW Department of Health, 2008, 2010), and for gastritis and duodenitis among persons born in Continental Europe and Asia (AIHW, 2004).

Refugee and asylum seeker populations in particular have a number of recognised health needs including (Chaves et al., 2016):

- psychological disorders such as post-traumatic stress disorder, anxiety, depression and psychosomatic disorders
- direct physical consequences of war or torture such as musculoskeletal pain, injury or deafness
- under-recognised and under-managed chronic non-communicable conditions such as hypertension, diabetes, vision deficit or chronic pain
- poor oral health, secondary to nutrition and diet, lack of fluoridated water, poor dental hygiene practices and/or limited dental care access
- infectious diseases including tuberculosis, intestinal parasites, malaria and chronic hepatitis B, often due to inadequate immunisation
- delayed growth or development in children and poorly managed disability
- vitamin and nutritional deficiencies, such as vitamin D deficiency or anaemia
- sexual health issues, including female genital mutilation and stigma related to infections
- poor maternal and infant health, particularly among refugees, across multiple mental and physical health outcomes (Hadgkiss & Renzaho, 2014; Vang et al., 2015).

2.3.2 Factors affecting CaLD health

2.3.2.1 Migration and settlement

Immigration and settlement can impact adversely on the physical and/or mental health of both individuals and communities. This is particularly evident for refugee entrants, however there is evidence that factors associated with immigration and settlement in a different country can negatively affect the health status of all migrants (NSW Department of Health, 2008). The vulnerabilities of refugee groups in Australia result from pre-settlement experiences including physical and psychological trauma and torture and the deprivation of food, clean water, sanitation, shelter, education and health care in their countries of origin or transit. Exacerbating factors once in Australia include limited English proficiency, cultural differences, lack of knowledge of the health system, racism and discrimination, and socioeconomic disadvantage. Initial arrivals from a region face extra challenges, including social isolation and inaccessibility of interpreters, without the support networks associated with established migrant communities.

2.3.2.2 Health service access and equity

CaLD populations, particularly those from refugee and asylum seeker backgrounds, face recognised barriers in accessing and using health services, further contributing to health inequities and adverse healthcare events (Day, 2016; Department of Health & Human Services, 2015; Murray & Skull, 2005; Sheikh-Mohammed, Macintyre, Wood, Leask, & Isaacs, 2006). Issues include those related to physical access, in particular: increased distance to healthcare services and decreased access to transport; perceived or actual cost of healthcare; unfamiliarity with services; competing life priorities such as securing or maintaining employment; and denial of access to Medicare for certain asylum seeker

visa categories. A further barrier is the cultural inappropriateness of some health services, relating to institutionalised or overt racism, time constraints, and staff who may be inadequately trained in culturally appropriate care or the health needs of CaLD groups. Health staff may tend not to use interpreters when appropriate leading to miscommunication, misdiagnosis, under- or over-use (longer hospital stays, readmissions, non-attendance at appointments), dissatisfaction with treatment services, and risk of adverse events (Health Research and Educational Trust, 2011). These factors can ultimately lead to increased costs for the health system. Diverse health beliefs, mistrust of government related to historical experiences and reduced health literacy can affect attitudes to health, health care and expectations of the health system.

The recent National Health Survey suggested reduced use of health services by migrants overall, particularly those recently arrived, across services including general practices, specialists, dentists and admissions to hospital. Of those arriving in Australia between 2009 and 2015, 74 per cent had visited a GP in the last 12 months and less than 8 per cent had been admitted to hospital, compared to the Australian born population total of 86 per cent and 12 per cent, respectively (ABS, 2017e). In addition, 46 per cent of people who spoke a language other than English at home had private health insurance, compared to 59 per cent of English speakers at home (ABS, 2017f).

There is evidence for reduced use of screening services for preventative care within CaLD populations. BreastScreen Australia found a lower rate of participation (49%) in mammograms in CaLD women compared to 55 per cent in the population overall over the period 2012-2014 (AIHW, 2016). In addition, there is evidence that cervical cancer screening uptake was lower in multiple migrant groups (ABS, 2002; Taylor, Mamoon, Morrell, & Wain, 2001). Reduced cancer screening uptake was a particular problem for those speaking another language at home and for those recently arrived, with only 28.6 per cent of those speaking a language other than English at home tested for any type of cancer in the last two years, compared to 48 per cent for the English speaking population, and 21.8 per cent of people arriving in Australia between 2009 and 2013 tested for any type of cancer in the last two years, compared to 48 per cent in the Australian-born population (ABS, 2017f).

There is evidence of varying use of mental health services and treatment for CaLD groups. Of people who spoke English at home, 7.9 per cent accessed a Medicare Benefits Schedule subsided mental health-related service in 2011, compared to 5.8 per cent of those speaking a language other than English at home (ABS, 2016d). The rate of dispensing of prescription medications for mental-health related medications also varied by country of birth and language spoken at home (ABS, 2016d).

2.3.2.3 Individual behavioural factors

Individual level behaviours affect health by either increasing risk or being protective for developing 'lifestyle-related' diseases. Commonly considered behaviours include diet, smoking, alcohol and physical activity. Individuals from some CaLD populations might have a reduced risk due to a healthier diet (e.g. traditional Mediterranean diet). Conversely, some CaLD populations have been found on average to have higher rates of smoking, physical inactivity or being overweight or obese. For example, the 2014–2015 National Health Survey found that 73.8 per cent of people who spoke a language other than English at home did no or low levels of exercise, compared to 65 per cent for those that spoke only English at home. In the same survey, the level of smoking varied by country of birth: 4.8 per cent of those born in Southern and Central Asia undertook daily smoking compared to 18.1% of those born in 'other Oceania' countries and 16.4% of those born in Australia (ABS, 2015). In general, evidence suggests behavioural risk factor patterns trend closer to the non-migrant population with increasing time since arrival in Australia (National Health and Medical Research Council, 2005).

It is important to note caution should be exercised in interpreting the ABS results as it is unclear whether results have been age standardised.

2.4 Cultural and linguistic diversity data collection in the healthcare setting

Accurate and consistent identification of persons from CaLD backgrounds in the health setting is important to ensure culturally appropriate service delivery and to address disparities in health outcomes occurring in particular CaLD groups. It is even more essential in the context of poor reporting and research related to CaLD communities across Australia in recent years. Effective and efficient services require that policy makers and program managers identify and measure the impact of policies and programs on potentially disadvantaged groups of the population, including those with various language and cultural characteristics. Organisations can more effectively undertake meaningful assessments of the impact of programs on individual cultural and language groups if they collect and analyse standardised data.

The importance of consistent collection and recording of CaLD information is relevant across multiple levels of healthservice delivery. There are implications at the individual worker level for clinical contact, planning and treatment; at the service level for handover, strategic planning and development; and at the broader whole-of-sector level for policy, strategic planning and resource allocation. Quality CaLD data may assist specifically with:

- organisation of appropriate interpreting services and avoiding under- or over-use of interpreters
- assessing client needs and responding effectively and in a culturally appropriate way to individuals and assisting to avoid medical errors
- assisting with diagnosis of disease more common in certain ethnic groups
- understanding usage patterns for interpreters and translated materials
- assisting with identifying unmet need
- assessing client satisfaction with services
- profiling of clients and service usage patterns
- assessing and measuring the impact of policies and programs on different CaLD groups
- recognising where discrimination, institutional racism and marginalisation may be arising
- enabling comparisons in rates of disease or access issues across groups or geographical areas via use of population data gathered through the ABS Census of Population and Housing
- providing a starting point to address likely issues with longer hospital stays, more frequent readmissions, increased rates of tests due to difficulty with communication, or more frequent non-attendance at appointments
- assessing and addressing increased potential for adverse events
- designing services to target recognised health disparities in particular CaLD populations
- assisting with designing a culturally appropriate service more broadly
- informing CaLD-specific policy, budgets, programs and research
- assisting with funding provision, in particular the move to Activity Based Funding may include a payment attached to the measure Interpreter Service Required
- improving staff cultural competency, including appropriate workforce recruitment and training.

The importance of CaLD data collection should be considered from multiple healthcare quality perspectives including: effective care; equity; patient-centred care, consumer knowledge and empowerment; resource efficiency and financial impacts; safety, avoiding adverse events and managing legal risk; regulatory and accreditation implications; and public perception of the health service (Victoria University, 2009). Overall, CaLD data can benefit CaLD patients, their families, and health service staff members.

3 Aim and Objectives

3.1 Aim

Review current data collection practices in relation to cultural and linguistic diversity across WA health services and make recommendations for improvement.

3.2 Objectives

1. Understand data variables available for measuring CaLD
2. Review recognised CaLD data quality issues
3. Understand current CaLD-related variables collected by relevant WA health services
4. Explore important CaLD data quality issues present in WA
5. Identify opportunities and priorities for improvement in health service CaLD data collection

4 Methodology

The project was undertaken part-time over a 12-month period as an initial review into health service CaLD data collection practices in WA. The methodology was designed to focus on identifying key existing issues and options for improvement.

Three steps were undertaken to address the project's objectives. These were:

- literature review of CaLD data variables and data quality issues
- desktop and key stakeholder review of large core data sets held by the WA Department of Health
- stakeholder consultation of health services seeking feedback on current CaLD collection practices and priorities for improvement.

The scope of the project was narrowed consistent with resources. The project focused on the variables for measuring CaLD background rather than comprehensive assessment of data quality, processing or use issues. Review of the feasibility and process for implementation of recommendations was beyond the scope of the analysis. The project focused on public health services funded and provided by the WA government, and data collection relevant to healthcare access and health outcomes rather than risk factors or determinants of health.

4.1 Literature review

A review of the literature was undertaken to compare the range and relative effectiveness of variables for measuring cultural and linguistic diversity relevant to health service provision. Preliminary assessment was also made of recognised data quality issues.

The peer-reviewed literature was searched via databases Pubmed, Google Scholar and Cochrane Database for international reviews and Australian studies using combinations of search terms including: 'CALD', 'cultural', 'minority', 'diversity', 'refugee', 'ethnic', 'migrant', 'migration', 'race', 'Australia', 'data collection', 'monitoring', 'health' and 'culturally and linguistically diverse'. Google Scholar results were not reviewed beyond the third page of results due to decreasing relevance to search terms.

Searching was otherwise limited to identifying Australian federal, Victorian, New South Wales and Western Australian government department reports and publications via contacting existing intergovernmental networks or searching government websites. Website searching was conducted via sites listing all government agencies and departments and selecting those dealing specifically with CaLD populations, language services and/or health. Departments or agency websites were searched for relevant publications using the search terms 'CaLD', 'cultural', 'ethnic', 'data collection' or 'data standards' or 'monitoring' and browsing the websites for relevant publications.

Inclusion criteria

Articles were included if they were in the English language and were Australian or international literature, or academic peer-reviewed Australian studies or international reviews published since 1998.

Exclusion criteria

Academic articles were excluded if the title and abstract clearly did not confirm a focus on CaLD-related data collection relevant to healthcare provision. Literature was excluded if title, executive summary or contents pages did not include specific reference to CaLD data collection practices.

4.2 Review of WA health system core data sets

Large aggregated data sets currently held by WA Department of Health were examined for current CaLD-related data collection practices. These large or 'core' data sets were examined because they represent the predominant data collections relating to the majority of the healthcare provision provided by the WA Government (and coincidentally WA private hospitals). The core data sets represent a total population record across the WA population of important health events including emergency department presentations, hospital admissions, births and deaths. The core data sets act as the predominant data source for researching and studying disease outcomes in the WA population and for evaluating and planning healthcare services and policy. It is therefore important that the core data sets act as the initial focus for CaLD data collection improvement.

CaLD data practices relating to the following core data collections were assessed:

- Hospital Morbidity Data Collection
- Emergency Department Data Collection
- Mental Health Information System
- Non-Admitted and Patient Waitlist Data Collection
- Midwives Notification System
- WA Cancer Register
- Mortality Data Set.

WA Department of Health Epidemiology Branch core data set records were examined via desktop review for the selection, completion rate and accuracy of CaLD-related variables where possible. It should be noted; however, that while the core data sets are primarily maintained by data custodians within the Data Collections Directorate of the WA Department of Health, this is not the only process for data collection so the records may not reflect the entirety of core data set data collected by WA Department of Health.

Manuals developed by the WA Department of Health guide the primary data collection for most of the core data sets. These manuals, where available, were examined to assess the quality of guidance given for primary data collection and understand any potential reasons for the selection, quality and completeness of core data set records.

Observation of primary data collection at clinical sites and verbal consultation with WA Department of Health data custodians was conducted to understand any discrepancies and collation processes between primary collection and WA Department of Health records and potential reasons for incomplete or inaccurate data. Observation of primary data collection was limited to one service using each of the commonly used WA health system software packages 'The Open Patient Administration System' (TOPAS) and 'Web-based Patient Administration System' (webPAS). Data custodians were verbally consulted in relation to the Hospital Morbidity Data Collection, Emergency Department Data Collection, Mental Health Information System and Non-admitted and Patient Waitlist Data Collection.

4.3 Stakeholder consultation

4.3.1 Development

A survey of a wide range of relevant health services and professions was conducted with the aim of identifying common CaLD data collection practices and gathering suggestions for improvement.

The target audience was defined as health programs or services with a key focus on CaLD clients and health services likely to have a significant proportion of CaLD users. In addition, the survey was designed for completion by staff with a variety of roles relevant to CaLD data collection or use including data custodians, clerks and coders, and clinical staff.

The content of the survey was informed by the Project's literature review and expert consultation with the Survey Team of the Epidemiology Branch of the WA Department of Health. Consideration was given to appropriate length, language, question and answer content, sequence, and mandatory versus optional responses. The survey was piloted with multiple Epidemiology Branch and clinical staff.

The final survey had a total of 21 question stems, relating to both current data collection features and problems, and future priorities and suggestions for improvement. Questions specifically focused on:

- data collection tools or software packages
- variables collected
- use of CaLD data collected
- data quality issues
- quality assurance processes
- suggestions for additional CaLD variables for collection
- improvements that would result from better data collection
- priorities for improvement of CaLD data collection
- general comments
- suggestions for improvement of CaLD data collection by other WA health services.

A full copy of the survey appears in Appendix 1.

4.3.2 Conduct

Programs were invited to participate in the survey either via direction arising from requests to Health Service Provider Chief Executives or via direct email invitation to a collated list of relevant services. To increase uptake, written support from the Acting Assistant Director General was added to invitations. Each program was asked to provide a small number of representative responses across different staff roles.

Given the resources available, invited services were narrowed to those within the WA health system and the Mental Health Commission. The exception was St John Ambulance and the Disability Services Commission which was asked to complete the survey given the likely system changes occurring with the introduction of the National Disability Insurance Scheme.

The consultation was initially opened for a period of five weeks, but was extended to seven weeks to facilitate additional responses. Direct email invitations were sent one reminder email if they had not completed the survey within a week of the initial closing date. Health Service Providers, programs from important health areas with a poorer response, and organisations external to the WA health system were contacted via phone if a response was not received following an initial letter or email.

Two question explanations were adjusted slightly during the survey period to counter a significant proportion of people not completing their personal details and poorly entering responses to question 1.

4.3.3 Analysis

Results were analysed using Microsoft Excel and SAS statistical package software. A frequency table of responses was run for each question. Free text responses were assessed qualitatively to identify themes and important points. Answers that fell out of scope for a particular question were omitted from further analysis of that question, but where possible were included in the analysis of a more appropriate question. Further analysis was beyond the scope of the review.

5 Results

5.1 Literature review

Overall, the key word search of the academic literature yielded 500 articles. Thirty-three articles were reviewed further based on title or abstract and excluded if they:

- did not directly relate to data collection and analysis of CaLD
- did not directly relate to healthcare provision or CaLD service provision in general
- focused on collection of ethnicity data in clinical trials or self-report instruments.

Nine academic articles were relevant for final inclusion, including just four at the Australian level, with the remainder being narrative reviews on ethnicity data collection practices in Europe, the United Kingdom, the USA or more generally. The four Australian studies were primary research examining data collection across health data sets generally or in relation to specific fields of health, for example mental health and type 2 diabetes (Abouzeid, Bhopal, Dunbar, & Janus, 2014; Blignault & Haghshenas, 2005; Minas et al., 2013). One Australian study examined the reliability of country of birth as a proxy measure for self-identified ethnic group in maternal women (Porter, Todd, & Zhang, 2016).

Twenty literature documents were eligible for inclusion, with 15 from the Australian setting and the rest from the United Kingdom or the USA.

The following text summarises common themes and important findings contained within the reviewed academic and literature, including in relevant policies and standards.

5.1.1 Measuring cultural, linguistic and ethnic diversity

Measuring cultural and linguistic diversity is complex. The core tenet of ethnicity has multiple definitions and is a subjective, composite construct. Ethnicity has been described as the 'group a person belongs to, and either identifies with or is identified with by others, as a result of a mix of cultural and other factors including language, diet, religion, ancestry, and physical features…' (Bhopal, 2004), but there are various definitions. Adding to the definitional complexity, some of these dimensions used to conceptualise ethnicity are themselves multidimensional, including the importance of nationality or ancestry (Abouzeid et al., 2014; ABS, 1999). For example, someone may be an Australian citizen or consider themselves to bear an Australian identity, but this sense of identity may also be shaped by the ethnic background of ancestors prior to migration, for example Italian, Somalian or Chinese. Competing concepts of ethnicity may thus create measurement difficulty in recording a single ethnicity. Ethnic identity may also change over the life course.

Ethnic data collection is an ongoing global issue. There have been calls for more development of the theoretical understandings of ethnicity and for a standardised international approach for how best ethnicity should be conceptualised and measured in explaining ethic inequalities in health (Abouzeid et al., 2014; Aspinall, 2001). However, standardising the approach internationally would be difficult because the various ways healthcare providers and researchers currently collect, interpret and use culture and ethnicity data of patient populations is very much dependent on the local context (Abouzeid et al., 2014). Historical and political contexts, administrative and political structures, welfare regimes and immigration histories can all affect relevant measures of diversity (Rechel, Mladovsky, & Devillé, 2012). In general, chosen measures are used to inform the basis for equitable distribution of power and resources (ABS, 1995).The Indian Census, based on the most multicultural society in the world, measures language, religion and caste. European data collection is variable across countries. Country of birth and citizenship only is collected in France and ethnicity is avoided in Germany due to historical concerns around categorisation (Rechel et al., 2012). In the UK and the USA, federal health data collection includes explicit variables on race and/or ethnicity, while in Australia the predominant practice has been to make use of one or more surrogate variables such as country of birth and languages spoken but not generally 'ethnicity' (aside from Indigenous Status) per se (ABS, 1999; Borrie, Beaurepaire, Kiosoglous, Sheldrake, & Zubrzycki, 1984; NHS, 2005; Wallman, Evinger, & Schechter, 2000).

It is widely agreed that in Australia there are many elements to cultural and language diversity which must be considered to provide an accurate measure of cultural and language background or ethnic diversity. By definition the term CaLD itself reflects a diverse range of groups. Therefore it is generally recognised that in the Australian context no single measure is adequate, but rather a combination of variables is likely to be more useful in answering particular policy questions in terms of advantage or disadvantage related to cultural and language background (ABS, 1999).

Prior to 1999 in Australia, non-English speaking background (NESB) was used as a broad measure of culturally related need or disadvantage. There was an official shift away from NESB as an appropriate measure of culturally related disadvantage in terms of access to government services for a variety of reasons, including but not limited to (ABS, 1999):

- the term has many conflicting definitions
- it groups people who are relatively disadvantaged with those who are not disadvantaged
- it is unable to separately identify the many cultural and linguistic groups in Australian society
- it has developed negative connotations.

5.1.2 Variable options

In 1999, the ABS developed Standards for Statistics on Cultural and Language Diversity in response to a widely recognised need for a nationally consistent framework for the collection and dissemination of data on cultural and language diversity (ABS, 1999). The aim was to increase ability of all government agencies to capture a common core set of cultural indicator data which would allow a more precise and meaningful assessment of service uptake by different cultural groups across a number of different portfolio services, as well as a comparative assessment across agencies. The ABS was engaged by government to cost, develop and pilot a data collection instrument and trial it in a number of government agencies (ABS, 1999).

Following an extensive testing process, the Standards for Statistics on Cultural and Language Diversity (the Standards) recommended a set of statistical standards which were designed to collect the cultural and language information considered necessary for consistent and accurate measurement of cultural diversity in Australia. The thorough development work was intended to lead to long-term cost savings by providing an 'off the shelf solution' for most data collections. It proposed a "Standard Set" of variables to measure cultural and language diversity which could be used in all administrative and service provision settings, and a "Minimum Core Set" of variables from the full standard set that would effectively replace NESB (ABS, 1999). This recognised that precise measurement of cultural and language diversity, and related advantage or disadvantage, requires a combination of variables which produces a range of data about a person's background.

The Minimum Core Set consists of four variables:

- Country of Birth of Person
- Main Language Other Than English Spoken at Home
- Proficiency in Spoken English
- Indigenous Status (forms part of the core set for those collections which are not specifically aimed at migrants to Australia; not considered further under the definition of CaLD used in this report).

The full Standard Set includes the Minimum Core Set variables and the following additional variables:

- Ancestry
- Country of Birth of Father
- Country of Birth of Mother
- First Language Spoken
- Languages Spoken at Home
- Main Language Spoken at Home
- Religious Affiliation
- Year of Arrival in Australia.

The ABS advised that any of these additional variables can be added to the Minimum Core Set variables to collect other relevant data to meet particular information needs (ABS, 1999).

In addition to the ABS Standards, the Australian Institute of Health and Welfare (AIHW) has developed a number of additional CaLD variables within the National Health Data Dictionary (AIHW, 2012). The National Health Data Dictionary is based on endorsed metadata standards as listed on the Metadata Online Registry (METeOR) website and specifies standardised collection parameters for variables accepted for use across the Australian health sector (AIHW, 2017a). The Dictionary includes the CaLD variables Interpreter Service Required and Preferred Language, in addition to the ABS Standard Set CaLD variables of Country of Birth of Person, Proficiency in Spoken English and Main Language Other Than English Spoken at Home.

Measurement strengths and limitations of variables listed in the Standards, in the National Health Data Dictionary and otherwise frequently considered in the literature are outlined in Appendix 2. There is a vast range of other variables that can potentially identify CaLD service users and seekers – along with their specific cultural and linguistic service needs – but which are rarely mentioned or considered in the literature, and are therefore not included in Appendix 2. These include country of origin, nationality/citizenship, preferred written language and level of literacy. Measures related to more particular service needs beyond language proficiency, such as diet or dietary restrictions, preferred gender of interpreter or care-giver, culture-specific values, and activities and events related to cultural traditions were also rarely discussed in the reviewed literature and were thus considered beyond the scope of this review.

The ABS Standard Set contains several different language variables which each measure a different concept associated with language usage and therefore have different practical implications for service needs and delivery, health promotion and policy. The language variable, Main Language Other Than English Spoken at Home is included as part of the Minimum Core Set because it was identified by the ABS, following extensive consultation with users of language data, as the most useful general purpose language variable. It is the language variable used in the ABS Census of Population and Housing and its use therefore enables service data to be directly compared or integrated with census data (ABS, 1999). The ABS suggests organisations should use other language variables in addition to Main Language Other Than English Spoken at Home if additional language data is required. The choice of additional language variables, the order in which they are asked and which language variable is used as a filter to Proficiency in Spoken English or equivalent depends on particular information requirements (ABS, 1999).

The ABS undertook a significant scope of work to investigate the feasibility of collecting a variable directly measuring ethnicity in the period prior to the 1986 Census (Borrie et al., 1984). Although this report is now more than 30 years old, a number of findings are likely to have ongoing relevance today. The report found two broad sub-concept choices within the multi-dimensional concept of 'ethnicity'. These were the sub-concepts of ethnicity relating to current self-perceived group versus that which is more historically determined based on ancestral origins. Despite stakeholder preference for the former, piloting of possible questions identified responder confusion around current self-perceived ethnicity and therefore significant risk of invalid and unreliable data and undercount of ethnic group numbers if such a question were to be included in the Census. It was suggested further assessment of potential current self-perceived ethnicity questions be assessed via an ABS survey (Borrie et al., 1984); however, it is unknown whether this was pursued. Ancestry was a concept better understood than 'ethnicity' in pilot questions (Borrie et al., 1984).

Subsequent reports since the Borrie Report confirmed the major benefit of including a question on Ancestry in the Census is identification of ethnic groups not identified by Country of Birth of Person or languages spoken (Castles, 1991; Kunz & Costello, 2003). This includes ethnic groups that exist across country borders or as a sub-population within countries, such as Kurds from Iraq or Turkey or Pacific Island peoples from New Zealand. Ancestry is likely therefore to have ongoing relevance for identifying refugees arriving as persecuted ethnic groups from regions within a country or who may have been born within refugee camps outside their family's country of origin.

The ABS has however continued to highlight that Ancestry in the Australian context is complex as there are many Australians with origins and heritage that do not, in practice, relate to their current ethnic identity, which may be based on nationality. It advises that "Ancestry data alone, therefore, is not considered a good measure of service needs or the extent to which persons from certain backgrounds are associated with advantage or disadvantage. When Ancestry data is used alone, it should only be done to represent a broad measure of cultural diversity". The ABS strongly advises that the Ancestry variable be used in conjunction with country of birth variables, Indigenous Status, Religious Affiliation, and language variables in order to identify particular ethnic origin or cultural groups (ABS, 2017c).

5.1.3 Variable combinations

In general, each agency will collect data for different reasons and has unique administrative and management practices. The aim of data collection and how data is collected will depend on the:

- core business of the agency or program
- target demographic of the agency or program
- purpose of the data (for example, service provision, performance indicators, needs analysis, community profiles and/or research) (ABS, 1999).

The purpose of data collection can be further considered in terms of aiming to use data about individuals versus aiming to use data about community groups more generally. Data about individuals can be used to determine the facilities and services directly required by an individual, such as interpreter services. Conversely, aggregate data collected from individuals to create community group profiles is used more for the purposes of policy setting, service monitoring, analysis and thematic reporting. Ideally variables collected are useful for both individual and group uses but this may be difficult to achieve. This was one of the drawbacks with NESB due to the potential for making unfounded assumptions about individuals and their service needs on the basis of the characteristics of the community group to which they belong (ABS, 1999). As a further example, Country of Birth of Person used in conjunction with Year of Arrival in Australia can be useful at a group level to determine which community groups have the most difficulty, or take the longest time, to adapt to Australian society. It cannot, however, be used to assume a recently arrived individual's particular service needs e.g. need for an interpreter.

The ABS advises that agencies are encouraged to collect the Minimum Core Set and to identify and include additional Standard variables as appropriate. Where particular collections currently do not include one of the four Minimum Core Set, but include other cultural diversity variables which are better suited to the collection, it is not intended that a core variable necessarily replace a variable currently collected. Rather it is recommended that the other Minimum Core Set variables be added to the collection. The ABS suggests the general principle of asking as wide a range of questions as possible to provide a comprehensive picture of an individual's origins and characteristics.

Resulting combinations of variables collected can be useful for particular purposes. For example (ABS, 1999; Victoria University, 2009):

- Combining Main Language Other Than English Spoken at Home with Proficiency in Spoken English can help identify the need for language services and to inform marketing and promotional strategies.
- Using Country of Birth of Person with a language variable provides a cultural dimension which is likely to indicate a person's familiarity with Australian institutions, labour market, etc. If the person is a child, using mother or father's country of birth might provide information on potential for accessing services.
- A number of variables, such as Country of Birth of Person, Main Language Other Than English Spoken at Home and Year of Arrival in Australia might be needed when developing more detailed client profiles for planning or evaluation purposes.
- Combining Ancestry with Country of Birth of Person and language variables may provide more information about a person's cultural identity than Ancestry alone.
- The use of Country of Birth of Person, a language variable and Religious Affiliation will accurately identify most cultural and ethnic groups.

Documenting a range of variables is likely to be useful to both improve servicing to individuals based on accurate and reflective data (rather than a generic measure such as NESB), and more widely for creating aggregate data to create community group profiles for the purposes of policy setting, service monitoring, analysis and thematic reporting (ABS, 1999).

5.1.4 Current data collection practice by the Australian Bureau of Statistics

In practice, the most recent 2016 Census of Population and Housing collected all of the variables of the Minimum Core Set (Country of Birth of Person, Main Language Other Than English Spoken at Home, Proficiency in Spoken English) and a selection of the additional variables included in the Standard Set (Year of Arrival in Australia, Ancestry, Religious Affiliation, Country of Birth of Mother and Country of Birth of Father). All these variables were also collected in the 2011 Census, however Country of Birth of Mother and Country of Birth of Father were updated in 2016 to facilitate entry of specific countries, rather than the sole answer options of "Australia" or "Overseas" as included in the 2011 Census. Census data is generally presented and available on the ABS website by geography, with total counts at local government, state and national levels. Census CaLD data is important for informing denominator data for calculating and comparing health status or health service use rates. Census data is presented by the ABS either by the individual variables listed above or by combinations of variables, such as Born Overseas, Year of Arrival in Australia and Main Language Spoken at Home, or Country of Birth and Main Language Other Than English Spoken at Home.

The ABS also conducts a number of other collections, including the National Health Survey and the Patient Experience Survey. The 2014–2015 National Health Survey included data collection on the variables Country of Birth of Person, Year of Arrival in Australia, Main Language Spoken at Home, English Proficiency, Ancestry and Country of Birth of Parents. Data is predominantly published on the ABS website by Country of Birth of Person, Year of Arrival in Australia and Main Language Spoken at Home (ABS, 2017f). The Patient Experience Survey collects Country of Birth of Person and Year of Arrival in Australia; however, this is not published on the ABS website (ABS written correspondence, 2016).

5.1.5 Current data collection practice by Australian health services

Despite the Standards guidelines published by the ABS in 1999, evidence indicates the Standards have not been implemented as intended and considerable variation and data insufficiency continues to exist in the measurement of ethnicity in Australian health data collections (Abouzeid et al., 2014; Minas et al., 2013; Thow & Waters, 2005). This appears consistent with the situation internationally; despite many national guidelines for collecting ethnicity data, considerable variability remains if and when ethnicity is measured, in diverse arenas including national censuses (Morning, 2008) and published biomedical literature (Ma, Khan, Kang, Zalunardo, & Palepu, 2007).

Although the Standards stipulate that using "a single standard variable, such as country of birth...is inadequate", in practice this has been the widespread level of implementation of the standards in the health sector. A 2005 review of health sector national data collections found (Blignault and Haghshenas 2005):

- Of seven national data dictionaries using ABS standards and classifications, only one included all the ABS Minimum Core Set indicators for CaLD.
- Of 17 national data sets reviewed, 12 included Country of Birth of Person, two included Preferred Language and one included Main Language Other Than English Spoken at Home.
- Two data sets (injury surveillance and cancer) did not include any data concerning the client's cultural background.

Several studies since 2005 have confirmed a lack of meaningful data collection beyond Country of Birth of Person in many health data settings across Australia, including mental health, drug and alcohol services, maternal health, and aged care (AIHW, 2014; Donato-Hunt & Grima, 2009; Minas et al., 2013; Porter et al., 2016). In an assessment of databases containing information from which ethic group-specific estimates of type 2 diabetes burden could be gleaned, Abouzeid et al. (2014) found of 32 relevant databases, birthplace was recorded in 27 and only nine recorded other aspects of self-perceived race/ethnicity, aside from Indigenous Status. A 2014 Australian Institute of Health and Welfare (AIHW) review concluded that out of 15 aged care data sets reviewed, 10 collected the ABS measures Country

of Birth of Person and Main Language Spoken at Home/Main Language Other Than English Spoken at Home, although lack of standardisation to the ABS data collection method reduced the comparability of data (AIHW, 2014). AIHW also found very few aged care settings collect data to provide insight into specific service needs related to CaLD characteristics such as diet, culture-specific values, activities or events related to cultural traditions, or preferences for care such as doctor gender requirements (AIHW, 2014). In a 2009 survey of drug and alcohol workers across Australia on mandatory CaLD data collection fields, 86 per cent collected Country of Birth of Person, 79 per cent collected Preferred Language, and 56 per cent collected Interpreter Service Required. Only 40 per cent collected Main Language Spoken at Home and 27 per cent collected English proficiency (Donato-Hunt & Grima, 2009).

The National Health Data Dictionary maintained by the AIHW is used to construct health problem or service specific 'National Minimum Data Sets', or minimum sets of data elements agreed for mandatory collection and reporting at a national level. Although the National Health Data Dictionary lists standards related to Interpreter Service Required, Preferred Language, and all of the variables in the ABS Minimum Core Set, few translate to being included in the National Minimum Data Set requirements. For example, the National Minimum Data Set relating separately to each of emergency department presentations, hospital admissions and community mental health services only requires Country of Birth of Person. The Alcohol and other Drug Treatment Service National Minimum Data set only collects CaLD indicators Country of Birth of Person and Preferred Language. A 2009 study reviewing CaLD data collection practices across drug and alcohol services across a number of states found most services were collecting Country of Birth of Person and Preferred Language, the mandatory items for drug and alcohol service reporting, but not all, including a number of respondents representing services in Western Australia (Donato-Hunt & Grima, 2009). This suggests that the few variables included in the National Minimum Data set requirements may not be enforced or in practice.

The health sectors occasional additional focus on alternative language variables of Preferred Language and Interpreter Service Required as opposed to the two suggested by the ABS in the Minimum Core Set (Main Language Other Than English Spoken at Home and Proficiency in English) likely reflects historical circumstance but also the health system's additional requirement to focus specifically on individual service needs. The ABS' suggested Minimum Core language measures are likely more targeted towards aggregate statistical measurement of cultural and linguistic diversity in a survey or questionnaire setting.

The ABS has not published an update of general guidelines relating to CaLD data collection since 1999, although Queensland Health reported a review was being untaken by the ABS in 2011 (Queensland Health, 2012). Verbal discussion with the ABS suggests there are currently no plans to review the variables included in the Standards.

A number of recommendations and changes have been suggested or implemented by the health sector as a result of evidence of poor CaLD data collection practices within the health setting. A 2014 review of aged care data sets by AIHW included a consultation with AIHW and Department of Social Services staff to rank 41 different measures of CaLD according to a set of in-house developed criteria (AIHW, 2014). These criteria related to the concepts of utility, technical performance, measurement feasibility, compliance with ABS standards, simplicity and completeness. Their final working paper stated, "ABS CALD measures provide for statistical measurement of cultural and linguistic diversity, but not the associated service needs." This paper recommends that:

- Data sets should employ, as a minimum, the ABS measures Country of Birth of Person and Main Language Spoken at Home, augmented with Interpreter Service Required, Preferred Sex of Interpreter and Preferred Language, where the Main Language Spoken at Home is not English.
- Proficiency in Spoken English and Year of Arrival in Australia, along with three linked measures that are associated with spirituality, are also recommended for supplemental inclusion.
- Data sets should ensure they comply with ABS data collection methods.

In making these recommendations it is important to note that it appears that AIHW has not recognised the subtle but important difference between Main Language Other Than English Spoken at Home and Main Language Spoken at Home.

In 2007, Queensland Health endorsed a multicultural minimum data set to be collected across Queensland Health's major data collections from 1 July 2007 (Queensland Health, 2012). It comprised:

- Country of Birth
- Preferred Language
- Interpreter Service Required
- Religion (religion collected locally but not centrally extracted).

A submission by Queensland Health to a reported 2011 ABS review of the ABS Standards stated that the current standards are inadequate to capture the cultural diversity and needs in the Queensland population. Specifically, Queensland Health's three priority CaLD population groups, namely refugees, Australian-born South Sea Islander people and New-Zealand born Pacific Islander and Maori people, were very poorly identified by Country of Birth of Person and main language other than English. Their submission recommended that the ABS Standards be redesigned to add a variable to better identify ethnic groups. The proposed variable was for addition was Ancestry (Queensland Health, 2012). A subsequent 2012 Queensland Health internal review went further in recommending that Queensland Health work with the Australian Bureau of Statistics to consider collection of ethnicity (Ancestry) in the CaLD national Minimum Core Set instead of Country of Birth of Person, and in the interim continue to collect the Queensland Health multicultural minimum data set endorsed in 2007 (Queensland Health, 2012).

5.1.6 Current data linkage projects

The ABS has undertaken a number of data linkage projects in recent years, which has increased the capacity to present CaLD data. This has included the statistical publication of a project that integrated 2011 Census data with information from the Department of Immigration and Citizenship's Settlement Data Base using probabilistic linking to produce an enhanced dataset, the Australian Census and Migrants Integrated Dataset, 2011. The data allows for the settlement outcomes of recent migrants (language capability, employment, income and education), however generally not health outcomes, to be cross-classified by their entry conditions, such as visa stream, whether they applied onshore or offshore, and whether they were a main or secondary applicant. This data is particularly useful for identifying particular visa-sub groups, such as refugees and women at risk, by their standard Census information and is generally available at the level of small geographic areas of Statistical Area Level 2 (ABS, 2014) (ABS written correspondence, 2016).

In addition, there is an ongoing Multi-Agency Data Integration Project collaboration between government departments including the federal Department of Health and ABS. The collaboration resulted in a project linking Medicare Benefits Schedule and Pharmaceutical Benefits Data with Census Data and explored the cultural and linguistic characteristics of people using Medicare Benefits Schedule (MBS) subsidised mental health-related services and Pharmaceutical Benefits Scheme (PBS) subsidised mental health-related medications in 2011 (ABS, 2016b). No other data linkage projects were identified or discussed in the reviewed literature.

5.1.7 Data collection quality

In Australia, instructions for standard definitions, question phrasing, mode of assessment and standard coding classifications are included in the ABS' Standards for Statistics on Cultural and Language Diversity (ABS, 1999). The Standards are supplemented by standards on each of the individual variables which are reviewed and updated on a five-yearly or less basis depending on a need (ABS,

2016a)(ABS written correspondence, 2016). Guidelines for variable classification and coding are also updated periodically to reflect locally relevant categories and demographic changes (ABS, 2016c). All the Standard variables are designed to be self-reported answers to questions administered by personal interview, questioning of a responsible adult (representing a household or family) or via self-enumeration questionnaires. ABS reports that its statistical standards are "finalised following thorough and rigorous development of concepts, definitions, questions, classifications, and processing and dissemination procedures". ABS also states their standards are designed to harmonise, as far as possible, with established Australian and international practices so as to facilitate comparability internationally, across collections, across time, across agencies and within a given subject area (ABS 1999).

Beyond that, there is limited discussion in the Australian literature about the quality of CaLD data (Donato-Hunt & Grima, 2009; Minas et al., 2013). An AIHW 2014 review of aged care services data collection concluded that those data sets collecting the ABS concepts of Country of Birth of Person and Main Language Spoken at Home were subject to variable methods of implementation. The review went on to say, "As a result, data collected both within a single data set and across data sets are not always directly comparable or appropriate to aggregate into a single collection. There could also be difficulties in comparing data extracted from these data sets because ABS population statistics are based strictly on ABS measures and collection methods." (AIHW, 2014). A survey of drug and alcohol services nationally identified varying determination methodology between self-report, general communication with the client, and worker discretion (Donato-Hunt & Grima, 2009). This contrasts to the methodology recommended by the ABS for CaLD data variables.

Some studies have highlighted the explicit exclusion of people with limited or no English proficiency from surveys (Blignault & Haghshenas, 2005; Minas et al., 2013). This creates bias limiting the generalisability of the results and may fail to represent potential disadvantage in CaLD groups associated with socioeconomic disadvantage. Reported challenges with collecting and presenting CaLD data from surveys include poor resourcing for interpreter use to enable participation, and small CaLD sample sizes in many smaller geographic areas.

It is recognised completeness and accuracy of data can be affected by various policy mechanisms, whether it be legislation, regulation, code of practice, such as mandatory fields and contract conditions, or organisational culture or policy. Donato-Hunt noted "Victoria has a relatively well-developed program and legislative infrastructure for supporting cultural diversity and sound community relations", correlating with a higher percentage of data collection in Victoria compared to Western Australia and New South Wales in a study examining CaLD data collection across drug and alcohol services (Donato-Hunt & Grima, 2009). Legislated acts in relation to CaLD populations in both NSW and Victoria add policy focus and create an environment for increased requirements for data collection. For example, the *Multicultural Victoria Act 2011* mandates government departments to report annually to the Minister for Multicultural Affairs and to Victorian Parliament on their achievements in multicultural affairs over the previous financial year. At 30 June 2015, four departments had cultural diversity plans (CDPs) in place and three departments were in the process of updating or developing new CDPs. Departments had reported significant progress in cultural diversity planning and implementation including on the use of data to improve service planning and delivery. Departments were also required to report on the use of translating and interpreting services (Victorian Government, 2015). In WA, data in relation to use of interpreter and translating services by health services is mandatorily reported under the Western Australian Health Language Services Policy (Department of Health, 2017). However, the data collection techniques are not standardised across health services. An overarching Act in WA in relation to multicultural affairs more generally is also lacking. Comprehensive review of the literature relating to policy mechanisms to improve CaLD data practices was beyond the scope of this review.

5.2 Review of WA health system core data sets

Review of data collection practices across the WA health system large core data sets revealed a number of useful findings. It is important to firstly understand that most of the core data sets collate data from healthcare services via a large number of 'feeder' software packages used for primary data collection. These 'feeder' packages vary between public and private services, between some hospitals and between different types of services. The main package used by the WA public hospital system is the 'Patient Administration System' (PAS). The older version of PAS called TOPAS ('The Open Patient Administration System') and another separate software package 'HCARe CMS' are being progressively transitioned across hospital services to a newer web-based version of PAS, namely 'webPAS'. The PAS system is frequently used to auto-populate between other primary data collection packages including to specific clinics or services beyond or within inpatient, outpatient and community services, and makes up the primary source for the CaLD fields recorded by multiple different core data sets.

The webPAS and TOPAS software packages have a standardised demographic page that includes CaLD variables, namely Country of Birth of Person, Religious Affiliation, and Preferred Language (only meant to be entered if the patient requires an interpreter according to the TOPAS manual). WebPAS also includes a separate field for Interpreter Service Required. Details entered on the demographic page are meant to be checked and updated by clerks each time the patient presents for a new occasion of service: however, stakeholders state this may not occur and there is no automated recording of the history of field updates.

The core data sets are maintained centrally by data custodians located within the WA Department of Health, as part of the Department's 'system manager' role. Core data collection manuals are also produced by the WA Department of Health to guide primary data collection by health service providers. The manuals represent documentation of the relevant 'statistical standards'. Statistical standards are defined as a set of components which, when used together, produce consistent and high quality statistical output across collections and over time. For each variable, a statistical standard generally specifies a standard name, definition, question, classification, coding procedure and output category. It is generally the responsibility of frontline health service providers to ensure quality of collection according to the core data collection manuals. The WA Department of Health only enforces CaLD data field entries for mandatory fields or algorithm-identified nonsensical entries, Preferred Language not entered if Interpreter Service Required, for example. Further data core data set completion and quality issues are explored individually under the separate core data sets below.

CaLD variables collected by each of the core data sets are included in Table 1, with specific CaLD data issues for each data set discussed below.

Table 1. Variables collected by the core data sets according to available data manuals

Core Data Collection	CaLD Variable					
	Country of Birth of Person	Interpreter Service Required	Preferred Language	Ethnicity (aside from Indigenous Status)	Year of Arrival in Australia	Religious Affiliation
Hospital Morbidity Data Collection	✔	✔	✔			
Emergency Department Data Collection	✔	✔				
Mental Health Information System	✔				✔	✔
Non-Admitted and Patient Waitlist Data Collection	✔	✔				
Midwives Notification System		✔	✔	✔		
WA Cancer Register	✔					
Mortality Data set	✔					

5.2.1 Hospital Morbidity Data Collection

Data transmits to the Hospital Morbidity Data Collection (HMDC) from various feeder systems including TOPAS and WebPAS.

Data collection in relation to hospital admissions is guided by conditions specified in the Hospital Morbidity Data System (HMDS) Reference Manual (Department of Health, 2015a). The Reference Manual outlines collection relating to three CaLD-related elements: Country/State of Birth, Interpreter Service and Language.

Assessment of rate and quality of completion of these fields below refers to the all WA hospital admissions (both public and private) discharged between 1 July 1999 and 13 June 2017, equalling to a total of 14,874,066 separations.

Country/State of Birth

Country or Australian state of birth of the patient is designated a mandatory collection requirement in the HMDS. In addition to missing data, two input options relate to unknown data; 'Not Stated' or 'Inadequately Described'.

A total of 1.98 per cent (n=294 192) of all Country of Birth of Person values were either missing (n=179) or designated 'Not Stated' (n=280 629) or 'Inadequately Described' (n=13 384) out of all WA inpatient admissions (n=14 874 066) over the specified period.

It was beyond the scope of the project to investigate the accuracy of the Country of Birth of Person data collection, but it is recognised there may be data error related to both data entry error and country coding changes over time.

Interpreter Service

The HMDS manual defines the 'Interpreter Service' field as 'whether an interpreter service is required by or for the patient'. It is a mandatory collection requirement. The HMDS creates ambiguity as to how this is determined because it does not specify between self-report or judgement by the clerk. Indeed, discussion with staff confirmed this may be determined by a judgement of the clerk or entered based on a referral form, rather than via a direct question to the patient. This is in contrast to the National Health Data Dictionary which specifies, "Recommended question: Do you [does the person] require an interpreter?" The data field guidance included in the HMDS manual also suggests this field is misunderstood by those developing the manual as a retrospective determination of whether an interpreter was used. This includes, "This data item should only have a value of (1) Yes if an official paid interpreter service is used" and a data quality edit flag 'Language required when an interpreter was used'. However, observation indicates front-line staff are completing this field prospectively, for example whether an interpreter is needed at the time of admission, which is the correct interpretation of the field specified by the National Health Data Dictionary.

The Interpreter Service field generally had a very high level of completion with only 32 (0.000002%) missing values out of a total of 14 874 066 admissions. Only 0.39 per cent (n= 57 630) of all admissions were recorded as requiring an interpreter service.

Language

This field is defined in the HMDS manual as "the language (including sign language) most preferred by the person for communication" and is a filter question from the Interpreter Service field, specified as "required if used interpreter; blank otherwise". The 'Guide for Use' section for this field states, "This data item should only be completed if an official paid interpreter service is used". Again, several statements suggest retrospective determination of Interpreter Service.

This field was complete in 0.8 per cent (n= 113 878) of cases, approximately double the number 'requiring an interpreter'. This demonstrates data inconsistency as this field is only meant to be completed if an interpreter is required.

Of those recorded as requiring an interpreter service (n= 57 630), 4.5 per cent (n= 2 605) of patients did not have a language listed; however, all of these occurred prior to July 2006, when new ABS Australian Standard Classification of Languages 1st Edition coding was implemented.

Of those admissions listed as not requiring an interpreter service (n=14 816 404), 58 853 (0.4%) had a Preferred Language listed. In the period since July 2006, 6 607 admissions not requiring an interpreter had a language listed, however most of these (n=4 501, 68.12%) were 'Not stated'. This leaves 2,106 admissions (0.01 per cent of all admissions over the period 1 July 2006–13 June 2017) recorded as not requiring an interpreter but with a recognised Preferred Language other than English documented. This raises the question whether a portion of these patients actually required an interpreter.

5.2.2 Emergency Department Data Collection (EDDC)

Data recorded as part of the Emergency Department Data Collection (EDDC) contains data on emergency department (ED) activity in WA's public hospitals as well as ED activity in private hospitals under financial contract to the WA Government. This data collection commenced in 2002 and has had various interventions to improve data collection processes overtime. Data is collated from multiple feeder systems: Emergency Department Information Systems (EDIS) (all metropolitan public hospitals, the Joondalup Health Campus and Bunbury Regional Hospital), WebPAS, progressively transitioning from HCARe (all rural hospitals except Bunbury Regional Hospital and Peel Health Campus, and St John of God Midland), and Medtech, transitioning from Electronic Patient Administration System (ePAS) (Peel Health Campus). It is understood there is an auto-populating function between EDIS and the broader Patient Administration System.

According to the EDDC Data Manual Version 1 (2007) currently in use, Country of Birth and Interpreter Service Required are the only CaLD-related variables collected. There are no items relating to language(s) spoken or religion (Department of Health, 2007).

Some fields within the EDDC are routinely targeted as priorities for completion. In general, these relate to performance indicators such as timelines to being seen by various staff, and disease coding, rather than others including CaLD variables. Recurrent abnormalities in CaLD data may be noticed ad hoc and investigated but are not routinely monitored.

Assessment of rate of completion of these fields below relates to data records for the period July 2005–June 2017 inclusive.

Country of Birth

Country of Birth is deemed a mandatory variable in the relevant data manual. Out of a total of 10 797 393 presentations, 140 455 (1.3%) did not have a country of birth recorded. In addition, 0.32 per cent (n=33 847) were entered as 'not stated' and 0.07 per cent (n= 7 585) were entered as 'inadequately described'.

Interpreter Service Required

It is not specified whether 'Interpreter service required' is a mandatory or optional collection field in the Data Manual, and there is no guidance as to how this is actually determined. Response options are 'Yes', 'No' or 'Not specified/Unknown', the latter being outside of permissible values for this variable as listed in the National Health Data Dictionary.

Review of Interpreter Service Required field revealed 0.3 per cent of presentations were listed as 'requiring an interpreter service', 89.2 per cent as 'interpreter service not required' and 10.9 per cent as 'not specified/unknown'. There were also a small number of presentations erroneously listed with specific languages.

5.2.3 Mental Health Information System

The Mental Health Information System (MHIS) records data on all ambulatory mental health services provided by public mental health facilities in WA including services provided to interstate and overseas visitors. Data feeds from a number of systems including TOPAS, HCare and a further package specific to mental health entitled 'PSOLIS' (Psychiatric Online Information System). Matching mandatory demographic information between TOPAS or webPAS and PSOLIS flows through to the PSOLIS client demographics primary collection page. Non-matching mandatory demographic information fields are able to be 'value added' in the client demographics fields in PSOLIS.

According to the MHIS Ambulatory Data Dictionary (Version 4 updated August 2010), the CaLD items collected are Country of Birth, Year of Arrival to Australia and Religion (Department of Health, 2010). No fields flow through for collection to the core data set in relation to requirement for interpreter or preferred language however PSOLIS does have these fields as optional entries on the primary demographic page.

Currently there are concerns regarding the quality of CaLD information collected by PSOLIS. Interpreter Service Required, Language, Year of Arrival and Religion are optional and therefore poorly completed. There is also no function to record update history, so it is not known whether the interpreter requirement or religious affiliation has been updated for the current admission or event recently. Reportedly, there has been no state or national mandate to collect this information, and as such comprehensive policies and guidelines have never been developed to inform recording practices, resulting in incomplete and inaccurate recording of this information in PSOLIS. Previous requests for preferred language and interpreter requirement data in order to evaluate services have been denied given the poor quality of the data.

The MHIS is being replaced with a new data collection system (MIND) in 2017. It has been deemed that the MIND collection will not include variables collecting preferred language, interpreter requirement or year of arrival because of incomplete and inaccurate recording of CaLD information in PSOLIS. Advice was received that should there be a mandatory requirement to collect this information in MIND, PSOLIS will need to be amended to include these fields as mandatory and feeding through to the MIND core data set.

The WA Department of Health is working to develop the MIND Collection to include a data quality assurance process. This will include working on basic validation rules for fields to meet national data submissions, such as Indigenous Status. The WA Department of Health will unlikely validate Country of Birth until further discussions and agreement with health services has occurred (maybe in 2018).

Country of Birth

Country of Birth is specified as a mandatory data field collection in the MHIS Data Dictionary. The MHIS data set reviewed related to the period 2006–2016 inclusive, equating to 7,998,020 occasions of service. There were no missing entries in relation to the Country of Birth field. 'Not stated' was entered for 1.36 per cent of entries (n=108,604) and 'Inadequately Described' was recorded for just 25 entries.

Year of Arrival

This field is described in the data manual as 'Year of Arrival in Australia from overseas' and listed as collected between 1966 and 2003 and not collected by the MHIS since PSOLIS started. The PSOLIS primary data collection package does however include this field as an optional input.

Religion

This field is described in the data manual as 'religious affiliation at admission' and collected between 1967–present. It states 'missing data increases overtime (maximum 60% missing in 2003)'.

5.2.4 Non-admitted and Patient Waitlist Data Collection

The Non-admitted Patient Activity and Wait List Data Collection (NAPAAWL DC) includes patient-level non-admitted activity data predominantly from the following feeder systems:

- The Open Patient Administration System (TOPAS)
- Health Care And Related Information Systems / Ambulatory, Other Patient and Domiciliary (HCARe / AOD)
- Web-based Patient Administration System (webPAS).

Variable data requirements are specified on the AIHW METeOR website page entitled "WA Health Non-Admitted Patient Activity and Wait List Data Collection (NAPAAWL DC) 2016–17" (AIHW, 2017b). Country of Birth of Person and Interpreter Service Required are the only CaLD variables specified for collection.

Feedback from data custodians indicates 'Interpreter Service Required flag' (presumably based on Interpreter Service Required) and Country of Birth of Person are collected from some systems but these are not of high quality.

The existing data set held by the Epidemiology Branch covers the period 1 July 2008–28 February 2010 and is not routinely updated. It does not contain any variables related to CaLD characteristics.

5.2.5 Midwifes Notification System

The Midwives Notification System (MNS) has been in operation since 1975 and receives information from midwives about births they attend in Western Australia (WA). Notifications are for all births where the infant is of a gestational age of 20 weeks or more or a birth weight of 400 grams or more if gestation is unknown. Data collection is guided by documents including the 'Guidelines for Completion of the Notification of Case Attended Health Act (Notification by Midwife) Regulations Form No.2' (Department of Health, 2006), and 'Guidelines for Midwives Notification of Case Attended' (Department of Health, 2015b). The 'ethnic origin of the mother' has been collected since 1975: however, Interpreter Service Required and Preferred Language were only added for collection in July 2016. Country of Birth of Mother is not collected. The Ethnicity field is described as the "Self reported ethnic origin of the woman giving birth. A woman who identifies herself as more than one of the listed descriptions can be reported as 'Other'; however, where Aboriginal or Torres Strait Island is included report as Aboriginal and/or TSI". Input options include: Caucasian – Includes all people of Caucasoid (European) heritage such as Maltese, Lebanese and Italian; Aboriginal, TSI or Aboriginal and TSI- Includes persons of Australian Aboriginal and/or Torres Strait Islander (Australoid) heritage; Asian; Indian; African/Negroid; Polynesian; Maori; Other. The 'ethnic origin of the mother' is not listed by the ABS or AIHW as a recognised standard variable for large data set collections.

The data set reviewed included only ethnicity in terms of Indigenous or non-Indigenous Status, which had been re-coded according to an algorithm developed by the Data Linkage Branch.

5.2.6 Mortality database

The mortality database summarises data on cause of death, and is processed centrally by the ABS before being shared with the WA Department of Health. The mortality data set reviewed contained only the CaLD variable 'place of birth', which in practice is Country of Birth of Person.

Place of Birth

For the period 1983 to 2015 inclusive, 7197 (1.9%) out of 369 862 deaths did not have a place of birth recorded, as well as, 0.3 per cent of entries were 'inadequately described' or 'not stated'.

5.2.7 WA Cancer Register

The data set reviewed included only the CaLD variable Country of Birth for the three individual registers relating to each of cancer incidence, cancer mortality and cancer in-situ.

Country of Birth of Person

The register recording cancer incidence for the period between 1982 and 2015 inclusive had 776 missing (0.29%) Country of Birth values out of a total of 270 575 entries. Country of Birth was recorded as 'not stated' or 'inadequately described' in 4 485 (1.66%) entries.

The register recording cancer in-situ in the period 1982–2015 inclusive had 1184 (1.8%) missing Country of Birth values out a total of 65 624 entries. Country of Birth of Person was recorded as 'not stated' or 'inadequately described' in 2138 entries (3.26%).

The mortality register relates to all cancer-related deaths occurring between 1990 and 2015, inclusive. Only three deaths had Country of Birth data missing out of a total of 86 763 (0.003%). Country of Birth was listed as 'not stated' or 'inadequately described' in 208 (0.24%) deaths.

5.3 Stakeholder consultation

5.3.1 Respondents

A total of 72 responses to the survey were received. Survey respondents were from a range of services and programs across the WA health system and those falling under the jurisdiction of the Mental Health Commission. A full list of programs as entered by respondents appears in Appendix 3. Several services had multiple staff members complete the survey. In addition, a number of organisations replied directly via email to advise of important information regarding their current CaLD data collection practices. This included a response from St John's Ambulance Western Australia stating that their service, 'doesn't record the religion, race, language or ethnicity of callers (or patients)'.

Programs with staff responding to the survey included a number of those targeting specifically CaLD clients and many larger services with significant CaLD constituencies, including primary care health services, hospitals (inpatient, emergency and outpatient), health support services, and community services including screening services and allied health programs. A broad representation across city, rural and regional services was also achieved. Multiple responses were received from those involved with data input such as data coders and clerks, and data custodians, in addition to clinical or front-line staff and managers.

5.3.2 Data collection tools

The 72 respondents listed 92 entries for data collection tools in current use. A full list of data tool responses by health service organisation is contained in Appendix 3.

Table 2. Frequency of data collection tools reported by survey respondents

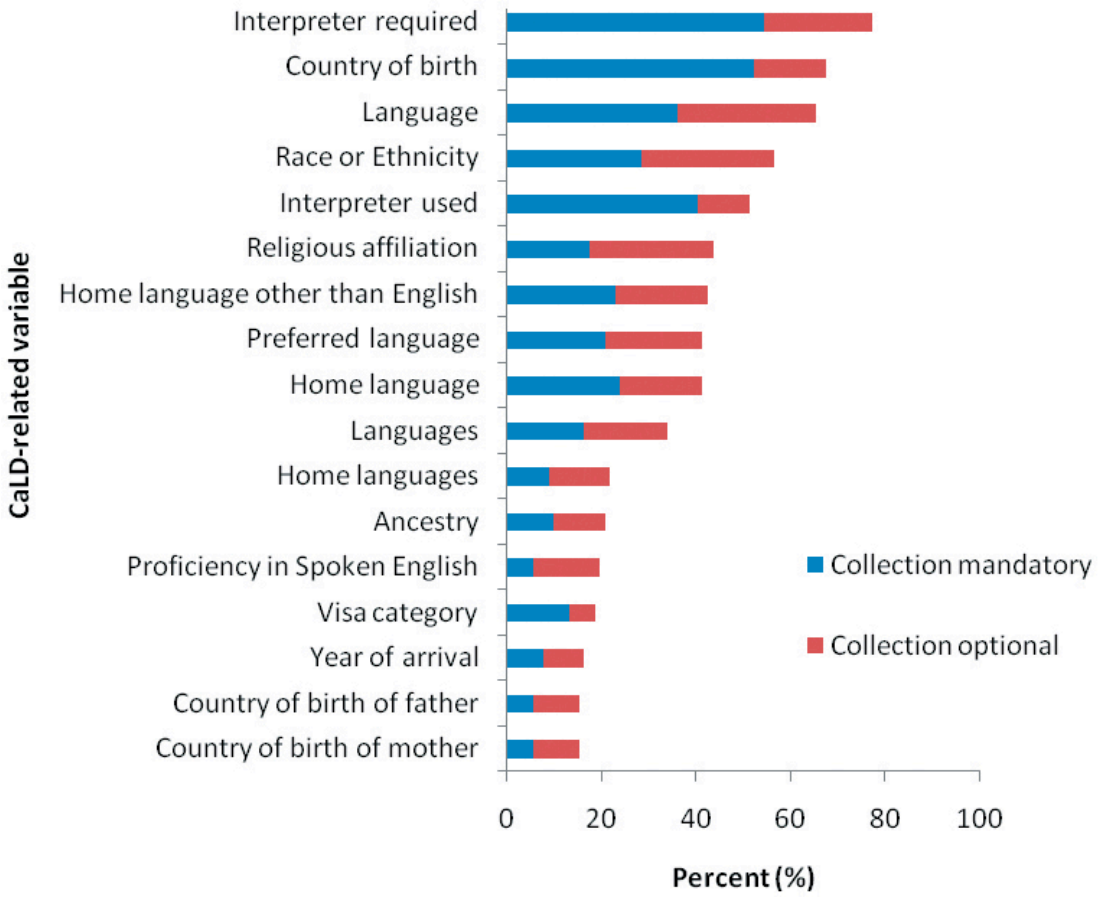
Data collection tool	Frequency of response	Main service type
WebPAS	19	Hospital services
TOPAS	16	Hospital services
BOSSNET	6	Multiple
Language Services Systems (LASS)	6	Language services
Burns Injury Management System	3	Burns
DENIM	3	Dental health services
Psychiatric Online Information System (PSOLIS)	3	Mental health
iSOFT	2	Pathology
Medtech	2	Specialised outpatient services
Oracle	2	Multiple
SHaRE	2	Cancer services
SHIP	2	Sexual health services
SIMS	2	Drug and alcohol services

Table 2 lists those data collection tools which were listed by more than one respondent. Many respondents also listed a unique data tool (n=47) not used by any other respondents to the survey. Several participants listed multiple data collection tools within the one answer so the total sum of the frequency of data tools reported does not total the number of entries. Most often the multiple data collection tools listed within one entry were WebPAS or TOPAS and a more specific software package, likely reflecting the auto-populating function of PAS into other software packages.

5.3.3 Variables in current use

For each data collection tool entered, respondents were asked to nominate whether collection was optional, mandatory, not collected or they were unsure for each CaLD variable in a pre-specified list (Figure 1). A response for each variable was mandatory. The most frequently collected item reported across all data collection tool entries was Interpreter Service Required (n= 71, 77.2%), followed by Country of Birth of Person (n= 62, 67.4%). Main Language Other Than English Spoken at Home and Proficiency in Spoken English were reported as recorded by 42.4 per cent and 19.6 per cent of data tools respectively. It is suspected however there was some confusion in entering language variable answers as multiple respondents entered multiple options and it is unlikely any data tool would collect more than one language data variable. Relatively fewer tools collected Year of Arrival in Australia, Country of Birth of Parents, Ancestry or Visa Category. Many data tools collected Race or Ethnicity but this is presumed to relate predominantly to Indigenous Status. All variable options were collected by at least one data collection tool. Three data collection tool entries listed other variables collected beyond survey answer options, which were 'country of refuge before arriving in Australia', 'financial election status' (overseas country, reciprocal country etc), 'GP use' (and whether GP uses an interpreter or is bilingual), 'form of interpreter' (onsite or telephone), 'parental education levels', 'parental literacy levels', 'what families do with letters if they receive in English and can't translate themselves'.

Figure 1. Frequency of collection of CaLD-related variables across data collection tool entries



Six data tools reportedly collected no CaLD data or collected 'race' only, presumably related solely to Indigenous Status. These included the WA Cervical Screening Program Register, the School Based Immunisation Program Database and the HCARe software package used by Child and Adolescent Community Health Service high school nurses, enuresis nurses, refugee health nurses and education support nurses to record occasions of service. Stork was also a data tool reported to have no CaLD data, however this is known to be the feeder system to the MNS which collects ethnicity, so this is incorrect.

Respondents were given the opportunity to comment on the variables their data tools were collecting. A total of 10 people provided comment, for which three themes emerged.

Theme 1: Practicality of CaLD-data variable collection

Analysis of comments identified a theme relating to recognised circumstances where it was not always possible or practical to enter CaLD data. There included sensitive clinical settings, and data entry completion based on paper referral forms which may be submitted with missing fields.

> "…provides an anonymous and confidential telephone service so demographic data (including CaLD related information) is only recorded if gathered as part of the conversation, i.e., we do not ask clients a series of questions as a matter of course prior to them being able to engage with the service."

Theme 2: Requirement to record CaLD-variable data

A further theme related to policy or regulatory requirements to submit CaLD data, whether that be lacking or existing requirements. Existing collection requirements identified included statutory requirements or those related to collection according to a national minimum data standard.

> " …. Registry receives data from health care providers and laboratories.....The Register has the ability to record the ABS code for non-English language, however it is not a requirement for health care providers and laboratories to record and transmit these data to the register. Further, there is no method for providers and laboratories to record this information".

Theme 3: Issues in relation to selection of CaLD variables in use

Respondents also identified issues in relation to the range of CaLD variables in current use. Issues included the lack of 'ethnicity' or 'race' collection beyond Indigenous Status, and lack of refugee identification.

> "There is no flag for refugee status or visa status on WebPas, so makes identification (e.g. for data linkage or analyses purposes) very difficult. Interpreter utilisation is also poorly documented in wider notes (part of mandatory organisation collection), but essential for data collection (and optimal patient care). Maternal country of birth does not tell you necessarily about refugee status (especially for those with prolonged transit) or socioeconomic status".

5.3.4 Current use of data

Participants were asked how their program had used their CaLD data collected over the last five years. A response was mandatory and respondents were asked to select all options that applied. Most services used the data in multiple ways (Table 3); of those respondents reporting their program had used the data in at least one of the ways listed, the median number of uses reported was five. Three respondents reported using the data in some 'other' way, with one specifying that use to be 'providing CaLD data to other programs'.

Table 3. Reported WA health services use of CaLD-related data in the last five years

Use	Number of survey respondents	Percentage of survey respondents (%)
Service reporting	38	52.8
Informing individual clinical care	32	44.4
Informing a culturally-appropriate service	31	43.1
Interpreter service provision	31	43.1
Targeting services or interventions	29	40.3
Program monitoring or evaluation	28	38.9
Research	18	25
Policy planning	18	25
Shared data with another service	17	23.6
Informing larger data set	12	16.7
Not used	10	13.9
Not sure	4	5.6
Other	2	2.8

Respondents were given the option to provide comment on their selected answers. Ten respondents offered comments about their program's current use of data. Two themes emerged in these comments.

Theme 1: Other specific uses of CaLD data

Several respondents identified using data currently collected to inform the development of translated health information and brochures in appropriate languages. A further use listed related to analysing the demographics of those completing surveys on health behaviours.

Theme 2: Core data set use of CaLD data

Responders who were central core data set custodians made comments demonstrating current or desired use of core data set data to inform service delivery.

> "Data Requests for Language and Interpreter Service Required have been previously requested to evaluate services required, however given that these fields are not mandatory in …., … did not disclose this information...."

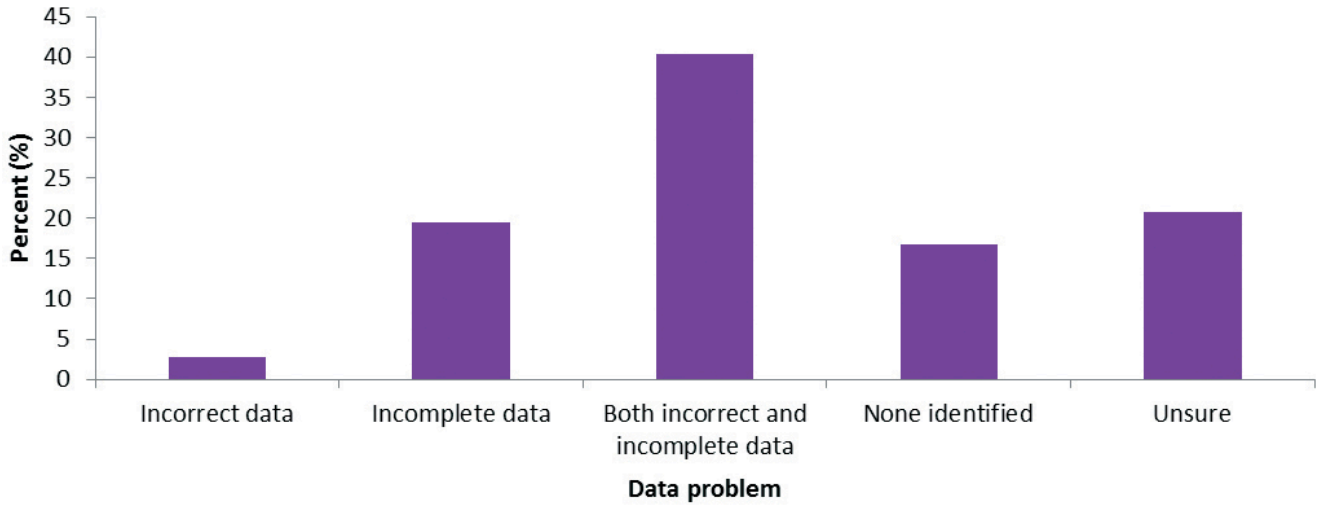
> "While we specifically do not use the data for purpose other than service reporting, we do provide CaLD data to other programs/areas that use the data for some of the above"

5.3.5 Barriers to CaLD data collection

5.3.5.1 Incorrect or incomplete data

Respondents were asked whether there were any problems with incorrect or incomplete CaLD-related data currently collected by their program. A response was mandatory. A total of 40.3 per cent of respondents (n=29) said both incorrect and incomplete data was a problem (Figure 2): 19.4 per cent of respondents said 'incomplete data was a problem' and just 2.8 per cent(n=2) said only incorrect data was a problem. This suggests incomplete data was a significant issue, but incorrect data was also a problem for some services.

Figure 2. Reported problems with CaLD-related data collected in WA health services by survey respondent frequency



Respondents stating problems with incorrect and/or incomplete data were asked to outline the problem, including any reasons that might explain the issue. Thirty-seven participants entered responses. Three themes emerged.

Theme 1: Patient-related factors

Respondents commented on factors related to the patient as reasons for either incomplete or incorrect data. Reasons included deliberate non-disclosure from patients related to mistrust or wishes to remain confidential, and accidental misinformation related to self-reported fields.

> "As an anonymous and confidential service, incomplete data will always be an issue; not all clients will disclose all information and we will not ask for information beyond what callers are willing to discuss."

> "On occasions demographic details may first be collected over the phone or at first face-to-face contact. Some families and young people may be unwilling to share all of their information or provide incorrect information which on most occasions will be updated if they engage with services. If they choose not to engage however there can be gaps in information collected."

Theme 2: Systemic or institutional factors

A large number (n=18) of respondents listed a systemic or institutional factor as reason for incomplete or incorrect data. Examples ranged from more general reasons such as lack of institutional and management prioritisation of CaLD data collection to more specific examples such as lack of referral form requirements to complete CaLD information (which form the basis for initial data entry) or insufficient options for entering languages related to emerging or local dialects or multiple languages. Further specific examples included lack of a mandatory collection requirement for some variables, inconsistent monitoring of entry, poor software capability to auto-populate across packages and insufficient updating of the interpreter required field over time as a family's English improves.

> "Thoroughness to request data and to carefully enter this data has not been monitored consistently."

> "Due to the nature of the integration between the Patient Administration System and our databases, patient information is not automatically updated therefore often incorrect. Many of the related data are not mandatory in our system and therefore are sometime missed or unknown at the time of entry"

Theme 3: Staff factors

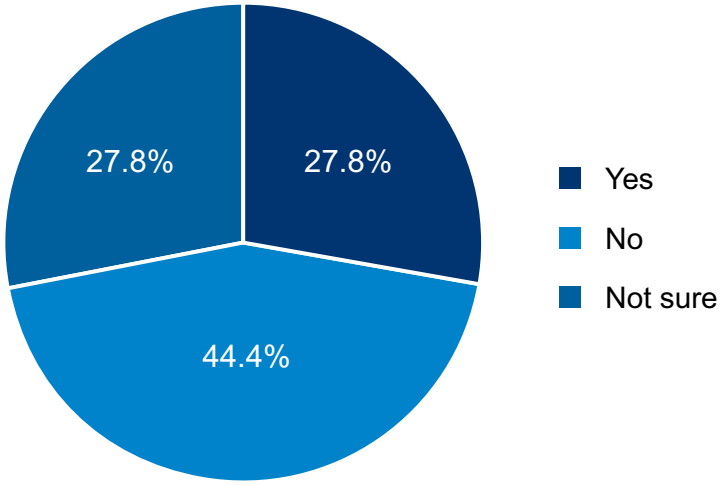
Respondents commented on staff factors as reasons for incorrect or incomplete entry. Specific staff issues nominated were 'lack of diligence' or 'attention to detail', human error, staff turnover, poor training, poor supervision, substandard compliance monitoring, staff overload and poor cultural awareness.

> "Incomplete data collected at time of registration of patients causes issues with providing appropriate services from commencement of treatment. This is possibly due to staff being focused on the particular information required by their area and lack of knowledge around complementary services...incorrect data in relation to country of birth, language spoken and requirement for interpreter. This can be due to unavailability of new and emerging languages in TOPAS language list so selections are often vague at best, if not completely incorrect."

5.3.5.2 Recent quality assurance processes

Respondents were asked to nominate whether their program had undertaken a quality assurance process in the last five years to review the correctness and completeness of CaLD-related data collection (Figure 3). A response was mandatory.

Figure 3. Proportion of respondents reporting program completion of a quality assurance process of CaLD-related data collection in the last five years



Respondents answering yes to completing a quality assurance process were asked on an optional basis to provide a brief description. Sixteen responses were entered. Multiple themes emerged in these responses.

Theme 1: Frequency of quality assurance processes is variable

Responses included comment on the frequency of quality insurance processes. Those mentioned varied widely from ongoing, regular (e.g. monthly, annually), to random review or audit of data.

Theme 2: Specific quality assurance program content

Respondents described varying specific quality assurance content details. Examples included annual audit of client records for each clinician (including 'interpreter needs identified, appropriately flagged and acted upon'), CaLD data assessment for accuracy and completeness or verification against medical records, audit on the use of interpreters for consent or otherwise, and evidence that patient files or outpatient head sheets had been marked as requiring an interpreter based on PAS entry.

Theme 3: Measures to improve data quality

Services reported on specific initiatives they had undertaken to improve CaLD data collection. These included increased education of clerical and clinical staff with regard to completion of CaLD data, and regular reminders to services to review all demographic-related data to ensure correctness.

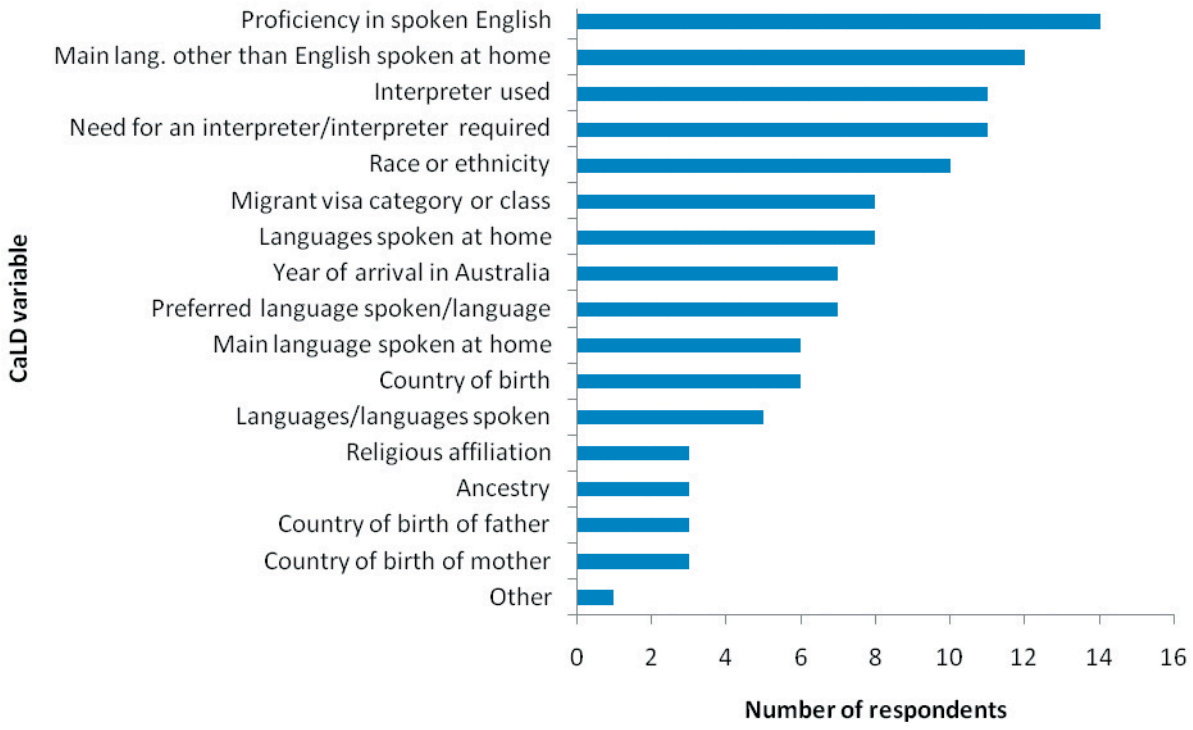
5.3.6 Future data collection

5.3.6.1 Variable suggestions

Respondents were asked whether there were any CaLD-related variables that would be useful to add to their current data sets. Completion was mandatory. A total of 38.9 per cent of respondents answered 'yes', 44.4 per cent 'no or not particularly' and 16.7 per cent 'not sure'.

Respondents who answered 'yes', were then asked to select which variable(s) would be useful to add (Figure 4). For the 'other' response option, the variable specified was 'English literacy'.

Figure 4. CaLD variables desired to be added to current collections



Respondents were asked to explain why each variable they had selected would be a useful addition to their CaLD data collection. Two themes emerged in the comments.

Theme 1: Reasons for selected variable additions

Reasons given for each variable are included in the table below (Table 4).

Table 4. Reasons given as to why variables would act as useful additions to current CaLD data collection

Variable	Reason for useful addition
English language proficiency	Consenting for research trials; informing need for interpreter (including if interpreter is 'unavailable or delayed') or other services; assisting in determining a format or level of language that can be understood by the client
Interpreter Service Required	A mandatory field to say whether they require (+/- used) an interpreter or not would be helpful for clinical service provision and for expenditure and service monitoring
Ethnicity	Identifying increased rates of particular diseases in some populations; identifying cultural differences associated with declines for screening
Migrant visa category	Financial and billing status; determining uptake of screening across visa groups; ensuring a universal screening and diagnosis system if some services were not covered by Medicare for particular CaLD clients
Language at home	Important for clinicians to have information to provide a better service
Year of Arrival in Australia	Financial billing; proxy indicator of likely proficiency in English and need for interpreter; identifying CaLD clients; identifying services to support client or general CaLD needs
Languages spoken	Assisting to inform languages of translated brochures or target services to underserved language groups; helping with patient care
Country of Birth of Mother/Father	Assisting with research and service review
Literacy	Assisting with patient information translation, appointments, English language acquisition rates; monitoring participation; targeting strategies and resources; evaluating recruitment to screening

Theme 2: Additional variable collection should be justifiable additions in terms of resource use

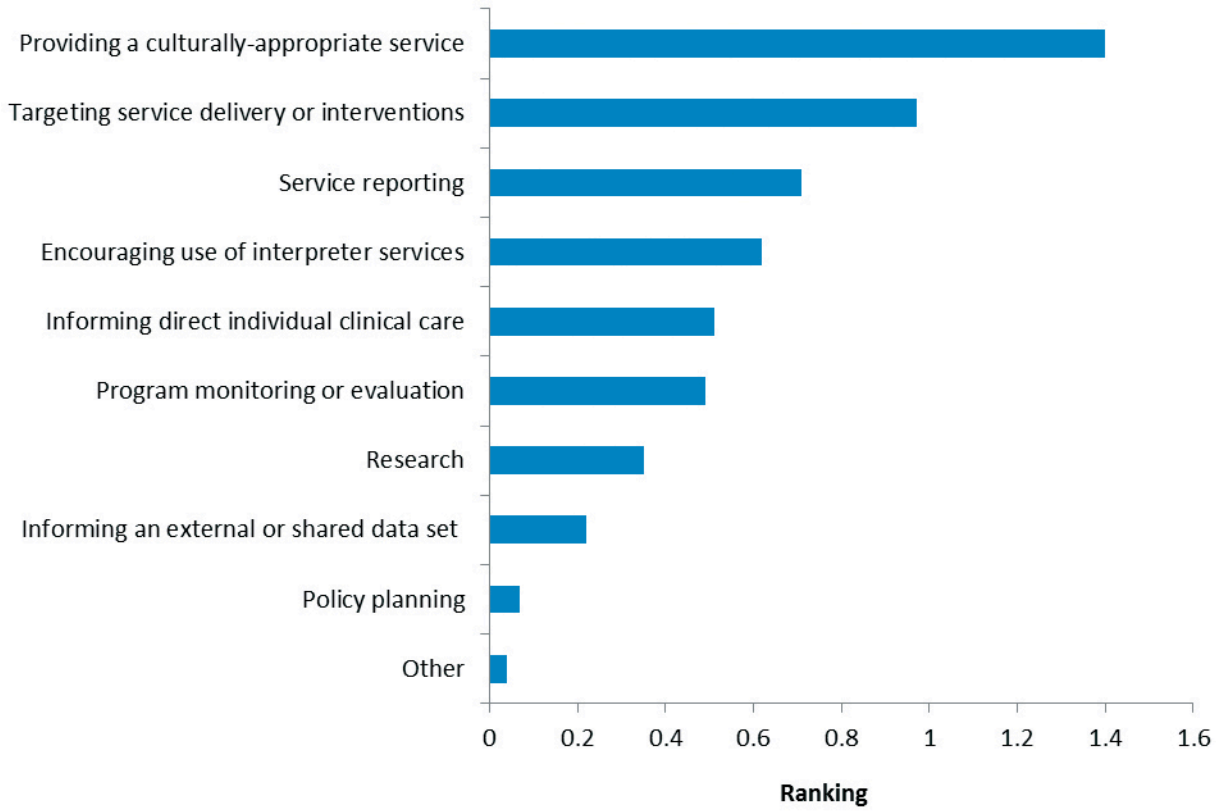
Comments pointed out the poor collection or use of existing variables and stated that further variables would need to be of significant benefit to justify addition.

> "Country of Birth, Race, Year of Arrival, Language and Interpreter Service Required are already in the system but are seldom used. Additional collection of data should have clinical meaning and improve service provided to patients, not just for reporting / research purposes as such requirements are costly (for system enhancement and clinician/administrative burden to collect the information)."

5.3.6.2 Potential program benefits of improved CaLD data collection

Respondents were asked to rank how an improved CaLD-related data collection could most assist their program (Figure 5). A response was optional. Ranking question components allowed respondents to prioritise a set of items from one to three, with one being the highest priority. The survey software then combined rankings into an overall score that could be used to determine the overall popularity of that item. This score was calculated using a combination of weighting for ranking position (first place has the highest weighting) and the number of respondents choosing each ranking position. This number was the item's ranking (Figure 5). A total of 66 out of 72 respondents completed the ranking. Options to answer 'not sure' and 'not applicable' were also given.

Figure 5. Program assistance that could be offered by improved CaLD data collection by ranking



Respondents were given the opportunity to provide further details on their rankings. Four comments were entered in response. One theme was identified.

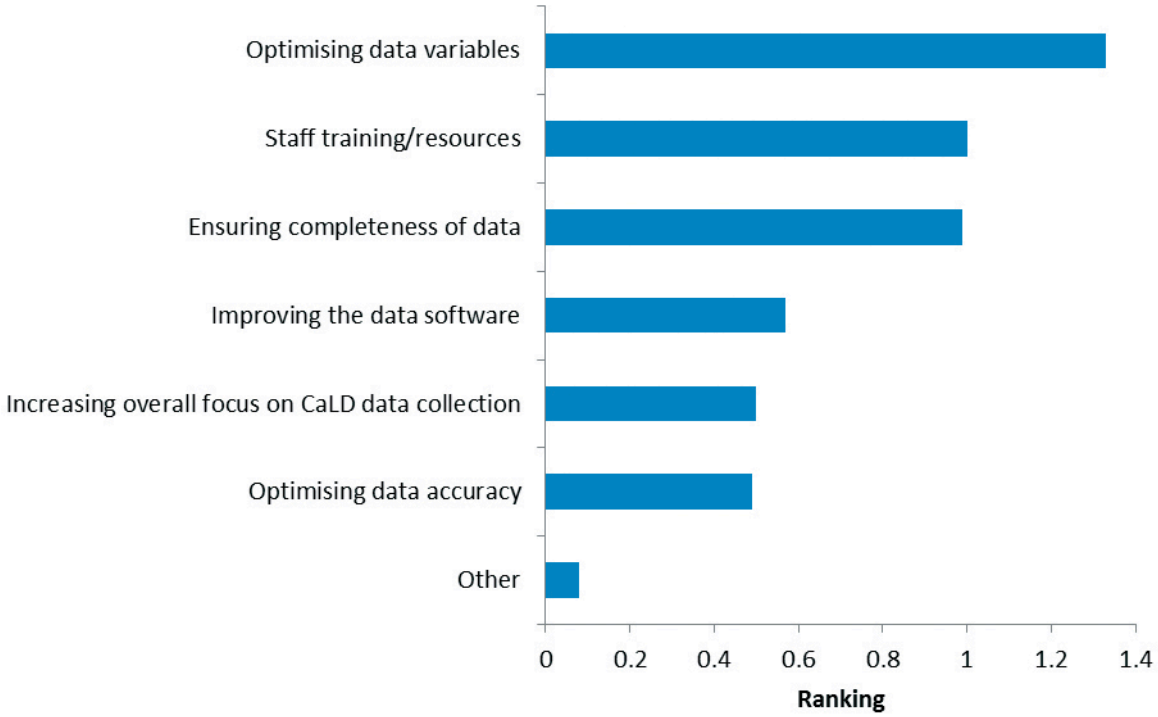
Theme: Improving CaLD data would provide program assistance in many ways

Comments suggested that improving CaLD data would assist programs across the many options listed in the question, beyond the top three priorities able to be selected.

5.3.6.3 Program priorities for improvement of CaLD data collection practices

Respondents were asked about their priorities for improving CaLD-related data collection practices in their program, should there be an opportunity. A response was optional. Response options included ranking the top three priorities out of seven options (Figure 6) or selecting 'not sure' or 'they are already optimal'. Attaining an overall rank for each of the items was conducted according to the ranking calculation process described previously above. Of the 61 respondents who completed a ranking, six selected 'not sure' and two 'they are already optimal'. 'Other' respondent nominated priorities included 'increasing automated integration' and 'interrogating the CaLD data we collect more than we currently do'.

Figure 6. Program priorities for improvement of CaLD data collection practices by ranking



5.3.7 General comments

5.3.7.1 General comments on CaLD-related data collection practices in their program

Respondents were asked for any other general comments they would like to make about CaLD-related data collection within their program. Nineteen participants provided a response across four themes.

Theme 1: Overall importance of CaLD data collection

Comments suggested overall a strong level of support for CaLD data collection.

> "CaLD is a difficult area to prospectively collect data. Collecting demographic data is essential. Any effort to automate the data collection greatly appreciated."

> "It is important we continue to monitor the demographics of young people and families accessing services to ensure we meet varying demands."

> "The data collection process for this programme is not fit-for-purpose and this has been raised with management along with the associated Clinical Incident Management System. Having worked internationally in this field, I am very aware of the lack of appropriate resources we provide for CaLD populations, and am very keen to overcome this. Any support would be welcome."

Theme 2: Specific desired improvements in CaLD data collection practices

Several comments centred around specific suggestions for improvement in CaLD data collection practices, such as improved auto-populating capacity across software packages to avoid need for duplicate data collection, increasing cultural competency audits of organisations, and engaging staff in data collection when they are not under time pressure.

> "Consider collecting data using/engaging staff when they are not compromised with regard to after-hours management limiting or discouraging proactive data collection."

Theme 3: CaLD data collection is sometimes determined by national or system-wide requirements

Comments alluded to necessities for some programs to collect specific variables according to wider system requirements. These included national population-wide screening programs for which minimum demographic collection is specified at a national level.

Theme 4: Funding as an important determinant of data collection priority

Further comments related to the incentive for complete data created by funding linkage to particular data variables. It was noted that CaLD variables are not currently linked to funding models, unlike examples given such as Indigenous Status and postcode. Comment was made that this may change in the future.

> "A National Costing study was undertaken recently by the Independent Hospital Pricing Authority (IHPA) and Health Policy Analysis (HPA) to identify potential drivers that increase the cost of ED episodes in hospitals. The study will be used to develop a new cost model for ED episodes. One of the items included in the study was a flag of whether the patient was unable to communicate in English. Analysis of the data that was collected is currently underway, and the investigation will include looking at whether patients that are unable to communicate in English has any impact on the cost/resourcing of the episode, and whether the data should be included in the model. Should this flag be included in the new model, it will be imperative that the EDDC ensure that this data is collected for all WA sites with a high degree of completeness and accuracy".

5.3.7.2 Suggestions for improvement in CaLD-related data collection practices across health services in WA

Finally, respondents were given the option to provide general suggestions for improvement of CaLD-related data collection by other WA health services either internal or external to the WA health system. A total of 24 responses were entered for this question, corresponding to five themes.

Theme 1: Increasing data sharing across organisations

Eleven respondents commented on increasing capacity to share data across services via suggestions including standardisation of variables, increasing auto-populating functions between software packages to avoid duplication of collection, increasing staff access to other databases to access different CaLD data, and increasing data linkage capability.

> "Improve the integrated services from the PAS for all systems. PAS is the foundation system that collects patient demographical information amongst other information. Collect once, use for many."

> "The ability to harness data from other collections, to supplement our own through Data linkage."

Theme 2: Innovation in data collection and use

Respondents commented on capacity to link or make use of innovative tools beyond the WA health system. This included a suggestion to use Google translate, My Health records for refugee clients, and a single central data set containing data from both government and non-government services to understand challenges facing all health services not just those publicly administered.

Theme 3: Benefits of CaLD data collection

Respondents recognised the benefits that would likely result from improved data, including reduced non-attendance at appointments and appropriate use of interpreters.

> "If wider groups collected this data, it would have the potential to decrease non-attendance costs (and lost health opportunities), improve communication with families, ensure interpreters were used appropriately (e.g. using interpreters to contact Limited English Proficiency families to notify of upcoming appointments rather than English letter in mail which they can't read and then will miss appointment). Also many CALD families are relatively itinerant, so phone contact with interpreter is more useful as letters often don't get re-directed."

Theme 4: Specific individual suggestions for improvement

Many respondents suggested unique measures for data improvement. These included statewide databases for specific disease-based programs that link to databases of other medical modalities, ceasing the use of written records, using an alert sticker for Interpreter Service Required on admission head-sheets and making fields mandatory on the first occasion of service.

Theme 5: Priority given to collecting CaLD data

Several comments related to the priority of collecting various CaLD data. Some comments alluded to the need for further priority, such as treating CaLD data collection as 'more than a tickbox exercise' and considering the imperative to collect data given the diversity of WA's population. Other comments suggested a considered balance between the benefits of CaLD data collection and the extra resources required to improve or change collection.

> "Consider a balance between administrative burden, i.e. clinician/admin time in recording the information vs reporting needs. Core system enhancements to add new fields are a costly exercise that needs funding. Perhaps prioritise:
>
> - providing training, additional guide for use, useful tools for clinicians to collect the information.
> - improving collection of country of birth and other CALD-related information that already exists in the system and improve the collection of this information prior to adding new data collection requirements."

6 Discussion

6.1 Use of CaLD data

Results demonstrate WA health system front-line services make varied use of CaLD data currently collected. Most respondents to the stakeholder survey listed multiple uses for data currently collected (median n=5). The most frequently nominated uses of 'service reporting' and 'to inform individual clinical care' suggest current data collection is being collected to inform both population level assessment of trends and individual service delivery. A focus on population level assessment of trends is consistent with stakeholder reports of ongoing requests to the WA Department of Health for summarised or raw CaLD core data.

Results of the survey suggest health services would make further use of data if data collection practices were improved. A vast majority of survey respondents completed the ranking of expected program benefits of improving CaLD data, with 'providing a more culturally appropriate service' achieving the highest ranking. However, conclusions that can be drawn from this are limited because quantitative response options were broad and few respondents entered further qualitative detail. Some comments in the survey overall alluded to specific benefits expected to result from improved collection such as a reduction in non-attendance at appointments or increased use of appropriate interpreters.

6.2 CaLD variables

Despite the recognised benefits of CaLD data collection, results demonstrate the configuration of CaLD variables collected across WA health system services is inconsistent and generally suboptimal. Review of the large data sets demonstrated most are collecting Country of Birth of Person, some Interpreter Service Required and only two were collecting any variable related to language. The Midwives Notification System was the only core data collection collecting 'ethnicity' directly, beyond Indigenous Status. Both Year of Arrival in Australia and Religious Affiliation were only collected by the Mental Health Information System. The core data sets were in general collecting the ABS Minimum Core Set in terms of Country of Birth of Person and Indigenous Status, but none were collecting the other variables Main Language Other Than English Spoken at Home or Proficiency in English.

Survey results were in general consistent with these findings, with most respondents reporting their program collected Country of Birth of Person and Interpreter Service Required. However, many reported collecting a language variable; this varies compared to the large data sets and presumably relates to a language variable appearing in primary software such as WebPAS but not feeding through as a core data collection requirement. Far fewer respondents reported collecting Ancestry, Year of Arrival in Australia, Visa Category or country of birth of either parent.

While the survey gave an overall indication of the frequency of variables collected, variables collected by specific databases were not assessed because of some suspected confusion with entering response options by some participants. This included possible difficulty in choosing between 'mandatory collection', 'optional collection', 'not collected' and 'unsure if collected' given the complexity of database functioning, and the multiple language variables offered. It is unlikely any database would collect more than one language data variable however multiple respondents selected multiple language variables.

The predominance of collection of Country of Birth of Person is consistent with findings in the literature for other Australian health jurisdictions. Despite the work by the ABS published in 1999 recommending collection of four variables as part of a Minimum Core Set (Country of Birth of Person, Main Language Other Than English Spoken at Home, Proficiency in English and Indigenous Status), evidence indicates considerable variation continues to exist in the measurement of ethnicity in Australian health data

collections and often Country of Birth of Person is the only variable collected. ABS Minimum Core Set Variables of 'Main Language Other Than English' and 'Proficiency in English' are rarely collected. Failing to include the Minimum Core Set variables beyond Country of Birth of Person severely limits the capacity to calculate population-level CaLD trends because numerator data is in a non-compatible format in comparison to denominator CaLD data obtained from the ABS Census.

Literature findings relating to health services across Australia are also consistent with the WA health system's additional focus on alternative language variables Interpreter Service Required and Preferred Language, rather than the two suggested by the ABS in the Minimum Core Set: Main Language Other Than English Spoken at Home and Proficiency in English. Practice in WA likely reflects historical circumstance but also the health system's requirement to focus on service needs, beyond the ABS's suggested measures. In addition, although the ABS states that the Minimum Core Set was piloted extensively across service settings prior to 1999, the questions for Main Language Other Than English Spoken at Home and Proficiency in English respectively, "[Do you] [Does the person] [Does (name)] speak a language other than English at home?" and "Do you consider [you speak] [(name) speaks] English very well, well, or not well?", may be difficult to practically implement in healthcare settings such as emergency departments and admissions to hospital. Concern exists that Main Language Other Than English Spoken at Home will tend to be erroneously collected as Main Language Spoken at Home, and Proficiency in English will not sufficiently directly inform the need for an interpreter. Overall, the number and type of ethnicity variables collected likely reflects both the pragmatic and logistical issues of minimising administrative and respondent burden, and the different primary purposes of health databases.

There is strong front-line support for assistance to review and improve the CaLD data variables currently collected. Survey respondents ranked 'reviewing and optimising CaLD data variables collected' as the highest priority for improvement of data collection over a range of other measures to increase data completeness or accuracy, improve data software or increase overall program focus on CaLD data. Survey respondent suggestions for specific variable additions to current collections should be interpreted with the knowledge that respondents from front-line services will likely be more familiar with the direct service use implications of data than broader statistical use. As an example, it is interesting that respondents ranked Proficiency in Spoken English as the highest priority for adding to existing variables collected. Comments suggest this is related to the variable's potential use in the service setting, such as to inform capacity to consent or understand information or the need for an interpreter (including if an interpreter is unavailable or delayed), rather than its potential use for aggregated statistical assessment. Main Language Other Than English Spoken at Home was the second highest ranking variable for addition to current collections; however, qualitative reasons were not specified. It was beyond the scope of the project to comprehensively compare additional variable suggestions with currently collected variables in survey-entered data collections tools, and identify opportunities for standardisation.

In the Australian setting no single measure is adequate for measuring CaLD, but rather a combination of variables is likely to be more useful. However, it is clear from the results that combinations of variables collected across health jurisdictions nationally variably differ from ABS recommendations. Currently a standardised quality approach in the health sector is an unresolved issue nationally with no stand-out or agreed combination of variables for collection across health settings. Generally the combination chosen by individual health organisations currently reflects the priorities, resources and logistics of the organisation, and for this reason particular variable issues relevant to WA are further discussed within the recommendations. Organisations that increase the number of CaLD variables collected can clearly achieve more with the data (since each variable generally measures a different aspect of CaLD background), but this must be balanced with administrative and responder burden and cost. One strategy to increase the number of variables possible to collect is smart use of filters or conditional questions based on prior responses, for example Preferred Language variable appears only if 'yes' is entered to Interpreter Service Required. In addition, variables that do not change over time, such as Year of Arrival in Australia,

in theory only need to be collected once and auto-populating functions across software packages could prevent re-collection. Data linkage across databases is a further mechanism to increase data variable availability and reduce duplicate collection.

6.3 CaLD data quality

Results suggest several issues with the quality of the CaLD data currently collected by the WA health system, both in terms of completeness and accuracy. A total of 59.7 per cent of survey respondents cited issues with their program's data completeness and 43.1 per cent cited issues with their program's data accuracy. Review of the large data sets confirmed several issues with CaLD data collection quality.

Final core data set records provide objective evidence of underlying issues with both completion and accuracy. Country of Birth of Person and Interpreter Service Required generally had a high rate of completion across all core data sets but Interpreter Service Required had data inconsistencies pointing to accuracy issues. For example, Interpreter Service Required was more than 99.9 per cent complete for HMDS and EDDC; however, the EDDC had 10.9 per cent of entries listed as 'not specified/unknown', likely to be higher than the true figure. Only 0.39 per cent of hospital admissions were listed as requiring an interpreter, but 0.8 per cent of admissions had a listed language; this demonstrates a data collection deficiency because language is meant to be a conditional field only completed if an interpreter is required. Other variables had issues with completeness: according to the MHIS Ambulatory Data Dictionary (Version 4 updated August 2010) 'Year of Arrival in Australia if from overseas' is only collected between 1966 and 2003 and 'religious affiliation at admission' is collected between 1967–present but 'missing data increases overtime (maximum 60% missing in 2003)' (Department of Health, 2010).

Several issues were identified within the statistical standards guiding core data set primary data collection, reducing the comparability of data. An important example relates to data collection associated with hospital admissions (the Hospital Morbidity Data Collection). The HMDS data dictionary does not specify a standardised question for determining 'requirement for interpreter', unlike the variable as listed in the National Health Data Dictionary on METeOR (AIHW's Metadata Online Registry), and the definition is vague. This is creating confusion, with evidence suggesting this variable is understood to be based on a retrospective determination (was an interpreter used) by data set users rather than a prospective determination (will the patient require an interpreter during this occasion of service?) by front-end data enterers.

Further statistical standard issues relate to the mandatory or optional nature of variable collection, and the classification of answer coding options by various core data sets. While Country of Birth of Person is generally dictated as a mandatory field in most of the core data sets, other fields were optional or no condition was specified. For example, Interpreter Service Required is not specified as either a mandatory or optional field in the EDDC, creating opportunity for incompletion. It is also important coding options are of high quality; having 'unknown' as an option for Interpreter Service Required would seemingly encourage erroneous entry of such as option, and perhaps options should be limited to 'yes', 'no', or 'patient not able to answer' in the emergency department setting. The classification of 'Ethnicity' within the Midwifes Notification System also does not align with the current Australian Standard Classification of Cultural and Ethnic Groups developed by the ABS (ABS, 2016b) limiting comparability of 'ethnicity' across data sets. Survey respondents also commented on insufficient entry options existing for emerging or local dialects and inability to enter multiple languages.

It should be noted that standardised data collection to achieve both highly reliable and valid data may not always be possible in the CaLD setting, rather it may be a trade-off between the two. For example, in the case of Interpreter Service Required, if standardisation was implemented to be based on a self-report question alone 'do you require an interpreter?' (creating reliability), this may miss patients who answer 'no' based on responder bias or misunderstanding related to language barriers, and therefore be less valid (Donato-Hunt & Grima, 2009).

Several issues were identified with primary data collection practice. The survey and key stakeholder engagement revealed data entry is dependent on the skills, staff turnover rate, attention to detail, staff overload, motivation and training of data entry clerks. One example of poor training relates to the suggestion by a consulted ward clerk that if a family could provide interpreter services during the admission of a relative (not recommended practice), this would result in a 'no' for interpreter Services Required. Multiple data collection software package settings across the WA health system (and within private hospital feeder systems to HMDS and EDDC) and Mental Health Commission services are likely to lead to variable data collection quality. In particular, some features of the current TOPAS, WebPAS and PSOLIS system (and likely other primary software packages) facilitate poor data quality. In these packages, CaLD data is currently collected as part of the demographic field and while in theory this is checked on each admission, the auto-populating process from previous admissions combined with lack of an automated field update history, creates uncertainty as to whether fields such as Interpreter Service Required or Preferred Language are accurate for the occasion of service in question, or even recently updated. Given the PAS auto-populating function it is likely that some clerks skip over these fields leaving existing entries in place.

Multiple fields are also optional within some primary data collection software packages resulting in incomplete data entry. One such example is PSOLIS; Interpreter Service Required, Preferred Language, Year of Arrival in Australia and Religion are all optional. These software issues and the resulting poor data quality feeding to the MHIS core data set have played a large part in the decision of the MHIS to withdraw most CaLD data fields in the upcoming MIND data set replacement for the current MHIS. Survey respondents also suggested other reasons for incomplete data, such as lack of a requirement to specify need for interpreter on referral forms or patient incapacity or unwillingness to respond in some healthcare situations.

Finally, it is at the discretion of health services how they implement the data standards for primary data collection. This can create variability as to whether the data statistical standard is actually followed and whether managerial and staff priority is given by any particular service to ensure accurate data collection. Survey findings identified lack of institutional prioritisation as a key theme in explaining poor CaLD data quality.

There is limited Australian literature on CaLD data accuracy and completeness issues. The local findings discussed above are consistent with AIHW identifying inconsistent implementation of associated ABS or other standards for many of the CaLD variables being collected across aged care service settings. Difficulties associated with collecting Main Language Other Than English Spoken at Home, including erroneous interpretation of this question as Main Language Spoken at Home even by researchers and AIHW (AIHW, 2014; Donato-Hunt & Grima, 2009), were not an issue locally as this variable is not currently collected, but are considered in making recommendations.

6.4 CaLD data sharing

Project results have important implications for maximising the use of currently collected CaLD data. The large core data sets collate data from healthcare services via a large number of 'feeder' software packages used for primary data collection. The PAS system is the predominant primary data collection package used in the public hospital system, and has an auto-populating function across a number of other primary data collection packages. Thus PAS directly or indirectly constitutes the predominant source data for a number of the core data sets. PAS currently collects Country of Birth of Person, Religious Affiliation, Interpreter Service Required and Preferred Language; however, not all of these variables feed through to the various core data sets. Notable amongst these is the current lack of Preferred Language being collected by the EDDC. It would therefore be relatively straightforward to standardise collation of Country of Birth of Person, Interpreter Service Required and Preferred Language across all of the core data sets feeding from PAS. PAS is also used by multiple specialised programs within or strictly outside of ED

presentations, inpatient or outpatient services, so implementing a standardised and mandatory set may assist data collection more widely or for specific health issues.

Sharing data across organisations was a predominant theme in general suggestions for improvement in CaLD-data collection practices across WA health services. Specific suggestions included improving the auto-populating function of PAS or other software packages to minimise duplicate collection, increasing staff access to other databases to access different CaLD data, and increasing data linkage capability.

6.5 CaLD data priority

Overall, results suggest there has traditionally been a low priority given to CaLD data collection across most of the WA health services assessed. Many survey respondents reported lack of institutional prioritisation as a reason for poor quality data and less than three in 10 respondents said their service had completed some form of quality assurance check on their data in the previous five years. Quality assurance processes were variable in their format and level of rigour. There were a few exceptions, with a small number of services with a sizeable CaLD clientele having significant commitment to quality CaLD data collection. At a system level, some data custodians reported there had been no state or national mandate to ensure collection of CaLD information into particular core data sets, and as such comprehensive policies and guidelines had never been developed to inform recording practices, resulting in incomplete and inaccurate recording of this information.

Literature suggested the topic has also been given a low priority in other jurisdictions. Despite a reported review of CaLD data collection practices by the ABS in 2011, there have been no ABS publications on the topic overall since the 1999 document *Standards for Statistics on Cultural and Language Diversity*. The AIHW review of CaLD data collection demonstrated variable implementation of the Standards across aged care collections (AIHW, 2014). At a state level, Queensland Health established a minimum core set in 2007 with a review paper in 2012 (Queensland Health, 2012). This is seemingly consistent with the situation internationally; despite many national guidelines for collecting ethnicity data, considerable variability remains if and when ethnicity is measured, in diverse arenas including national censuses (Morning, 2008) and published biomedical literature (Ma et al., 2007).

6.6 Limitations

The limited resources available to this project meant that comprehensive assessment of data quality issues and implementation processes of potential solutions was beyond the scope of the study. In addition, the project focused on large public health services funded and provided by the WA Government, and data collection relevant to healthcare access and health outcomes rather than risk factors or determinants of health.

Stakeholder consultation was largely limited to survey feedback, which is likely to have limited the capacity to make thorough assessment of what is a complex topic with complex solutions. Limited understanding of a complex topic was evident in some survey responses, where respondents entered inconsistent or no responses at times, limiting the extent of analysis possible. In-depth interviews, further observation of data collection practices or workshop-style consultation was beyond the scope of the project.

7 Recommendations

While implementation of improved CaLD data collection will involve resources, it will also improve the quality of the information collected and its comparability with data collected by other organisations. Better data can add significantly to an organisation's effectiveness.

The healthcare sector is likely to undergo significant change in coming years with the increasing use globally of 'big data' and electronic patient records. Getting the right practices in place now will assist any future efforts to maximise the use of CaLD data. There is an opportunity for the WA health system to be a leader in CaLD data collection, but this will take resource commitments and consideration of opportunity costs associated with spending in other areas.

7.1 CaLD variables

Statistical measurement of CaLD, which is generally derived from the ABS Standard recommendations, provides for broad analysis of diversity in a population but is seemingly less practical for assessing the associated service needs. With service needs particularly relevant in the health sector, ABS recommendations cannot be easily applied.

1. **It is recommended that the WA health system develops its own interim minimum core set for initial implementation across the Patient Administration System and the large core data sets. It is proposed this includes:**
   - **Country of Birth of Person**
   - **Need for interpreter**
   - **Preferred Language**
   - **Indigenous Status**
   - **Main Language Other Than English Spoken at Home** – this should be added subject to piloting. It should act as a filter question to Interpreter Service Required, which could then filter to Preferred Language.

Additional optional variables that could be added to the WA health system minimum core set according to program needs and in order of preference are:

- **Year of Arrival in Australia**
- **Country of birth of parents and Year of Arrival in Australia of parents**
- **Ancestry**
- **Religious Affiliation**
- **Proficiency in Spoken English.**

This minimum core set recommendation recognises both the value of Interpreter Service Required and Preferred Language in determining individual service needs and the significant effort and cost it would require implementing the ABS Minimum Core Set variable of Proficiency in English as an alternative to Interpreter Service Required. Concern exists that Proficiency in English may fail to accurately identify those needing an interpreter service in a significant number of clients (Donato-Hunt & Grima, 2009), despite ABS testing prior to the 1986 Census reportedly showing that it is a good identifier of people who are likely to need assistance in the form of interpreter services (ABS, 1999). Proficiency in English is also likely to be less practical than Interpreter Service Required in a health service setting, given the relative complexity of the written question, and the verbal question, "Do you consider [you speak][name speaks] English very well, well or not well?". The alternative ABS suggestion to simply add their Minimum Core

Set variables to existing collections may be impractical and an unjustifiable use of resources in the case of collecting Proficiency in English in addition to Interpreter Service Required.

Main Language Other Than English Spoken at Home is included because it would provide a broader measure of CaLD status, and very importantly is included in the ABS Standards for Statistics on Cultural and Language Diversity, allowing for comparability and calculation of rates using Census data. It is recommended that Main Language Other Than English Spoken at Home is piloted before roll out to entire data sets or software packages. Some literature suggests health service confusion and non-standard implementation of this variable, with some services changing the variable to instead be, Main Language Spoken at Home. While similar, they are not the same and not comparable. It would be important the standard question is clearly visible to staff to avoid confusion for Main Language Spoken at Home: '[Do you] [Does the person] [Does (name)] speak a language other than English at home?'

Country of Birth of Person is recommended as it is currently in wide use, aligns with ABS recommendations, is relatively objective, stable and comparable, and gives some indicator of cultural background. In the perinatal setting Country of Birth of Mother/Father is comparable and should be implemented. Indigenous Status is included as part of the Minimum Core Set in line with ABS recommendations and comprehensive identification of ethnic background, but it is not relevant to the 'CaLD' definition used in this report.

Supporting evidence for the additional optional variable recommendations is included in Appendix 4.

Although the recommendation is for initial implementation focused on the widely used Patient Administration System and the WA Department of Health large core data sets, further efforts could target wider uptake of the minimum core set across health services.

2. **It is recommended the WA health system work with other jurisdictions over coming years to develop a standardised national approach to health-related CaLD data collection.**

The recommendation above for a minimum core set for WA health services is based on the author's knowledge gained during this project, rather than a systematic analysis of the best combination of variables. AIHW and the ABS have each conducted such systematic processes via consultation with regard to a number of criteria and/or piloting previously, but these were targeted at determining the ideal variables for aged care servicing and more general CaLD statistics respectively (ABS, 1999; AIHW, 2014). There is a need both within WA and nationally to comprehensively determine the ideal CaLD variable combination for collection across the health sector. This could include piloting of variables such as Main Language Other Than English Spoken at Home and Proficiency in English, in comparison with Interpreter Service Required and Preferred Language. The WA health system should seek to cooperate nationally or with other states to determine standardised recommended or mandatory minimum CaLD indicators for use in most health settings. Alternatively WA could develop a local process to examine and compare variables systematically.

In general, measures should be chosen within the wider context of informing an equitable distribution of power and resources and in keeping with national ABS standards. They should ideally 'value-add' to choose measures that together will provide both statistical measurement of cultural background and inform direct service delivery.

Specifically, selection should include consideration of the following factors:

- Overall agreed aims of CaLD data collection in the WA healthcare setting
- Identifying language needs to inform clinical care
- Identifying CaLD groups within Western Australia potentially at particular disadvantage in terms of access and health outcomes, for example refugees, those recently arrived and those from particular regions.
- Ethnicity-related hereditary genetic diseases
- Capacity for data linkage and consistency across sectors and with ABS Standards
- Likely future activity-based funding scenarios e.g. extra allocation for services to CaLD clients.
- Patient-centred care, for example Service users may prefer Preferred Language over Main Language Other Than English Spoken at Home, or capacity to decline an interpreter, rather than on the subjective Proficiency in Spoken English.
- Patient and community acceptability of measures, for example Entry visa category, may not be considered appropriate
- Practicality and ease of use
- Likely quality of data collected including accuracy and reliability
- Existing variables being collected either within primary software data collection tools or by the core data sets, and the associated cost of changing variables.

7.2 CaLD data quality

Given evidence of significant data quality issues within existing CaLD data collections, it is prudent to consider data quality improvements in conjunction with any efforts to review or expand CaLD variables.

3. **It is recommended that specification of standards for collection of CaLD variables are improved within each of the relevant core data set dictionaries.**

For each variable, this should include an explicit standard: name, definition, question(s), classification of response, coding procedure and output category. Where possible, this should align with the Standards used by the ABS or the National Data Dictionary to allow data comparability. Response categories should minimise capacity to enter ambiguous responses such as ‘unknown’.

4. **It is recommended WA health system minimum core set variables are made mandatory rather than optional across the core data sets.**

Existing incomplete data is likely largely due to optional entry of responses for many of the CaLD variables. Making minimum core set variables mandatory across collection for the core data sets and eventually the WA health system more generally will increase the institutional priority given to data completion.

5. **It is recommended existing primary data collection tools (software packages) are improved to incorporate an automatic history of field updates function.**

Incorporating an automatic field update history function will encourage routine checking of data on each occasion of service, and enable auditing of data collection quality.

6. **It is recommended staff have comprehensive training to understand correct data collection technique and underlying reasons for CaLD data collection.**

7.3 Maximising the use of collected data

There are many potential opportunities to capitalise on existing or future data collection.

7. **It is recommended that the auto-populating function of PAS to other user interfaces is expanded and primary data collection 'feeder' systems are improved to maximise collation of primary data to core data sets.**

WebPAS (progressively taking over from TOPAS) currently has good functionality to auto-populate many other linked patient IT systems in the public hospital system but this should be expanded to minimise duplicate collection and expand the range of data sets available with comparable CaLD data. Feeder systems should maximise the collation of primary data collections.

8. **It is recommended consideration be given to increasing the data linkage capability for CaLD data or identifying CaLD groups by other innovative means.**

Consideration should be given to developing an algorithm to identify CaLD persons using data collected across the various core data sets, similar to that developed for Indigenous Status. Data linkage between children and parents could be particularly useful in the CaLD population as many refugees arrive as young adults of child-bearing age, or as a member of a young family. The majority of refugees arriving in Western Australia visit the WA health system's Humanitarian Entrant Health Service on arrival and get allocated a Unique Record Number. This same Unique Record Number will apply should the patient subsequently present to public hospital services in WA. A project could therefore examine hospital outcomes for the refugee population in comparison to the non-refugee population.

7.4 Other steps

9. **It is recommended a working group is established and staff time dedicated to progress the improved collection of CaLD data.**

Tasks could include progression of above recommendations and consideration of data collection issues that were beyond the scope of this project. This could include ongoing review, revision and refinement of CaLD variables, standards and implementation across services, with associated support and tracking of progress on a system-wide basis. It could also include review of CaLD data collection in relation to WA health system surveys, CaLD risk factor and preventative health data, health service outcomes (e.g. was an interpreter used) and consideration of variables related to specific needs associated with cultural background such as diet, gender of caregiver etc. A working group could have an ongoing role to identify opportunities to implement changes via IT system reforms, tendering of services, redesign of client referral forms, and input into program reviews. There could also be a role to engage senior leadership to ensure collection and provide support for staff training. The working group could also contribute to developing wider system prioritisation of servicing to CaLD clients via a number of policy mechanism options, as undertaken in other states; data collection would then become part of a more comprehensive approach.

8 Conclusion

The WA health system is committed to equal opportunity and diversity. It is essential health services are collecting appropriate data to identify, measure and address the level of disadvantage in CaLD individuals and groups. This report represents an initial review into CaLD data collection practices across key WA health services. It includes comparison to recommended guidelines and the practices of other health jurisdictions. The project found a number of issues with current data collection in terms of variables in use and the completeness and accuracy of collected data. The report makes a number of recommendations for improvement including development of a 'minimum core set' of variables for collection across health services, several measures to enhance the quality of collection, maximising the use of existing data collection, and provision of resources to implement recommendations and make other improvements.

9 References

Abouzeid, M., Bhopal, R. S., Dunbar, J. A., & Janus, E. D. (2014). The potential for measuring ethnicity and health in a multicultural milieu–the case of type 2 diabetes in Australia. Ethnicity & Health, 19(4), 424-439.

ABS. (1995). *Ethnic and Cultural Diversity in Australia*. Canberra.

ABS. (1999). *Standards for Statistics on Cultural and Language Diversity*. Australian Bureau of Statistics, Canberra.

ABS. (2002). *National Health Survey: Summary of Results*. ABS Cat No. 4364.0, Australian Bureau of Statistics, Canberra.

ABS. (2014). 3417.0.55.001 - Microdata: Australian Census and Migrants Integrated Dataset, 2011. from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3417.0.55.001Main+Features12011

ABS. (2015). 4364.0.55.001 - National Health Survey: First Results, 2014–15. from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4364.0.55.0012014-15?OpenDocument

ABS. (2016a). 1200.0.55.003 - Religious Affiliation Standard, 2016. from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1200.0.55.003Main+Features12016?OpenDocument

ABS. (2016b). *1249.0 Australian Standard Classification of Cultural and Ethnic Groups (ASCCEG), 2016*. Australian Bureau of Statistics, Canberra: Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1249.02016?OpenDocument

ABS. (2016c). 1266.0 - Australian Standard Classification of Religious Groups, 2016. from http://www.abs.gov.au/ausstats%5Cabs@.nsf/0/08D67CE7C3A715C5CA2570D700130F04?Opendocument

ABS. (2016d). 4329.0.00.001 - Cultural and Linguistic Characteristics of People Using Mental Health Services and Prescription Medications, 2011. from http://www.abs.gov.au/ausstats%5Cabs@.nsf/0/F7F2748508863DFECA257FC60018866D?Opendocument

ABS. (2017a). 2016 Census: Western Australia [Press release]. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/mediareleasesbyReleaseDate/27CAE1A5F7F2BC90CA258148000A4331?OpenDocument

ABS. (2017b). 2071.0 - Census of Population and Housing: Reflecting Australia - Stories from the Census, 2016. from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2071.02016?OpenDocument

ABS. (2017c). 1200.0.55.009 - Ancestry Standard, 2014, Version 2.1. from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/1200.0.55.009Main%20Features12014,%20Version%202.1?opendocument&tabname=Summary&prodno=1200.0.55.009&issue=2014,%20Version%202.1&num=&view=

ABS. (2017d). 2071.0 – Census of Population and Housing: Reflecting Australia – Stories from the Census, 2016, Cultural Diversity Data Cubes. Retrieved 18/01/2018. from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2071.02016?OpenDocument

ABS. (2017e). 4364.0.55.002 - Health Service Usage and Health Related Actions, Australia, 2014-15 from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4364.0.55.0022014-15?OpenDocument

ABS. (2017f). 3415.0 - Migrant Data Matrices, 2017; Health. from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/3415.0Main%20Features92017?opendocument&tabname=Summary&prodno=3415.0&issue=2017&num=&view=

AIHW. (2004). *Australia's Health 2004. The ninth biennial report of the Australian Institute for Health and Welfare*. AIHW Cat No AUS44. Australian Institute of Health and Welfare, Canberra.

AIHW. (2012). *National Health Data Dictionary Version 16*. Australian Institute of Health and Welfare, Canberra: Retrieved from https://www.aihw.gov.au/getmedia/a90ff140-8abf-4e30-8861-da2ebf368d35/14306.pdf.aspx?inline=true

AIHW. (2014). *Cultural and linguistic diversity measures in aged care. Working paper*. Australian Institute of Health and Welfare, Canberra.

AIHW. (2016). *Australian's Health 2016*. Australia's health series no. 15. Cat. no. AUS 199. Canberra: Australian Institute of Health and Welfare.

AIHW. (2017a). About METeOR. from http://meteor.aihw.gov.au/content/index.phtml/itemId/276533

AIHW. (2017b). WA Health Non-Admitted Patient Activity and Wait List Data Collection (NAPAAWL DC) 2016-17. *Australian Institue of Health and Welfare, Metadata Online Registry*. from http://meteor.aihw.gov.au/content/index.phtml/itemId/648949

Aminisani, N., Armstrong, B. K., Egger, S., & Canfell, K. (2012). Impact of organised cervical screening on cervical cancer incidence and mortality in migrant women in Australia. *BMC Cancer, 12*, 491.

Anikeeva, O., Bi, P., Hiller, J. E., Ryan, P., Roder, D., & Han, G. S. (2010). The health status of migrants in Australia: a review. *Asia-Pacific Journal of Public Health*, 22(2), 159-193.

Aspinall, P. J. (2001). Operationalising the collection of ethnicity data in studies of the sociology of health and illness. *Sociol Health Illn*, 23(6), 829-862. doi: 10.1111/1467-9566.00277.

Australian Human Rights Commission. (2014). Face the facts: Aboriginal and Torres Strait Islander Peoples, available at: https://aifs.gov.au/cfca/sites/default/files/publication-documents/ps6.pdf

Australian Institute of Family Studies (2011): Working with Indigenous children, families, and communities: Lessons from practice, available at https://aifs.gov.au/cfca/publications/working-indigenous-children-families-and-communities

Bhopal, R. (2004). Glossary of terms relating to ethnicity and race: for reflection and debate. *Journal of Epidemiology and Community Health*, 58(6), 441-445. doi: 10.1136/jech.2003.013466.

Blignault, I., & Haghshenas, A. (2005). Identification of Australians from culturally and linguistically diverse backgrounds in national health data collections. *Australian Health Review*, 29(4), 455-468.

Borrie, W. D., Beaurepaire, B., Kiosoglous, J., Sheldrake, P., & Zubrzycki, J. (1984). *The Measurement of Ethnicity in the Australian Census of Population and Housing (The Borrie Report)*. Report to the Australian Statistican by the 1986 Population Census Ethnicity Committee, Australian Bureau of Statistics, Canberra.

Castles, I. (1991). *Multicultural Australia, Catalogue No. 2505.0*. Canberra: Australian Bureau of Statistics.

Chaves, N., Paxton, G., Biggs, B., Thambiran, A., Smith, M., Williams, J., . . . Davis, J. (2016). Recommendations for comprehensive post-arrival health assessment for people from refugee-like backgrounds.Australasian Society for Infectious Diseases and Refugee Health Network of Australia Guidelines writing group.

Cultural and Indigenous Research Centre Australia. (2017). Consumer health information needs and preferences: perspectives of cuturally and linguistically diverse and Aboriginal and Torres Strait Ilander people. Sydney: ACSQHC.

Davidson, N., Skull, S., Burgner, D., Kelly, P., Raman, S., Silove, D., . . . Smith, M. (2004). An issue of access: Delivering equitable health care for newly arrived refugee children in Australia. *Journal of Paediatrics and Child Health, 40*(9-10), 569-575. doi: 10.1111/j.1440-1754.2004.00466.x.

Day, G. E. (2016). Migrant and Refugee Health: Advance Australia Fair? *Australian Health Review, 40*(1), 1-2.

Department of Health. (2006). *Guidelines for Completion of the Notification of Case Attended Health Act (Notification by Midwife) Regulations Form No.2*. Department of Health, Government of Western Australia, Perth: Retrieved from http://www.health.wa.gov.au/healthdata/docs/8971-GUIDELINES-A6-21602.pdf

Department of Health. (2007). *Emergency Department Data Collection Data Dictionary Version 1.0* Department of Health, Government of Western Australia, Perth: Retrieved from http://www.health.wa.gov.au/healthdata/docs/eddc_dictionary.pdf

Department of Health. (2010). *Mental Health Information System (MHIS) Ambulatory Data Dictionary (1966 – Present)* Department of Health, Government of Western Australia, Perth.: Retrieved from https://www.datalinkage-wa.org.au/sites/default/files/MHIS_ambulatory_data_items_1.pdf

Department of Health. (2015a). *Hospital Morbidity Data System, HMDS Reference Manual July 2015*. Perth: Department of Health, Government of Western Australia..

Department of Health. (2015b). *Guidelines for Completing Notification of Case Attended for Birth*. Department of Health, Government of Western Australia, Perth: Retrieved from http://ww2.health.wa.gov.au/~/media/Files/Corporate/general%20documents/Data%20collection/PDF/Guidelines_Completion_of_NOCA_201607.pdf

Department of Health. (2017). *WA health system Language Services Policy*. Perth: Retrieved from http://www.health.wa.gov.au/circularsnew/pdfs/13359.pdf

Department of Health & Human Services. (2015). *Refugee and asylum seeker health services.Guidelines for the community health program*. Melbourne: Victorian Government.

Department of Home Affairs. (2017). Fact sheet – Australia's Refugee and Humanitarian programme. from https://www.homeaffairs.gov.au/about/corporate/information/fact-sheets/60refugee

Department of Human Services Victoria. (2006). Doing it with us not for us, Strategic Direction.

Donato-Hunt, C., & Grima, K. (2009). Counting diversity: Indicators of cultural diversity used by drug and alcohol workers in Australia. Sydney: The Drug and Alcohol Multicultural Education Centre.

Drevdahl, D. J., Philips, D. A., & Taylor, J. Y. (2006). Uncontested categories: the use of race and ethnicity variables in nursing research. *Nursing Inquiry*, 13(1), 52-63.

Garrett, P. W., Dickson, H. G., Whelan, A. K., & Whyte, L. (2010). Representations and coverage of non-English-speaking immigrants and multicultural issues in three major Australian health care publications. *Australia and New Zealand Health Policy, 7*(1), 1. doi: 10.1186/1743-8462-7-1.

Gushulak, B. D., Pottie, K., Hatcher Roberts, J., Torres, S., DesMeules, M., Canadian Collaboration for, I., & Refugee, H. (2011). Migration and health in Canada: health in the global village. *CMAJ Canadian Medical Association Journal, 183*(12), E952-958.

Hadgkiss, E. J., & Renzaho, A. M. (2014). The physical health status, service utilisation and barriers to accessing care for asylum seekers residing in the community: a systematic review of the literature. *Australian Health Review, 38*(2), 142-159.

Harawa, N. T., & Ford, C. L. (2009). The foundation of modern racial categories and implications for research on black/white disparities in health. *Ethn Dis, 19*(2), 209-217.

Health Research and Educational Trust. (2011). *Improving Health Equity through Data Collection and Use: A Guide for Hospital Leaders*. Chicago: Health Research and Educational Trust.

Kunz, C., & Costello, L. (2003). 2001 Census: Ancestry – Detailed Paper. ABS Census Paper 03/01a.

Ma, I. W. Y., Khan, N. A., Kang, A., Zalunardo, N., & Palepu, A. (2007). Systematic review identified suboptimal reporting and use of race/ethnicity in general medical journals. *Journal of Clinical Epidemiology, 60*(6), 572-578. doi: 10.1016/j.jclinepi.2006.11.009.

Minas, H., Kakuma, R., San Too, L., Vayani, H., Orapeleng, S., Prasad-Ildes, R., . . . Oehm, D. (2013). Mental health research and evaluation in multicultural Australia: developing a culture of inclusion. *International Journal of Mental Health Systems, 7*(23).

Morning, A. (2008). Ethnic Classification in Global Perspective: A Cross-National Survey of the 2000 Census Round. *Population Research and Policy Review, 27*(2), 239-272. doi: 10.1007/s11113-007-9062-5.

Murray, S. B., & Skull, S. A. (2005). Hurdles to health: immigrant and refugee health care in Australia. *Aust Health Rev, 29*(1), 25-29.

National Health and Medical Research Council. (2005). *Cultural Competency in Health: A guide for policy, partnerships and participation*. Canberra: Australian Government.

Neckowaya, R., Brownlea, K., & Castellana, B. (2007). Is attachment theory consistent with Aboriginal parenting realities? First Peoples Child & Family Review: A Journal of Innovation and Best Practices in Aboriginal Child Welfare Administration, Research, Policy & Practice, 65.

NHS. (2005). *A Practical Guide to Ethnic Monitoring in the NHS and Social Care*. London.

NSW Department of Health. (2008). *The Health of the People of New South Wales – Report of the Chief Health Officer*. Sydney.

NSW Department of Health. (2010). *The Health of the People of New South Wales – Report of the Chief Health Officer*. Sydney.

Office of Management and Budget. (1997). *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. Federal Register:62: No.210, October 30.

Office of Multicultural Affairs. (2017) *Victorian Government Standards for Data Collection on Interpreting and Translating Services*. Retrieved from http://www.multicultural.vic.gov.au/images/stories/pdf/victorian-government-standards-for-data-collection.pdf

Office of Multicultural Interests. (2010). *Implementing the Principles of Multiculturism Locally, Department of Local Government*. Perth.

Phillips, J., & Simon-Davies, J. (2017). Migration to Australia: a quick guide to the statistics. from https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp1617/Quick_Guides/MigrationStatistics

Porter, M., Todd, A. L., & Zhang, L. Y. (2016). Ethnicity or cultural group identity of pregnant women in Sydney, Australia: Is country of birth a reliable proxy measure? *Women & Birth: Journal of the Australian College of Midwives, 29*(2), 168-171.

Queensland Health. (2012). *Literature review: Best practicce multicultural policy implementation*. Attachment B to the Queensland Health guideline for multicultural health policy implementation, Brisbane.

Rao, D. V., Warburton, J., & Bartlett, H. (2006). Health and social needs of older Australians from culturally and linguistically diverse backgrounds: issues and implications. *Australasian Journal on Ageing, 25*(4), 174-179. doi: 10.1111/j.1741-6612.2006.00181.x.

Rechel, B., Mladovsky, P., & Devillé, W. (2012). Monitoring migrant health in Europe: A narrative review of data collection practices. *Health Policy, 105*(1), 10-16. doi: http://dx.doi.org/10.1016/j.healthpol.2012.01.003

Sheikh-Mohammed, M., Macintyre, C. R., Wood, N. J., Leask, J., & Isaacs, D. (2006). Barriers to access to health care for newly resettled sub-Saharan refugees in Australia. *Med J Aust, 185*(11-12), 594-597.

Supreme Court of Western Australia. (2017). 7.1.1. Population. Retrieved 08/01/2018, from http://www.supremecourt.wa.gov.au/equaljustice/P/population_print.aspx

Taylor, R., Mamoon, H., Morrell, S., & Wain, G. (2001). Cervical screening in migrants to Australia. *Aust N Z J Public Health, 25*. doi: 10.1111/j.1467-842X.2001.tb00551.x.

Thow, A., & Waters, A.-M. (2005). *Diabetes in culturally and lingusitically diverse Australians: identification of communities at high risk*. Canberra.

Vang, Z., Sigouin, J., Flenon, A., & Gagnon, A. (2015). The Healthy Immigrant Effect in Canada: A Systematic Review. *Population Change and Lifecourse Strategic Knowledge Cluster Discussion Paper Series, 3*(1).

Victoria University. (2009). Review of Current Cultural and Linguistic Diversity and Cultural Competence Reporting Requirements, Minimum Standards and Benchmarks for Victoria Health Services Project. In P. Higgins (Ed.). Melbourne: Victorian Government Department of Health.

Victorian Government. (2015). *Victorian Government Report in Multicultural Affairs 2014-2015*. Melbourne.

Wallman, K. K., Evinger, S., & Schechter, S. (2000). Measuring our nation's diversity: developing a common language for data on race/ethnicity. *Am J Public Health, 90*(11), 1704-1708.

10 Appendices

Appendix 1

Copy of online survey

WA Health Services' Data Collection Relating to Culturally and Linguistically Diverse (CaLD) Persons

Overview

You are invited to provide feedback on how your organisation or program is collecting data relating to culturally and linguistically diverse (CaLD) clients. *For the purposes of this consultation, the term "culturally and linguistically diverse" (CaLD) refers to "groups and individuals who differ according to religion, race, language and ethnicity, except those whose ancestry is Anglo Saxon, Anglo Celtic, Aboriginal or Torres Strait Islander".*

The Western Australian Department of Health is seeking to review data collection practices to inform and improve health service provision for CaLD communities.

Why we are consulting

WA has a high level of cultural diversity, with 15% of people born in non-main English speaking countries at the time of the 2011 census. CaLD populations can have difficulty in accessing and using healthcare.

Relevant information about CaLD persons can help to ensure we provide culturally appropriate services and address health inequalities.

There is a limited amount known about current CaLD data collection practices employed across WA Health* services and other relevant organisations, and whether there is capacity to coordinate, consolidate or improve data collection practices.

**WA Health refers to Western Australia's public health system. It consists of the Department of Health, five Health Service Providers (Child and Adolescent Health Service, North Metropolitan Health Service, South Metropolitan Health Service, East Metropolitan Health Service and WA Country Health Service) and Health Support Services.*

Completing the consultation

Please complete the consultation on behalf of your organisation or program (eg. a hospital

department or community program). You may wish to seek input from colleagues such as clinical staff and those involved with data, otherwise select 'unsure' if required.

The consultation consists of up to 21 questions, and will take between 10-15 minutes to complete, depending on responses. Responses will be de-identified to inform a final report of recommendations.

The questionnaire may be completed over more than one session by clicking the option at the bottom of each page 'Save and come back later'.

The survey will close at 11.59pm on **Friday 12 May 2017** (extended).

Thank you for participating in the consultation. Your feedback will be used to assist planning of CaLD data collection across WA health services.

Please click on the Online Survey Link below to access the consultation.

Current CaLD data collection

We are interested to know about current data collection practices within your health organisation or program that capture information on cultural and linguistic diversity.

1 Which WA Health organisation or program* are you representing in completing this survey?

*This might for example be a particular hospital emergency department or inpatient service, or an outpatient or community program eg. King Edward Memorial Hospital Visiting Midwife Service.

'Organisation or program' will be called 'program' from now on.

Program (Required)

2 Please name the main data collection tool^ used by your program.

^Data collection tools are used to record patient information and might include specific computer software programs or packages, databases, patient surveys, or initial clinic or hospital visit forms not otherwise entered into a database.

Data collection tool (Required)

3 Keeping that data collection tool in mind, select if EACH of the CaLD-related data items# below is collected on an optional or a mandatory basis, or not collected at all. If you don't know, select 'Unsure if collected'.

A comment box is provided if you wish to expand on your answers.

#"CaLD-related data items" refers to data items or parameters that measure cultural or language background (eg. country of birth, language spoken etc.).

(Required)	Collection optional	Collection mandatory	Not collected	Unsure if collected
Country of birth *Please select only one item*	○	○	○	○
Country of birth of mother *Please select only one item*	○	○	○	○
Country of birth of father *Please select only one item*	○	○	○	○
Main language other than English spoken at home *Please select only one item*	○	○	○	○
Main language spoken at home *Please select only one item*	○	○	○	○
Languages spoken at home *Please select only one item*	○	○	○	○
Languages/languages spoken *Please select only one item*	○	○	○	○
Preferred language spoken *Please select only one item*	○	○	○	○
Language *Please select only one item*	○	○	○	○
Proficiency in spoken English *Please select only one item*	○	○	○	○

Need for an interpreter/ interpreter required *Please select only one item*	○	○	○	○
Interpreter used *Please select only one item*	○	○	○	○
Ancestry *Please select only one item*	○	○	○	○
Religious affiliation *Please select only one item*	○	○	○	○
Year of arrival/ time since arrival in Australia *Please select only one item*	○	○	○	○
Migrant visa category or class *Please select only one item*	○	○	○	○
Race or ethnicity *Please select only one item*	○	○	○	○

Other CaLD-related data item collected - specify

Comments

4 Are there any other data collection tools used by your program?

(Required)

Please select only one item

◯ Yes ◯ No

You will be able to enter up to three further data collection tools before being directed to Question 13.

Current CaLD data collection 2

5 Please name a further data collection tool used by your program.

Data collection tool (Required)

6 Keeping that data collection tool in mind, select if each of the data items below is collected on an optional or a mandatory basis, or not collected at all.

A comment box is provided if you wish to expand on any of your answers.

(Required)	Collection optional	Collection mandatory	Not collected	Unsure if collected
Country of birth *Please select only one item*	◯	◯	◯	◯
Country of birth of mother *Please select only one item*	◯	◯	◯	◯
Country of birth of father *Please select only one item*	◯	◯	◯	◯
Main language other than English spoken at home *Please select only one item*	◯	◯	◯	◯
Main language spoken at home *Please select only one item*	◯	◯	◯	◯

Language/languages spoken *Please select only one item*	○	○	○	○
Preferred language spoken *Please select only one item*	○	○	○	○
Language *Please select only one item*	○	○	○	○
Need for an interpreter/interpreter required *Please select only one item*	○	○	○	○
Interpreter used *Please select only one item*	○	○	○	○
Ancestry *Please select only one item*	○	○	○	○
Religious affiliation *Please select only one item*	○	○	○	○
Year of arrival/ time since arrival in Australia *Please select only one item*	○	○	○	○
Migrant visa category or class *Please select only one item*	○	○	○	○
Race or ethnicity *Please select only one item*	○	○	○	○

Other CaLD-related data item collected - specify

Comments

7 Are there any other data collection tools used by your program?

(Required)

Please select only one item

◯ Yes ◯ No

Current CaLD data collection 3

8 Please name a further data collection tool used by your program.

Data collection tool (Required)

9 Keeping that data collection tool in mind, select if each of the data items below is collected on an optional or a mandatory basis, or not collected at all.

A comment box is provided if you wish to expand on any of your answers.

(Required)	Collection optional	Collection mandatory	Not collected	Unsure if collected
Country of birth *Please select only one item*	◯	◯	◯	◯
Country of birth of mother *Please select only one item*	◯	◯	◯	◯

Country of birth of father *Please select only one item*	○	○	○	○
Main language other than English spoken at home *Please select only one item*	○	○	○	○
Main language spoken at home *Please select only one item*	○	○	○	○
Language/languages spoken *Please select only one item*	○	○	○	○
Preferred language spoken *Please select only one item*	○	○	○	○
Language *Please select only one item*	○	○	○	○
Proficiency in spoken English *Please select only one item*	○	○	○	○
Need for an interpreter/ interpreter required *Please select only one item*	○	○	○	○
Interpreter used *Please select only one item*	○	○	○	○
Ancestry *Please select only one item*	○	○	○	○
Religious affiliation *Please select only one item*	○	○	○	○
Year of arrival/ time since arrival in Australia *Please select only one item*	○	○	○	○
Migrant visa category or class *Please select only one item*	○	○	○	○

Race or ethnicity
Please select only one item

Other CaLD-related data item collected - specify

Comments

10 Are there any other data collection tools used by your program?

(Required)

Please select only one item

Yes No

Current CaLD data collection 4

11 Please name a further data collection tool used by your program.

Data collection tool (Required)

12 Keeping that data collection tool in mind, select if each of the data items below is collected on an optional or a mandatory basis, or not collected at all.

A comment box is provided if you wish to expand on any of your answers.

(Required)	Collection optional	Collection mandatory	Not collected	Unsure if collected
Country of birth	☐	☐	☐	☐
Country of birth of mother	☐	☐	☐	☐
Country of birth of father	☐	☐	☐	☐
Main language other than English spoken at home	☐	☐	☐	☐
Main language spoken at home	☐	☐	☐	☐
Languages/languages spoken	☐	☐	☐	☐
Preferred language spoken	☐	☐	☐	☐
Language	☐	☐	☐	☐
Proficiency in spoken English	☐	☐	☐	☐
Need for an interpreter/interpreter required Language	☐	☐	☐	☐

Interpreter used	☐	☐	☐	☐
Ancestry	☐	☐	☐	☐
Religious affiliation	☐	☐	☐	☐
Year of arrival/ time since arrival in Australia	☐	☐	☐	☐
Migrant visa category or class	☐	☐	☐	☐
Race or ethnicity	☐	☐	☐	☐

Other CaLD-related data item collected - specify

Comments

Current CaLD data use

13 Thinking about the last five years, how has your program used the CaLD-related data that it collects?

"CaLD-related data" refers to data items that measure cultural and language background (eg. country of birth, language spoken etc.)

Select all those that apply. A comment box is provided to expand your answers.

(Required)

Please select all that apply

- [ ] Service reporting
- [ ] Research
- [ ] Sent to inform a larger dataset
- [ ] Shared data with another service
- [ ] Targeting service delivery or interventions
- [ ] To better provide a culturally-appropriate service
- [ ] To encourage use of interpreter services
- [ ] Policy planning
- [ ] Program monitoring or evaluation
- [ ] To inform individual clinical care
- [ ] Not used
- [ ] Not sure
- [ ] Other

If other, please specify

Comments

Barriers to CaLD data collection

14 Thinking about the CaLD-related data collected by your program, are there any problems with incorrect or incomplete data?

Incorrect data problems refers to the extent to which final information records do not reflect true answers to intended questions. Incomplete data refers to the degree to which missing information is a problem.

(Required)

Please select only one item

- ◯ Yes, incorrect data is a problem
- ◯ Yes, incomplete data is a problem
- ◯ Yes, both incorrect and incomplete data are problems
- ◯ No
- ◯ I'm not sure

If yes, please briefly outline the problem(s), including any reasons you think might explain the incomplete and/or incorrect data.

Barriers to CaLD data collection 2

15 In the last 5 years, has your program undertaken a quality assurance process to review the correctness or completeness of CaLD-related data collection?

(Required)

Please select only one item

◯ Yes ◯ No ◯ I'm not sure

If yes, please provide a brief description of the quality assurance process (2-3 sentences).

Future CaLD data collection

Best practice data collection can help to provide culturally appropriate services and address health inequalities. We are interested in finding out your thoughts on improving data collection for your program.

16 Are there any CaLD-related data items that would be useful to add to your program's current data collection tools? eg. country of birth, proficiency in spoken English, ancestry etc.

(Required)

Please select only one item

◯ Yes ◯ No or not particularly ◯ I'm not sure

If yes above, select which data item(s) from the list, then explain why below.

Please select all that apply

☐ Country of birth ☐ Country of birth of mother ☐ Country of birth of father
☐ Main language other than English spoken at home
☐ Main language spoken at home ☐ Languages spoken at home
☐ Languages/languages spoken ☐ Preferred language spoken/language
☐ Proficiency in spoken English ☐ Need for an interpreter/interpreter required
☐ Interpreter used ☐ Ancestry ☐ Religious affiliation
☐ Year of arrival/time since arrival in Australia ☐ Migrant visa category or class
☐ Race or ethnicity ☐ Other

Other- please specify

Explanation

17 How could an improved CaLD-related data collection most assist your program?

Rank your top three. There is an opportunity for you to provide further details if you wish.

	1	2	3
Service reporting *Please select only one item*	○	○	○
Research *Please select only one item*	○	○	○
Targeting service delivery or interventions *Please select only one item*	○	○	○
To better provide a culturally-appropriate service *Please select only one item*	○	○	○
To encourage use of interpreter services *Please select only one item*	○	○	○
Policy planning *Please select only one item*	○	○	○
Program monitoring or evaluation *Please select only one item*	○	○	○
To inform direct individual clinical care *Please select only one item*	○	○	○
To inform a larger or shared data set with other health services *Please select only one item*	○	○	○
Other *Please select only one*	○	○	○

Other- please specify

☐ Not sure

☐ Not applicable

Details

Future CaLD data collection 2

18 If there was an opportunity to improve CaLD-related data collection practices in your program, what would be your main priorities?

Rank your top three.

	1	2	3
Reviewing and optimising the data items (or variables) collected eg. need for interpreter etc. *Please select only one item*	○	○	○
Staff training or resources *Please select only one item*	○	○	○

Improving the data software *Please select only one item*	○	○	○
Increasing your program's overall focus on CaLD-related data collection *Please select only one item*	○	○	○
Ensuring completeness of data entry for existing data items and/or making data item entries mandatory *Please select only one item*	○	○	○
Optimising data correctness by improving question wording or answer entry guidelines *Please select only one item*	○	○	○

If other- please specify

☐ Not sure

☐ They are already optimal

Improving the data software *Please select only one item*	○	○	○
Increasing your program's overall focus on CaLD-related data collection *Please select only one item*	○	○	○
Ensuring completeness of data entry for existing data items and/or making data item entries mandatory *Please select only one item*	○	○	○
Optimising data correctness by improving question wording or answer entry guidelines *Please select only one item*	○	○	○

If other- please specify

☐ Not sure

☐ They are already optimal

Future CaLD data collection 3

19 Please provide any other comments you would like to make about CaLD-related data collection within your program.

Comments

CaLD data collection across health services in WA

We are interested to know your thoughts on CaLD-related data collected by other health organisations, services or databases in WA.

20 Please provide any suggestions you have for improvement of CaLD-related data collection by other WA health services either internal or external to WA Health.

This could for example relate to the data items collected or opportunities to collect the same data items or use the same data collection tools across services to allow grouping or comparison.

Suggestions

Final details

21 What is your name and current position title?

This information will assist the project team in contacting you should any questions arise in relation to responses, however completion is optional. Personal details entered will not be used to attribute individual responses to any one person in any reports that are compiled.

Name

Position title

Contact phone number(s)

Appendix 2

Strengths and limitations of variables measuring cultural and linguistic diversity

Variable Source		Variable	Strengths	Limitations
AIHW National Data Dictionary	ABS Standard Set (Minimum Core Set)	**Country of Birth of Person**	One of the most common demographic items used in Australia (Donato-Hunt & Grima, 2009), enabling comparison with local and overseas data. Objective, stable and comparable. Can collate to geopolitical region, 'born overseas' or non-main English speaking countries to increase sample size. Identifies people who have moved to Australia, and indicates what their cultural background might be. Ideally should be considered in combination with other indicators such as Main Language Spoken at Home, Ancestry or self-identified cultural background which can enable identification of subgroups within a migrant population (ABS, 1999; Donato-Hunt & Grima, 2009).	Does not clearly identify a connection to another culture. People may be born in one country but live most of their life in another. Porter, Todd and Zhang (2016) found poor correlation between Country of Birth of Person and self-identified ethnic group in a sample of pregnant Australian women. May not reflect ethnic identity based on religion or neglect subgroups within countries, across borders or outside of their country of origin (Porter et al., 2016). Makes no distinction between migrants who may have arrived long ago or more recently. Does not identify categories of migrants such as refugees, or culture in descendants of migrants (Rechel et al., 2012). Collating to region or migrant may reduce discriminatory capacity. Country borders change overtime, making coding not directly comparable.
		Indigenous Status	One of the ABS's Minimum Core Set for collecting data on cultural diversity (not further considered here as not relevant to the definition of CaLD used in this report).	
		Main Language Other Than English Spoken at Home	Provides information on the number of people who speak English only at home and the main non-English language spoken at home. Captures a language other than English where the main language spoken is English but a language other than English is still spoken in the home. This maximises numbers for more established migrant communities (ABS, 1999). Consistent with the ABS Census, allowing for comparability of language data.	May not reflect complete language use e.g. Will miss languages still spoken outside the home or miss languages used but not spoken in the home. May capture those who have learnt another language and occasionally use it at home or those forced to use English at home. May be confusing and therefore poorly implemented by services and get interpreted as Main Language Spoken at Home (Donato-Hunt & Grima, 2009)

Variable Source		Variable	Strengths	Limitations
AIHW National Data Dictionary	ABS Standard Set (Minimum Core Set)	**Main Language Other Than English Spoken at Home (cont.)**	Chosen for its strength as a measure of language use to be included in the Minimum Core Set. Extensive consultation with users of language data revealed it is the "most useful general purpose language variable" (ABS, 1999). When used in combination with Proficiency in Spoken English, regarded as the best measure for identifying service needs and the potentially disadvantaged (ABS, 1999) Gives an indication of frequency of language use and retainment of culture and provides useful information to assist in the provision of culturally appropriate services.	Alone may not measure disadvantage related to English proficiency.
		Proficiency in Spoken English	Designed to be a filter from Main Language Other Than English Spoken at Home for non-native English speakers (ABS 1999). Key indicator in terms of disadvantage in accessing services. Analysis of 1996 Census showed clear association between this variable and socioeconomic disadvantage (ABS, 1999). Subjective, however ABS testing prior to the 1986 Census showed that is a good identifier of people who are likely to need assistance in the form of interpreter services or English language classes. Those people likely to need assistance will generally respond 'Not well' or 'Not at all' (ABS, 1999). May assist to determine unmet need for interpreting service for those that speak English 'not well' or 'not at all'. Donato-Hunt and Grima (2009) found a lower rate of clients recognised as 'needing an interpreter' compared to those identified as speaking English 'not well' or 'not at all'.	English proficiency is an important indicator to collect however this information could be collected by Interpreter Service Required, which may be more aligned to service delivery (Donato-Hunt & Grima, 2009) Does not assess proficiency in reading or writing English May be difficult to ask in some health settings with the standardised verbal question being, "Do you consider [you speak] [(name) speaks] English very well, well, or not well?
		Preferred Language spoken	Targeted towards more immediate service provision compared to Main Language Spoken at Home (Donato-Hunt & Grima, 2009). Better for service provision of interpreters compared to Main Language Spoken at Home for patients having a different Preferred Language to that spoken at home.	Is not included by the ABS in the Standard Set of Cultural and Language Indicators. Not collected by any national health surveys or censuses (Blignault & Haghshenas, 2005) May not be asked of a proficient English speaker although they may prefer a language other than English.

Variable Source		Variable	Strengths	Limitations
AIHW National Data Dictionary		**Preferred Language spoken (cont.)**	Currently collected by the Hospital Morbidity Data System as filter question from Interpreter Service Required. May be easier for staff to collect than Main Language Other Than English Spoken at Home.	In terms of service delivery involving family engagement, Main Language Spoken at Home may be more useful (Donato-Hunt & Grima, 2009). May include people who can speak fluent English.
		Interpreter service required	Important variable to collect in service delivery settings, informs booking and facilitation of interpreter services. Can assist funding allocation and staff training in interpreter use. Easy to categorise as binary 'yes' or 'no' response. May be comparable with Proficiency in Spoken English based on response options: Yes=not at all or not well, No= well or very well (ABS, 1999), however this needs further assessment (Donato-Hunt & Grima, 2009). May be more practical in some healthcare settings for sufficiently trained staff to determine the answer than for the alternative variable Proficiency in Spoken English. May be more accommodating of patient preference than interpreter provision based on Proficiency in English.	Does not determine specific language groups, needs an additional question on language group. If lower rates than Preferred Language spoken or those with poor English proficiency can suggest unmet interpreter need (Donato-Hunt & Grima, 2009). Literature provides minimal discussion on data collection process e.g. whether self-determined or otherwise. May not identify true need for interpreter if client declines based on misunderstanding on role of interpreter or concerns regarding confidentiality or competence. Is not included by the ABS in the Standard Set of Cultural and Language Indicators.
	ABS Standard Set (Additional Optional Variables)	**Ancestry**	The Ancestry variable provides a self-assessed measure of ethnicity and cultural background by identifying a person's origins and heritage. Can be used in combination with other variables as a measure of the extent to which people retain the ethnicity and culture of their forebears or determining first and second generation migrants' ethnic backgrounds. Useful for determining ethnic background beyond first and second generation migrants (ABS verbal consultation). The major advantage of the Ancestry variable is that it is able to measure an association with an ethnic origin and cultural groups which cannot be measured by questions relating to Country of Birth of Person or language variables (ABS, 1999).	Ancestry in the Australian context is problematic as there are many Australians with origins and heritage that do not, in practice, relate to their current ethnic identity (ABS, 1999). The question can be perceived as relating to measures of self-identification of ethnic or cultural group affiliation or nationality as well as of descent from one or more particular groups. Ancestry data alone, therefore, is not considered a good measure of service needs (ABS, 1999). When Ancestry data is used alone, it should only be done to represent a broad measure of cultural diversity. May be complicated to classify if multiple ancestral heritages.

Variable Source	Variable	Strengths	Limitations
ABS Standard Set (Additional Optional Variables)	**Ancestry (cont.)**	Important in identifying particular cultural groups such as Pacific Islanders born in New Zealand and those spread across country borders e.g. Kurds from Turkey or Iraq (ABS, 2017b). Extensive work conducted by ABS found Ancestry likely to be more valid and reliable than self-perceived current 'ethnicity' (Borrie et al., 1984).	
	Country of Birth of Mother/Country of Birth of Father	Indicates a person's migrant background. Provides an indicator of possible cultural connection to parent's country of birth. Useful in collections relating to perinatal or child outcomes. Migrants and their descendants have been identified by program providers as forming population groups likely to experience relative disadvantage or advantage (ABS, 1999). Can be used in association with other variables to determine extent to which second generation Australians retain their parents' culture, ethnicity or language.	No certain link between parent's country of birth and child's culture or language. Some parents are born in long-term refugee camps outside a family's country of origin.
	First Language Spoken	Surrogate measure of ethnicity (ABS, 1999). Can be a filter question to Proficiency in English (ABS, 1999). Good indicator of likely current language use (ABS, 1999).	May be more than one first language spoken, however of little significance in Australian collections (ABS, 1999). May not be the language of greatest competence (ABS, 1999).
	Languages Spoken at Home	Provides data on the overall stock of languages actively spoken in Australian homes (ABS, 1999). Useful in determining active ethnicity and intergenerational language retention (ABS, 1999).	Does not reflect languages spoken outside the home or written language use. Does not put restrictions on the number of entries or determine the frequency of language use, potentially capturing languages used infrequently or spoken with limited proficiency. Does not provide a reliable filter to Proficiency in English.

Variable Source		Variable	Strengths	Limitations
	ABS Standard Set (Additional Optional Variables)	**Main Language Spoken at Home**	Provides a good indicator of the language with which a person will be most at ease (Office of Multicultural Affairs, 2017) Indicator of active ethnicity. May be useful over other language variables when aim is to measure active ethnicity or likely disadvantage experienced by persons whose usual language is not English.	May not be relevant for second or third generation Australians who identify with a particular cultural group but do not speak the language. Tends to underestimate current community language usage amongst the longer standing migrant groups who now mainly use English at home. Does not provide family language use if person speaks English most often. ABS reports Main Language Other Than English Spoken at Home is a better filter to Proficiency in Spoken English (ABS, 1999). Misses Preferred Language if a different language is spoken at home.
		Religious Affiliation	Provides additional data for identifying specific ethnic or cultural groups when used in conjunction with other cultural and language variables. Certain religions may require particular service approaches. Collects information on a factor that may influence health behaviours not captured by other CaLD-related variables.	In Australian society, religious affiliation does not clearly indicate cultural attachment (Donato-Hunt & Grima, 2009).
		Year of Arrival in Australia	Indicates how familiar a migrant is likely to be with local practices, services and the English language and how long it may have taken to overcome difficulties (ABS, 1999). May assist as proxy indicator of language proficiency or unmet need for interpreter. Clear input options. Can collect once only Can study disease trends or health policy related to regency of arrival May assist in elderly and young migrant population to determine likely primary language and cultural exposure.	Only indicative of particular issues, rather than a direct measure of service need or ethnic background.

Variable Source	Variable	Strengths	Limitations
Other Academic or Grey Literature	**All languages spoken**	May provide an overview of the linguistic diversity of clients. Useful if interpreter not available in particular language.	All inclusive and does not indicate the frequency of languages spoken, proficiency or cultural attachment.
	Migrant visa category	Enables assessment of health outcomes and servicing needs in skilled migrant, family migrant, refugee and asylum seeker populations, which is currently lacking (Hadgkiss & Renzaho, 2014) and likely to be very variable across visa types. Could be a filter question from Country of Birth of Person.	Was assessed by ABS and found it would be difficult to implement in particular administrative settings (ABS 1999). May have poor acceptability by some participants and the community in some settings.
	Race	Collected widely in the UK, and the USA under minimum standards provided by the Federal Office of Management and Budget (Office of Management and Budget, 1997). Can track conditions for populations that historically have experienced discrimination and most at risk of ill-health in order to target services (Harawa & Ford, 2009). May be useful for assisting diagnoses related to genetic diseases more common within particular ethnic groups.	Of less relevance to Australia in terms of ethnic disadvantage; Indigenous Status asked separately. Some may not 'fit' in categories offered, especially if increasing diversity. May be considered offensive and have poor acceptability by the community.
	Ethnicity	Gives an indication of self-identified cultural attachment. Self-identified ethnicity collected widely in USA and UK (over country of birth) most often in association with race. Umbrella term which can avoid multiple other proxy measures. May evolve over time, more accurately reflecting current cultural affiliations. Contextual, potentially influenced by political forces and acculturation, can be beneficial in terms of behavioural risk identifier. May be useful for assisting diagnosis related to genetic heritage (Abouzeid et al., 2014). Abouzeid et al. (2014) recommends considering self-identified ethnicity in Australian's set of health indicators and as a core component of the national census. Collected by comparable countries, increasing capacity for international comparison.	Definitional complexity and inconsistent understanding, including between researchers and participants, creates difficulty for accurate collection (Borrie et al., 1984; Drevdahl, Philips, & Taylor, 2006). Potential questions piloted by ABS prior to 1986 Census with poor data outcomes (Borrie et al., 1984). Can include identification with multiple elements e.g. identification with nationality, race, current community engagement with various ethnic groups, ancestry, extent of maintenance of traditional customs, creating difficulty with answer categories (Abouzeid et al., 2014). May evolve over time and is contextual, potentially influenced by political forces and acculturation (Abouzeid et al., 2014). May be confusing to classify if mixed ancestral heritage or intercultural unions.

Appendix 3

Survey responses – programs and data collection tools

The table represents the list of health service programs represented by staff as entered by survey respondents, and the corresponding data collection tool(s) used by each program.

Table 1. Health service programs represented as entered by survey respondents, and corresponding data collection tool(s).

Service or program	Data collection tool		
Patient Information Services	TOPAS		
EMHS Coding Unit	webPAS / TOPAS / Codefinder		
Dental Health Services	In-house developed Patient Management System named DenIM		
Armadale Health Service	TOPAS, iSOFT, AHS Stats		
WA Tuberculosis Control Program	MedTech 32		
Epidemiology and Surveillance Program, Communicable Disease Control Directorate	WANIDD and HIV Notifications Database (separate databases)	Several "enhanced surveillance" datasets for specific diseases and outbreak investigation tools specific for investigation of particular outbreaks.	
Princess Margaret Hospital for Children Refugee Health Service	Initial PMH RHS first visit assessments proformas		
Burns unit FSH	BIMS / TOPAS		
Sexual Health Service	SHIP	webPAS	
SMHS clinical coding	webPAS		
Clinical Coding FSH	BOSSnet and webPAS		
Outpatients	Patient Administration System (webPAS)	BossNET	Cardiobase
In and outpatient, elective and emergency General Surgical Service	webPas		
FSH breast service	SHaRE		
Hospital Emergency Department	BOSSnet Clinical Program, EDIS, iSoft, webPAS		
Fiona Stanley Hospital Burns Unit	webPAS and BIMS		

Service or program	Data collection tool		
Social Work	webPAS		
Dental Health Services	Patient Surveys, Reports from Information system		
Burns	Burn Injury Management System - digital records + systematic outcome measurement battery AND BOSSnet	BOSSnet	
FSH SRS	FIM, Admission info		
Humanitarian Entrant Health Service	Medtech 32		
Complex Care Co-ordination Team (CoNeCT)	Excel spreadsheet patient data base		
RITH	RITH database / webPAS / TOPAS		
FSH Allied Health Medical Specialties	BOSSnet		
Home Link	Home Link referral form		
Midwives Notification System	MNS	Abortion Notification System	
RGH Medical Assessment Unit	Inpatient Journey doc (RGMR121)	TOPAS	
Fiona Stanley Hospital	webPAS		
SMHS CSP&PH	Community Physiotherapy Services Database	webPAS	
BreastScreen WA	Clinic client detail registration form		
Mental Health Commission	Qualtrics		
SMHS CSP and PH RITH	RITH interdisciplinary initial screen form		
Alcohol and Drug Support Service (part of the Mental Health Commission)	Alcohol and Drug Support Service (ADSS) Database	ADSS Call Back Referral Form	
Genetic Services of Western Australia	Kintrak and Progeny patient information management software	TOPAS and webPAS	
Western Australian Register of Developmental Anomalies	Western Australian Register of Developmental Anomalies Database		

Service or program	Data collection tool		
Community Alcohol and Drug Service	SIMS	Evaluation	
King Edward Memorial Hospital / Language Services	LaSS program		
Royal Perth Hospital Sexual Health Service	TOPAS; SHIP		
Royal Perth Hospital Language Services	TOPAS and LaSS (Language Service System)		
Language Services – Royal Perth Hospital	TOPAS (The Open Patient Administration System) and LaSS (Language Services System)		
Prevention and Control Program, Communicable Disease Control Directorate	WA Vaccine Safety Surveillance		
Sir Charles Gairdner Hospital	Language Services System		
Osborne Park Hospital	TOPAS		
SCGH Allied Health	Initial referral form		
SCGH Physio	Physio specific Ax and Rx forms		
Prevention and Control Program, Communicable Disease Control Directorate	School Based Immunisation Program Database		
OPH	TOPAS		
Fiona Stanley Language Services	LaSS and webPAS		
MHC-Next Step	SIMS		
WA Cancer and Palliative Care Network	SHaRE, DISC	webPAS / TOPAS	
Bentley Health Service Community Rehabilitation program	Community Rehabilitation Database		
Population Health	webPAS / Written Records		
WA Newborn Hearing Screening Program	Newborn hearing screening forms for each baby	Stork birth list	Raw newborn hearing screening data
North Metropolitan Health Services; Youth Mental Health Services	PSOLIS		

Service or program	Data collection tool		
Non-admitted Data Collections Unit	NAPAAWL		
Patient Liaison Service Sir Charles Gairdner Hospital	Datix CFM	Datix CFM	
SCGH Chaplaincy and Spiritual Services	AHS (which is constructed using TOPAS information)		
Albany Health Campus admissions	webPAS		
Child and Adolescent Community Health	CDIS (Child Development Information System) – Used by child health and primary school services		
Child and Adolescent Community Health	HCARe non-registered clients (used by high school nurses, enuresis nurses, refugee health nurses and education support nurses to record occasions of service)	HCARe Registered Client – Client Master Index (used by the Refugee Health Team)	
Child and Adolescent Community Health Immunisation Services and Central Immunisation Clinic	WinVacc	Vaccination record card – hard copy	
Child and Adolescent Mental Health Service	Patient Administration System (PAS)	PSOLIS	
WA Cervical Cancer Prevention Program (Women and Newborn Health Service)	WA Cervical Screening Register (Oracle database)		
SCGH	LaSS	TOPAS	
ACAP	My Aged Care		
Warwick CAMHS Cross Cultural Clinician	PSOLIS	Registration form, medical records and standard documentation	
Mental Health Data Collection, Purchasing and System Performance	Mental Health Data Collection (Source System: Psychiatric Online Information System, PSOLIS)		
Emergency Department Data Collection	Oracle		

Appendix 4

Supporting evidence for additional optional CaLD variable recommendations

Supporting evidence for additional optional CaLD variables beyond a minimum core set for collection by WA health services is described below.

Year of Arrival in Australia is recommended because it would be an objective, stable and comparable measure, could be collected just once, and would give an indicator of familiarity with Australian society and institutions, and the English language. It could be used to assess temporal health effects such as identifying groups taking longer to access appropriate health services or those making disproportionate use of particular services after arrival or on an ongoing basis. It may allow assessment of diseases or behaviours associated with recency of changes in physical or cultural environments, such as asthma, particular malignancies or dietary behaviours. It would also allow assessment of likely primary language and childhood or adult societal experiences for elderly patients reverting to their primary language. It is already collected by PSOLIS software and included in the MHIS, and is included in the ABS Standards (ABS, 1999).

Country of birth of mother and/or father and year of arrival in Australia of mother and/ or father would be particularly useful in the paediatric, perinatal and maternal health setting, and could be considered as an alternative for 'Ethnicity' in the Midwives Notification System. It is stable and comparable, and Country of Birth of Mother and Country of Birth of Father is included in the ABS Standards. It would be an important addition as poor CaLD health outcomes are recognised particularly in the maternal and perinatal health settings. It would provide an indication of families' likely familiarity with Australian society and institutions and the English language.

Ancestry is recommended for its capacity to identify diversity beyond Country of Birth (of Person or parents), language variables and religion. It would be useful for identifying ethnic groups occurring within countries or across borders e.g. Pacific Islanders from New Zealand or Kurds from Turkey or Iraq. It is included in the ABS Standard Set. It could assist with diagnosis and research relating to genetic diseases. It is however likely to be less useful for detecting current disadvantage or service needs for some groups whose ancestral origins do not in practice relate to their current ethnic identity. Current self-perceived 'ethnicity' is not recommended due to complexity of definition, poor response to pilot questions and the lack of a standard developed by the ABS (Borrie et al., 1984).

Religious Affiliation is recommended because it is already collected by the major primary collection packages (WebPAS, TOPAS and Psolis) and reportedly will assist to identify most cultural groups in combination with Country of Birth of Person and a language variable (ABS, 1999). It may also assist to identify current service needs or preferences. The value of this variable should ideally however be further assessed via consultation with service providers.

Proficiency in Spoken English is recommended for consideration where resources allow as a potential comparison measure to Interpreter Service Required. The ABS includes this measure in the ABS Minimum Core Set allowing for population level rate assessment and comparisons, and supposedly answers 'not well' or 'not at all' align well with the need for an interpreter. Conversely, speaking English 'well' or 'very well' should correlate with absence of need for an interpreter. In this sense, it could detect unmet need for interpreters or poor training of staff in recognising the need for interpreters.

This document can be made available in alternative formats on request for a person with disability.

Produced by Epidemiology

health.wa.gov.au

EPI-013424 FEB'18