Heterogeneity/Granularity in Ethnicity Classifications outside the United States (HGEC project)

A report to the Robert Wood Johnson Foundation
February 2016
By:

Nazmy Villarroel
University of Limerick

Emma Davidson
University of Edinburgh

Pamela Pereyra-Zamora
University of Alicante

Allan Krasnik
University of Copenhagen

Dr. Donna Cormack
Investigator, New Zealand

Dr. Tahu Kukutai
Investigator, New Zealand

Dr. Kelsey Lucyk
Investigator, Canada

Dr. Karen Tang
Investigator, Canada

Professor Hude Quan
Investigator, Canada

Dr. Liv Stubbe Østergaard
Investigator, Denmark

Dr. Peter Aspinall
Investigator, United Kingdom

Dr. Inez Zsófia Koller
Investigator, Hungary

Dr. Shyamala Nagaraj
Investigator, Malaysia

Dr. Chiu Wan Ng
Investigator, Malaysia

Raj Bhopal
University of Edinburgh

This is one in a series of six research reviews supported by grants from The Robert Wood Johnson Foundation to the authors' universities or organizations. The opinions are those of the authors, not the Foundation or PolicyLink.

Heterogeneity/Granularity in Ethnicity Classifications outside the United States (HGEC project)

Professor Raj Bhopal, CBE, MD, DSc (hon), MPH, University of Edinburgh

Emma Davidson, MBChB, MPH, FNZCPHM, University of Edinburgh

Nazmy Villarroel, MPH, PhD, University of Edinburgh

Pamela Pereyra-Zamora, BSc(Stats), MSc, PhD, University of Alicante

Professor Allan Krasnik, MD, MPH, PhD, University of Copenhagen

Edinburgh Migration, Ethnicity and Health Research Group (EMEHRG), University of Edinburgh

Report to: The Robert Wood Johnson Foundation (RWJF)

THE UNIVERSITY *of* EDINBURGH

Usher Institute
Population Health Sciences & Informatics

Structure of this report

Executive Summary

Chapter 1: Background to the Robert Wood Johnson Foundation (RWJF) project and some conceptual background.

Authors: Nazmy Villarroel, Emma Davidson, Pamela Pereyra-Zamora, Allan Krasnik and Raj Bhopal.

Chapter 2: Overview of ethnicity data sources throughout the European Union (EU).

Authors: Nazmy Villarroel, Emma Davidson, Pamela Pereyra-Zamora, Allan Krasnik and Raj Bhopal.

Chapter 3: Ethnic group classification in Aotearoa New Zealand.

Authors: Donna Cormack and Tahu Kukutai.

Chapter 4: Ethnic group classification in Pluri-National State of Bolivia.

Author: Pamela Pereyra-Zamora.

Chapter 5: Ethnic group classification in Canada.

Authors: Kelsey Lucyk, Karen Tang and Hude Quan.

Chapter 6: Country of birth classification in Denmark.

Authors: Liv Stubbe Østergaard and Allan Krasnik

Chapter 7: Ethnic group classification in Great Britain

Author: Peter Aspinall

Chapter 8: Ethnic group classification in Hungary.

Author: Inez Zsófia Koller.

Chapter 9: Ethnic group classification in Malaysia.

Authors: Shyamala Nagaraj and Chiu Wan Ng.

Chapter 10: Overall discussion of findings and recommendations

Executive summary

Background

The changing nature of global migration and increasing diversity of populations have transformed the social landscape of many countries. Such complex social formations have challenged not only public health but also other private/public agendas (e.g. cultural tailoring, diversity in the workforce). Demographic data that capture population heterogeneity (e.g. by ethnicity) are required to understand how collective identities are produced; to identify the health needs of diverse groups; to detect and address inequities in healthcare provision and outcomes. However, little is known about current methods of ethnic classification internationally and, in countries where ethnicity data is collected, about what level of granularity is employed in their ethnicity categorization. Thus, this project aimed to explore and provide an overview of how EU-28 countries and four countries outside Europe approach the collection of granular ethnic classifications.

Methods

For the overview of EU-28[i] countries, data were obtained primarily from official population censuses or registers. For each country these data sources were examined for their approach to ethnicity. When ethnic information was not gathered, country of birth (CoB) and/or parents' CoB, language spoken, religion and national identity were examined as a proxy for ethnicity. The granularity of approaches to ethnic classification were assessed using the OMB Standards for the Classification of Federal Data on Race and Ethnicity, with those countries collecting more than six ethnic categories being considered granular.

Seven in-depth country report were also undertaken, in collaboration with international experts in the field, in countries identified as potentially having valuable lessons in their approaches to ethnic classification. This included three EU countries (Great Britain[ii], Hungary and Denmark) and four countries outside Europe (Aotearoa New Zealand, Bolivia, Canada and Malaysia). A convening of these expert was subsequently held to discuss the findings and distil overarching principles.

[i] EU-28 includes the countries within the United Kingdom (England, Wales, Scotland and Northern Ireland) Therefore, a total of 31 European countries were included in this project. (EU-27 plus four countries)

[ii] For the purpose of the project, we consider UK as part of the EU. For the UK to leave the EU it has to invoke an agreement called Article 50 of the Lisbon Treaty.

Results

Overview of EU-28 countries

For the overview of EU-28 countries, granular approaches to ethnicity data collection were found in eight countries: the United Kingdom (England, Wales, Northern Ireland, and Scotland), Republic of Ireland, Hungary, Poland and Slovakia, which collected more than six categories. We found that Estonia, Lithuania, Croatia, Bulgaria, Republic of Cyprus and Slovenia paid some attention to granularity, collecting one to six categories. Information on ethnicity with only a free text option was found in Latvia, Romania and Czech Republic. The inclusion of a free text option may provide the most granular approach for collection of ethnicity data, but only if granularity is subsequently retained in the analyses and reporting of data. There were also 14 countries who collected proxy variables instead of ethnicity, for example also CoB (individual and parents), nationality, religion and language (mother tongue).

Within the EU, we found that ethnicity is conceptualised in different ways and diverse terminology is employed for census/population register questions and the categorization of responses. For the eight countries with the most granular approach, there is also variation in the focus of disaggregating categories. For example, in Scotland the categories are based on a mixture of colour, nationality and ethnic origins and there is an emphasis on exploring heterogeneity within the 'White' and 'Asian' categories. In Poland, categories also include, and are disaggregated according to language and religion.

Overall, the extent to which ethnicity data are collected within the EU and the approaches to classification appear to be strongly influenced by political rights and legislation; historical events; ideology and sensitivity towards cultural identity; and ongoing migration patterns.

In-depth country reports

Aotearoa New Zealand

Authors: Donna Cormack and Tahu Kukutai

Aotearoa New Zealand has long-standing and embedded practices of ethnic enumeration, although these have shifted over time with changes to broader political and social contexts. Early approaches following colonisation reflected assimilationist policies towards Māori, New Zealand's indigenous peoples, and state interest in delineating access to resources and rights. Early censuses asked about country of birth, introducing a question about 'race' in 1916. Official approaches to ethnicity shifted over time to a 'degrees of blood' conceptualisation, then to self-identified ethnic affiliation in the 1990s.

In New Zealand, Statistics New Zealand is the agency responsible for the official standard for ethnicity that outlines the official definition, standard ethnicity question, and classification system. The current standard was released in 2005 and applies to all-of-government. Administrative and survey collections routinely collect ethnicity data, including the population census, vital registrations, official surveys, and many administrative collections in education, justice, health and other sectors. It has been compulsory to collect ethnicity data in the health sector since the 1990s. Although issues with quality have been documented, ethnicity data is included in key health sector collections and are routinely used for monitoring, planning, and funding purposes.

The official approach to ethnicity data in New Zealand supports granularity in that the standard question allows people to self-identify with multiple groups and to provide free text responses. The official classification system has four levels, from least to most detailed, with more than 230 ethnic categories at the most detailed level. In practice, however, granularity is often restricted in approaches to data collection, recording and output. Many systems do not collect or record ethnicity data at the most detailed level or do not capture all ethnicities reported by an individual. Data are often aggregated for analysis and reporting, with official data routinely reported for broad ethnic groupings (e.g. European, Pacific, Asian).

The health sector has comprehensive coverage across administrative and survey collections, that is critical for the measurement and monitoring of ethnicity and ethnic health inequities. However, disaggregation at levels of detail greater than broad ethnic groups remains relatively uncommon. This limits understanding of communities' diverse realities and priorities, and potentially also masks health need and ethnic health inequities. Changes to ethnicity data systems in the future, including proposed changes within the health sector to increased disaggregation in recording ethnicity data, may contribute to improved granularity of over time.

Bolivia

Author: Pamela Pereyra-Zamora

Current ethnic classification in Bolivia began with the Spanish conquest of the Inca Empire. Although in less quantities than in other Latin American countries, Europeans, Arabs and Asians established in Bolivia during the colonial times and the 19th and 20th centuries. Therefore, the population composition of Bolivia is formed by indigenous people, former slaves, colonial settlers and more recent immigrants.

Since independence from Spain in 1825 eleven censuses have been completed and since the 1950 a question on language has been included as a proxy to ethnic identity, although it has evolved across time. In the 2001 census for the first time an ethnic self-identification question was introduced (Quechua, Aymara, Guaraní, Chiquitano, Mojeño and 'other native'). The result was that 39 indigenous peoples were later recognised in the 2009 constitution. The most significant feature of the 2012 census

is the great increase in the granularity of formerly marginal peoples, indigenous and Afro-Bolivian. The question had a free text option and the outcome was over a hundred ethnic self-identifications.

According to the 2012 census there were 10,059,856 inhabitants, and 41.7% has declared to belong to any of the peasant, aboriginal, indigenous or Afro-Bolivian people or nation. Of this percentage, 43.7% are Quechua, 38.1% are Aymara, 3.5% are Chiquitano and 2.3% Guaraní. In the 2012 census, the percentage of people belonging to an indigenous people decreased compared to 2001. This controversial decrease could be related to technical aspects such as the question, the previous filter, and also may be due to the abandonment of indigenous identities in urban contexts.

According to the new constitution Bolivia is a Pluri-National State in which indigenous nations and peasant nations are guaranteed their rights. This entails their right to home rule, to manage their land, to apply their own justice, to use their language, and to live according to their own *cosmovision*. A related use of the census granularity, perhaps the most important, is the allocation of seats in the parliament to different indigenous peoples according to the percentages that they obtain in the census. It is therefore a period of implementation of the political consensus on ethnicity in Bolivia (recording, granularity and reporting) reflected in the 2009 constitution as well as its insertion in areas such as health.

Canada

Authors: Kelsey Lucyk, Karen Tang and Hude Quan

Historically, Canada's changing policies on multiculturalism and immigration have influenced the collection of ethnic group data to become more inclusive and granular. Important periods include early attempts at nation-building during the late 19th century, social changes post–World War II (WWII) and the introduction of multiculturalism into federal policy, and present-day efforts in ethnic classifications for research purposes and for preserving cultural diversity. As such, the collection of ethnicity and cultural data in Canada has evolved to include further granularity for common identities reported in national data sources, as well as to provide multi-cultural examples for respondents to more accurately capture their ethnic origins. There are four main sources of ethnicity data in Canada: 1) Provincial health insurance registries, 2) Canadian Health Measures Survey, 3) Canadian Community Health Survey, and 4) Census. Of these, ethnicity data are most limited in the provincial health insurance registries, flagging only Aboriginal status. The other three data sources are nationally administered, with all asking individuals to select, out of 11 categories, self-identified racial or ethnic groups. The questions on ethnic origin for the 2016 Census included citizenship, place of birth, immigration status, language, ancestry, ethnic origin, and Aboriginal status.

There are some lessons to be learned outside of the health field regarding the collection of ethnic data: ethnic or racial origins reported by individuals may not necessarily be the ethnic group with which they

identify; the length of time that individuals have been in Canada affects the strength of their identity with ethnic ancestry; and a large proportion of Canadians have multiple ethnic groups of origins, which supports the need for follow-up questions to better understand how individuals best identify. In sum, there exists the need for greater granularity in ethnic classifications to reflect the diversity of the Canadian population. Because ethnicity is a socio-cultural concept, consideration should be made to incorporate questions about sense of belonging with the identified ethnic ancestry, rather than relying solely on reported ethnic origin and race.

Denmark

Authors: Liv Stubbe and Allan Krasnik

Since 1850 Statistics Denmark has been the central authority on Danish statistics including demographic data. Statistics Denmark does not register ethnicity directly, but does identify origin based on country of birth and ancestry (parents' country of birth and citizenship). The definitions and classifications of immigrants and descendants are solely Danish definitions and the data used in these statistics derives from the Central Person Registry (CPR Registry). Statistics Denmark's definition of origin divides the population into three groups: Persons of Danish origin (country of birth does not matter, but at least one parent holds a Danish citizenship and is born in Denmark), Immigrants (born abroad, none of the parents hold Danish citizenship and are born in Denmark), Descendants (born in Denmark, none of the parents hold Danish citizen-ship and are born in Denmark). Furthermore, country of birth is divided into Western/Non-Western countries. National registers on disease and healthcare usage in Denmark do not routinely include data on ethnicity or country of birth. However, these registers can be linked to other registers in Statistics Denmark by using the personal ID number, the CPR-number. Linking between population registers and registers on disease and healthcare provides relatively good opportunities for studying relations between migration, ethnicity (defined as country of birth or parents' country of birth) and health. Research on ethnic minorities and health in Denmark largely use Statistics Denmark's definitions of persons of Danish origin, immigrants and descendants. In sub-sequent categorizations and variables, the focus on country of birth, ancestry and citizenship varies, but country of birth (own or parents') grouped in broad categories is most often used as the main variable.

Measures such as country of birth will always be rough proxies of complex mechanisms; hence, in combination with other data such as language and socio-economic status, it may present a more valid measure of ethnicity. However, country of birth is mainly used because it is easily accessible and easy to make operational, whereas routinely collected data on ethnic affiliation is considered difficult, costly and time consuming, and like religious affiliation is considered as sensitive data which is not included in national registers because of the risk of abuse, discrimination and violation of privacy.

Great Britain

Author: Peter Aspinall

In the second half of the twentieth century migrant flows were largely related to Britain's colonial past. The marked increase in immigration since the early 1990s, for reasons of asylum-seeking, education and work, and family migration and from an increasing number of countries, has transformed the country's ethnic diversity and ushered in an era of *super-diversity*, challenging census ethnic group categorisation as never before. This diversity has been captured in three decennial censuses which have listed subgroups in the Asian and Black pan-ethnicities and, since 2001, in the White and Mixed groups. However, the major contribution of the decennial census to granularity has been through the analysis of free-text responses, the extensive release of detailed country of birth data, and the use of cross-tabulation in the cultural question set. The use of granular ethnicity categories in health datasets is more limited, comprising the NHS Personal Demographics Service Birth Notification Data Set, the Family Origin Question in antenatal settings, and Medical Read and SNOMED CT ethnic origin codes in general practice. Most of the 40 or so routine health datasets still use the 2001 Census ethnic group classification, to the exclusion of the new groups added in 2011. The main set of granular ethnicity categories outside the health field is the Department for Education's 'Extended Categories' list, containing around 100 ethnic categories.

The main barriers to the introduction and use of granular ethnicity categories in official health datasets are organisational, involving complex bureaucratic processes and substantial costs. Further, in Britain there has been no strong advocacy or leverage for greater granularity from professional bodies in medicine, the NHS, and Public Health England. There are competing data priorities for the NHS, including the demands of the public sector's Equality Duty under the Equality Act 2010. Finally, there are technical issues with granular data itself, starting with the selection of these categories and the complex process of cross-mapping fine-grained categories back to census ethnic group categories for reporting. Ongoing developments across government to obtain greater granularity in ethnicity classifications are limited. They include continuing re-view of the case to add the new 2011 Census categories to routine health datasets and consideration of new ethnic categories for the upcoming 2021 Census.

Hungary

Author: Inez Koller

The Hungarian definition on ethnicity differs in specifications from the general definition of the project. According to the Hungarian version a person belongs to a certain ethnic group – which is called now "nationality" (1) as a general term – if oneself identifies with it, so it is a self-determined action. Furthermore, identification focuses on culture, language and historical traditions in the territory of

Hungary, which limits the number of the accepted nationalities in the country. Finally, it lacks physical features and third party determination also for historical reasons and avoids the usage of the term race which is considered as a discriminating and humiliating word in political and cultural contexts. Ethnic minority groups are all accepted nationalities except from the Hungarians according to law but common usage call them minorities just as all other ethnic groups living also in the country who are also called migrant communities. In Hungary there is also development in gaining more granularity in data collection on ethnicity.

Malaysia

Authors: Shyamala Nagaraj and Chiu Wan Ng

Malaysia is a multi-ethnic society. Historically, the country is home to a multitude of indigenous tribal groups. The country's geographical position in the middle of maritime trade routes between the east and the west, as well as British colonial policies of bringing in migrant workers from countries in the region to work in rubber plantations and tin mines, helped set the scene for increasing ethnic diversity over the past two centuries. Malaysia has also seen an increasing presence of migrant workers in agriculture, construction and services mostly from Indonesia, but also from Nepal, Bangladesh and the Philippines, often through inter-governmental arrangements. Different from earlier British policy, these migrants are required to return home after a fixed period. Economic opportunities have also made Malaysia a magnet for illegal economic migrants from neighbouring countries with which it shares borders.

Public agencies, in particular the Department of Statistics Malaysia, take the lead in efforts to accurately measure ethnic diversity for purposes of policy formulation and evaluation. Ethnicity is essentially self-reported and only one ethnic category is recorded per person. The granularity and identification of ethnic categories have changed and improved over time in line with changes in size of a group or its importance to public policy. Though data capture is often granular, information on ethnicity is mainly reported by only a few broad ethnic groups: Bumiputera, Chinese, Indian and others. Malaysia provides a unique example of the impact of public policy and concerns on ethnicity classification. The ethnic category Bumiputera (translation: princes of the soil) is a result of the New Economic Policy (NEP) first introduced in 1971 that provides special benefits to Malays, the largest ethnic sub-group in the Bumiputera category and to selected indigenous groups.

Malaysia has a welfare based health system. Health policies have been aimed at reducing health disparities between sub-populations which may or may not coincide with ethnic classification. The poor health status of rural communities has been a policy focus in the past but policy attention is shifting towards health needs of the urban communities. Health data by ethnicity captured by public agencies, in particular the Ministry of Health Malaysia, are often quite granular. Studies consider ethnicity a socio-political construct that can be used essentially as a social determinant of health.

Conclusions

Our findings demonstrate a diversity of approaches to ethnic group classifications internationally which follow a complex pattern. Within Europe, the UK, Hungary, Poland, Slovakia, Latvia, Romania and the Czech Republic, appear to have the most disaggregated approaches. Outside Europe, Malaysia, Canada, New Zealand and Bolivia also have disaggregated approaches to classification. Overall, we found multiple variations in the way in which ethnic classification is undertaken including differences in the underlying concept of ethnicity; the number of categories used; the way in which questions are phrased; the format of responses permitted; to what level responses are analysed; and whether the questions are compulsory.

Although lessons garnered in this report are primarily from within the health field, some examples were identified from the in-depth country reports of granular data being collected outside this field (e.g. education, and policy). Again, there is great diversity in classifications used and no particularly exemplary practice found which unquestionably warrants adoption within the health arena. One finding from both within and outside the health field, however, is that the greater the granularity within the classification, the more adaptable it is to different settings and to being analysed for differing purposes.

From the in-depth country reports we found a broad range of explanations as to why data are not being collected, analysed and reported at a more granular level. These included organisational factors; for example, the logistics and cost of designing and implementing new categories (e.g. UK, Denmark). There were also methodological reasons; a lack of advocacy for greater granularity; fear of stigma for particular ethnic minority groups (e.g. Hungary); political reasons; administrative barriers; potential for harm (e.g. Canada); and in countries where data is actually collected in a granular manor, methodological tendencies towards still aggregating data at the point of analysis (e.g. Aotearoa New Zealand).

It is apparent across the overview of EU countries and the in-depth country reports that the great variation in approaches to ethnic classification, and granularity of data, appear to be contingent on contextual factors unique to each country, including the country's social, political, economic, historical and geographical circumstances. These factors are therefore important to bear in mind when identifying practices and considering if these are generalizable to another context, such as within a US setting. It is therefore problematic to specify an ideal way that data should be collected, analysed and reported. We can identify good practices, such as the four-tier classification system of analysis in Aotearoa New Zealand classification which offers more than 230 ethnic categories at the most detailed output level, but most likely these practices will need to be adapted in another context. Therefore, in place of recommending specific practices, a set of considerations/principles for developing classifications for use within the health field have been developed from this work. The adoption of such principles may

assist in the development of country-specific ethnic systems of classification and also could support comparisons of data being made over time and across countries.

Considerations/Principles:

1. Ethnicity is predominantly a social construct, there is no global consensus about the concept and definition of ethnicity, and it varies across contexts – it should be made explicit what underlying concepts are being applied when developing a classification.

2. Operationalising ethnicity as a concept to classify population groups needs to take into account country contexts which greatly influence the feasibility of implementing granular ethnic enumeration.

3. Clarifying the 'objective' of collecting ethnicity data is paramount to determining what should be collected. It should be made explicit what the data are to be used for, how granularity can operate in classifications given the outcomes in mind and what the societal objectives are (e.g. health equity).

4. We need to consider how ethnicity data are collected in relation to the outcomes in mind. For example in the health field where self-assigned ethnicity is the standard as it correlates with health behaviours and risks. However, there are instances where socially assigned ethnicity has been demonstrated to correlate with quality of healthcare received and inequitable outcomes – (see appendix 10)

5. The number and order of ethnic group categories needs to be considered to determine the usability and equity of a classification. (e.g. using a long list of categories might be confusing and particular ordering of groups may favour dominant groups).

6. Allowing free text responses as well as multiple response may assist in achieving greater granularity and accommodating the increasing population of people identifying themselves as *mixed-origin* (although this may present challenges in data processing and analysis).

7. Developing a flexible hierarchical categorisation which can be expanded or collapsed and enables a very high degree of granularity when appropriate, may be the most advantageous approach to pursue for analysing data from free text responses (as seen in New Zealand).

8. Developing protocols to guide the analysis of granular ethnic group may assist in appropriately utilising available data and motivating the collection of granular data.

9. Processes to enable community involvement in the development of ethnic group categories and in decision-making on analysis of and inclusion in statistical surveys should be considered.

Table of contents

Chapter 1- Background to the Robert Wood Johnson Foundation project

Nazmy Villarroel, Emma Davidson, Pamela Pereyra-Zamora, Allan Krasnik and Raj Bhopal

The United States perspective leading to the Robert Wood Johnson Foundation project

The Robert Wood Johnson Foundation (RWJF) is undertaking a multi-component research project to identify and develop strategies and opportunities for disaggregating ethnic/racial group data in the United States (US). This project is prompted by the critical importance of obtaining and utilising high-quality group data to understand and eliminate racial and ethnic health disparities. Current ethnic/racial classifications in the US official statistics are usually published in six race/ethnicity categories: White, Hispanic or Latino, Black or African American, Asian, Native Hawaiian or other Pacific Islander, Native American and Alaska Natives. These are extremely broad categories, especially Asian, which comprises all countries east of Istanbul through to Japan, encapsulating a large and heterogeneous population (Hunt and Bhopal 2004). Ethnic disparities can exist within, and be masked by, the aggregation of such heterogeneous populations and therefore, in the US, there is a move to explore the use of more granular ethnic classifications which recognise the heterogeneity within traditionally broad categories. It is envisaged that improving the quality and granularity of data may assist in providing appropriate services, develop culturally tailored programs, producing sound policies and achieving equality (Bhopal 2006).

The Heterogeneity/granularity in ethnicity classifications project outside the US

The RWJF recognises that countries vary greatly in ethnic/racial composition and in their histories of migration. Policies and practices for the collection and analysis of ethnic/racial data have developed in heterogeneous ways throughout the world. Approaches range from no systematic collection of data (e.g. France), to the collection by country of birth (e.g. The Netherlands), or to a more fine-grained approach utilising disaggregated ethnicity data and country of birth (e.g. United Kingdom). This wide variation in approaches to data collection may relate to the degree of political and societal acceptability of acknowledging ethnic diversity. Consequently, international collaboration and debate about terminology and systems of classification is both desirable and potentially beneficial. To supplement its US work, the RWJF invited the Heterogeneity/Granularity in Ethnicity Classifications Project (HGEC) to gather international contexts and perspectives.

Aim and objectives of the HGEC project

General aims of the HGEC project were:

1. To explore approaches and lessons from selected countries with exemplary models outside the US, related to collecting, analysing and reporting disaggregated data using granular ethnic classifications.
2. To use the results to inform the US projects, and in turn to be discussed and refined in collaboration with US colleagues.

Specific objectives of the HGEC project:

- To identify up to seven countries outside the US that collect, analyse and report data for racial and ethnic groups which go beyond broad categorizations, as well as assessing the degree of heterogeneity/granularity of their ethnic classifications, especially in censuses and population registers.
- To seek to understand how and why heterogeneity/granularity in these classifications has developed in terms of the social, historical and political context in the selected countries.
- To understand why disaggregated data is not being collected, analysed or reported more often even though the field generally agrees that this is critical to understanding disparities.
- To identify some examples amongst the selected countries of how disaggregated data has been used and the impact it has had on policies, programmes and population health outcomes.
- To assess if there are lessons relating to heterogeneity/granularity of ethnic and racial classifications that can be learnt from outside the health field.
- To synthesize lessons and develop findings to support US and international efforts.
- To explore global lessons for how data should ideally be analysed and reported given these heterogeneity/granularity ethnic group classifications, to complement the US project.

Elements of the HGEC project

To address these specific objectives, three elements of the HGEC project were developed and will be reported on. These are:

1. An overview of European data sources conducted by the core HGEC research team.
2. Seven in-depth country reports undertaken by international experts in the field.
3. A meeting of the core team and the above experts in Edinburgh to critically reflect on our findings to generate lessons and recommendations.

Conceptual background

Globalisation has resulted in increasing ethnic diversity worldwide, and the structure of societies is being transformed socially and culturally. Understanding social processes such as assimilation, social mobility, and the cultural construction of ethnic groups has become important to comprehend what constitutes the dynamics of ethnic identity. Consequently, approaches to adequately describe and interpret complex multicultural societies are required (Simon and l'intolérance 2007, Rodríguez-García 2010, Simon 2012). However, both the terminology and the concepts underpinning classification of ethnicity/race are problematic and practices vary greatly, including geographically and across disciplines (Bhopal 2013).

Definitions

Throughout this report the following definitions will be used and authors of the country reports will state when and how their definitions may differ:

Ethnicity

"The social 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 traditionally associated with race. Increasingly, the concept is being used synonymously with race but the trend is pragmatic rather than scientific" (Bhopal 2004, Bhopal 2013).

Ethnic minority group

"Usually, but not always, this phrase is used to refer to a population other than white. Alternatively, it may be used to describe a specific identifiable group, for example, gypsy travellers, and less commonly, Irish in the UK. Some people consider the phrase inaccurate and prefer minority ethnic group, but the two phrases are used synonymously" (Modified definition) (Bhopal 2004, Bhopal 2013).

Ethnic group and Ethnic category

"If by 'group' we mean a mutually interacting, mutually recognizing, mutually oriented, effectively communicating, bounded collectivity with a sense of solidarity, corporate identity and capacity for concerted action, or even if we adopt a less exigent understanding of 'group', it should be clear that a category is not a group. It is at best a potential basis for group-formation or 'groupness'" (Brubaker 2004).

Ethnocentrism

"The tendency to perceive and interpret from the standpoint of one's own culture. In epidemiology the tendency is reflected in the practice of using the White population as the norm or standard" (Bhopal

2004, Bhopal 2013). *Some have argued that the use of the term 'non-white' may be interpreted as an example of this* (Aspinall 2008).

Race

"By historical and common usage the group (sub-species in traditional scientific use) a person belongs to as a result of a mix of physical features such as skin colour and hair texture, which reflect ancestry and geographical origins, as identified by others or, increasingly, as self-identified. The importance of social factors in the creation and perpetuation of racial categories has led to the concept broadening to include a common social and political heritage, making its use similar to ethnicity. Race and ethnicity are increasingly used as synonyms causing some confusion and leading to the hybrid terms race/ethnicity" (Bhopal 2004, Bhopal 2013).

Granular ethnicity

"Granularity means a fine level of detail; the greater level of granularity the more finely detailed the data category is" (McFadden, Nerenz et al. 2009).

Cultural hybridity

"Hybridity refers to the process of the emergence of a culture, in which its elements are being continually transformed or translated through irrepressible encounters. Hybridity offers the potential to undermine existing forms of cultural authority and representation." (Scott and Marshall 2009).

Nationality

"The state of being a citizen or subject of a particular country" (e.g. British nationality) (Law 2015).

Super-diversity

"Immigrant super-diversity is distinguished by a dynamic interplay of variables, including their country of origin (comprising a variety of possible subset traits such as ethnicity, language, religious tradition, regional and local identities, cultural values and practices" (Vertovec 2006).

Methods to gather information on ethnic group classification

To develop a suitable ethnic group classification it is essential to understand the context of the country such as its conception of ethnicity, the ethnic composition of the population and migration patterns (Bhopal 2013). The method by which data are collected should also be considered, in particular whether responses are self-determined (to a group someone truly identifies with) or assigned (identified with a group by others). Internationally, four different approaches to racial or ethnic classifications are commonly used:

- "Self-identification by the person concerned" (Simon 2004).

- "Identification by a third party based on indicators such as birthplace or nationality" (Simon 2004).
- "Identification by a third party based on visual observation (e.g. by company manager or a school's administrative staff, survey conductor)" (Simon 2004).
- "Identification by community members (e.g. In the US this method is used for classing American Indians in 'federally recognised tribes', in Hungary by prominent members of the minority group)" (Simon P, 2004).

In Europe, the first two methods predominate (Simon 2004, Ringelheim 2011). Self-identification is applied in the UK and Ireland. For instance, within the healthcare setting and other organizations, individuals are given a list of pre-established ethnic categories. They are asked to declare which group they consider themselves part of, or to add an option not mentioned on the list. Even with this procedure, cognitive research by Britain's Office for National Statistics (ONS) indicates that the process of selection is not always clear-cut/definitive, i.e. it may be the option the person most closely identifies with though it may still be regarded as unsatisfactory and falling short of the 'groupness' test (i.e. 'true identification with') (Aspinall and Song 2013). Additionally, in practice, individuals may not always be asked, and identification by a third party on visual observation still occurs. In the UK and Hungary censuses individuals can identify their own ethnic group, although one member of a family may complete on behalf of others so that again this may not represent true identification. In countries like the Netherlands and Sweden, people are classified by the population registers into pre-defined categories according to their country of birth (CoB)/parents' CoB (Ringelheim 2011).

As this brief conceptual background illustrates, obtaining informative data on ethnicity/race for official statistics is complex. Every country requires appropriate methods to record, analyse, and report these data according to its context (Simon 2012). The concepts underpinning data must be defined, appropriate ethnic categories chosen, and the approach for the collection of data should be decided. These considerations will be further explored for a variety of contexts within our European overview (see chapter 2) and country reports (see chapters 3-9).

Chapter 2- Overview of data sources within European Union countries

Nazmy Villarroel, Emma Davidson, Pamela Pereyra-Zamora, Allan Krasnik and Raj Bhopal

Abstract

Background: The changing nature of global migration and increasing diversity of populations have transformed the social landscape of European countries. Such complex social formations have challenged not only public health but also other private/public agendas (e.g. cultural tailoring, diversity in the workforce). To understand how collective identities are produced, and identify the health needs of diverse groups, demographic data that capture population heterogeneity (e.g. by ethnicity) are needed. Little is known about the granularity of current methods of ethnic categorization in Europe. Thus, we aimed to explore and provide an overview of how EU-28 countries approach the collection of granular ethnic classifications.

Methods: Data were obtained primarily from official population censuses or registers. For each country these data sources were examined for their approach to ethnicity. When ethnic information was not gathered, country of birth (CoB) and/or parents' CoB, language spoken, religion and national identity were examined as a proxy for ethnicity.

Results: Granular ethnicity data were found in eight countries: England, Wales, Republic of Ireland, Northern Ireland, Scotland, Hungary, Poland and Slovakia, which collected more than six categories. We found that Estonia, Lithuania, Croatia, Bulgaria, Republic of Cyprus and Slovenia paid some attention to granularity, collecting one to six categories. Information on ethnicity with only a write-in option was found in Latvia, Romania and Czech Republic. Most countries collected CoB (individual and parents), nationality, religion and language (mother tongue).

Conclusions: 1) There was a lack of data collection disaggregated by ethnic categories within EU-28 countries and in countries where ethnicity data was collected, the number of categories varied widely 2) Ethnicity is conceptualised in different ways and diverse terminology was employed. 3) Categories are influenced by political rights and legislation; historical events, ideology and sensitivity towards cultural identity; and ongoing migration patterns. 4) A free text option may provide the most granular approach for ethnic enumeration, but only if granularity is retained in the analyses and reporting of data.

Background to the Overview

The EU perspective on the collection of ethnic/racial data

Prior research programmes have described the availability of official data sources for the identification of migrant populations and ethnic categories. A global study on ethnic classifications using the 2000 censuses, found that 63% of the national censuses included ethnic categorization (Morning 2008).

Recent studies within Europe have examined official data sources for migrants and ethnicity/racial information. These are summarised in table 2.1.

Table 2.1: Summary of existing European studies of official data sources for ethnicity/racial information

Research project	Duration	Objectives	Main findings	URL
Migrant and Ethnic Health Observatory (MEHO)	1 January 2007 to 1 January 2010	"To construct an inventory of existing data sources on migrant health across EU member states and to develop migrant and ethnic-specific indicators within different areas, including health care utilization" (Rafnsson and Bhopal 2009)	• Lack of registry data in 16 EU countries • Diversity in the definition of migrant status hampers cross-national comparisons • Calls for urgent establishment of registries, expansion of the existing registry information, and adoption of a common, generally acceptable definition and identification method of migrants across the EU.	MEHO
Adapting European health systems to diversity (ADAPT)	15 December 2011 to 31 July 2016	"To identify obstacles to translating into action the existing body of knowledge concerning health and healthcare inequalities among migrants and ethnic minorities, and to propose ways of overcoming these obstacle"(Ingleby 2011) "To connect researchers working on specific topics with each other, who could then undertake work together- to create "think tanks" (Verbal communication from project leader Ingleby, D.)	• Established 'think tanks' • Developing policy briefs. One of these is about Remedying the shortcomings of existing data on health, migration and ethnicity – but this work is still on going.	ADAPT

This existing body of work demonstrates a mixed landscape throughout Europe. Although data sources were identified, and used to study inequalities in health, it was apparent that both a lack of data and the variability in categorisations were obstacles to effective analysis and use of this information. Within Europe, there exists much debate surrounding the benefits and risks of the collection and use of ethnic/racial data within official statistics which has resulted in such diverse practice. The Race Equality

Directive, European Commission (URL accessed) the Act CLXXIX/2011 on the Rights of Nationalities in Hungary (URL accessed) and the UK's Race Relations (Amendment) Act 2000 (URL accessed), highlight the potentially crucial role played by statistics in setting in motion anti-discrimination policies and developing their capacity to ensure social cohesion and promote diversity and equality (Parsons, Godfrey et al. 2004). However, this argument does not offer sufficient justification for the construction of ethnic group categories in official statistics throughout Europe, and there is much contention about the collection and use of such sensitive data (Jacobs and Rea 2012).

In France, for example, the debate arising from a proposal to construct ethnic group categories in statistics and the lack of research using the concept of ethnicity reflect the fact that this is a volatile political matter (Holst 2012, Jacobs and Rea 2012). It is feared that political parties and coalitions could potentially misuse these data to maintain strained relationships between majority and minority population groups (Holst 2012), and that ethnic categorization might be directly or indirectly influenced by ethnocentrism; contributing to the targeting of prejudice towards particular groups (e.g. immigrants) (Volpato, Durante et al. 2010). Indeed, recent complaints on immigration and the assumption that this issue was causing strains for public services was used as one of the main key strategies that led to the UK leaving the EU ("Brexit") in the referendum vote (McKee and Galsworthy 2016). This situation might potentially influence future categories in the current census (e.g. Polish).

Ethnic group categorization can act as an instrument for excluding members of neglected minority groups (Krizsán 2001), by not encompassing them in the established categories. This absence of population information is detrimental to assessing and addressing health and social care needs. For example, the Roma people (who are Europe's largest ethnic minority group) have an estimated population of just over 11 million and have some of the greatest health needs, but there is great uncertainty over their exact number (Parekh and Rose 2011). Roma people might not have documentation and may be reluctant to identify themselves for fear of stigmatisation, but the foremost reason for this lack of information is that many European countries do not collect data about their Roma population (Masseria, Mladovsky et al. 2010).

Collection of ethnicity data, therefore, remains inadequate in most of Europe, either because countries do not collect any systematic ethnic data, or because they collect inadequate data for research, policy, and practical purposes (Krizsán 2001). In this chapter we provide a more in-depth overview of the current situation within EU-28 countries, especially in relation to the granular/heterogeneity of ethnicity classifications.

Research questions of the overview

The research questions for the overview were:

1. What official sources of national data are there within EU-28 countries regarding the ethnic/racial composition of their population? (For example, census and population registers).

2. From the official sources of data identified, which countries collect, analyse, and report data for racial and ethnic groups with attention to actual or potential heterogeneity/granularity?

General aim of the overview

We aimed to explore and provide an overview of how selected EU-28 countries approach the collection of granular ethnic classifications.

Methods of the Overview

Literature review

A scoping review focused on census data has identified information reporting the current situation of ethnic/racial data collection within the EU-28 countries. For this purpose we consulted two expert librarians (Marshall Dozier and Stuart MacDonald) who advised on the data sources, particularly the databases and grey literature research.

Systematic searches were carried out on government websites and other sources, including identification of grey literature, using the following key words and databases.

- Key words: (ethnic* or race) AND (classif* or categor*)
- Databases: ASSIA (Applied Social Sciences Index and Abstracts); IBSS (International Bibliography of the Social Sciences); Race Relations Abstracts; SocINDEX with Full Text;
- International Micro-data bases: IPUMS; Data without borders
- International Meta-data bases: Eurostat
- Data Archives: CESDA; UK-data archive
- Surveys: World health survey; European Interview health survey; European Social Survey.
- Official statistical sources: United Nations Statistical Division (UNSTAT); Office for National Statistics (ONS)

The following criteria were used to select relevant information:

- Reports/papers/data sources reporting on EU-28 countries
- Sources that identified data on ethnic/racial composition of country
- Sources that adequately described ethnic/racial classifications used in these official data sources

Analysis and interpretation of data

The standards for the classification of federal data on race in the US census bureau and in the office of management and budget have a minimum of six ethnic/racial categories (URL, data accessed). We classified country results based on this. We considered countries with one to six 'tick box' categories as having some granular approach, while countries with more than six 'tick box' categories and countries that had a write-in options, were considered as having a granular classification. (see appendix 2)

Consultation with investigators and other relevant individuals and universities

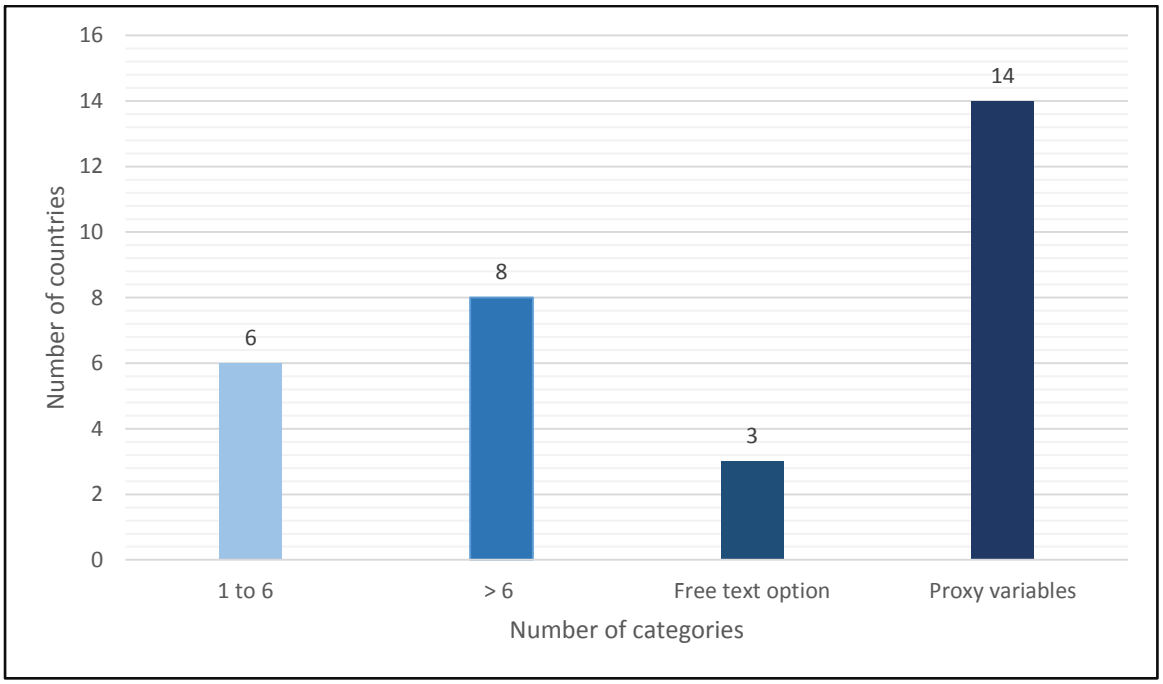
We consulted a group of international investigators (Dr. Donna Cormack and Dr. Tahu Kukutai (New Zealand); Dr. Pamela Pereyra-Zamora (Bolivia); Prof. Hude Quan, Kelsey Lucyk, Dr. Karen Tang (Canada); Prof. Allan Krasnik and Dr. Liv Stubbe (Denmark); Dr. Peter Aspinall (UK), Dr Inez Koller (Hungary) and Dr. Shyamala Nagaraj and Dr. Chiu Wan Ng (Malaysia) through various meetings, teleconference and a final meeting held in Edinburgh in May 2016, who prepared the country reports for each country and who advised on the overview (see appendix 2).

Results of the Overview

Heterogeneity/granularity of approaches in censuses or population registers

The results of the overview for the EU-28 countries are illustrated in Figure 2.1 and discussed below.

Figure 2.1 Number of ethnic group categories found in EU 28 censuses or population registers and number of countries with a free text option and proxy variables.



Countries using the ethnicity concept

We found six EU countries whose official census has classifications of one to six ethnic categories (see table 2.2). Three of these countries included a write-in option as one of these categories (e.g. other ethnicity). Estonia and the Republic of Cyprus had the most categories in this group; Bulgaria, Lithuania and Slovenia had four categories and Croatia gave only two options. Most countries, particularly Bulgaria, Estonia, Lithuania and Croatia, appear to base their classifications primarily on the main nationalities that make-up their population and the immigrants from neighbouring countries. Strikingly, the Republic of Cyprus took a slightly different approach to categorization as they included ethno-religious groups such as 'Maronite' and 'Latin'. Slovenia used fewer defined categories and included the option of non-disclosure and, uniquely, it was the only country to provide an 'ethnic indeterminate' category. Some examples of the text are included in Box 2.1:

Box 2.1 Examples of approaches to categorization in countries using 1-6 categories

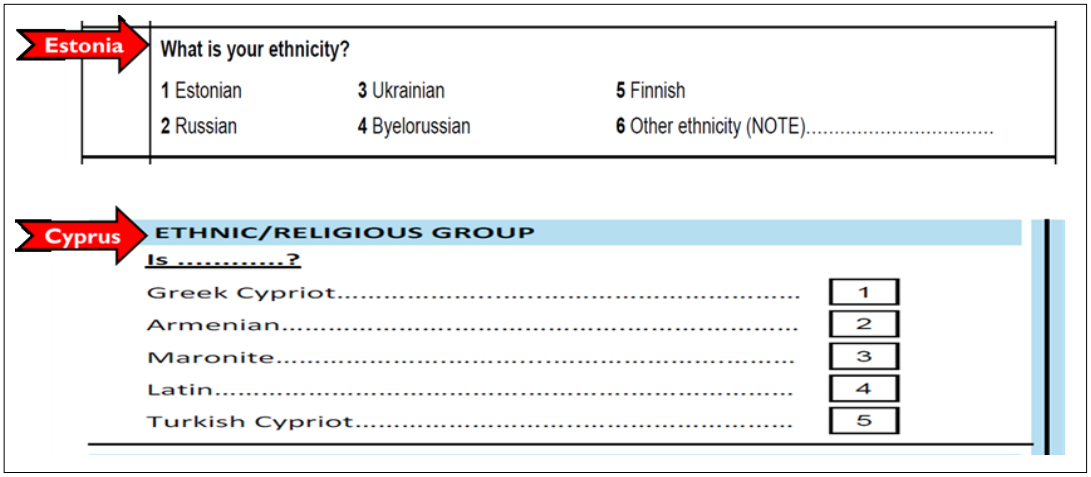


Table 2.2 Countries with one to six ethnic categories

COUNTRY PROFILE	CATEGORIES
Geographical region: Eastern Europe	
Estonia† census year: 2011 number of categories: 6	1. Estonian 2. Russian 3. Ukrainian 4. Byelorussian 5. Finnish 6. Other ethnicity (please note)
Lithuania† Census year: 2011 Number of categories: 4	1. Lithuanian 2. Polish 3. Russian 4. Other (please note)
Geographical region: South-Eastern Europe	
Republic of Cyprus Census year: 2011 Number of categories: 5	1. Greek Cypriot 2. Armenian 3. Maronite 4. Latin 5. Turkish Cypriot
Bulgaria‡ Census year: 2011 Number of categories: 4	1. Bulgarian 2. Turkish 3. Roma 4. Other (please note) Not declared.
Croatia† Census year: 2011 Number of categories: 2	1. Croat 2. Other, state which
Geographical region: South-Central Europe	
Country: **Slovenia†‡** Census year: 2011 Number of categories: 4	1. Slovenian 2. Italian 3. Hungarian 4. Other nationality/ethnicity (please note) I'm nationally/ethnically indeterminate; I don't wish to answer this question; Answer is not possible, as there is no P3/NV statement for the absent person

† Note: Countries which include one or more free text option.
‡Note: Countries which include a not compulsory answer
Source: World Population and Housing Census Program

There were eight EU-28 countries with more than six categories within their census (see table 2.3), four being within the United Kingdom: Scotland, England, Northern Ireland and Wales. These countries, in particular, have a wide range of options for disaggregating the White population group and base their classification on a concept of ethnicity with elements of ancestry, culture and skin colour predominating. An extract of the Scottish census is provided in Box 2.2. Some countries outside the UK based their categories on national identities (Koller 2014), (Hungary) and on ethno-religious groups (Poland and Slovakia).

The Scottish, English and Welsh census categories also provide a relatively disaggregated classification for Asian and African, Caribbean or Black groups. In contrast, the Republic of Ireland and Northern Ireland provide limited options for these population groups. There is a rapidly increasing population of people globally who identify as mixed-ethnicity and this has been included as a category in Scotland, England and Wales. Scotland's mixed category included a write-in option, offering more freedom for responses, but potentially creating difficulties for analysis and interpretation; compared to England and Wales which included disaggregated 'tick boxes' within the Mixed/multiple ethnic group category (e.g. White and Black Caribbean; White and Asian).

Most UK countries provide a Gypsy/Traveller category and Hungary, Poland and Slovakia include an option for the Roma population group (see table 2.3). However, these categories may still include people with very diverse ethnic backgrounds, for example, the Roma population sub-groups (e.g. English Romanichals, Iberian Kale, Welsh Kale, Scandinavian Kale, Central European Sinti, Hungarian-Slovak Bashalde, Romanian Ludar and Boyash, and East European and Vlach-speaking Roma or Vlach Rom) (Parekh and Rose 2011), and sub-groups in Hungary (e.g. Romungro, Beás and Lovári).

Table 2.3 Countries with more than six ethnic categories

COUNTRY PROFILE	CATEGORIES	
Geographical region: Northern Europe		
Scotland† Census year: 2011 Number of categories: 19	A. **White:** 1. Scottish 2. Other British 3. Irish 4. Gypsy / Traveller 5. Polish 6. Other white ethnic group, please write in B. **Mixed or multiple ethnic groups:** 7. Any mixed or multiple ethnic groups, please write in C. **Asian, Asian Scottish or Asian British:** 8. Pakistani, Pakistani Scottish or Pakistani British 9. Indian, Indian Scottish or Indian British 10. Bangladeshi, Bangladeshi Scottish or Bangladeshi British 11. Chinese, Chinese Scottish or Chinese British. **12.** Other, please write in	**D. African:** 13. African; African Scottish or African British 14. Other, please write in. E. **Caribbean or Black:** 15. Caribbean, Caribbean Scottish or Caribbean British 16. Black, Black Scottish or Black British 17. Other, please write in F. **Other ethnic group:** 18. Arab, Arab Scottish or Arab British **19.** Other, please write in
England† Census year: 2011 Number of categories: 18	A. **White:** 1. English / Welsh / Scottish / Northern Irish / British 2. Irish 3. Gypsy or Irish Traveller 4. Any other White background, write in B. **Mixed / multiple ethnic groups:** 5. White and Black Caribbean 6. White and Black African 7. White and Asian 8. Any other Mixed / multiple ethnic background, write in	C. **Asian/Asian British:** 9. Indian 10. Pakistani 11. Bangladeshi 12. Chinese 13. Any other Asian background, write in D. **Black / African / Caribbean / Black British:** 14. African 15. Caribbean 16. Any other Black/African/Caribbean background, write in E. **Other ethnic group:** 17. Arab 18. Any other ethnic group, write in

Wales† Census year: 2011 Number of categories: 18	A. **White:** 1. Welsh / English / Scottish / Northern Irish / British 2. Irish 3. Gypsy or Irish Traveller 4. Any other White background, write in B. **Mixed / multiple ethnic groups:** 5. White and Black Caribbean 6. White and Black African 7. White and Asian 8. Any other Mixed / multiple ethnic background, write in	C. **Asian / Asian British:** 9. Indian 10. Pakistani 11. Bangladeshi 12. Chinese 13. Any other Asian background, write in D. **Black / African / Caribbean / Black British:** 14. African 15. Caribbean 16. Any other Black/African/Caribbean Background, write in E. **Other ethnic group:** 17. Arab 18. Any other ethnic group, write in
Northern Ireland† Census year: 2011 Number of categories: 11	1. White 2. Chinese 3. Irish Traveller 4. Indian 5. Pakistani	6. Bangladeshi 7. Black Caribbean 8. Black African 9. Black Other 10. Mixed ethnic group, write in 11. Any other ethnic group, write in
Republic of Ireland† Census year: 2011 Number of categories: 8	A. **White:** 1. Irish 2. Irish Traveller 3. Any other White background B. **Black or Black Irish:** 4. African 5. Any other Black background	C. **Asian or Asian Irish:** 6. Chinese 7. Any other Asian background D. **Other, including mixed background:** 8. Other, write in description

† Note: Countries which include one or more free text option.
‡Note: Countries which include a not compulsory answer
Source: World Population and Housing Census Program

Table 2.3 Continued

COUNTRY PROFILE	CATEGORIES	
Geographical region: Central Europe		
Hungary†‡ Census year: 2011 Number of categories: 19	1. Hungarian 2. Bulgarian 3. Gipsy (Roma) 4. Greek 5. Croatian 6. Polish 7. German 8. Armenian 9. Romanian 10. Ruthenian	11. Serbian 12. Slovakian 13. Slovenian 14. Ukrainian 15. Arabian 16. Chinese 17. Russian 18. Vietnamese 19. Other (please note) Do not wish to answer
Poland† Census year: 2011 Number of categories: 15	1. Polish 2. Belarussian 3. Czech 4. Karaitic 5. Lithuanian 6. Lemko 7. German	8. Armenian 9. Romany 10. Russian 11. Slovakian 12. Tatar 13. Ukrainian 14. Jewish 15. Other (please note)
Slovakia† Census year: 2002 Number of categories: 15	1. Slovak 2. Ruthenian 3. German 4. Serbian 5. Moravian 6. Hungarian 7. Ukrainian	8. Polish 9. Russian 10. Bulgarian 11. Roma 12. Czech 13. Croatian 14. Jewish 15. Other (please note)

† Note: Countries which include one or more free text option.
‡Note: Countries which include a not compulsory answer
Source: World Population and Housing Census Program

Box 2.2 Extract from Scottish census demonstrating disaggregation of White and Asian categories

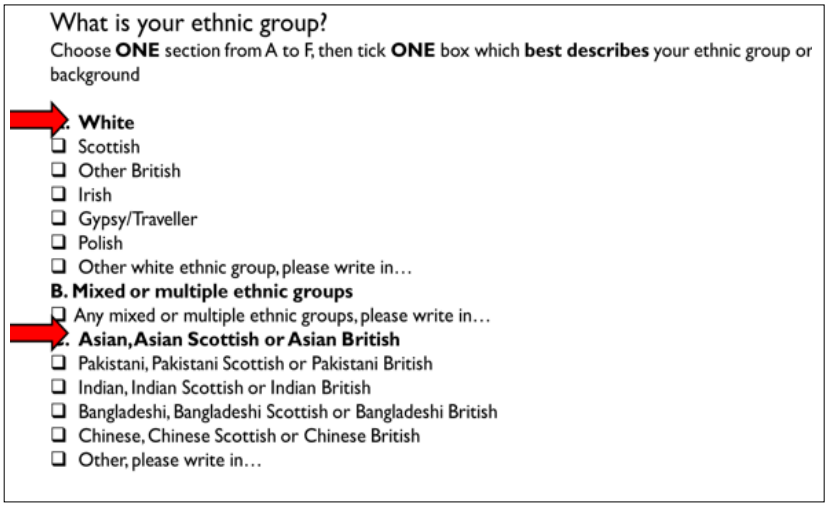


Table 2.4 shows the three EU-28 countries which provide only a write-in option and outlines the ethnicity question they pose. The advantage of a write-in option is that it provides the greatest flexibility for people to self-identify with the group they most affiliate with, increasing the granularity of the census. The difficulty is that this information may be challenging to collect and analyse, and it might require aggregation into higher categories by a third party.

Table 2.4 Countries with exclusively a free text option for ethnicity

Geographical region	Countries with write-in option	Census year	Question
Eastern Europe	Latvia	2011	What is your ethnicity?
South-Eastern Europe	Romania‡	2011	What ethnic group does the person consider he/she belongs to?
Central Europe	Czech republic‡	2011	Ethnicity (You may state two ethnicities)

Source: World Population and Housing Census Program
‡ Countries which include a not compulsory answer

Countries with potential proxy variables for ethnicity

The fourteen countries which did not collect ethnicity data, did include a proxy of ethnicity. Table 2.5 shows that most of these countries collected information on CoB and nationality in their population registers and censuses, while two collected other proxy variables such as mother tongue and religion. Some of these proxy variables can be used to assign ethnicity, as is practiced in Denmark (such as Western/Non-Western or more specific categories), and this is discussed further in the country report in Chapter 6.

Table 2.5 Countries with potential proxy variables for ethnicity

Geographical region	Countries	Census/population register year	Proxy variables for ethnicity
Northern Europe	Denmark	2011	CoB, and/or parents' CoB
	Sweden	2011	CoB and/or parents' CoB
	Finland	2010	CoB, nationality and language
Central Europe	Germany	2011	CoB, citizenship, religion, immigrant ancestry and religion
	Austria	2011	Place of birth and country of citizenship
Western Europe	Netherlands	2011	CoB
	France	2008	CoB, nationality
	Luxembourg	2010	CoB, nationality
	Belgium	2011	CoB, citizenship
Southern Europe	Malta	2011	CoB and citizenship
	Spain	2011	CoB, nationality and parents' CoB
	Italy	2001	CoB, citizenship
	Greece	2001	CoB, citizenship
	Portugal	2011	Nationality

Source: World Population and Housing Census Program

Non-compulsory status of ethnicity/ethnicity proxy variable in census and population registers

We observed that in three of the EU-28 countries it was not compulsory for individuals to disclose their ethnicity status. This is likely due to sensitivities around the misuse of data, stigmatisation and possible discrimination which have historically affected continental Europe.

Discussion of the overview

Previous research has documented the availability of official data sources for the identification of migrant populations and ethnicity classification (Morning 2008, Nielsen, Krasnik et al. 2009, Ingleby 2011). However, these studies have not focused on evaluating the granularity of ethnic categorization. In this work, we have provided both an overview of the EU-28 countries which collect ethnicity data in official data sources, and an assessment of the granularity of classifications within these data sets.

We found that: 1) There was a lack of data collection by ethnic categories within many EU-28 countries and, in countries where ethnicity data was collected, the number of categories varies widely. 2) Ethnicity is conceptualised in different ways for the selected countries and, consequently, diverse terminology was employed. 3) Categories seem to be influenced by political rights and legislation; historical events, ideology and sensitivity towards cultural identity; and ongoing migration patterns. 4) A write-in option may provide the most granular approach for ethnic enumeration, but only if granularity is retained in the analyses and reporting of data.

Lack of data collection and variation in number of categories

Only eight out of thirty–one countries (four included in the UK) provided a granular approach with more than six ethnic categories. Most of these countries were in the UK and the others were in Central Europe. The lack of collection of ethnicity data within other EU-28 countries might reflect attitudes to data and interpretation of legislation protection. For countries which collected ethnicity data, the number of categories ranged from two (Croatia), to nineteen (Scotland and Poland). Examining countries with comparable numbers of ethnic categories, there was no consistent approach to categorisation, except across countries within the UK. These diverse approaches relate in part to an individual country's perspective on conceptualising and operationalising ethnicity.

Conceptualising ethnicity and diverse terminology

EU-28 censuses and population registers measure ethnicity in various ways. The concept of ethnicity may encompass elements of nationality, CoB, parents' CoB and language, in combination with national/geographical origin, or ancestry, religion and racial group. (Statistics 2003) Ethnic group categorization might also include characteristics that are grounded in biological differences, such as skin colour (Omi 2001, O'Hearn 2008). For example, we found that in England, the Republic of Ireland, Northern Ireland, Scotland and Wales, the censuses used skin colour terms in combination with nationality/ancestry (Aspinall 2012). However, European statistical guidelines have suggested that skin colour should not be used in ethnic group categories and the majority of EU-28 countries used nationality to categorize groups (see table 2.5). To add to the complexity, the concept of nationality also differs between countries and is awarded according to varying criteria. UK nationality is related to citizenship and in some countries nationality is related to both ethnicity and citizenship (Morning 2008).

Furthermore, in countries like Bulgaria and Hungary, nationality often denotes ancestry, national origin or cultural nationality.

EU-28 countries also use combinations of concepts. An example of this is the Republic of Cyprus (see table 2.2 and Box 2.1), where we found in the census two ethno-religious groups, the Maronite and Latin. Also, in Poland, the census categories include a combination of nationality (Polish, Belarussian, Czech, Lithuanian, German, Armenian, Russian, Slovakian and Ukrainian), ethnicity (Karaitic, Lemko), language/religion (Tatar, Jewish), and ethnic minority groups (Romany). Lastly, the 2011 census of Slovenia uniquely gave the option I'm nationally/ethnically indeterminate and I don't wish to answer this question.

Operationalisation of the concept of ethnicity was variable, as reflected in the questions asked. In general, the questions asked for ethnicity combined with other terms such as ethnic affiliation (Poland), ethnic nationality (Hungary), ethno-religious group (The Republic of Cyprus), ethnic group (UK and the Republic of Ireland) and ethnicity (Czech Republic). The concept may vary when translated into English although we interpreted the findings with the assistance of experts from each country, to check that the meaning was not lost or altered in translation.

Census Classification faces even greater challenges due to the rise of new mixed or multiple identities within population groups (Aspinall 2012). Some countries (e.g. England, Scotland), include a *mixed group* category, but further work needs to develop approaches to this categorisation (Aspinall 2003, Bhopal 2004). The intersection between ethnic categories, religion and nationality was seen in the Polish Tartar group (Tatar, Muslim and Pole, at the same time), which is a hybridisation of national, ethnic and cultural identity (Cieslik and Verkuyten 2006).

Factors influencing categories used in different countries in the EU

The degree of granularity and the terminology of ethnic categorization were extremely heterogeneous. Potential reasons for these findings relate to the contextual influences of political rights and legislation, history, and sensitivity towards cultural identity, and ongoing migration patterns. For example, categories may interlink with economic and policy choices. This could occur both because political leaders may actively pursue policies that impact on perceived ethnic diversity and because citizens may choose their identity differently in response to political and economic conditions (Alesina and Ferrara 2005). Hence, future political decisions could modify the granularity of ethnic group categorization. The main contextual considerations identified in our background and European overview are illustrated in Figure 2.1 which summarises the complex interplay of these influences on the development of ethnic/racial group classifications. We discuss each of these influences below.

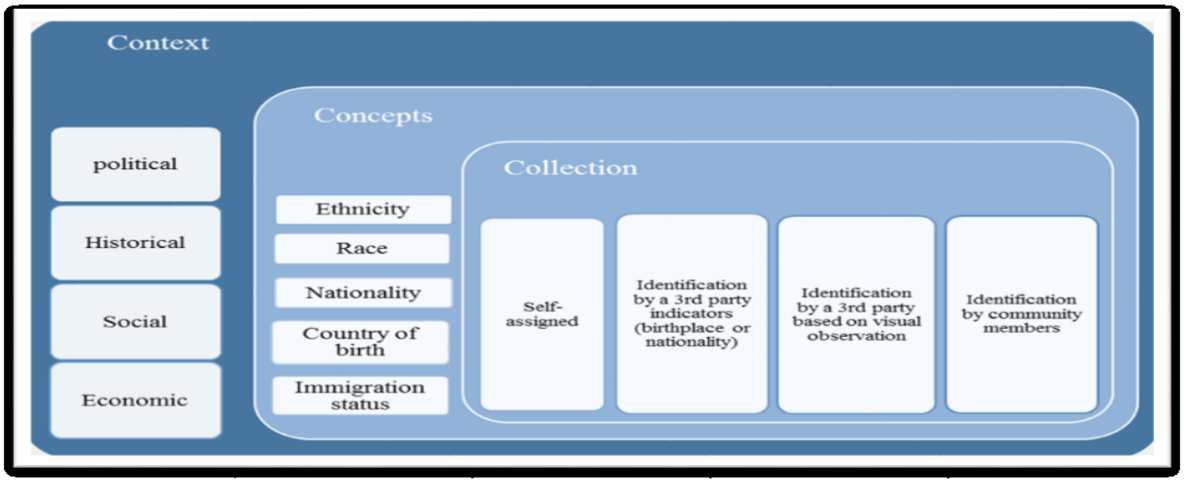
Political rights and legislation

Within the EU, government bodies will have to respond to a variety of civil rights movements and recognize active ethnic groups within their societies. For example, there may be national legislative

processes, involving community consultation, which leads to the development of official lists of ethnic groups that are recognized (Jenkins 2000, Morning 2008). During the development or revision of such lists, community groups may lobby for inclusion and may or may not achieve, or fulfil criteria for, recognition. For example, the Kashmiri group, who undertook considerable lobbying but after consideration by the Office of National Statistics (ONS) in the UK were not included as a discrete category for the 2011 census (Statistics 2009, Aspinall 2013). These legislative processes enable ethnic groups to mobilise and influence decision-making.

However, there is also legislation which deters the collection of ethnicity data at all, and particularly at a granular level. EU data-protection laws on monitoring and use of sensitive personal data, including ethnic identity and religion, have been interpreted in some countries as legally prohibiting the collection of data, but in others not so (Krizsán 2001). This contributes to the lack of data we find in mainland Europe, with collection noticeably absent in some EU-28 countries such as Germany, the Netherlands, Austria and Luxembourg, among others. Some trends are shifting, however, and in the 2011 census we observed that some countries, such as Germany, included items such as religion and immigrant ancestry. Regulations regarding sensitive data protection in Europe are contained in the 1981 Council of Europe Convention No. 108 (protection of electronic processing of personal data) and in the EU Directive 95/46/EC (protection of free movement of personal data) (Simon and l'intolérance 2007). These decrees do, however, contain a clause which enables countries to collect sensitive data to address equal rights initiatives and, as such, permit all European countries to collect granular ethnicity data for these purposes (Ringelheim 2011).

Figure 2. Contextual considerations for developing ethnic racial classifications[3]



[3] Ethnicity incorporates all the elements as defined in our conceptual background

Historical and social influences

A country's ideology towards the expression of cultural or national identities may fundamentally influence the acceptability of developing and using ethnic classifications. For example, ethnic classification in France has been influenced by its conception of 'nation' as a unified civic core. This means dividing the population between those who were inherent to the nation and those who were not (foreigners) (Blum 2002). The legacy of distinct historical events also has a profound impact. Historically, German ethnic categorisation was based on language (Kertzer and Arel 2001). However, Labbé shows that this country *"…introduced racial categorisations based on documented lineage – in the 1939 census, respondents had to indicate whether one of their grandparents was Jewish – it constituted a break not only from the German census tradition of categorizing identity by self-professed language, but also from the entire European census practice of rejecting race as a category…"* (Labbé 1998). The implications of such historical events, and documented abuses of gathering ethnicity data, engender historical sensitivity within the ethnicity question, which persist to this day and propagate a reluctance to request and collect such information (Morning 2008, Ringelheim 2011). The result of this is that countries may utilise proxy data which are perceived as more acceptable to collect (see table 2.5), or have an ethnicity question but not require it to be answered.

Migration patterns

Most countries collected CoB, nationality and citizenship either as proxy measures for ethnicity or in addition to ethnicity. The collection of this information can be informative to understand migration flows and the formation of multi-ethnic societies. This information can also assist in identifying generation of migration (e.g. second generation of migrants living in a country) which, in relation to health, can influence health behaviours and risks of disease.

Data may assist with the integration of new migrant and ethnic minority groups into European society, in particular in understanding the individual needs of these heterogeneous ethnic groups, such as healthcare, education, political, social, economic and cultural requirements. The recent influx of new migrant groups into Europe may create new challenges for collecting official statistical data (Aspinall 2009). On the one hand, this would be an ideal occasion to improve ethnic coding, as more detailed information will be needed, but on the other hand, the social and political reaction to this situation may create an atmosphere which is not conducive to people declaring such data.

The exclusively free text option

In Latvia, Romania and the Czech Republic (see table 2.4.), individuals identified their ethnic group in a free text option rather than using predefined ethnic categories. This creates the opportunity of collecting granular data. However, a free text option for ethnic origin can create difficulties for data interpretation (Aspinall 1998). Thus, this method will only be valid if granularity is maintained when data are analysed, reported and used to inform health, health care and policy for these population groups.

Strengths and weaknesses of this overview

The strengths of the HGEC work include the breadth of the overview of official data sources and the description of granularity of ethnic group classifications across EU-28 countries. There are variations among ethnic categorization and important differences in comparisons of diverse ethnic groups have been confirmed. Ethno-religious groups are rarely described in EU studies, but this work has provided descriptive information, of three countries: The Republic of Cyprus, Poland and Slovakia.

Information from sources outside the census and population registers (e.g. health surveys etc.) would have augmented the findings, but was too complex to comprehensively access for the number of countries covered and was therefore outwith the scope of the current work. Furthermore, there may be some granular approaches to ethnicity classifications in parts of Europe outside EU-28.

Conclusions from the overview

There is clearly a need for more consistency and a more granular approach to the collection of ethnic group data within Europe. Some census countries revealed considerable heterogeneity within their classifications of ethnic groups, but others do not collect ethnicity at all. Sometimes there are potential proxies for ethnicity but they are not collected for every country. The diversity of approaches reflect multiple influences including political rights and legislation; historical events, ideology and sensitivity towards cultural identity; and ongoing migration patterns. For this reason, a unified approach may not be possible and this will continue to create obstacles for international comparisons.

We will now look at selected countries within Europe and internationally, to understand how their granular classification systems have developed and what lessons they offer.

References

Alesina, A. and E. L. Ferrara (2005). "Ethnic diversity and economic performance." Journal of economic literature **43**(3): 762-800.

Aspinall, P. and M. Song (2013). Mixed race identities, Springer.

Aspinall, P. J. (1998). "Describing the "white" ethnic group and its composition in medical research." Social Science & Medicine **47**(11): 1797-1808.

Aspinall, P. J. (2003). "The conceptualisation and categorisation of mixed race/ethnicity in Britain and North America: identity options and the role of the state." International Journal of Intercultural Relations **27**(3): 269-296.

Aspinall, P. J. (2008). ""Non-white": a candidate for the lexical room 101." Journal of epidemiology and community health **62**(10): 875-875.

Aspinall, P. J. (2009). "The future of ethnicity classifications." Journal of Ethnic and Migration Studies **35**(9): 1417-1435.

Aspinall, P. J. (2012). "Answer formats in British census and survey ethnicity questions: does open response better capture 'Superdiversity'?" Sociology **46**(2): 354-364.

Aspinall, P. J. (2013). "Do the 'Asian'categories in the British censuses adequately capture the Indian sub-continent diaspora population?" South Asian Diaspora **5**(2): 179-195.

Bhopal, R. (2004). "Glossary of terms relating to ethnicity and race: for reflection and debate." J Epidemiol Community Health **58**(6): 441-445.

Bhopal, R. (2006). "Race and ethnicity: responsible use from epidemiological and public health perspectives." The Journal of Law, Medicine & Ethics **34**(3): 500-507.

Bhopal, R. S. (2013). Migration, ethnicity, race, and health in multicultural societies, OUP Oxford.

Blum, A. (2002). "Resistance to identity categorization in France." Census and identity: The politics of race, ethnicity, and language in national censuses: 121-147.

Brubaker, R. (2004). Ethnicity without groups, Harvard University Press.

Cieslik, A. and M. Verkuyten (2006). "National, ethnic and religious identities: Hybridity and the case of the Polish Tatars." National Identities **8**(2): 77-93.

Holst, F. (2012). Ethnicization and identity construction in Malaysia, Routledge.

Hunt, S. M. and R. Bhopal (2004). "Self report in clinical and epidemiological studies with non-English speakers: the challenge of language and culture." Journal of Epidemiology and Community Health **58**(7): 618-622.

Ingleby, D. (2011). Adapting European health systems to diversity. Brussels, COST European Cooperation in Science and Technology.

Jacobs, D. and A. Rea (2012). ""Allochthones" in the Netherlands and Belgium." International migration **50**(6): 42-57.

Jenkins, L. (2000). Separate Tables: The Census and Social Classifications. Census Workshop, Watson Institute, Brown University.

Kertzer, D. I. and D. Arel (2001).

Koller, I. Z. (2014). Ethnic Minorities and Censuses. Minority politics within the Europe of regions. H. István and T. Márton, Editura ISPMN**:** 319-333.

Krizsán, A. (2001). Ethnic monitoring and data protection: The European context, Central European University Press.

Labbé, M. (1998). "«Race» et «Nationalité» dans les recensements du Troisième Reich. De l'auto-déclaration au diagnostic racial." Histoire & mesure: 195-223.

Law, J. (2015). A dictionary of law, OUP Oxford.

Masseria, C., P. Mladovsky and C. Hernández-Quevedo (2010). "The socio-economic determinants of the health status of Roma in comparison with non-Roma in Bulgaria, Hungary and Romania." The European journal of public health **20**(5): 549-554.

McFadden, B., D. R. Nerenz and C. Ulmer (2009). Race, Ethnicity, and Language Data:: Standardization for Health Care Quality Improvement, National Academies Press.

McKee, M. and M. J. Galsworthy (2016). "Brexit: a confused concept that threatens public health." Journal of Public Health **38**(1): 3-5.

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.

Nielsen, S. S., A. Krasnik and A. Rosano (2009). "Registry data for cross-country comparisons of migrants' healthcare utilization in the EU: a survey study of availability and content." BMC Health Services Research **9**(1): 1.

O'Hearn, C. C. (2008). Half and half: Writers on growing up biracial and bicultural, Pantheon.

Omi, M. A. (2001). "The changing meaning of race." America becoming: Racial trends and their consequences **1**: 243-263.

Parekh, N. and T. Rose (2011). "Health inequalities of the Roma in Europe: a literature review." Central European journal of public health **19**(3): 139.

Parsons, C., R. Godfrey, G. Annan, J. Cornwall, M. Dussart, S. Hepburn, K. Howlett and V. Wennerstrom (2004). Minority ethnic exclusions and the race relations (amendment) Act 2000, DfES Publ.

Rafnsson, S. B. and R. S. Bhopal (2009). "Large-scale epidemiological data on cardiovascular diseases and diabetes in migrant and ethnic minority groups in Europe." The European Journal of Public Health **19**(5): 484-491.

Ringelheim, J. (2011). "Ethnic categories and European human rights law." Ethnic and Racial Studies **34**(10): 1682-1696.

Rodríguez-García, D. (2010). "Beyond Assimilation and Multiculturalism: A Critical Review of the Debate on Managing Diversity." Journal of International Migration and Integration / Revue de l'integration et de la migration internationale **11**(3): 251-271.

Scott, J. and G. Marshall (2009). A dictionary of sociology, Oxford University Press, USA.

Simon, P. (2004). "Comparative Study on the collection of data to measure the extent and impact of discrimination within the United States, Canada, Australia, Great-Britain and the Netherlands." Luxembourg: Office for Official Publications of the European Communities.

Simon, P. (2012). "Collecting ethnic statistics in Europe: a review." Ethnic and Racial Studies **35**(8): 1366-1391.

Simon, P. and C. e. c. l. r. e. l'intolérance (2007). " Ethnic" statistics and data protection in the Council of Europe countries: study report, Council of Europe Strasbourg.

Statistics, O. f. N. (2003). Ethnic Group Statistics: A Guide for the Collection and Classification of Ethnicity Data, The Stationery Office London.

Statistics, O. f. N. (2009). Kashmiri Research Project: Final Report. C. Programme.

Vertovec, S. (2006). The emergence of super-diversity in Britain (centre on Migration, policy and Society, Working Paper No. 25), Oxford: University of Oxford.

Volpato, C., F. Durante, A. Gabbiadini, L. Andrighetto and S. Mari (2010). "Picturing the other: Targets of delegitimization across time." International journal of conflict and violence **4**(2): 270.

Appendix 2

Protocol project RWJF

Heterogeneity/Granularity in ethnicity classifications (HGEC) outside the U.S

Professor Raj Bhopal (P.I.)

Dr. Emma Davidson (Co-P.I.)

Dr. Nazmy Villarroel (Postdoctoral Research fellow)

Dr. Pamela Pereyra-Zamora (Investigator-Bolivia)

Dr. Peter Aspinall (Investigator-United Kingdom)

Dr. Liv Stubbe and Prof. Allan Krasnik (Investigators-Denmark)

Dr. Inez Koller (Investigator-Hungary)

Dr. Donna Cormack and Dr. Tahu Kukutai (Investigators-New Zealand)

Prof. Hude Quan, Dr. Kelsey Lucyk and Dr. Karen Tang (Investigators-Canada)

Dr. Shyamala Nagaraj and Dr. Chiu Wan Ng (Investigators-Malaysia)

PROTOCOL PROJECT

Heterogeneity/Granularity in ethnicity classifications (HGEC) outside the U.S

Table of contents

1. Background to the Robert Wood Johnson Foundation (RWJF) project

Current ethnic classifications and the U.S. perspective leading to this project

Current racial and ethnic classifications, mostly based on the census questions, tend to be more suitable for social and planning purposes than for scientific ones. For example, in the United States official statistics are often published in six race/ethnicity categories – White, Hispanic or Latino, Black or African Americans, Asian, Native Hawaiian or other Pacific Islander, Native American and Alaska Natives. These are extremely broad categories, especially Asian, which comprises the countries east of Istanbul in Turkey through Japan, so the *Asian Label* does encapsulate a large and heterogeneous population.[1,2] One of the biggest obstacles to scientific work is this heterogeneity within the population described by current categories. For example, ethnic disparities in health (status or outcomes) can exist within, and be masked by, broad categories; therefore, there is value in collecting and utilizing data incorporating a more fine-grained approach.[3] In the U.S., race has historically been the preferred concept. However, recent findings also indicate that utilizing a more detailed measure that conceptually separates race and ethnicity, and that provides respondents with adequate flexibility to identify themselves both racially and ethnically, decreases missing data and misclassification and, as such, may increase validity.[4,5]

Granular ethnicity categories (heterogeneity within broad categories)

Obtaining and utilising high-quality race and ethnicity data are critically important in our efforts to understand and eliminate racial and ethnic health disparities and therefore, in the U.S., there is momentum to explore the use of more granular ethnic classifications; classifications which recognise the heterogeneity within traditional broad categories.[6] Moreover, increasing demands of inclusiveness and identity visibility require that classifications are headed in the direction of greater complexity.[7]
For example, the Subcommittee on *Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement* recommends a separate question to collect data on granular ethnicity—defined as "a person's ethnic origin or descent, 'roots,' or heritage, or the place of birth of the person or the person's parents or ancestors…"[8] as the concept to adopt for healthcare quality improvement. Hence, granular ethnicity is this more fine-grained ethnic subgroup data (e.g., Vietnamese, Haitian, Cape Verdean) .[6] The idea of granularity is U.S. driven but echoed in countries where heterogeneity within ethnic groups has been emphasized.

Public health researchers, practitioners and health care organizations must determine an approach to collecting granular ethnicity data that allows all individuals, if they desire, to self-identify and that at the same time is feasible, given that the population of their service area may include hundreds of granular ethnicities. Self-identification enables entities to learn about the composition of their service population so they can decide which ethnicity categories will yield the most appropriate information on

which to base decisions on service provision, and where to target interventions.[9] "Being able to focus interventions at the more granular level has been posited as a way to use resources most efficiently to reduce disparities". [6,10] Additionally, such individualized data collection has the potential benefit of preserving small subgroup identities that might be of interest for analytic studies (assuming preservation of the specific identifiers during data transfer) at the state, health plan, or national level but that might prove too small to reveal any group-specific quality issues at the local level (e.g., higher cancer mortality among persons of Samoan descent). [9]

Challenges to collecting/analysing and reporting granular ethnicity data: mixed race/ ethnicity populations as an example

Despite the clear rationale for collecting more granular ethnicity data, there are few examples of this approach being undertaken internationally and many challenges to arriving at an agreed framework for how ethnicity data should ideally be collected, analysed and reported. The terminology supporting both the concepts and classification is itself problematic yet, despite the difficulties of the task, progress towards an internationally agreed vocabulary is prerequisite for progress. [2]

An additional complexity is presented by the category of 'mixed ethnicity'. The USA ethnic lobbies have played a stronger part in determining the census's ethno-racial categories than in Europe. In the lead-up to the 2000 Census, they were responsible for obtaining the enumeration of the mixed race ('two or more races') population. [7]

However, some researchers have argued that the main challenge to the sustainability of ethnicity classifications will be the 'mixed' group, as increasing rates of inter-ethnic union formation and their offspring challenge classifications based on the option to tick one box only. [7] The total self-reported mixed race/mixed ethnicity group remains small though growing: 1.2 million persons of 'mixed/ multiple' ethnicity in England and Wales in 2011, or 2.2% of the total population; nine million persons of 'two or more races' in the USA in 2010, or 2.9% of the total population.[11,12] Combinations of racial and ethnic status are necessarily heterogeneous. Also, the development of terminology and categories suitable for use with mixed populations is challenging and there are no inherently 'correct' terms. The body of existing evidence on mixed populations' health remains scattered and small, impeding the creation of well-informed public health policies. These challenges need to be overcome since the limited evidence suggests mixed race/ethnicity populations have distinct health needs. [11] For example, the Scottish Health and Ethnicity Linkage Study (SHELS) shows high relative risks of lung cancer among mixed populations. [13]

Components of the U.S. Project

As the Robert Wood Johnson Foundation has identified, ethnic classifications tend to be insufficiently developed to fulfil their potential to advance our understanding of the impact of ethnicity on health, or to adequately inform the development of equitable healthcare policy and practice. In order to address

some of these challenges and identify how data should ideally be collected, analyzed, and reported, the RWJF has commissioned an analysis of the disaggregation of health data by racial/ethnic groups. US project components aims to answer the following four overarching questions:

1. How do some major surveys collect, analyze, or report data for ethnic/racial groups that go beyond those five categories?

2. Are there any lessons that can be learned from outside the health field?

3. Why are disaggregated data not being collected, analyzed, or reported more often if the field generally agrees that this is critical to understanding disparities?

4. Ideally, how should data be analyzed and reported given the health outcomes that the Foundation is interested in?

The project contributes to the RWJF's vision towards a Culture of Health (COH) that 'enables all members of our diverse society to lead healthy lives, now and for generations to come'.

The RWJF also recognises that, within the global setting, countries vary greatly in the ethnic/racial composition of their populations and in their histories of migration. Policies and practice for the collection and analysis of data for ethnic/racial groups internationally, have developed in heterogeneous ways and range from no systematic collection of data (particularly in developing nations), the collection of solely country of birth (e.g. The Netherlands), to a more fine-grained approach utilizing disaggregated ethnicity data and country of birth (e.g. United Kingdom). Consequently, international collaboration and debate around the terminology and systems of classification is both desirable and has the potential to be very beneficial. The RWJF has, therefore, commissioned a project outside of the U.S. to complement their work and provide international perspectives to these four questions.

2. Aim and Objectives of the non-U.S. (HGEC) project

General aim:

To explore approaches and lessons from selected countries with exemplary models, outside the U.S., related to collecting, analysing, and reporting data disaggregated, granular ethnic classifications; to inform the U.S. project and to be discussed and refined in collaboration with our U.S. colleagues.

Specific objectives:

1. To identify up to seven countries, outside the U.S., that collect, analyse, and report data for racial and ethnic groups which go beyond broad categorizations, and to assess the degree of heterogeneity/granularity of their ethnic classifications (e.g. census and population registers, national health services data, national health surveys).

2. To seek to understand how and why heterogeneity/granularity in these classifications has developed in terms of the social, historical and political context in our selected countries.

3. To understand why disaggregated data is not being collected, analysed, or reported more often if the field generally agrees that this is critical to understanding disparities.

4. To identify some examples, amongst our selected countries, of how disaggregated data has been used, and the impact it has had on policies, programmes, and population health outcomes.

5. To asses if there are lessons relating to heterogeneity/granularity of ethnic and racial classifications that can be learnt from outside the health field.

6. To synthesize lessons and develop findings to inform U.S. and international efforts.

7. To explore global lessons for how data should, ideally, be analysed and reported given these heterogeneity/granularity ethnic group classifications, to complement the U.S. project.

3. Components and methods of the non-US (HGEC) project

We propose four components:

1. Overview of data sources within European Union Countries and selected countries outside Europe

2. In-depth analysis of selected countries (country reports)

3. Investigator's meetings

4. Synthesis of findings

3.1. Overview of data sources within European Union Countries and selected countries outside Europe

3.1.1 Research questions for Overview of data sources

1. What sources of data are there within European Union Countries, and selected countries outside Europe, about the ethnic/racial composition of their population? (e.g., census and population registers, national health services data, and national health surveys)

2. From the sources of data identified, which countries collect, analyse, and report data for racial and ethnic groups with attention to actual or potential heterogeneity/granularity.

3.1.2 Methods for overview of data sources

Literature review

A literature review (e.g. scoping review) will be carried out to identify information reporting the current situation of ethnic/racial data collection within the European Union and selected countries outside Europe. Much work has already been done e.g. MEHO and ADAPT projects.

Databases

Systematic searches will be carried out on government websites and other sources, including identification of grey literature

Selection of relevant literature

The HGEC Research Fellow (NV) will screen the literature and select relevant papers utilising the following criteria:

- reports/papers/data sources of European Union countries and selected countries outside Europe
- Identify sources of data on ethnic/racial composition of country
- Adequately describe classifications used in these data sources

Search strategy:

Literature data bases: (ethnic* or race) AND (classif* or categor*) Databases:

ASSIA (Applied Social Sciences Index and Abstracts) IBSS

(International Bibliography of the Social Sciences) Race Relations

Abstracts

SocINDEX with Full Text

International Micro-data bases:

IPUMS

Data without borders

International Meta-data bases:

Eurostat

Data Archives:

CESDA

UK-data archive

Surveys:

European Social Survey

Official statistical sources:

United Nations Statistical Division (UNSTAT)

Office for National Statistics (ONS)

Consultation with investigators and other relevant individuals and organizations:

We have consolidate a group of investigators internationally which to date includes:

- Dr. Peter Aspinall from the United Kingdom
- Professor Allan Krasnik and Dr. Liv Stubbe Østergaard from Denmark
- Dr. Inez Koller from Hungary
- Dr. Donna Cormack and Dr. Tahu Kukutai from New Zealand
- Professor Hude Quan, Dr. Kelsey Lucyk and Dr. Karen Tang from Canada
- Dr. Pamela Pereyra-Zamora from Bolivia
- Dr. Shyamala Nagaraj and Dr. Chiu Wan Ng from Malaysia

Information identified in the HGEC project by the investigators group will be complementing the literature review. In addition, we will organize, telephone interviews and skype calls with the following umbrella organizations in order to seek advice, such as:

- ADAPT members
- The Ethnic Health List led by M.Johnson
- The Migrant and Health Section from EUPHA

This information will be used to supplement the literature review on ethnic/racial data collection within Europe and outside the European setting.

Report on overview of data sources: timescales

An initial draft report will be prepared on the overview of data sources within European Union Countries and selected countries outside Europe for the US project team for the end of February 2016.

3.2 In-depth analysis of selected countries (country reports)

3.2.1. Research questions for in-depth analysis of selected countries

We envisage country reports of about four pages i.e. about 1600 words (with appropriate tables) covering:

1. Briefly, what is the demographic background of this country in terms of the development of its ethnic/racial composition, its history of migration and how it became an increasingly multi-ethnic society?

2. In detail, what sources of data are available in each country about the ethnic/racial composition of their population, especially inside the health field , and what level of heterogeneity/granularity is used in their collection, analysis and reporting?

3. Are there any lessons to be learned from outside the health field relating to the heterogeneity/granularity of ethnic classifications?

4. How has the heterogeneity/granularity in these classifications developed historically in these countries? (It may be that such development are only for some groups e.g. in the UK, we have granularity for White groups but not for Chinese, in the Netherlands the same applies to Surinamese (Hindu/Creole), while in many countries attention is on indigenous populations)

5. Why has the heterogeneity/granularity in these classifications developed in these countries in terms of their social, historical and political context?

6. Why disaggregated data are not being collected, analysed, or reported more often if the field generally agrees that this is critical to understanding disparities?

7. Are there any examples in these countries of how the disaggregated data (collected, analysed and reported) has been used to impact on policies, programmes, and population health outcomes.

8. Are these ongoing developments (policies, movements, lobbies) in these countries to alter/improve their current systems of classification?

3.2.2 Methods for in-depth analysis of selected countries

These questions will be addressed by the Investigators group examining each of the country reports,

with the assistance of the Edinburgh-based team. We will meet by telephone and skype calls during this process. The information gathered from all country reports will then be synthesised for sharing with our U.S. colleagues.

3.3 Investigators group meeting(s)

The investigators group meetings will be convened for participatory analysis of all information gathered in the scoping review and country reports, and to identify common principles and generalizable examples of good practice in relation to each country's approach to the collection, analysis and reporting or ethnic/racial data.

3.3.1 Research questions for Investigators group meeting(s)

1. What findings/lessons are there from the information gathered which have the potential to inform U.S. and international efforts?

2. What are the global lessons emerging for how data should, ideally, be analysed and reported?

3.3.2 Methods for Investigators group meeting(s)

Meetings

Meetings of the Investigators group will commence during the literature review (e.g. scoping review) which will be conducted with a collective effort of our investigators.

Subsequently, following completion of the country reports, at least one meeting will be held in Edinburgh with all investigators participating either in person (European investigators) or by teleconference (e.g. NZ/Canada/Latin America).

-Methods will be developed further.

Report following Investigators group meeting(s)

A draft report on the in-depth analysis will be prepared for the US project team for the end March 2016.

3.4 Synthesis of material

Following the submission of the draft reports, there will be consultation between our investigators group and the US team to discuss and distil any global lessons arising from the entire project findings, including at least one meeting in the U.S. which the HGEC PI (Professor Raj Bhopal) or deputy(ies) will attend.

The results from all preceding project components, including findings arising from the US consultation, will be synthesised into the final report. It is our intention that the investigators group will publish academic articles from this work.

4. Timelines

The following timelines have been developed in order to work in parallel with the US project:

Timeline	2015			2016								
	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Stage 1 Overview												
European												
Other countries												
Stage 2 In-depth analysis												
Stage 3 Draft Reports												
Overview												
In-depth report												
Stage 4 Consultation												
Meet US team												
Draft final report												
Complete report												

5. References

1. Bhopal et al. Heterogeneity of coronary heart disease risk factors in Indian, Pakistani, Bangladeshi, and European origin populations: cross sectional study. BMJ. 1999 Jul 24;319(7204):215-20.

2. Bhopal RS. Migration, Ethnicity, Race, and Health in Multicultural Societies. Second Edi. United Kingdom: Oxford University Press; 2014.

3. IOM (Institute of Medicine). Race, ethnicity, and Language data Standardization for Health Care Quality Improvement. Washington, DC: The National Academies Press.; 2009.

4.Valles SA. Heterogeneity of risk within racial groups, a challenge for public health programs. Prev Med. 2012 Nov;55(5):405-8. doi: 10.1016/j.ypmed.2012.08.022. Epub 2012 Sep 8.

5. Eisenhower A, Suyemoto K, Lucchese F, Canenguez K. "Which Box Should I Check?": Examining Standard Check Box Approaches to Measuring Race and Ethnicity. Health Serv Res [Internet]. 2014 Jun 3 [cited 2015 Apr 4];49(3):1034–55. Available from: http://doi.wiley.com/10.1111/1475-6773.12132.

6. Webster PS, Sampangi S. Did We Have an Impact ? Changes in Racial and Ethnic Composition of Patient Populations Following Implementation of a Pilot Program. 2014;00(0):1–9.

7. Aspinall PJ. The Future of Ethnicity Classifications. J Ethn Migr Stud [Internet]. 2009 Nov [cited 2015 Apr 4];35(9):1417–35. Available from: http://www.tandfonline.com/doi/abs/10.1080/13691830903125901.

8. U.S. Bureau of the Census. U.S. Bureau of the Census [Internet]. 2008. Available from: www.census.gov/population/ancestry/

9. MD A for HR and Q (AHRQ) R. Chapter 3: Defining Categorization Needs for Race and Ethnicity Data Race, Ethnicity, and Language Data: Standardization for Health Care Qu Need for Locally Relevant Granular Ethnicity Categories [Internet]. Available from: http://www.ahrq.gov/research/findings/final- reports/iomracereport/reldata3a.html.

10. Agency for Healthcare Research and Quality (AHRQ). Evidence of disparities among ethnicity groups: Race, ethnicity, and language data: Standardization for Healthcare [Internet]. 2010. Available from: www.ahrq.gov/research/findings/finalreports/

11. Valles S a., Bhopal RS, Aspinall PJ. Census categories for mixed race and mixed ethnicity: impacts on data collection and analysis in the US, UK and NZ. Public Health [Internet]. The Royal Society for Public Health; 2015 Mar [cited 2015 Apr 4];129(3):266–70. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0033350614003552

12. Aspinall PJ, Song M. Mixed race identities. Basingstoke: Palgrave Macmillan; 2013.

13. Bhopal RS, Bansal N, Steiner M, Brewster DH. Does the "Scottish effect" apply to all ethnic groups? All-cancer, lung, colorectal, breast and prostate cancer in the Scottish Health and Ethnicity Linkage Cohort Study. BMJ Open. 2012. p. e001957– e001957.

6. Appendix

HGEC research team

Core investigators team	**Role**
Raj Bhopal	Principal investigator (PI)
Emma Davidson	Co-Principal investigator (Co-Pi)
Nazmy Villarroel	Research fellow (RF)

International Expert Researchers **Role:** to deliver the country report	**Assisting International Expert Researchers*** **Role:** assisting in delivering the country report	**Countries**
Peter Aspinall		United Kingdom
Liv Stubbe	Allan Krasnik	Denmark
Inez Koller		Hungary
Hude Quan	Kelsey Lucyk and Karen Tang	Canada
Donna Cormack	Tahu Kukutai	New Zealand
Pamela Pereyra		Bolivia
Shyamala Nagaraj	Chiu-Wan Ng	Malaysia
**Please complete your assisting collaborator if needed*		

Overview advisers **Role:** to provide broad advice and comment on draft reports
Allan Krasnik
Raj Bhopal (PI); Emma Davidson (Co-Pi) and Nazmy Villarroel (RF)
Pamela Pereyra-Zamora

Chapter 3: Ethnic group classification in Aotearoa New Zealand

Donna Cormack, Tahu Kukutai

Abstract

Aotearoa New Zealand has long-standing and embedded practices of ethnic enumeration, although these have shifted over time with changes to broader political and social contexts. Early approaches following colonisation reflected assimiliationist policies towards Māori, New Zealand's indigenous peoples, and state interest in delineating access to resources and rights. Early censuses asked about country of birth, introducing a question about 'race' in 1916. Official approaches to ethnicity shifted over time to a 'degrees of blood' conceptualisation, then to self-identified ethnic affiliation in the 1990s. In New Zealand, Statistics New Zealand is the agency responsible for the official standard for ethnicity that outlines the official definition, standard ethnicity question, and classification system. The current standard was released in 2005 and applies to all-of-government. Administrative and survey collections routinely collect ethnicity data, including the population census, vital registrations, official surveys, and many administrative collections in education, justice, health and other sectors. It has been compulsory to collect ethnicity data in the health sector since the 1990s. Although issues with quality have been documented, ethnicity data is included in key health sector collections and are routinely used for monitoring, planning, and funding purposes.

The official approach to ethnicity data in New Zealand supports granularity in that the standard question allows people to self-identify with multiple groups and to provide write-in responses. The official classification system has four levels, from least to most detailed, with more than 230 ethnic categories at the most detailed level. In practice, however, granularity is often restricted in approaches to data collection, recording and output. Many systems do not collect or record ethnicity data at the most detailed level or do not capture all ethnicities reported by an individual. Data are often aggregated for analysis and reporting, with official data routinely reported for broad ethnic groupings (e.g. European, Pacific, Asian).

The health sector has comprehensive coverage across administrative and survey collections, that is critical for the measurement and monitoring of ethnicity and ethnic health inequities. However, disaggregation at levels of detail greater than broad ethnic groups remains relatively uncommon. This limits understanding of communities' diverse realities and priorities, and potentially also masks health need and ethnic health inequities. Changes to ethnicity data systems in the future, including proposed changes within the health sector to increased disaggregation in recording ethnicity data, may contribute to improved granularity of over time.

Introduction

Aotearoa New Zealand (NZ) has a long tradition of collecting ethnicity data. The first official census of the indigenous Māori people was undertaken in 1857 (Fenton 1859), less than two decades after the 1840 Treaty of Waitangi was signed between Māori chiefs and agents of Queen Victoria.[1] Since then, state practices of ethnic enumeration have changed many times, reflecting broader shifts in the country's political and social context. This report summarises key features of ethnic enumeration in Aotearoa NZ, with a particular focus on the collection, use and dissemination of ethnicity data in relation to health. First a brief history of the country's ethnic context and ethnic classification system is provided.

Colonial and ethnic context

As in the other CANZSUS settler states (Canada, NZ, Australia, United States) (Ford 2012), the colonisation of Aotearoa NZ involved the social, political, cultural and demographic domination of the precursor peoples. In 1840 the ratio of settlers to Māori was about one to 40; by 1874 Māori were outnumbered ten to one. This dramatic reversal was largely due to rapid migration from Britain but also reflected increased Māori mortality as a consequence of colonisation (Pool and Kukutai 2014). Māori began to recover demographically from the early 1900s but their health outcomes were very poor by comparison with the health of the settler descent population (Pool 1991, Woodward and Blakely 2014) and substantial health inequities persist to the present day (Ministry of Health 2015).[2]

Mass migration from the United Kingdom, coupled with immigration policies excluding or restricting the entry of people from outside the British Commonwealth, ensured that non-Māori Aotearoa NZ remained overwhelmingly white. Although Chinese and Indian peoples have had a long-standing presence in Aotearoa NZ, they were few in number and subject to multiple forms of racial discrimination (Ip 2003, Li 2007/2008). From the 1970s the ethnic composition of Aotearoa NZ began to change rapidly, first with increasing numbers of migrants from around the Pacific (e.g. Samoans, Tongans), then with sweeping changes to migration policies which removed source country preferences. In the 2013 census, 25 per cent of the total population was foreign-born, higher than either the United States (13 per cent) or Canada (21 per cent), with the majority of migrants originating from countries outside the Commonwealth (Statistics New Zealand 2014). The country's largest city, Auckland, is now considered to be *super-diverse*, with more than 40 per cent of the population born overseas (Spoonley 2015).

[1] The Treaty of Waitangi set out an agreement for engagement between Māori as tangata whenua and the Crown. Contestations have arisen due to differences between the meanings of the English and Māori versions, as well as repeated breaches of The Treaty of Waitangi by the Crown (for further reading, see for example: Kingi TK. The Treaty of Waitangi: a framework for Māori health development. New Zealand Journal of Occupational Therapy. 2007;54(1):4-10, Orange C. The Treaty of Waitangi. Wellington: Bridget Williams Books; 1987.

[2] There is some recent narrowing of the gap in life expectancy, although progress is slow (Woodward A, Blakely T. The healthy country? A history of life and death in New Zealand. Auckland: Auckland University Press; 2014).

In terms of ethnic composition, people identifying with European ethnic groups still constitute the majority in Aotearoa NZ, but their dominance has declined from 91 per cent in 1981 (Didham and Bedford 2004), to 74 per cent in 2013 (Statistics New Zealand 2014). In 2013 Māori comprised just under 15 per cent, with Asian and Pacific ethnic groups at 11.8 and 7.4 per cent respectively. Those identifying with Middle Eastern, Latin American and/or African ethnicities were a very small proportion, at 1.2 per cent, while *Other* ethnicities represented 1.7 per cent of the population (Statistics New Zealand 2014).[3] The combined ethnic shares exceed 100 per cent, a function of the total count methodology allowing individuals to report and be counted in multiple ethnic groups.

The history of ethnic enumeration

Early census reports in Aotearoa NZ reflected the state's assimilationist approach to Māori incorporation and its interest in determining access to, or exclusion from, rights and resources. Māori were counted in a separate census until 1956, after which time it was deemed they had reached a sufficient level of development to be included with the rest of the population (Kukutai 2012). For the general (i.e. non-Māori) population, early census questions asked about country of birth, with a *race* question introduced in 1916. This, and subsequent questions, reflected the prevailing racial logics underpinning notions of difference as well as the state's interest in monitoring potential 'problem' groups.[4] Over time the notion of race-based difference evolved into very detailed reporting and classification of the population by 'degree of blood'. From the 1970s there was a shift away from the language of race to the concept of ethnic origin although the reporting of blood fractions in the census persisted until 1981. The 1986 census was a watershed moment as it dispensed with fractional reporting and the reference to origins was dropped altogether in the 1991 census. The census question simply asked: *'Which ethnic group do you belong to? Tick the box or boxes that apply to you'.*

Although racial-ethnic data had long been part of official collections, the 1975 Statistics Act made it compulsory for ethnicity data to be collected for the whole population. In 1993, following a major review, the national statistical office Statistics New Zealand introduced an official ethnicity standard that included an official definition and a standard hierarchical classification system to categorise ethnic groups in official statistics (Department of Statistics 1993). The current official standard was released in 2005 and reviewed in 2009 (Statistics New Zealand 2005, Statistics New Zealand 2009). It contains the current official definition (see appendix 3.1), a standard question (see figure 3.1) and a hierarchical

[3] The majority of those people in the "Other" category were respondents who identified with "New Zealander" in their response to the ethnicity question (Statistics New Zealand. 2013 Census QuickStats about culture and identity [cited 2016 February 28]. Available from: http://www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-culture-identity.aspx).

[4] For examples of wording of historical Census ethnicity questions, see for example: Public Health Intelligence. Monitoring ethnic inequalities in health Wellington: Ministry of Health, 2001; Statistics New Zealand. 2006

classification of four levels (from least to most detailed) (see appendices 3.2–3.5).

Figure 3.1: Aotearoa NZ census ethnicity question 2013

Source: Statistics New Zealand. Ethnicity - questionnaire module [cited 2016 29 February]. Available from: http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity/questionnaire-module.aspx.

Key sources of ethnicity data

Ethnicity data are routinely collected in administrative and survey collections. The five-yearly census is the most detailed source of nationally representative ethnicity data. It provides the basis for ethnic population estimates and projections and the denominator used for calculating population rates. Aotearoa NZ is one of only a small number of countries worldwide that asks numerous ethnicity related questions in the census including country of birth (and time in NZ), languages spoken, Māori descent, and iwi/tribe (Statistics New Zealand 2013). The collection of Māori descent data is required under the Electoral Act for the purposes of defining Māori electoral boundaries (Kukutai 2004).

Ethnicity data are also embedded in vital registrations. Prior to 1995, ethnicity on birth and death certificates was only recorded for Māori and Pacific peoples using a 'degrees of blood' approach, resulting in substantial under-reporting of Māori and Pacific births and deaths, and a lack of data for other ethnic groups (Public Health Intelligence 2001, Ajwani, Blakely et al. 2004, Tan, Blakely et al. 2010). The alignment of vital registrations with the Statistics NZ ethnicity question since the mid-1990s led to significant improvements, and the standardisation of collection and recording approaches has helped to reduce numerator/denominator bias over time (Public Health Intelligence 2001, Ajwani, Blakely et al. 2004, Tan, Blakely et al. 2010).

Beyond the census and vital registrations, ethnicity data are also collected as part of routine survey programmes, and in administrative collections across the social sector including schools, the police and courts. While the official statistical standard is a whole-of-government standard, there is variable uptake

across different sectors and ongoing issues with standardisation and ethnicity data quality (Cormack 2010).

The collection of ethnicity data in the health sector has been mandatory since the 1990s. Ethnicity is recorded on the National Health Index[5] (Ministry of Health), and is collected (or derived) as part of national collections, public health surveillance, surveys, and administrative data collections (Cormack and Harris 2009, Cormack and McLeod 2010). Ethnicity data are used to monitor and address health inequities between population groups, to inform planning and the setting of health priorities, for funding allocation at local and national levels (Ministry of Health 2004, Statistics New Zealand 2009, Ministry of Health 2014), and as part of Treaty of Waitangi obligations and imperatives for Māori (Robson, Reid et al. 2001, Cormack and McLeod 2010). In 2004 *Ethnicity data protocols for the health and disability sector* were released and provide guidance on the minimum requirements expected of the health sector in relation to ethnicity data, which includes the use of the standard census ethnicity question (Ministry of Health 2004). Issues with data quality in the sector are well-documented (Kilgour and Keefe 1992, Harris, Robson et al. 1997, Bramley and Latimer 2007, Cormack and McLeod 2010), although coverage, standardisation and completeness of data have improved over time (Cormack and McLeod 2010).

Granularity of ethnicity data in the official statistics system

In Aotearoa NZ the focus has tended to be on the quality and completeness of ethnicity data, rather than the level of detail at which ethnicity is captured and analysed. Ethnic granularity can be expressed in a number of ways including via the categories used, the use of write-in responses, the ability to report multiple ethnic affiliations, and processes of data aggregation and disaggregation. While ethnicity data collection in Aotearoa NZ has become more granular over time, there remains a tendency to revert to the use of aggregated categories at the analysis and reporting stage (see next section).

The standard ethnicity question intended for use across the whole-of-government has eight ethnic group tick boxes (see figure 3.1) and also allows for write-in responses. Individual responses are then aggregated according to the four-tier classification standard. In the current standard, the most detailed level of the classification (level four) has over 230 separate categories (see appendices 3.2–3.5). Where appropriate, changes are made (e.g. categories added) to the classification as part of regular reviews of survey responses by Statistics NZ (Statistics New Zealand). Groups included in the current

[5] The National Health Index (NHI) holds information on individuals using health services in New Zealand, including ethnicity data. Each individual is assigned a unique NHI number, with coverage around 95% of the population (Ministry of Health. National Health Index; 2012 [updated 10 May 2012; cited 2016 28 February]. Available from: http://www.health.govt.nz/our-work/health-identity/national-health-index).

classification reflect historical systems (Statistics New Zealand 2009), response patterns to the question, as well as the need to be able to identify specific groups for particular legal or policy interests (Allan 2001). Māori are the only group visible at all levels of the classification due to their special status as tangata whenua (original peoples) and Treaty partners. The other level one options are pan-ethnic categories (e.g., Asian, European) rather than specific ethnic groups (Allan 2001, Statistics New Zealand 2014).

The ability to report multiple ethnic affiliation in the census and other key collections allows for greater detail on self-identified ethnic affiliation. Historically, multiple affiliations were captured as 'degrees of blood' or 'proportions of descent'. With the shift to a cultural affiliation concept, the official approach to multiple ethnicity has allowed for individuals to report multiple responses and for them to be counted in all groups reported (Kukutai and Callister 2009), with no requirement for individuals to prioritise ethnic group identification. Statistics NZ currently records up to six ethnic responses per individual, although this level of detail is not generally outputted. The proportion of the population that identifies with two or more major ethnic groups has increased over time, from 5 per cent in 1991 (Kukutai and Callister 2009) to 11.2 in the 2013 census (Statistics New Zealand 2013).[6] Although the census captures up to six responses, in reality very few people report more than two groups (less than 2 per cent in 2013). The share reporting multiple ethnicities is much higher among Māori, 53.5 per cent in 2013 (Statistics New Zealand 2014), reflecting historically high rates of intermarriage with European/Pākehā (Harré 1968). While the term 'half-caste' was historically used by Europeans to draw distinctions between Māori, particularly in the colonial era, Māori themselves do not use notions of blood quantum as a meaningful social distinction, a key exemplar being the absence of 'blood' in designating tribal enrolment status.

In contrast to other sectors, the health sector protocols outline expectations for the sector and require ethnicity to be recorded at a minimum of level two of the standard (21 ethnic categories, four residual codes) and that systems are able to record at least three ethnicities per respondent (Health 2009). In some other sectors, for example education, multiple ethnicity is able to be recorded for students (up to three) but a single prioritised ethnicity is often used for outcome reporting (Ministry of Education 2014). While prioritisation is no longer a recommended output method by Statistics New Zealand, with 'total' or 'single/combination' methods supported, it remains a common method of analysis and output in health and other social sectors (Ministry of Health 2004, Statistics New Zealand 2009).[7]

[6] Per cent of those giving a valid ethnic response.

[7] Prioritisation order is based on a hierarchy with Māori first and the numerically dominant population last (e.g, at level one prioritisation order is: Māori->Pacific->Asian->MELAA->Other->European). It remains relatively common in the health sector, e.g. in funding formulae and in some types of analysis where mutually-exclusive categories are preferred. Statistics NZ recommend use of either 'total response' output (where an individual identifying with more than one ethnicity is counted in each of the reported groups, in practice usually after

Barriers to ethnic granularity in the official statistics system

Compared to many other countries, the ethnic enumeration system in Aotearoa NZ is relatively well developed. However, there are ongoing issues with respect to granularity, mainly relating to procedures and conventions for analysis and reporting of ethnicity in standard outputs, rather than to the ethnicity question or classification (Statistics New Zealand 2009). Aggregation at the point of capture in some systems remains a barrier to ethnic granularity. In health, despite a longstanding requirement for use of a standard question, not all collections follow this recommendation, and some may restrict granularity at collection by modifying the question, limiting possible response options, or asking a non-standard question. Inconsistencies across the official statistics system in granularity of data collection also impact on the ability to compare disaggregated data between health and other collections.

A further key barrier is slippage between granularity at collection stages and aggregate recording and output of ethnicity data. In practice, even where granular data are collected, many systems do not record or report at a granular level. Much data routinely reported in official statistics are presented for broad level one ethnic groupings (i.e. European, Māori, Pacific, Asian, Middle Eastern/Latin American/African, Other). Some disaggregated data are available, for example, within Statistics New Zealand's online data tools (e.g.(Statistics New Zealand)), in outputs and reports that have a specific focus on ethnicity, and through customised data requests (though these tend to be more for counts, rather than rates due to issues with precision and stability). Calls have been made for increasing availability and reporting of more granular data, for example, for Pacific and Asian ethnic groups, to take account of heterogeneity within these broad categories in exposures and outcomes (Walker 2001, Workshop Organising Team 2005, Statistics New Zealand 2009).

While the sector protocols require recording at a minimum of level two, most health data are aggregated and output at level one and sometimes, level zero (Māori, Pacific, and European/Other). Level zero continues to be used in the health sector, although it was retired from the Statistical Standard in 2005 (Ministry of Health 2004). Because most ethnicity data are aggregated into broad level one categories, the reporting of multiple ethnic groups are only captured across broad ethnic categories but not within them. For example, a person who checks tick boxes for Māori, Chinese and Indian tickboxes would be counted once in the Māori category and in the Asian category at level one.(Callister, Didham et al. 2007).

---

aggregation, with totals adding to more than 100%) or 'single/combination' output (where an individual is counted in either a single (e.g. sole Māori) or a combination ethnic group (e.g. Māori/Asian) based on their responses, from specified single and combination ethnic categories (Ministry of Health. Ethnicity data protocols for the health and disability sector. Wellington: Ministry of Health, 2004; Statistics New Zealand. Statistical Standard for Ethnicity Wellington: Statistics New Zealand).

In part, these reporting conventions may be artefacts of earlier collection and recording processes that restricted and shaped the data available. Small numbers (or small sample size for surveys) have also been cited as a reason for not reporting or utilising disaggregated ethnicity data at the lower levels of the classification, particularly at sub-national levels, due to issues such as potential identifiability of respondents and reliability and precision of estimates in the calculation of rates. (Bycroft 2011).

Implications for health equity

The use of pan-ethnic labels is problematic in so far as it masks the considerable cultural and socio-economic heterogeneity that exists within broad categories (Rasanathan, Craig et al. 2004). Within the health sector the use of pan-ethnic categories has implications for minority groups, particularly where those groups have poorer health outcomes (Rasanathan, Craig et al. 2004). It is possible that health priorities and inequities are being masked in the continued reliance on reporting health data with restricted categories and broad groupings (Workshop Organising Team 2005), as suggested by some recent evidence for Pacific and Asian ethnic groups (e.g.(Blakely, Richardson et al. 2009, Jatrana, Richardson et al. 2014)). This has impacts on the ability of communities to understand populations and plan for future health priorities, and on their rights to access to data in sufficient depth and detail for their needs (Jatrana, Richardson et al. 2014).

While disaggregated ethnicity data reporting remains relatively uncommon in Aotearoa NZ, some granular reporting occurs as part of specific focused outputs or analysis. In addition, some work has been undertaken to increase the sample size of groups other than NZ European (the dominant ethnic group) in surveys and research to allow for better estimates, stratification by ethnicity, and greater disaggregation (e.g. the NZ Health Survey) (Ministry of Health 2008).

Ongoing and future work and considerations

A number of changes have been proposed that have the potential to impact on the availability and granularity of population-level ethnicity data (Statistics New Zealand 2015). Like other National Statistical Offices, Statistics NZ has been making greater use of online technologies and aims to have 70 per cent of forms completed online for the 2018 census (Kukutai, Thompson et al. 2015). The flexibility provided by online technologies, such as no restrictions on space, predictive text, and immediate lookup, has the potential to improve recording of detailed ethnicity data. Systems designed by Statistics NZ to collect granular ethnicity data for their products may become available and/or integrated more broadly to facilitate improved granularity in other collections, which may improve issues with slippage in granularity between collection and recording. In the health sector, work is underway to revise the ethnicity data protocols, including recommendations to collect and record data

at level four (the most detailed level) and for up to six ethnicities per person. Should this be implemented, it could increase the potential to analyse and report ethnicity data at a far greater level of granularity (not withstanding issues with small numbers).

In the longer-term, Statistics New Zealand are looking at ways to replace the enumeration-based census with a fully administrative model based on an anonymised government database called the Integrated Data Infrastructure (Bycroft 2011, Statistics New Zealand 2015). The shift to an administrative census may have trade-offs for the granularity of ethnicity data as most other administrative systems (with the exception of births and deaths registrations) do not collect and record ethnicity at the level of detail of the census. Initial scoping work has revealed significant inconsistencies in how ethnicity is recorded and reported within the IDI (Gibb and Shrosbree 2014). A recent matching exercise between a subset of the IDI dataset and the 2013 census showed match rates for Māori of below 60 percent, the lowest for a major ethnic group except MELAA (Middle Eastern, Latin American and African) (Gibb 2015). Statistics NZ has recognised that future census transformation must meet Māori data rights and interests, given their Treaty responsibilities and the importance of data disaggregation for indigenous development and self-determination (United Nations Department of Economic and Social Affairs 2015).

Finally, Statistics New Zealand has indicated that it is prepared to consider replacing the current hierarchical ethnic classification with a system that allows for more flexibility in the ways that ethnic groups are added to the classification (e.g. through direct proposals or monitoring of responses, rather than as part of periodic review processes) and in the ways that ethnicities can be grouped for analysis in the absence of a hierarchy (Allan 2001).

Conclusion

Ethnicity data are institutionalised in Aotearoa NZ in official statistics and in the health sector. While there is some room for improvement, the health sector has established ethnicity data protocols and relatively comprehensive coverage across administrative and survey collections that facilitates the measurement and monitoring of ethnicity in relation to health. Although the standard official question provides for heterogeneity and the classification system is relatively granular, much ethnicity data are reported in aggregate forms or in relation to broad ethnic groupings only. Disaggregation at greater detail than Level 1 pan-ethnic categories is relatively uncommon in official statistics or in policy contexts, and very few official surveys have sample sizes that would allow for such disaggregation. This becomes problematic in terms of understanding the diversity of experience and outcome within these broad populations, and planning for population growth and priorities (Rasanathan, Craig et al. 2004). In health, it also has the potential to mask inequities, privilege, and health needs. At a broader

level, it contributes to the homogenisation of ‘minority’ or ‘Other’ groups (Rasanathan, Craig et al. 2004). While there are challenges with increased granularity, there remains much potential for improving the utility of ethnic statistics generally, and in health specifically.

References

Ajwani, S., T. Blakely, B. Robson, J. Atkinson and C. Kiro (2004). "Unlocking the numerator-denominator bias III: adjustment ratios by ethnicity for 1981-1999 mortality data. The New Zealand Census Mortality Study." NZMJ **116**(1175).

Allan, J. (2001). International concepts and classifications (Review of the Measurement of Ethnicity: main paper). Wellington, Statistics New Zealand.

Blakely, T., K. Richardson, J. Young, P. Callister and R. Didham (2009). "Does mortality vary between Pacific groups? Estimating Samoan, Cook Island Maori, Tongan and Niuean mortality rates using hierarchical Bayesian modelling " NZMJ **122**(1307).

Bramley, D. and S. Latimer (2007). "The accuracy of ethnicity data in primary care." NZMJ **120**(1264): U2779.

Bycroft, C. (2011). Future New Zealand censuses: implications of changing census frequency or adopting other models. Wellington, Statistics New Zealand.

Callister, P., R. Didham, D. Potter and T. Blakely (2007). "Measuring ethnicty in New Zealand: developing tools for health outcomes analysis." Ethnicity & Health **12**(4): 299–320.

Cormack, D. (2010). The politics and practice of counting: ethnicity data in official statistics in Aotearoa/New Zealand. Wellington, Te Rōpū Rangahau Hauora A Eru Pōmare.

Cormack, D. and R. Harris (2009). Issues in monitoring Māori health and ethnic disparaties: an update. Wellington, Te Rōpū Rangahau Hauora A Eru Pōmare.

Cormack, D. and M. McLeod (2010). Improving and maintaining quality in ethnicity data collections in the health and disability sector. Wellington, Te Rōpū Rangahau Hauora A Eru Pōmare.

Department of Statistics (1993). New Zealand Standard Classification of Ethnicity. Wellington, Department of Statistics.

Didham, R. and R. Bedford (2004). Peopling the future, interpreting the changing demography of New Zealand. Southbank Victoria, AU, Dunmore Press**:** 1-21.

Fenton, F. (1859). Observations on the state of the Aboriginal inhabitants of New Zealand. Auckland, New Zealand Government.

Ford, L. (2012). Locating Indigenous self-determination in the margins of settler sovereignty. Between Indigenous and settler governance. L. Ford and T. Rowse. Abingdon & New York, Routledge**:** 1-11.

Gibb, S. (2015). Population estimates and ethnicity: Recent findings from the IDI. Towards a future census. Paper presented at Population Association of New Zealand Conference. University of Waikato, Hamilton.

Gibb, S. and E. Shrosbree (2014). Evaluating the potential of linked data sources for population estimates: The Integrated Data Infrastructure as an example. Wellington, Statistics New Zealand.

Harré, J. (1968). Maori-Pakeha intermarriage. The Maori People in the Nineteen-sixties. A Symposium. E. Schwimmer. London, C. Hurst & Co**:** 118-131.

Harris, R., B. Robson, P. Reid and V. Keefe (1997). CHE cultural reviews: ethnicity data collections - error rates and training effectiveness (Report to the Central Regional Health Authority). Wellington, Te Rōpū Rangahau Hauora A Eru Pōmare.

Health, M. o. (2009) "Ethnicity data protocols codset appendices."

Ip, M. (2003). "Maori-Chinese encounters: indigine-immigrant interaction in New Zealand." Asian Studies Review **27**(2): 227-252.

Jatrana, S., K. Richardson, T. Blakely and S. Dayal (2014). "Does mortality vary between Asian subgroups in New Zealand: an application of hierarchical Bayesian modelling." PLOS ONE **9**(8): e105141.

Kilgour, R. and V. Keefe (1992). Kia piki te ora: the collection of Maori health statistics. Wellington, Department of Health.

Kukutai, T. (2004). "The problem of defining an ethnic group for public policy: who is Maori and why does it matter? ." Social Policy Journal of New Zealand **23**: 86-108.

Kukutai, T. (2012). Quantum Māori, Māori quantum: Representations of Māori identities in the census, 1857/8-2006. Counting stories, moving ethnicities: Studies from Aotearoa New Zealand R. McClean, B. Patterson and D. Swain. Hamilton, University of Waikato**:** 27-51.

Kukutai, T. and P. Callister (2009). "A 'main' ethnic group? Ethnic self-prioritisation among New Zealand youth." Social Policy Journal of New Zealand **36**: 16-31.

Kukutai, T., V. Thompson and R. McMillan (2015). "Whither the census? Continuity and change in census methodologies worldwide, 1985–2014." Journal of Population Research **32**(1): 3-22.

Li, W. (2007/2008). "Changing Asian immigration and settlement in the Pacific Rim." New Zealand Poluation Review **33/34**: 69-93.

Ministry of Education. (2014, November 2014). "Ethnic group codes." Retrieved 28 February, 2016, from https://www.educationcounts.govt.nz/data-services/collecting-information/code-sets-and-classifications/ethnic_group_codes.

Ministry of Health (2004). Ethnicity data protocols for the health and disability sector. Wellington, Ministry of Health.

Ministry of Health (2008). Methodology report for the 2006/07 New Zealand Health Survey. Wellington, Ministry of Health.

Ministry of Health. (2012, 10 May 2012). "National Health Index." from http://www.health.govt.nz/our-work/health-identity/national-health-index.

Ministry of Health (2014). Equity of Health Care for Māori: A framework. Wellington, Ministry of Health.

Ministry of Health (2015). Tatau Kahukura: Māori Health Chart Book 2015. Wellington, Ministry of Health.

Pool, I. (1991). Te iwi Maori: a New Zealand population, past, present & projected. Auckland, Auckland University Press.

Pool, I. and T. Kukutai. (2014, 14 October 2015). "Taupori Māori – Māori population change." Te Ara - the Encyclopedia of New Zealand Retrieved 28 February, 2016, from http://www.TeAra.govt.nz/en/taupori-maori-maori-population-change

Public Health Intelligence (2001). Monitoring ethnic inequalities in health Public Health Intelligence Occasional Bulletin No 4. Wellington, Ministry of Health.

Rasanathan, K., D. Craig and R. Perkins (2004). Is "Asian" a useful category for health research in New Zealand. Proceedings of the Inaugural International Asian Health Conference: Asian health and wellbeing, now and into the future Auckland, University of Auckland.

Robson, B., P. Reid and Te Röpü Rangahau Hauora a Eru Pömare (2001). Ethnicity matters: Mäori perspectives Review of the Measurement of Ethnicity in Official Statistics. Wellington Statistics New Zealand.

Spoonley, P. (2015). "New diversity, old anxieties in New Zealand: the complex identity politics and engagement of a settler society." Ethnic and Racial Studies **38**(4): 650-661.

Statistics New Zealand. "Classifications and Standards - review process." from http://www.stats.govt.nz/methods/classifications-and-standards/classifications-and-standards-review-process.aspx.

Statistics New Zealand. "NZ.Stat." from http://nzdotstat.stats.govt.nz/wbos/Index.aspx.

Statistics New Zealand. (2005). "Statistical Standard for Ethnicity." from http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx.

Statistics New Zealand (2009). Final report of a review of the official ethnicity statistical standard 2009. Wellington, Statistics New Zealand.

Statistics New Zealand. (2013). "2013 Census definitions and forms." from http://www.stats.govt.nz/Census/2013-census/info-about-2013-census-data/2013-census-definitions-forms.aspx.

Statistics New Zealand. (2013). "2013 Census information by variable." Retrieved September, 2016, from http://www.stats.govt.nz/Census/2013-census/info-about-2013-census-data/information-by-variable/ethnicity.aspx.

Statistics New Zealand (2014) "2013 Census QuickStats about culture and identity."

Statistics New Zealand (2015) "Census transformation – A promising future – Cabinet paper (redacted) ".

Tan, L., T. Blakely and J. Atkinson (2010). "Ethnic counts on mortality and census data 2001-06: New Zealand census-mortality study update." NZMJ **123**(1320): 37-44.

United Nations Department of Economic and Social Affairs (2015). State of the World's Indigenous Peoples. New York, United Nations Publications.

Walker, U. (2001). A question of ethnicity - one word, different people, many perceptions: the perspectives of groups other than Maori, Pacific peoples, and New Zealand Europeans. Palmerston North, New Zealand Federation of Ethnic Councils.

Woodward, A. and T. Blakely (2014). The Healthy Country? A History of Life and Death in New Zealand. Auckland, Auckland University Press.

Workshop Organising Team (2005). Issues and options paper: The use of the term 'Asian' in New Zealand and implications for research, policy development and community engagement. Auckland, University of Auckland.

Appendix 3

Appendix 3.1 : Current official definition of ethnicity in Aotearoa New Zealand

Statistics New Zealand is responsible for the official definition of ethnicity. The most recent definition is included in the *Statistical Standard for Ethnicity 2005*. It states:

> Ethnicity is the ethnic group or groups that people identify with or feel they belong to. Ethnicity is a measure of cultural affiliation, as opposed to race, ancestry, nationality or citizenship. Ethnicity is self perceived and people can belong to more than one ethnic group.
>
> An ethnic group is made up of people who have some or all of the following characteristics:
>
> - a common proper name
> - one or more elements of common culture which need not be specified, but may include religion, customs, or language
> - unique community of interests, feelings and actions
> - a shared sense of common origins or ancestry, and
> - a common geographic origin.
>
> This definition is based on the work of Smith (1986) (Statistics New Zealand 2005) p.1.

It was retained in the 2009 Review of Ethnicity (Statistics New Zealand 2009).

Appendix 3.2: Level one of the current official classification

Code	*Description*
1	European
2	Māori
3	Pacific Peoples
4	Asian
5	Middle Eastern/Latin American/African
6	Other
9	Residual Categories

Source: Statistics New Zealand. Ethnicity -- classification [cited 2016 28 February]. Available from: http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx.

Appendix 3.3: Level two of the current official classification

Code	*Description*
10	European nfd
11	New Zealand European
12	Other European
21	Māori
30	Pacific Peoples nfd
31	Samoan
32	Cook Islands Maori
33	Tongan
34	Niuean
35	Tokelauan
36	Fijian
37	Other Pacific Peoples
40	Asian nfd
41	Southeast Asian
42	Chinese
43	Indian
44	Other Asian
51	Middle Eastern
52	Latin American
53	African
61	Other Ethnicity
94	Don't Know
95	Refused to Answer
96	Repeated Value
97	Response Unidentifiable
98	Response Outside Scope
99	Not Stated

Source: Statistics New Zealand. Ethnicity -- classification [cited 2016 28 February]. Available from: http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx.

Appendix 3.4: Level three of the current official classification

Code	*Description*
100	European nfd
111	New Zealand European
121	British and Irish
122	Dutch
123	Greek
124	Polish
125	South Slav
126	Italian
127	German
128	Australian
129	Other European
211	Māori
300	Pacific Peoples nfd
311	Samoan
321	Cook Islands Maori
331	Tongan
341	Niuean
351	Tokelauan
361	Fijian
371	Other Pacific Peoples
400	Asian nfd
410	Southeast Asian nfd
411	Filipino
412	Cambodian
413	Vietnamese
414	Other Southeast Asian
421	Chinese
431	Indian
441	Sri Lankan

442	Japanese
443	Korean
444	Other Asian
511	Middle Eastern
521	Latin American
531	African
611	Other Ethnicity
944	Don't Know
955	Refused to Answer
966	Repeated Value
977	Response Unidentifiable
988	Response Outside Scope
999	Not Stated

Source: Statistics New Zealand. Ethnicity -- classification [cited 2016 28 February]. Available from: http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx.

Appendix 3.5: Level four of the current official classification

Code	*Description*
10000	European nfd
11111	New Zealand European
12100	British nfd
12111	Celtic nfd
12112	Channel Islander
12113	Cornish
12114	English
12115	Gaelic
12116	Irish
12117	Manx
12118	Orkney Islander
12119	Scottish
12120	Shetland Islander
12121	Welsh
12199	British nec
12211	Dutch
12311	Greek
12411	Polish
12500	South Slav nfd
12511	Croation
12512	Dalmatian
12513	Macedonian
12514	Serbian
12515	Slovenian
12516	Bosnian
12599	South Slav nec
12611	Italian
12711	German
12811	Australian
12911	Albanian

12912	Armenian
12913	Austrian
12914	Belgian
12915	Bulgarian
12916	Belorussian
12916	Corsican
12918	Cypriot nfd
12919	Czech
12920	Danish
12921	Estonian
12922	Finnish
12923	Flemish
12924	French
12925	Greenlander
12926	Hungarian
12927	Icelandic
12928	Latvian
12930	Maltese
12931	Norwegian
12932	Portuguese
12933	Romanian
12934	Gypsy
12935	Russian
12936	Sardinian
12937	Slavic
12938	Slovak
12939	Spanish
12940	Swedish
12941	Swiss
12942	Ukranian
12943	American
12944	Burgher

12945	Canadian
12946	Falkland Islander
12947	New Caledonian
12948	Afrikaner
12950	Zimbabwean
12999	European nec
21111	Māori
30000	Pacific Peoples nfd
31111	Samoan
32100	Cook Islands Maori nfd
32111	Aitutaki Islander
32112	Aitu Islander
32113	Mangaia Islander
32114	Manihiki Islander
32115	Mauke Islander
32116	Mitiaro Islander
32117	Palmerston Islander
32118	Penrhyn Islander
32119	Pukapuka Islander
32120	Rakahanga Islander
32121	Rarotongan
33111	Tongan
34111	Niuean
35111	Tokelauan
36111	Fijian
37111	Admiralty Islander
37112	Australian Aboriginal
37113	Austral Islander
37114	Palau Islander
37115	Bismark Archipelagoan
37116	Bougainvillean
37117	Caroline Islander

37118	Easter Islander
37119	Gambier Islander
37120	Guadalcanalian
37121	Chamorro
37122	Hawaiian
37123	Kanak
37124	Kiribati
37125	Malaitan
37126	Manus Islander
37127	Marianas Islander
37128	Marquesas Islander
37129	Marshall Islander
37130	Nauruan
37131	New Britain Islander
37132	New Georgian
37133	New Irelander
37134	Banaban
37135	Papua New Guinean
37136	Phoenix Islander
37137	Pitcairn Islander
37138	Rotuman
37139	Santa Cruz Islander
37140	Tahitian
37141	Solomon Islander
37142	Torres Strait Islander
37143	Tuamotu Islander
37144	Tuvaluan
37145	Ni Vanuatu
37146	Wake Islander
37147	Wallis Islander
37148	Yap Islander
37199	Pacific Peoples ned

40000	Asian nfd
41000	Southeast Asian nfd
41111	Filipino
41211	Cambodian
41311	Vietnamese
41411	Burmese
41412	Indonesian
41413	Laotian
41414	Malay
41415	Thai
41499	Southeast Asian nec
42100	Chinese nfd
42111	Hong Kong Chinese
42112	Cambodian Chinese
42113	Malaysian Chinese
42114	Singaporean Chinese
42115	Vietnamese Chinese
42116	Taiwanese
42199	Chinese nec
43100	Indian nfd
43111	Bengali
43112	Fijian Indian
43113	Gujarati
43114	Indian Tamil
43115	Punjabi
43116	Sikh
43117	Anglo Indian
43199	Indian nec
44100	Sri Lankan nfd
44111	Sinhalese
44112	Sri Lankan Tamil
44199	Sri Lankan nec

44211	Japanese
44311	Korean
44411	Afhani
44412	Bangladeshi
44413	Nepalese
44414	Pakistani
44415	Tibetan
44416	Eurasian
44499	Asian nec
51100	Middle Eastern nfd
51111	Algerian
51112	Arab
51113	Assyrian
51114	Egyptian
51115	Iranian/Persian
51116	Iraqi
51118	Israeli/Jewish
51118	Jordanian
51119	Kurd
51120	Lebanese
51121	Libyan
51122	Moroccan
51123	Omani
51124	Palestinian
51125	Syrian
51126	Tunisian
51127	Turkish
51128	Yemeni
51199	Middle Eastern nec
52100	Latin American nfd
52111	Argentinian
52112	Bolivian

52113	Brazilian
52114	Chilean
52115	Colombian
52116	Costa Rican
52117	Creole
52118	Ecuadorian
52119	Guatemalan
52120	Guyanese
52121	Honduran
52122	Malvinian
52123	Mexican
52124	Nicaraguan
52125	Panamanian
52126	Paraguayan
52127	Peruvian
52128	Puerto Rican
52129	Uruguayan
52130	Venezualan
52199	Latin American nec
53100	African nfd
53112	United States Creole
53113	Jamaican
53114	Kenyan
53115	Nigerian
53116	African American
53117	Ugandan
53118	West Indian
53118	Somali
53120	Eritrean
53121	Ethiopian
53122	Ghanaian
53199	African nec

61111	Indian
61112	Inuit
61113	North American Indian
61114	South American Indian
61115	Mauritian
61116	Seychellois
61117	South African Coloured
61118	New Zealander
61199	Other Ethnicity nec
94444	Don't Know
95555	Refused to Answer
96666	Repeated Value
97777	Response Unidentifiable
98888	Response Outside Scope
99999	Not Stated

Source: Statistics New Zealand. Ethnicity -- classification [cited 2016 28 February]. Available from: http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx.

Chapter 4: Ethnic group classification in Pluri-National State of Bolivia

Pamela Pereyra-Zamora

Abstract

Current ethnic classification in Bolivia began with the Spanish conquest of the Inca Empire. Although in less quantities than in other Latin-American countries, Europeans, Arabs and Asians established in Bolivia during the colonial times and the 19th and 20th centuries. Therefore, the population composition of Bolivia is formed by indigenous people, former slaves, colonial settlers and more recent immigrants.

Since independence from Spain in 1825 eleven censuses have been completed and since the 1950 a question on language has been included as a proxy to ethnic identity, although it has evolved across time. In the 2001 census for the first time an ethnic self-identification question was introduced (Quechua, Aymara, Guaraní, Chiquitano, Mojeño and 'other native'). The result was that 39 indigenous peoples were later recognised in the 2009 constitution. The most significant feature of the 2012 census is the great increase in the granularity of formerly marginal peoples, indigenous and Afro-Bolivian. The question had an open-ended answer and the outcome was over a hundred ethnic self-identifications.

According to the 2012 census there were 10,059,856 inhabitants, and 41.7% declared to belong to any of the peasant, aboriginal, indigenous or Afro-Bolivian people or nation. Of this percentage, 43.7% are Quechua, 38.1% are Aymara, 3.5% are Chiquitano and 2.3% Guaraní. In the 2012 census, the percentage of people belonging to an indigenous people decreased compared to 2001. This controversial decrease could be related to technical aspects such as the question, the previous filter, and also may be due to the abandonment of indigenous identities in urban contexts.

The new constitution declares that Bolivia is a pluri-national state in which indigenous nations and peasant nations are guaranteed their rights. This entails their right to home rule, to manage their land, to apply their own justice, to use their language, and to live according to their own cosmovision. A related use of the census granularity, perhaps the most important, is the allocation of seats in the parliament to different indigenous peoples according to the percentages that they obtain in the census. It is therefore a period of implementation of the political consensus on ethnicity in Bolivia (recording, granularity and reporting) reflected in the 2009 constitution as well as its insertion in areas such as health.

Introduction: Demographic background, development of ethnic/racial composition, and history of migration.

The current ethnic and racial composition of Bolivia began to form with the Spanish conquest of the Inca Empire and, subsequently, further territories towards the jungles of the east. A dichotomy was then born between *blancos* (whites) or Spanish conquistadores and *indios* (indigenous). The bringing of slaves[1] from Africa during the Spanish colonial times (16th to 19th centuries) added a new category, the *negro* (meaning black). The social disintegration of the indigenous world, and its cultural dislocation, forced forms of acculturation, and of social mobility within a stratified caste system, it created a social space for mestizo classes and groups within the colonial society. Intermarriage within these different groups secured its continuation up to the present.

Mestizo, a term originally applied to people of mixed blood (*indio* and *blanco*), also became applied to cultural mestizos. That is, those who combine cultural traits. *Mulato* - referring to offspring of *negro* and *blanco*, and *zambo* – offspring of *negro* and *indio*, were also used. This colonial schema and denomination system have been maintained by the *blanco* elites for centuries with different adaptations. The 18th century was particularly obsessed with the creation of a discriminatory denomination system (Iglesia and Schvarztman 1987) in which a multiple system of hierarchies intervened, states of the realm (*estamentos*), caste and cultural socio-economic strata in which the caste was the structuring matrix (Szeminski 1984). Among the terms are *cholo, criollo, birlocha, campesino, kunumi, t'ara, chota, medio pelo*. Some of these terms are still in use, although altered by generations of official discourses on equality.

The array of ethnic and racial denominations is even more complex as from the perspective of the indigenous peoples the denominations and the anthropological meaning behind them might be different - *qar'a, qolla, camba* are some examples. This would be the indigenous perspective and denominations of the same system of social stratification. The 19th and early 20th century were intensive in emigration from the old world. Although in less quantities than in other Latin-American countries, Europeans, Arabs and Asians established in Bolivia. During the second part of the 20th century this trend declined intensively. Recent immigrants are few, mostly Argentinians.

These processes have produced a population composed of people of indigenous origin, slave origin, colonial origin and recent immigrant origin and also individuals and communities with different degrees of assimilation, integration, creolization, cultural dislocation and unresolved cultural contact in areas of new colonization that were previously sparsely populated such as the Chapare and Alto Beni. Rather

[1] The bringing of black slaves to the Americas began short after 1492. In the case of Bolivia, different forms of slavery, unpaid work and serfdom were being abolished along the 19th and first half of the 20th century.

than discrete ethnic enclave communities, cities are mixed with widespread networks of contact. The city of Oruro, for instance, of heavy indigenous composition, is the centre of multiple commercial routes and networks built by indigenous women and reaching large parts of the country and beyond. Some observers (Cusicanqui Rivera 2007, Rivera Cusicanqui 2010) argue that ethnic maps, as cartographical representations of ethnic communities are not suitable metaphors of ethnic composition and distribution in Bolivia (see appendix 4.6), but cloth and fabric, instead, are a more accurate representation of the reality of ethnic distribution with multiple interactions, links, exchanges and crossing of all sorts of borders. This same author proposes a hypothesis for understanding *mestizaje* in Bolivia. According to her, the idea of *mestizaje* as an amalgam and a new kind of culture and citizenship overcoming indigenous or colonial elements is not a real representation of the contemporary nor the historical cultural process in Bolivia.

There are few spaces of melting pot and hybridity. She builds on other authors such as Larson (Larson 1998) and Barragán (Barragán 1992) to analyse *mestizaje's* conflictive economic base and complex internal articulation through which, far from disappearing, oppositions between western culture and native cultures, the two real focus of identity, are renewed. *Mestizaje* does not exist only as an amalgam it might also be a mixture without fusion. Nonetheless, those regarded as mestizos are still being marginalized in different ways.

Before 1952 indigenous and slave descendants did not really have full citizenship, after 1952 they got a second class citizenship. This was the time the term *campesino*[2] (peasant) spread as a substitute for terms such as *mestizo*, *indio* etc. The protests at the end of the last century and the beginning of this one, where indigenous activists have meant the core of the protest movement, began in the regions of the Chapare and Alto Beni, regions which are not inside the ethnic maps but areas that are in themselves zones of contact. All this led to the possibility of an indigenous person president (Evo Morales took office in 2006), important changes in the country's social structure and to the constitution of 2009. In the constitution, thirty-six ethnic groups (and maps) are officially recognized among the indigenous communities and the Afro-Bolivians of the Yungas region. The others are currently being nominally homogenized under the general non-ethnic all-encompassing denomination of 'Bolivian', a new space identified with the pluri-national state. Nevertheless, that population sector is not homogeneous nor has the mestizo dimension within it ever been.

According to the 2012 census (see appendix 4.1) there were 10,059,856 inhabitants, of which 41.7% have declared themselves to belong to any of the peasant, aboriginal, indigenous or Afro-Bolivian

[2] *Campesino* or peasant began to be used in order to avoid the ethnic a racial character of the social stratification in Bolivia.

people or nation. Of this percentage 43.7% are Quechua, 38.1% are Aymara, 3.5% are Chiquitano and 2.3% Guaraní, among the most numerous.

Although in the 2012 census under 15 year olds are included (unlike the 2001 census) the percentage of people identifying as belonging to an indigenous people has decreased. When considering this, it is also important to take into account that the question and the answers used in 2012 are not the same. In 2001 (see table 4.1) the percentage of people declaring themselves to belong to an indigenous people is 62.0% - 30.7% Quechua, 25.2% Aymara, 2.2% Chiquitano and 1.5% Mojeño, among the most numerous[3].

Table 4.1. Distribution of Bolivian population of 15 years of age and over according to self-identification with an indigenous nation or people – 2001

Population	Number	%
Total population in Bolivia	**8,274,325**	
Total amount of people to whom the question does not apply	3,198,074	
Total amount of people to whom the question is applicable	5,076,251	100.0
People who do not belong to any indigenous people	1,930,476	38.0
People belonging to any indigenous people	3,145,775	62.0
Quechua	1,557,689	30.7
Aymara	1,278,627	25.2
Chiquitano	112,271	2.2
Guarani	78,438	1.5
Mojeño	43,323	0.9
Other native	75,427	1.5
Total	**3,145,775**	**62.0**

Source: Own elaboration on the basis of census - 2001. INE (National Institute of Statistics).

There are 119,033 foreign born immigrants living in Bolivia, 1.2% of the population. 30.4% come from Argentina, 18.5% from Brazil, 8.6% from Spain, 7.8% from Peru and 7.1% from Mexico and the remaining 27.6% from different European, Asian or other Latin-American countries.

Key sources of ethnicity data

Although Bolivia has had in its history a variety of statistical projects, few of them recorded ethnicity (see appendix 4.2). As regards ethnicity, the most important projects are the 2001 census and that of 2012 (see below), the Demography and Health Survey (2003 and 2008) and the Household Continuous

3 For a view of all categories see Naciones Unidas/CEPAL/BIT. (2005). "Los pueblos indígenas de Bolivia: Diagnostico sociodemográfico a partir del censo 2001". pp 40. Retrieved 24 March, 2016, from http://www.cepal.org/es/publicaciones/3566-pueblos-indigenas-bolivia-diagnostico-sociodemografico-partir-censo-2001.

Survey (1999 to 2002) or Household Survey (2005 to 2013)[4]. The last census (2012) asks about ethnicity and language. It includes an open-ended question on language with a result of 36 indigenous languages plus the Castilian and several foreign languages and an open-ended question on belonging to a people with an outcome of the 39 peoples already contemplated in the 2009 constitution plus many more identified. The previous census (2001) also included a question on ethnicity with only five categories (Quechua, Aymara, Guaraní, Chiquitano, Mojeño) and 'other native'.

There is nevertheless a series of surveys that although in most cases do not record information on ethnicity, some do. Those carried out by the National Institute of Statistics (INE, Instituto Nacional de Estadística) collect information on ethnicity, language spoken in childhood or belonging to indigenous people since 1999, in some cases racial terms such as *raza negra* (black race) are included[5] but they disappeared shortly especially after the 2001 census.

There is also a whole range of surveys carried out from the private sector that contemplate ethnicity in different ways. The health surveys also include questions on ethnicity (2003 and 2008) and language (since 1989 to 2008). Hospital epidemiological surveillance registers do not record ethnicity. The most granular classification is the census. Bolivia has no register of mortality.

Lessons learned from the use of heterogeneity/granularity of ethnic classifications.

According to the new constitution Bolivia is a pluri-national state in which indigenous nations and peasant nations are guaranteed their rights. This entails their right to organize themselves in political autonomies (article. N° 2 and art. 289) to manage their land, to apply their own justice, to use their own language, and to live according to their own cosmovision. In the art. 30-I cosmovision is specifically mentioned while describing the characteristics of nations and indigenous peoples (see appendix 4.3 and 4.7). These autonomies are a part of the Bolivian state alongside departments, provinces and towns. A closely related use of the census granularity, perhaps the most important, is the allocation of seats in the parliament to different indigenous peoples. This consociational[6] distribution of parliament seats is calculated according to the percentages that indigenous peoples obtain in the census. Granularity is also taken into account in the allocation of ancestral lands to the indigenous communities for community management.

In indigenous areas the whole of the education is being taught through indigenous languages, Aymara, Quechua, Guaraní, and so on. Also, there are programs called 'certification of competencies' through

4 Encuesta de Demografía y Salud, Encuesta Continua de Hogares y Encuesta de Hogares.

5 For instance, the Household Continuous Survey of 1999 and 2000 (Encuesta Continua de Hogares).

6 Consociational refers to political strategies of power-sharing.

which the different productive vocations such as traditional professions[7] are certified by the Ministries of Education and Productive Development. This implies official recognition and certification of indigenous medicine practitioners such as the kallawaya doctors, and shamans such as the *yatiri*. There are also programs for the incorporation of traditional ancestral medicine and traditional pharmacopeia, into the National Health System (Asamblea Legislativa Plurinacional 2013).

Just as in education, there are developments regarding justice[8] within the indigenous communities and indigenous individuals. In a more general dimension, there are aspects of the Andean indigenous cosmovision which are entering the mainstream cultural world in Bolivia[9] (also in Ecuador) and its legislative and institutional bodies. One of them, affecting views of health but also many other areas is the concept of *buen vivir*[10] (good living) materialized in a *Ley Marco de la Madre Tierra y Desarrollo Integral para Vivir Bien* (Framework Law of the Mother Earth and Integral development for the Good Living) of Oct. 2012. It is considered a matrix principle from which other laws will be derived and it constitutes a complex concept involving the areas of sociology, politics, philosophy, economics or law, among others. The *buen vivir* or *sumak kawsay* in Quechua is a civilizational and cultural horizon based in Andean cosmovision and conceptions of the mother earth (the Pachamama), human life and our relationship with it. It has had implementations in the social and community managements of forests and environment, but the debate is ongoing for other developments.

Historical development of Bolivia's heterogeneity/granularity in ethnic classifications.

Ever since the Independence of Bolivia from Spain in 1825 eleven censuses have been completed[11] (1831, 1835, 1845, 1854, 1882, 1900, 1950, 1976, 1992, 2001, 2012) unregularly (INE 2015, INE 2016). Up to 1900 the aim of the censuses was more tributary than demographic (Mazurek 2012). Taxes were payed according to identity classification (see above) and the censuses also had an aim of ethnic identification with terms such as *blancos*, *criollos* (creole), *negros*, *mestizos* and *indios*.

In 1950 the first census population and households with a proper demographic character was completed. This census, assisted by the IASI (Inter-American Statistical Institute) meant a substantial improvement

7 These vocations might be culture specific and ethnic group specific and the curricula, methodologies, etc... in order to generate a real development impact in the communities and in keeping with the indigenist turn they are developed through the indigenous languages.

8 The community justice has the same rank as the ordinary justice and it is applied through local indigenous and culture specific institutions.

9 According to the sociologist Boaventura de Sousa Santos (2014), specialist in non-western epistemology, with the insertion of these paradigm into the legislation and practice of the state of Bolivia (and Ecuador) for first time in modern history concepts from outside the western cosmology and episteme are reaching official status at the highest levels (see Rivera Cusicanqui, S. and De Sousa Santos, B. (2016). "Conversa del Mundo". Retrieved 20 March, 2016, from https://www.youtube.com/watch?v=xjgHfSrLnpU).

10 The concept is present in the widest spread cultures of the region, among the Quechua under the term sumak kawsay, among the Aymaras under the term suma qamaña, and among the Guarani under the term teko porâ.

11 See for a whole relation of historical population registers, census, surveys, etc. INE – Instituto Nacional de Estadística (2016). "Hitos en la producción estadística". Retrieved 24 March, 2016, from http://www.ine.gob.bo/html/visualizadorHtml.aspx?ah=Cronologias.htm

in aims as much as in methodological accuracy. From 1950 onwards two more censuses have been completed on that century, 1976 and 1992 which had questions on linguistic issues.

In the 1976 census the question what Bolivian languages can you speak? was asked and 10 options were given: he/she does not speak yet or none, *Castellano*[12] (Spanish, Castilian), Aymara, Quechua, another (it does not allow to specify the others), Castilian-Aymara, Castilian-Quechua, Castilian with another, Aymara-Quechua, and Castilian-Aymara-Quechua. These last binary and trinary combinations might be seen as a substitute for the old mestizo denomination.

In the 1992 census the question about place of birth was kept and another was asked "what languages and/or dialect do you speak?" (only for people 6 years old or over) and the answers were: Castilian, Quechua, Aymara, Guaraní, other native (it does not allow to specify) and foreigner (*extranjero*) without the possibility of specifying either.

In the 2001 census a question on official identification and ethnic self-identification was introduced (only one identity could be chosen in 2001 and 2012 censuses). The question identified five major aboriginal or indigenous peoples (that are read out to the interviewee according to the instructions for the interviewer[13]) but it allowed self-identification. In the 1976 and 1992 censuses the emphasis was placed on self-identification with an aboriginal or indigenous people and in the answer five peoples are included (Quechua, Aymara, Guaraní, Chiquitano, Mojeño) and the possibility of specifying another native. Afro-Bolivians were not included in the question or in the answers given or in the definition of aboriginal or indigenous peoples used in the census instructions[14] as they were not considered as such. They were counted but in an indistinguishable way, moreover, we don't know how they identified themselves in the census. The 2009 constitution, nevertheless, is giving them equal rank as the other indigenous and aboriginal peoples.

The 20th century tried to fade identities out, the 21st was trying to emphasize them, but in a new way. Behind these changes there was a huge popular and indigenous mobilisation that has increased their demands in census and other surveys (CEPAL 2013).

The 2001 census kept the question on use and knowledge of language and options were Quechua, Aymara, Castilian, Guaraní, foreign, he/she does not speak, other native (this last one was free text option). It added another question on the language learnt in childhood with 8 closed answers (Quechua,

12 Castellano is the endonym of the Spanish language as it was originated in the region of Castile, in the centre-north of the Iberian Peninsula.

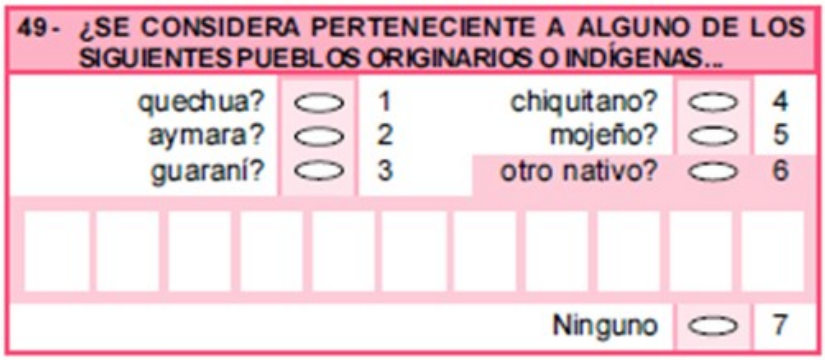
13 INE - Instituto Nacional de Estadística and CELADE - Centro Latinoamericano y Caribeño de Demografía. Manual del Empadronador 2001. Retrieved 20 March, 2016, from http://microdata.worldbank.org/index.php/catalog/449/related_materials.

14 The definition was: 'Se denominan pueblos originarios o indígenas a aquellos que vivían en américa antes de la llegada de los españoles'. That is: 'Aboriginal or indigenous peoples are considered those that lived in America before the arrival of the Spaniards'.

Aymara, Castilian, Guaraní, other native, foreigner, he/she does not speak). The aim of this question could be to identify the native speaker of the different languages. It would be worthwhile noticing the change of order of the answers in relation to 1992 and 1976 when Quechua and Aymara come before than Spanish. A change of mentality and power balance was thus being reflected on the census. This census introduced another question on birthplace with two possible answers, in or outside Bolivia. But in this occasion, the results can be crossed referenced with the use of native languages and other variables.

According to Xavier Albó, member of the 2001 Census Committee, the use or knowledge of languages does not equal ethnic self-identification (Albó 2012). This might explain the return to the ethnic denominations in the new question of the 2001 census (see box 4.1 and appendix 4.4).

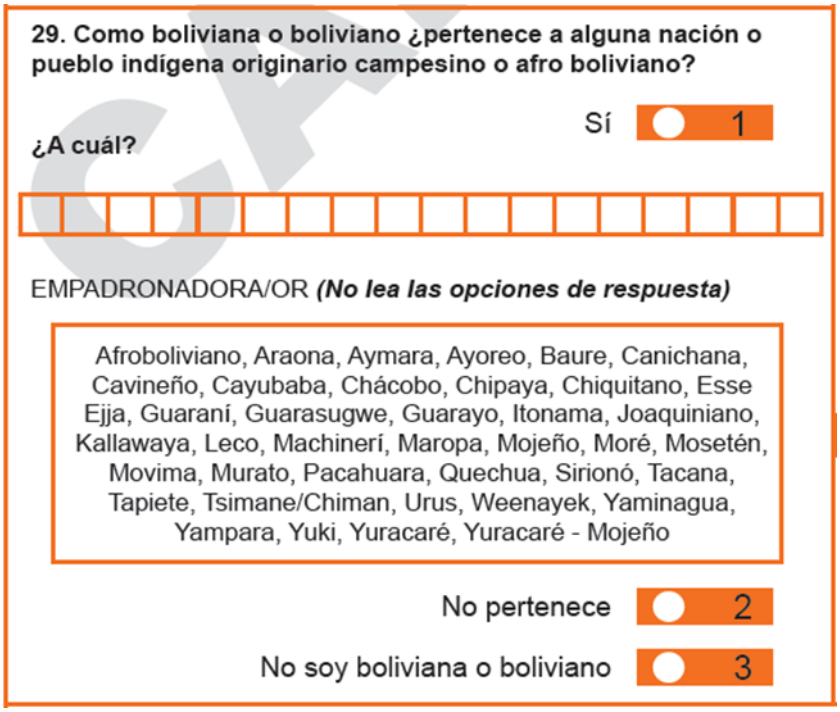
Box 4.1 2001 Census. Question on ethnicity



The result of this question was going to allow once again the vision of Bolivia as an indigenous country in a time when important political changes were about to start. The question 49 was "Do you consider yourself belonging to any of the following indigenous or aboriginal peoples?" and the options were Quechua, Aymara, Guaraní, Chiquitano, Mojeño, another native (free text option) or none, has been to a great extent, and according to many observers the key to the success of the 2001 census in relation to that of 2012.

In the 2012 census the question about language learnt in childhood remained but the answer, unlike in 2001 is completely open and another question is introduced about the languages spoken by the interviewee. The question about place of birth includes the year of arrival to Bolivia in case the interviewee has been born abroad. And finally, the most important difference with the 2001 census is the question 29 about ethnicity (see box 4.2 and appendix 4.5).

Box 4.2 2012 Census. Question on ethnicity



In 2012 it began by a filter, asking "As a Bolivian, do you belong to any nation or peasant aboriginal[15] indigenous or Afro-Bolivian people?" (echoes of the process of peasantization from the fifties are still at work, see above). If yes another question asks "Which one?" and a list of 39 names arranged in alphabetical order follows with an indication to the interviewer not to read the answers.
The results of this question give a granularity of 114 indigenous, aboriginal, peasant and Afro-Bolivian peoples or nations (see appendix 4.1). The paradox is that the percentage of people who declared to belong to one of these peoples in Bolivia decreased substantially. Moreover, this question, its results and its filter question, has unleashed the great ongoing national controversy on granularity and ethnicity recording with multiple accusations of methodological faults. Some experts even argued that the whole census should be put in quarantine.

This shows strong social processes as regards ethnic identity affecting the Bolivian populations. While the state makes efforts to support indigenous identity, culture and language, processes of linguistic substitution and abandoning of indigenous identities are occurring simultaneously, especially in the cities.

In any case, in the terminology used in the questions and granularity there clearly emerge the new visions of Bolivia and its peoples reflected in the 2009 constitution (see appendix 4.7). The rationale seems to be the dignifying and visibility of the peoples and communities previously regarded as

15 Pueblo indígena originario campesino in the original.

subalterns, outside the centres of power and invisible in the statistics, particularly the indigenous peoples and the Afro-Bolivian of the Yungas region. And at the same time a process of homogenization of the rest, including Mennonites, Arabs, Europeans, mestizos, etc. under the general non-ethnic population and also notable is the absence of the term of Bolivians.

The collection, analysis or reporting of data

Bolivia seems to be in the middle of a paradox as data are being collected with more intensity than it is being used. From the perspective of the wide-spread oral cultures of Bolivia this might not be the case as oral and popular cultures outside of the centres of power have also other ways to calculate, assess, deem and feel more inbuilt in their own cosmology and circumstances. This situation might change though as existence as indigenous communities validated by the census bestows rights.

In the final decades of the last century and the beginning of the current people have proved to be ahead of the state. Institutions are usually late in doing their task, i. e. in a moment of traditional medicine revival and expansion Bolivia has no valid health survey as the last one is from 2008. According to the PAHO/WHO-Bolivia (Pan-American Health Organization/World Health Organization) a National Demography and Health Survey has been completed in 2012, but data (and questionnaire) are not yet available.

Ongoing developments in the current systems of classification

Since the approval of the new constitution in 2009, Bolivia is in the middle of a profound change regarding public acknowledgement of ethnic identities and the formal and informal systems of ethnic classification. The current visibility of previously invisible and powerless communities is a consequence of the new political arrangements deriving from the outcome of the conflict at the turn of the century.

The alliance of trade unions, popular and mestizo classes, intellectuals, indigenous communities and activists and opposition parties has formed a historical bloc that rearranges ethnic relations and visibility of ethnic groups and communities, presence in the state and in the public administration, cosmovisions and forms of knowledge previously silenced, education, self-government, justice and forms of community and participatory democracy. It is, therefore, a period of implementation of the political and cultural consensus reflected in the 2009 constitution and further developments in the recording, granularity and reporting of ethnicity are expected as well as its insertion in areas such as health.

Acknowledgements

I would like to thank the following persons for their time and kindness in providing me with valuable information:

- Xavier Albó, Centro de Investigación y Promoción del Campesinado (Research Centre for the Promotion of the Peasantry).
- Marta Oviedo, Instituto de Estadística Teórica y Aplicada (Institute of Theoretical and Applied Statistics), Universidad Mayor de San Andrés.
- Alberto Camaqui, Viceministro de Medicina Tradicional e Interculturalidad del Ministerio de Salud del Estado Plurinacional de Bolivia (Vice-Minister of Traditional Medicine and Interculturality. Ministry of Health of the Pluri-national State of Bolivia).
- José M. Copete. Anthropologist and Research Collaborator. University of Alicante.
- Sergio Riveros. EURECNA Spa - Sucursal Bolivia.
- Hugo Rivera, OPS/OMS - Estado Plurinacional de Bolivia (PAHO/WHO- Pluri-National State of Bolivia).

References

Albó, X. (2012). "Censo 2012 en Bolivia: posibilidades y limitaciones con respecto a los pueblos indígenas." T'inkazos **32**: 33-45.

Barragán, R. (1992). Entre polleras, lliqllas y ñañacas. Los mestizos y la emergencia de la tercera república. Etnicidad, economía y simbolismo en los andes. Arze, S., Barragan, R.., Escobari L., Medinaceli X. La Paz, Institut français d'études andines, HISBOL, Sociedad Boliviana de Historia**:** 85-127.

CEPAL - Comisión Económica para América Latina y el Caribe. (2013). Mujeres indígenas en América Latina: dinámicas demográficas y sociales en el marco de los derechos humanos. Santiago de Chile, Naciones Unidas/CEPAL.

Cusicanqui Rivera, S. (2007). Violencia e interculturalidad: paradojas de la etnicidad en la Bolivia de hoy. Taller violencia e interculturalidad, Universidad Mayor de San Andrés.

Iglesia, C. and Schvarztman, J. (1987). Cautivas y misioneros. Mitos blancos de la conquista. Buenos Aires, Catálogos.

INE - Instituto Nacional de Estadística. (2015). Censo de Población y Vivienda 2012 - Bolivia. Caracteríticas de la población. La Paz, Instituto Nacional de Estadística.

INE - Instituto Nacional de Estadística. (2016). "Hitos en la producción estadística." Retrieved 24 March, 2016, from http://www.ine.gob.bo/html/visualizadorHtml.aspx?ah=Cronologias.htm.

Larson, B. (1998). Cochabamba, 1550-1900: Colonialism and Agrarian Transformation in Bolivia. New Jersey, Princeton University Press.

Mazurek, H. (2012). "Dialogo. El censo en Bolivia, una herramienta para el desarrollo." T'inkazos **15**(32): 9-31.

Asamblea Legislativa Plurinacional. (2013). "Ley 459 - Ley de Medicina Tradicional Ancestral Boliviana." Retrieved 22 March, 2016, from http://senado.gob.bo/legislativa/proyectos-ley/ley-459.

Rivera Cusicanqui, S. (2010). Violencias (re)encubiertas en Bolivia. La Paz, Piedra Rota.

Szeminski, J. (1984). La utopia tupamarista. Lima, Universidad Pondficia Católica del Perú - Fondo Editorial

Appendix 4

Appendix 4.1

Distribution of Bolivian population according to self-identification with an indigenous nation or people - 2012

Population	0-14		15+		Total	
	Number	%	Number	%	Number	%
Total population					**10,059,856**	**100.0**
People who do not belong to any indigenous people					5,859,879	58.3
People belonging to any indigenous people					4,199,977	41.7
A - Quechua	517,806	45.6	1,319,299	43.1	1,837,105	43.7
A - Aymara	377,243	33.2	1,221,564	39.9	1,598,807	38.1
B - Chiquitano	49,741	4.4	95,912	3.1	145,653	3.5
B - Guarani	31,713	2.8	65,129	2.1	96,842	2.3
B - Mojeño	12,925	1.1	29,168	1.0	42,093	1.0
B - Guarayo	8,706	0.8	15,204	0.5	23,910	0.6
B - Afroboliviano	6,475	0.6	16,855	0.6	23,330	0.6
B - Movima	5,405	0.5	13,474	0.4	18,879	0.4
B - Tacana	7,007	0.6	11,528	0.4	18,535	0.4
B - Tsimane Chiman	7,454	0.7	9,504	0.3	16,958	0.4
B - Itonoma	4,925	0.4	11,233	0.4	16,158	0.4
B - Leco	4,504	0.4	9,023	0.3	13,527	0.3
B - Kallawaya	3,887	0.3	7,775	0.3	11,662	0.3
B - Trinitario	2,547	0.2	4,526	0.1	7,073	0.2
B - Yuracaré	2,510	0.2	3,532	0.1	6,042	0.1
B - Weenayek	1,923	0.2	3,392	0.1	5,315	0.1
B - Maropa	1,434	0.1	3,071	0.1	4,505	0.1
B - Joaquiniano	1,243	0.1	2,980	0.1	4,223	0.1
B - Cavineño	1,615	0.1	2,269	0.1	3,884	0.1
B - Mosetén	1,388	0.1	2,128	0.1	3,516	0.1
B - Baure	933	0.1	2,395	0.1	3,328	0.1
B - Cayubaba	714	0.1	1,489	0.0	2,203	0.1
B - Ayoreo	628	0.1	1,561	0.1	2,189	0.1
B - Uru Chipayas	678	0.1	1,310	0.0	1,988	0.0
B - Esse Ejja	724	0.1	963	0.0	1,687	0.0
B - Chacobo	627	0.1	905	0.0	1,532	0.0
B - Urus	430	0.0	923	0.0	1,353	0.0
B - Ignaciano	405	0.0	602	0.0	1,007	0.0
B - Canichana	214	0.0	685	0.0	899	0.0
B - Sirionó	315	0.0	467	0.0	782	0.0
B - Yuracaré-Mojeño	273	0.0	460	0.0	733	0.0
B - Yuki	126	0.0	216	0.0	342	0.0
B - Yaminahua	123	0.0	136	0.0	259	0.0
B - Moré	88	0.0	167	0.0	255	0.0
B - Reyesano	54	0.0	198	0.0	252	0.0
B - Bésiro	77	0.0	166	0.0	243	0.0
B - Araona	50	0.0	178	0.0	228	0.0
B - Pacahuara	63	0.0	164	0.0	227	0.0
B - Murato	48	0.0	159	0.0	207	0.0
B - Tapiete	40	0.0	104	0.0	144	0.0
B - Guarasugwe	39	0.0	86	0.0	125	0.0
B - Loretano	16	0.0	77	0.0	93	0.0

B - Machineri	11	0.0	41	0.0	52	0.0
B - Javeriano	5	0.0	35	0.0	40	0.0
B - Uru-ito	0	0.0	2	0.0	2	0.0
C - Chichas	18,201	1.6	41,279	1.3	59,480	1.4
C - Yampara	2,323	0.2	4,727	0.2	7,050	0.2
C - Qhara Qhara	1,934	0.2	3,853	0.1	5,787	0.1
C - Qollas	1,161	0.1	3,783	0.1	4,944	0.1
C - Monkox	1,073	0.1	1,767	0.1	2,840	0.1
C - Coroma	617	0.1	957	0.0	1,574	0.0
C - Suyu Sura	386	0.0	1,138	0.0	1,524	0.0
C - Jacha Carangas	411	0.0	1,043	0.0	1,454	0.0
C - Lipez	404	0.0	865	0.0	1,269	0.0
C - Quillacas	256	0.0	675	0.0	931	0.0
C - Jach'a Pacajaqui	86	0.0	667	0.0	753	0.0
C - Pukina	232	0.0	519	0.0	751	0.0
C - Uchupiamonas	220	0.0	365	0.0	585	0.0
C - Aroma	79	0.0	504	0.0	583	0.0
C - Suyu Chuwi	205	0.0	277	0.0	482	0.0
C - Larecaja	84	0.0	365	0.0	449	0.0
C - Chayanta	37	0.0	298	0.0	335	0.0
C - Killacas	69	0.0	260	0.0	329	0.0
C - Chiriguano	31	0.0	296	0.0	327	0.0
C - Ayllu Jalka	90	0.0	221	0.0	311	0.0
C - Qhapaq Uma Suyu	102	0.0	198	0.0	300	0.0
C - Tinquipaya	38	0.0	183	0.0	221	0.0
C - Charagua	27	0.0	180	0.0	207	0.0
C - Totora Marka	24	0.0	178	0.0	202	0.0
C - Challapata	48	0.0	151	0.0	199	0.0
C - Corque	46	0.0	138	0.0	184	0.0
C - Ayllu Jila	38	0.0	138	0.0	176	0.0
C - Chullpas	19	0.0	128	0.0	147	0.0
C - Ayllu Porco	45	0.0	92	0.0	137	0.0
C - Huari	11	0.0	125	0.0	136	0.0
C - Ayllu Kharacha	37	0.0	94	0.0	131	0.0
C - Poroma	46	0.0	84	0.0	130	0.0
C - Yapacaní	33	0.0	93	0.0	126	0.0
C - Suyu Charcas	20	0.0	97	0.0	117	0.0
C - Chaqui	20	0.0	94	0.0	114	0.0
C - Salinas	13	0.0	95	0.0	108	0.0
C - Andamarca	14	0.0	89	0.0	103	0.0
C - San Juan	34	0.0	60	0.0	94	0.0
C - Sabaya	17	0.0	73	0.0	90	0.0
C - Jesús de Machaca	22	0.0	64	0.0	86	0.0
C - Tobas	35	0.0	51	0.0	86	0.0
C - Quila Quila	32	0.0	51	0.0	83	0.0
C - Curahuara de Carangas	8	0.0	66	0.0	74	0.0
C - Akarapis	12	0.0	56	0.0	68	0.0
C - Ayllu Yura	17	0.0	51	0.0	68	0.0
C - Huaylla Marka	16	0.0	49	0.0	65	0.0
C - Lagunillas	15	0.0	47	0.0	62	0.0
C - Pucara	14	0.0	43	0.0	57	0.0
C - Mataco	5	0.0	46	0.0	51	0.0
C - Ayllu Kacachaca	16	0.0	33	0.0	49	0.0
C - Orinoca	5	0.0	41	0.0	46	0.0

C - Turco	1	0.0	44	0.0	45	0.0
C - Pojos	3	0.0	41	0.0	44	0.0
C - Pampa Aullagas	9	0.0	32	0.0	41	0.0
C - Ayllu Jukumani	8	0.0	31	0.0	39	0.0
C - Mojocoya	6	0.0	33	0.0	39	0.0
C - Condo	10	0.0	27	0.0	37	0.0
C - Belen	4	0.0	31	0.0	35	0.0
C - Pati Pati	6	0.0	23	0.0	29	0.0
C - Charazani	6	0.0	21	0.0	27	0.0
C - Choquecota	4	0.0	21	0.0	25	0.0
C - Jatun Killacas	5	0.0	18	0.0	23	0.0
C - Urmiri de Quillacas	6	0.0	17	0.0	23	0.0
C - Moro Moro	1	0.0	18	0.0	19	0.0
C - Calcha	2	0.0	13	0.0	15	0.0
C - Huachacalla	10	0.0	3	0.0	13	0.0
C - Ucumasi	0	0,0	10	0.0	10	0.0
C - Pojpo	0	0.0	7	0.0	7	0.0
C - Layme	0	0.0	6	0.0	6	0.0
C - Non-specified indigenous or aboriginal	50,506	4.4	135,332	4.4	185,838	4.4
Total	**1,136,447**	**100.0**	**3,063,530**	**100.0**	**4,199,977**	**100.0**

Source: Own elaboration on the basis of the population census - 2012. INE (National Institute of Statistics).
http://datos.ine.gob.bo/binbol/RpWebEngine.exe/Portal?BASE=CPV2012COM&lang=esp
Legend: A: Major nations of peoples, B: Minor Nations of peoples, C: Other groups of population.

Appendix 4.2

Some examples of statistical sources that include some questions on indigenous self-identification

Statistical source	Responsible institution
▪ Censo Nacional de Población y Vivienda, 1976	Instituto Nacional De Estadística - Ministerio de Planeamiento y Coordinación
▪ Censo Nacional de Población y Vivienda, 1992	Instituto Nacional De Estadística - Ministerio de Planeamiento y Coordinación
▪ Censo Nacional de Población y Vivienda, 2001	Instituto Nacional De Estadística - Ministerio de Planificación para el Desarrollo
▪ Censo Nacional de Población y Vivienda, 2012	Instituto Nacional de Estadística - Ministerio de Planificación y Desarrollo / Dirección de Censos y Encuestas - INE Dirección de Informática y Cartografía e Infraestructura
▪ Encuesta Nacional de Demografía y Salud, 1989	Instituto Nacional de Estadística - Ministerio de Planeamiento y Coordinación
▪ Encuesta Nacional de Demografía y Salud, 1993-1994	Instituto Nacional de Estadística
▪ Encuesta Nacional de Demografía y Salud 1998	Instituto Nacional de Estadística
▪ Encuesta Nacional de Demografía y Salud, 2003	Instituto Nacional de Estadística - Ministerio de Salud y Deportes
▪ Encuesta Nacional de Demografía y salud 2008	Ministerio de Salud y Deportes - Instituto Nacional de Estadística
▪ Encuesta Continua de Hogares MECOVI, 1999	Instituto Nacional de Estadística - Ministerio de Planificación para el Desarrollo
▪ Encuesta Continua de Hogares MECOVI, 2000	Instituto Nacional de Estadística - Ministerio de Planificación para el Desarrollo
▪ Encuesta Continua de Hogares MECOVI, 2001	Instituto Nacional de Estadística - Ministerio de Planificación para el Desarrollo
▪ Encuesta Continua de Hogares MECOVI, 2002	Instituto Nacional de Estadística - Ministerio de Planificación para el Desarrollo
▪ Encuesta de Hogares, 2005	Instituto Nacional de Estadística - Ministerio de Planificación para el Desarrollo
▪ Encuesta de Hogares, 2006	Instituto Nacional de Estadística - Ministerio de Planificación para el Desarrollo
▪ Encuesta de Hogares, 2007	Instituto Nacional de Estadística - Ministerio de Planificación para el Desarrollo
▪ Encuesta de Hogares, 2008	Instituto Nacional de Estadística - Ministerio de Planificación para el Desarrollo
▪ Encuesta de Hogares, 2009	Instituto Nacional de Estadística - Ministerio de Planificación del Desarrollo
▪ Encuesta de Hogares, 2011	Instituto Nacional de Estadística - Ministerio de Planificación del Desarrollo
▪ Encuesta de Hogares, 2012	Instituto Nacional de Estadística - Ministerio de Planificación del Desarrollo
▪ Encuesta de Hogares, 2013	Instituto Nacional de Estadística - Ministerio de Planificación del Desarrollo
▪ Encuesta de Productividad de Empresas, 2007	David McKenzie - World Bank, Yaye Seynabou Sakho - World Bank
▪ Encuesta Nacional de Empleo I, 1996	Subdirección de Estadísticas Sociales - Instituto Nacional de Estadística / Departamento de Encuestas - Instituto Nacional de Estadística / Instituto Nacional de Estadística - Ministerio de Hacienda
▪ Encuesta Nacional de Empleo II, 1996	Subdirección de Estadísticas Sociales - Instituto Nacional de Estadística / Departamento de Encuestas - Instituto Nacional de Estadística
▪ Encuesta Nacional de Empleo III, 1997	Instituto Nacional de Estadística - Ministerio de Planificación para el Desarrollo
▪ Encuesta Nacional de Uso y Consumo de la Hoja de Coca en Hogares, 2009-2010	Instituto Nacional De Estadística - Ministerio De Planificación para el Desarrollo
▪ Encuesta piloto de Uso de Tiempo de los Hogares, 2010	Instituto Nacional de Estadística
▪ Latinobarómetro Survey, 2005	Corporación Latinobarómetro ONG
▪ Americas Barometer LAPOP, 2014	Latin American Public Opinion Project (LAPOP) - Vanderbilt University

Source: Own elaboration on the basis of the catalogs of INE and World Bank.
http://www.ine.gob.bo/anda/index.php/catalog
http://microdata.worldbank.org/index.php/catalog

Appendix 4.3

Ethnicity and cosmology

Cosmology, cosmogony, cosmovision or sometimes episteme refer to the general scheme of existence, the place of human beings on it and the forces engaged in the constitution and the generation of such a scheme. It could also be said that cosmologies are not only representations of world orders but practices so ingrained that they affect the dynamics of social formation and might even have implications in the structuring of social practices. In this regard, cosmologies could be seen as wider forms of knowledge in themselves or as an episteme.

In the case of the ethnic composition of Bolivia and the ethnic relations of its people it could be said that relations are not only relationships between ethnic groups but also between different cosmologies, epistemes and forms of knowledge. On the one hand, the western cosmology, anthropocentric, rational, bureaucratic, etc. and on the other, the Andean one, cosmo-centric, communitarian, nature respectful (human beings belong to the earth), etc. Some analysts, sometimes indigenous intellectuals, defend the concept of Indian cosmology rather than Andean because, they say, the Andean cosmology has been deeply affected by centuries of colonialism.

Even though official sources in Bolivia only recognise thirty odd indigenous nations, the afro-Bolivian slave-descended communities of the Yungas region and then the de-ethnicised and homogenised common Bolivians the truth is that *mestizaje* is a complex process that has developed along the centuries of cultural contact and colonialism. The many forms that creolisation and *mestizaje* takes in Bolivia might be related to processes of social mobility and cultural dislocation. Nevertheless, this process might operate in complex ways, indigenous Quechua or Aymara might abandon their language and costume, even identity to a homogenised Bolivian identity but they might retain rituals and beliefs shared by their cosmology. In the opposite direction, it is being perceived that European origin Bolivians have begun to chew coca, visit the traditional healer and to practice rituals and offers to the Pachamama (Mother Earth) in an open way. Many say that they used to do it in hiding. Although ethnic indigenous identity in Bolivia is strongly associated with Indian or Andean cosmology there are, at the same time, forms of unresolved cultural contact and creolisation taking place.

Appendix 4.4

2001 Census. Question on ethnicity

49. Do you consider yourself to belong to any of the following indigenous aboriginal peoples…

Quechua?	1
Aymara?	2
Guaraní?	3
Chiquitano?	4
Mojeño?	5
Other native?	6
None	7

Appendix 4.5

2012 Census. Question on ethnicity.

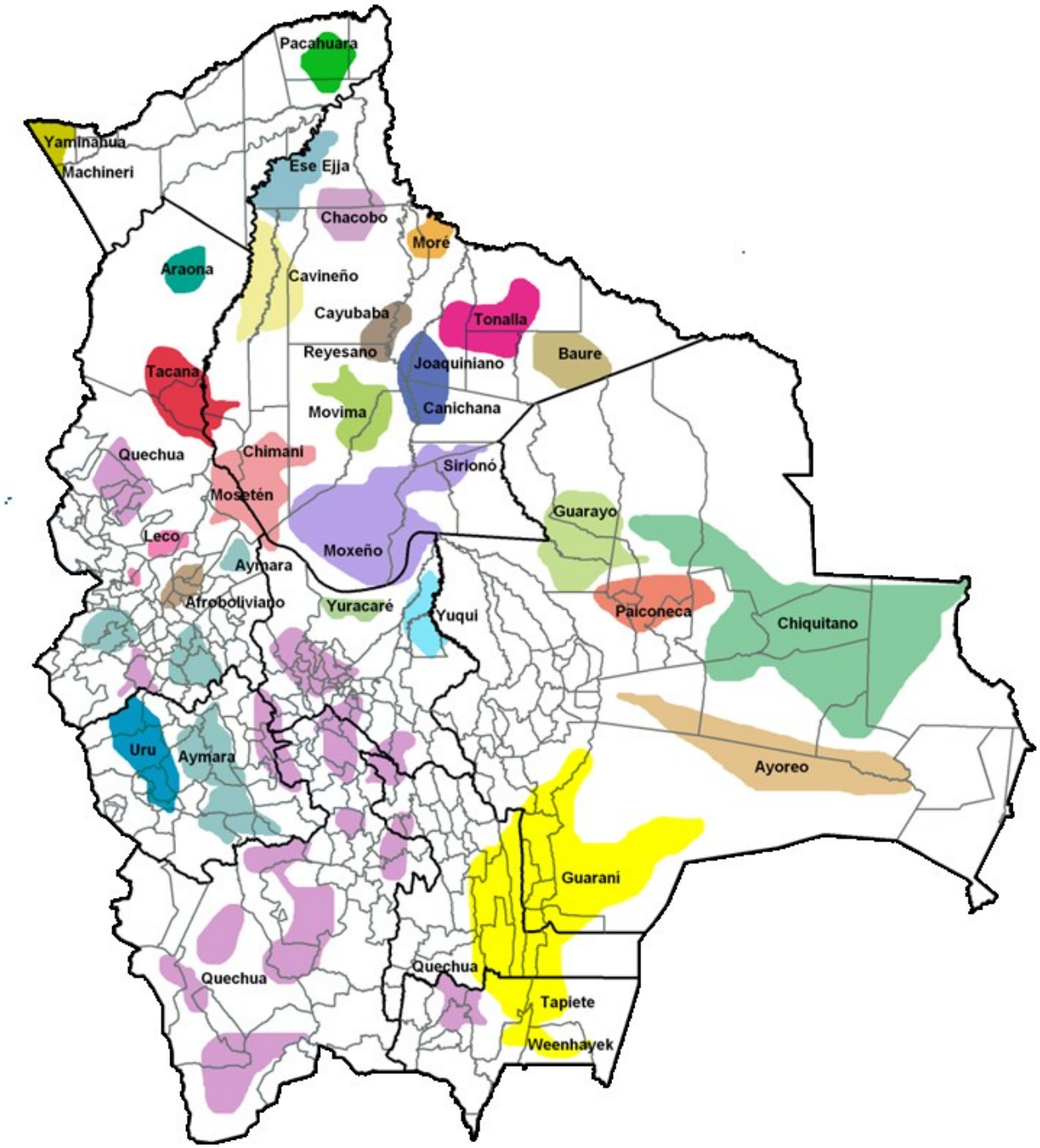
29. As a Bolivian, do you belong to any peasant aboriginal indigenous or Afro-Bolivian nation or people?
Yes
To which one?
Interviewer (do not read the answers)

(Names)

Do not belong
I am not a Bolivian

Appendix 4.6

Map of Bolivia's indigenous groups



Appendix 4.7

Constitution of Bolivia, section on indigenous and Afro-Bolivian peoples.

CHAPTER IV: Rights of the Nations and Rural Native Indigenous Peoples

Article 30
I. A nation and rural native indigenous people consists of every human collective that shares a cultural identity, language, historic tradition, institutions, territory and world view, whose existence predates the Spanish colonial invasion.

II. In the framework of the unity of the State, and in accordance with this Constitution, the nations and rural native indigenous peoples enjoy the following rights:

1. To be free.
2. To their cultural identity, religious belief, spiritualities, practices and customs, and their own world view.
3. That the cultural identity of each member, if he or she so desires, be inscribed together with Bolivian citizenship in his identity card, passport and other identification documents that have legal validity.
4. To self-determination and territoriality.
5. That its institutions be part of the general structure of the State.
6. To the collective ownership of land and territories.
7. To the protection of their sacred places.
8. To create and administer their own systems, means and networks of communication.
9. That their traditional teachings and knowledge, their traditional medicine, languages, rituals, symbols and dress be valued, respected and promoted.
10. To live in a healthy environment, with appropriate management and exploitation of the ecosystems.
11. To collective ownership of the intellectual property in their knowledge, sciences and learning, as well as to its evaluation, use, promotion and development.
12. To an inter-cultural, intra-cultural and multi-language education in all educational systems.
13. To universal and free health care that respects their world view and traditional practices.
14. To the practice of their political, juridical and economic systems in accord with their world view.
15. To be consulted by appropriate procedures, in particular through their institutions, each time legislative or administrative measures may be foreseen to affect them. In this framework, the right to prior obligatory consultation by the State with respect to the exploitation of non-renewable natural resources in the territory they inhabit shall be respected and guaranteed, in good faith and upon agreement.
16. To participate in the benefits of the exploitation of natural resources in their territory.
17. To autonomous indigenous territorial management, and to the exclusive use and exploitation of renewable natural resources existing in their territory without prejudice to the legitimate rights acquired by third parties.
18. To participate in the organs and institutions of the State.
III. The State guarantees, respects and protects the rights of the nations and the rural native indigenous peoples consecrated in this Constitution and the law.

Article 31
I. The nations and the rural native indigenous peoples that are in danger of extinction, in voluntary isolation and not in contact, shall be protected and respected with respect to their forms of individual and collective life.
II. The nations and the rural native indigenous peoples that live in isolation and out of contact enjoy the right to maintain themselves in that condition, and to the legal definition and consolidation of the territory which they occupy and inhabit.

Article 32
The Afro-Bolivian people enjoy, in everything corresponding, the economic, social, political and cultural rights that are recognized in the Constitution for the nations and the rural native indigenous peoples.

https://www.constituteproject.org/constitution/Bolivia_2009.pdf

Note: although the translation of the Bolivian Constitution used in this annex uses the expression 'rural native indigenous peoples' as a translation of *pueblo indígena originario campesino* in the report the expression used is 'peasant aboriginal indigenous people' which is deemed closer to the original.

Chapter 5: Ethnic group classification in Canada

Kelsey Lucyk, Karen Tang and Hude Quan

Abstract

Historically, Canada's changing policies on multiculturalism and immigration have influenced the collection of ethnic group data to become more inclusive and granular. Important periods include early attempts at nation-building during the late 19th century, social changes post-second war world II (WWII) and the introduction of multiculturalism into federal policy, and present-day efforts in ethnic classifications for research purposes and for preserving cultural diversity. As such, the collection of ethnicity and cultural data in Canada has evolved to include further granularity for common identities reported in national data sources, as well as to provide multi-cultural examples for respondents to more accurately capture their ethnic origins. There are four main sources of ethnicity data in Canada: 1) Provincial health insurance registries, 2) Canadian Health Measures Survey, 3) Canadian Community Health Survey, and 4) Census. Of these, ethnicity data are most limited in the provincial health insurance registries, flagging only Aboriginal status. The other three data sources are nationally administered, with all asking individuals to select, out of 11 categories, self-identified racial or ethnic groups. The questions on ethnic origin for the 2016 Census included citizenship, place of birth, immigration status, language, ancestry, ethnic origin, and Aboriginal status.

There are some lessons to be learned outside of the health field regarding the collection of ethnic data: (1) ethnic or racial orgins reported by individuals may not necessarily be the ethnic group with which they identify; (2) the length of time that individuals have been in Canada affects the strength of their identity with ethnic ancestry; and (3) a large proportion of Canadians have multiple ethnic groups of origins, which supports the need for follow-up questions to better understand how individuals best identify. In sum, there exists the need for greater granularity in ethnic classifications to reflect the diversity of the Canadian population. Because ethnicity is a socio-cultural concept, consideration should be made to incorporate questions about sense of belonging with the identified ethnic ancestry, rather than relying solely on reported ethnic origin and race.

Introduction: Demographic background of the country

Canada has a diverse population regarding its ethnic and racial composition, with 20.6% of its population born outside of the country and over 200 ethnic origins reported in 2011(Statistics Canada 2011). In Canada, ethnicity is captured according to the definition of *ethnic origin*, which refers to "the ethnic or cultural origins of the respondent's ancestors […that] are a reflection of each respondent's perception of their ethnic ancestry" (Statistics Canada 2015). This definition, which includes ethnic ancestry and ethnic group, implicitly includes race as a cultural classification. As shown later, racial group is often presented alongside ethnic group. This prevailing definition of ethnicity therefore differs slightly from the definition guiding this international report, as race is not specified as the physical features associated with certain populations. On the other hand, Canada does capture "visible minority" as a proxy for race, which refers to "persons, other than Aboriginal persons, who are non-Caucasian in race or non-white in colour" as designated by its *Employment Equity Act* (Statistics Canada 2015).

Prior to its colonization by European nations, Canada maintained a long history of diverse Aboriginal and Indigenous groups (Miller 2000). From the 16th to 18th Century, Europeans came into frequent contact with these groups for the purposes of trade, exploration, evangelism, and colonization (Miller 2000). By 1867, when Canada became a confederation, the country's population of 2.6 million was comprised predominantly of groups that had migrated from France, England, Wales, Ireland, and Scotland (Statistics Canada 2009). The implementation of the *Homesteading Act* in 1872 further facilitated migration from Europe and Scandinavia to Canada's agricultural lands, which resulted in ethnic bloc settlements across the central and Western provinces (Grenke 1991). As Canada industrialized during the late 19th Century, new groups (e.g., Chinese) migrated as temporary labourers to work in the country's industries (Grenke 1991). New cities also drew more Europeans and migrants from rural Canada into urban centres with the prospect of wage work.

As a young nation, Canada sought to build its identity as one that promoted the White race, and Anglo-Saxon, protestant values. The country's immigration policy reflected this; until the 1960s, immigration by non-Western Europeans was limited and individuals could be explicitly discriminated from immigrating to Canada on the basis of their ethnic identity (Verbeeten 2007). Within Canada, Aboriginal and Indigenous groups were subject to the government's systematic attempts to erase their ethnic identity, through restrictive policies that outlawed cultural practices and implemented assimilative residential schools (Miller 2000).

In the 1960s, Canada removed ethnicity as grounds for discrimination against prospective immigrants and implemented a points-based system whereby immigrants were assessed on their skills, education, and training (Verbeeten 2007). Since then, immigration policies have become increasingly inclusive. Groups from Southern Europe (e.g., Greece, Italy, Portugal) came to work in Canada's booming post-

war industries (e.g., manufacturing, mining) (Burnet J 2011). Since the 1970s, more immigrants from Asia have come to Canada more than any other group (Verbeeten 2007). Today, this trend continues. It is estimated that by 2031 up to 32% of the population will belong to a visible minority (non-White or non-Aboriginal) (Statistics Canada 2010).

Sources of data and their heterogeneity/granularity

Sources of data that report on the ethnic and racial composition in Canada include provincial health insurance registries, national health surveys, and Census data.

Provincial health registries do not ask specifically for race or ethnicity, but whether an individual is a new or returning resident to that province, and from which city and country he or she is moving from. Because the federal government fully funds health care for Aboriginal peoples, Aboriginal status is specifically flagged (Statistics Canada 2015).

On a national level, the Canadian Physical Measures Survey (CPMS) has collected demographic information and physical and laboratory measures of a representative sample biennially, since 2007 (Statistics Canada 2014). It excludes approximately 4% of the population, including residents of the territories or remote areas, on Aboriginal reserves, in institutions, and full-time members of the Canadian Armed Forces. Respondents are asked to choose up to four of the listed 11 racial or cultural groups to which they feel they belong (see table 5.1). They are also specifically asked whether they are Aboriginal, with further granularity as to whether they are First Nations, Metis, or Inuk.

The Canadian Community Health Survey (CCHS) is a cross-sectional survey of a nationally representative sample with approximately 65 000 respondents annually (Statistics Canada 2015). It includes residents from all provinces and territories, though maintains other exclusion criteria from CPMS. Interestingly, there are two separate questions regarding ethnicity and race. The first asks, “To which ethnic or cultural groups did your ancestors belong?”; there is no limit on the number of categories chosen. The second question asks, “You may belong to one or more racial or cultural groups on the following list. Are you…?”, with a maximum of four categories that can be chosen. Of note, there is much greater granularity, particularly for Caucasian ethnicities, for ancestral ethnicities compared with self-identified ethnic/racial groups.

Data about the racial and ethnic composition of Canada can also be obtained from the Census, performed every 5 years (Statistics Canada 2011). Prior to 2010, ethnicity was captured as part of the mandatory “Long Form Census”; although this form was eliminated in 2010. Since 2011, ethnicity has been captured in the “National Health Survey” (NHS) instead.

Table 5.1: Response categories for ethnic and ancestral groups from national health surveys and Census

Response Categories	Canadian Physical Measures Survey (2015)	Canadian Community Health Survey (2015)		Census (National Household Survey) 2011	
		Ethnic ancestry	Self-identified ethnic or racial group	Ethnic ancestry	Self-identified ethnic or racial group
White				Free text with the prompt to "Specify as many origins as applicable"	
South Asian					
Chinese					
Black					
Filipino					
Latin American					
Arab					
Southeast Asian					
West Asian					
Korean					
Japanese					
Other (specify with text)					
Aboriginal					
First Nations					
Metis					
Inuk					
Canadian					
French					
English					
German					
Scottish					
Irish					
Italian					
Ukrainian					
Dutch (Netherlands)					
Jewish					
Polish					
Portuguese					
Norwegian					
Welsh					
Swedish					

Unlike the Census, the NHS is voluntary (leading to potential for response bias), and does not include people living abroad or in collective dwellings. The NHS is completed by a random sample of nearly 1/3 of all households in Canada, or 4.5 million dwellings. It includes all provinces and territories, those living on Aboriginal reserves, permanent and non-permanent residents, and those on work or study permits (Statistics Canada 2011). Like the CCHS, the NHS asks about both the ethnic origins and self-identified ethnic/racial groups of participants. Though the NHS maintains the same level of granularity for self-identified ethnic/racial groups as the CCHS, the level of granularity for ethnic ancestry is far greater for the NHS, as there are no categories but rather space for respondents to write as many origins as applicable, with up to six origins are retained (Statistics Canada No date.). This information is then used to derive counts, percentage distribution, and relative ratios of ethnic groups for the analytical products put forth by Statistics Canada (Statistics Canada No date.). Data for ethnic groups are published if the count for a group is approximately 500 or higher (Statistics Canada No date.).

Lessons to be learned from outside the health field

The Ethnic Diversity Survey (EDS), was a one-time telephone interview survey conducted by Statistics Canada in 2003 with 42,476 respondents who were aged 15 years or older, lived in one of Canada's 10 provinces, and had completed the 2011 long form Census (Statistics Canada 2007). Persons living in Northern or remote areas or who claimed Aboriginal ethnic origins were not included. The objectives of the EDS was to understand how Canada's cultural, social, and economic life was influenced by people's background, and to better understand how individuals report their ethnicity (Statistics Canada 2007). The EDS revealed important lessons about ethnicity classifications, pertaining to how individuals identify with ethnicity (Statistics Canada 2003). While all individuals have ethnic ancestry, its importance to individuals varies substantially. For example, only half of adult respondents of the EDS described feeling a strong sense of belonging to their ethnic group, which also varied depending on the ethnic ancestry and length of time spent in Canada for individuals. Among first generation Canadians, a greater proportion of those who arrived in Canada more recently described a stronger sense of belonging with their ethnic group than those who arrived earlier.

In addition, ethnic classification is complicated by declarations of multiple ethnic groups of origin. Compared to first generation Canadians, a higher proportion of second, third, or more generations report multiple ethnic groups of origin (Jedwab 2008). One consideration, such as the method adopted by the EDS, is to ask individuals to rate the importance of each of their ethnic ancestries, to gain an understanding of the relative contributions of each ethnicity in shaping the identity of the individual.

In summary, lessons learned from outside the health field about ethnic classifications include:

1) Ethnic or racial origins reported by individuals may not necessarily be the ethnic group with which they identify. More relevant questions that better reflect ethnic identify should include their sense of belonging to ethnic group of origin and their views on the importance of passing on customs and traditions specific to these groups.
2) The length of time that individuals and their families have been in Canada affects the strength of their identity with ethnic ancestry; this information should therefore be collected.
3) Recognizing that a large proportion of the Canadian population reports having multiple ethnic groups of origin, follow-up questions to clarify the ethnic group with which individuals best identify, such as by having them rate the importance of each listed ethnic group, may help to simplify and streamline ethnic classifications of an increasingly ethnically-diverse population.

The development of heterogeneity/granularity in classifications

The first national Census of Canada was administered in 1871. The ninety-eight regional and colonial censuses that preceded it did not consistently enumerate ethnic classification; however, some did report on the race, religion, ethnic origin, and birthplace of respondents.(Statistics Canada 2015) The diversity of Aboriginal and Indigenous groups were recognized in the 1871 Census, but they were classified into just 4 races and 36 tribes (see table 5.2) (Statistics Canada 2015 [1876]).

Since 1871, information on ethnicity and race has been consistently collected (see table 2). Initially, this was recorded as the place of birth given by the person questioned, with 14 origins collected (Library and Archives Canada 2015). By 1881 it was recognized that a person's origin could be different from their country of birth (e.g., Indian) and so, beginning in 1891, the Census also enumerated province of birth and the birthplace of respondents' parents, as well as a new section on citizenship, nationality, and religion in 1901 (Library and Archives Canada 2015). In the year 1901 only, "Colour" was enumerated, which classified Canadians as White, Red, Black, and Yellow (Library and Archives Canada 2015).

Information on the birthplace of respondents and their parents, immigration, and language were asked throughout 1921 to 1971. In 1961, the Census began to include a question on ethnic/cultural group ancestry, which remained in place until 1981 when information specific to Aboriginal Registered/Treaty Indian was also enumerated (Stevens G 2015). Also in 1981, the Census introduced multiple write-in spaces for identifying ethnic origin, which has increased over time (i.e., 3 in 1986, 2 in 1991, and 4 in 1996 and 2001, 11 in 2006, and 4 in 2011) (Statistics Canada 2011, Statistics Canada No date.). Beginning in 1996, visible minorities were able to self-identify as such (Statistics Canada 2015).

In 2001, the Census updated and diversified the list of examples it used to define ethnic or cultural origins. Following this, in 2002 Statistics Canada and Canadian Heritage conducted the EDS, which collected information specific to the ethnicity and race of respondents, their parents and grandparents, and how they perceived various aspects of their ethnicity (Statistics Canada 2003). By 2011, the Census had evolved to include specific instructions for answering a question on ethnic origin (i.e., "What were

the ethnic or cultural origins of this person's ancestors?") and incorporated specific instructions and examples for Aboriginal and Indigenous respondents, in recognition of their diversity.(Statistics Canada 2011) Questions on ethnic origin (i.e., citizenship, place of birth, immigration status, language, ancestry, ethnic origin, Aboriginal status) remain in place for the upcoming 2016 Census. Respondents of the 2016 Census will also be asked ethnicity in a way that pertains to race. See table 5.2 for further details (Statistics Canada 2015).

Why has the heterogeneity/granularity in these classifications developed?

The heterogeneity and granularity of ethnic classifications has been substantially affected by the social, historical, and political context of Canada, particularly in the 19th and 20th Centuries. From Confederation until the WWII, while Canada underwent a period of nation building in an attempt to establish itself as a civilized country of white, Anglo-Saxon citizens instilled with British, Victorian, protestant cultural values (Mackey 1998). As reflected in government policies on immigration and Indians, attempts were made to assimilate and devalue alternative ethnic identities as inferior to the white Canadian.

Immigration increased in Canada following WWII, which brought new ethnic groups, intermarriages between groups, and the birth of multi-origin individuals. This diversity is reflected in changes to the Census at the time; 30 options for ethnicity were included in 1961 Census, while only 18 options had been available in 1951 (Stevens G 2015). Also during the post-war period, the Royal Commission on Bilingualism and Biculturalism in Canada was conducted, which brought the concept of multiculturalism to the forefront of the Canadian identity, before it was adopted as federal policy in 1971 (Burnet J 2011). Since the 1982 Charter of Rights and Freedoms, ethnic diversity – including the rights of Aboriginal and Indigenous groups – has been increasingly protected and celebrated (Burnet J 2011). For example, treaty rights (i.e., constitutionally recognized agreements between Canada and Aboriginal peoples regarding land, payments, and other commitments) were affirmed in the Constitution Act of 1982 (Bell 2006). Aboriginal groups have since challenged the Supreme Court of Canada to have their claims recognized in relation to ancestral or traditional lands, customary practices (e.g., marriage and adoption, fishing and hunting rights), and food harvesting, among other issues (Bell 2006). The adoption of multiculturalism brought the need to better understand the origins and ethnic diversity of Canadians, which resulted in "an explosion of multicultural research" (Burnet J 2011).This is reflected by the increasingly inclusive and expansive data that is collected in the Census and the administration of supplementary surveys that collect ethnic data. Today, multiculturalism informs the collection and analysis of ethnic data for research and decision-making; specifically, those that aim to reduce health inequalities between groups.

Table 5.2: History of Heterogeneity in National Data sources (primarily the Census)

Year	Data Source. Event
1666	Census. First census administered by Jean Talon
1765	Colonial Census. Records number of Acadians, settlers, Aboriginals, and Blacks living in New France
1844	Census of Lower Canada. Includes question asking the "birth place of the people"
1871	Census. First Census of the Dominion of Canada.
	Census. Collects information on place of birth (United Kingdom, Europe, Asia, Other [i.e., Native Indian, Inuit, Negroes]), with 14 origins collected
	Census. Bureau of Statistics compiles historical statistics on the population of Canada for the pre-1871 era from memoirs or traders, missionaries, and settlers
	Census. Groups Aboriginals into 4 races (i.e., Esquimaux or Innoït, Dénè-Dindjié, Algonquins, and Huron-Iroquois) and list 36 tribes throughout Canada, while recognizing that it would be impossible to make note of all tribes and sub-tribes
	Census. "Other" origin included Native Indian and Inuit, Negroes, and others
1881	Census. Recognizes that origin could be a country name different from one's place of birth
	Census. Indian included as an origin.
1891	Census. Collects country or province of birth for respondent, father, and mother.
1901	Census. Collects "Colour" of respondents: W – white (European descent), R – red (Native Canadians), B – black (African descent), Y – yellow (Japanese or Chinese descent). Children born of mixed European race were designated members of a non-white race.
	Census. Begins to collect "Citizenship, nationality and religion," which includes respondent's country or place of birth, year of immigration, year of naturalization, racial or tribal origin, and nationality
	Census. Terms "breed" and "half-breed" were used to refer to respondents of mixed Native ancestry.
	Census. Collects information on language(s) spoken at home.
1911	Census. Colour of respondents removed.
	Census. Information on language collected (English, French, or write-in alternate language).
1921	Census. No change from 1911
1951	Census. Question on origin
1956 (not national)	Census. No question on origin
1961	Census included questions on ethnic/cultural group ancestry
1966 (not national)	Census. No question on origin
1971	Census. Included questions on ethnic/cultural group
1981	Census. Included questions on ethnic/cultural group
	Census. New question included on Aboriginal Registered/Treaty Indian
	Census. Maternal ancestry (not just paternal) was collected for non-Aborigianl respondents.
	Census. Multiple write-in spaces were included to specify as many ethnic groups as applicable.

1986	Census. Write-in spaces for ethnic origins reduced to 3 additional classifications.
1991	Census. Write-in spaces for ethnic origins reduced to 2 additional classifications.
	Census. Question on ethnic origins includes the word "ancestor" (i.e., "To which ethnic or cultural group(s) did this person's ancestors belong?").
1996	Census. Write in spaces for ethnic origins increased to 4 additional classifications with up to 6 captured.
	Census. Included question on visible minority. Prior to this, language and origins was used to assign visible minority status by Statistics Canada.
Year	**Data Source. Event**
2001	Census. List of examples used to define ethnic origin was updated
2002	Ethnic Diversity Survey. Statistics Canada and Canadian Heritage conduct Ethnic Diversity Survey, which includes questions on: ethnic ancestry; heritage and background; immigration status; parents', grandparents', and spouse's background; family interaction, discrimination, language, and others.
2006	Census. Ethnic origin recorded in 11 segmented boxes.
2011	National Household Survey. Includes question on ethnic origins, with up to 4 lines available to write-in additional ethnic origins with 6 classifications captured.
2011	National Household Survey. Specifies difference between language group and ethnic group and ancestry.
2016	National Household Survey. Includes specific examples and instructions for respondents of Aboriginal ethnic origins.
	National Household Survey. Asks home language, mother tongue, first language spoke.
	Census. Includes reference guide specific to Aboriginal respondents, which includes questions on Aboriginal identity, Aboriginal group, Registered or Treaty Indian status, Membership in a First Nations or Indian band, Aboriginal ancestry (e.g., report 'Cree,' 'Mi'kmaq,' etc. and not 'Indian'), Area of residence.
	Census. Incorporates specific instructions for answering questions related to ethnic origin (e.g., 'report East Indian from India' not 'Indian').
	Census. Asks questions related to ethnic origins of respondents (e.g., born inside Canada or outside of Canada, citizenship or if they are an immigrant, languages spoken, ethnic or cultural origins of this person's ancestors)
2016	Census. Respondents are permitted to include as many additional write-in ethnic origins as applicable.
	Census. Includes questions specific to Aboriginal respondents, same as in 2011
	Census. Includes question where respondents are asked to list if they are: 1: White, 2: South Asian (e.g., East Indian, Pakistani, Sri Lankan, etc.), 3: Chinese, 4: Black, 5: Filipino, 6: Latin American, 7: Arab, 8: Southeast Asian (e.g., Vietnamese, Cambodian, Laotian, Thai, etc.); 9: West Asian (e.g., Iranian, Afghan, etc.), 10: Korean, 11: Japanese, 12: Other – specify.

Why are disaggregated data not being collected, analysed, or reported more often

Though disaggregated data on ethnicities are important to understanding disparities, there remain numerous barriers that hinder their collection. In Canada, health care is administered at the provincial/territorial level. Therefore, because there is no federal mandate to collect ethnic data at the health care organizational level, each province and territory must individually determine whether, and to what detail, these data are necessary. While certain jurisdictions collect ethnic data in clinical contexts or through health insurance registries, other jurisdictions oppose its collection, arguing that it infringes upon the population's rights and freedoms (Browne, Varcoe et al. 2014). The fragmentation of provincial administration and delivery of health care, along with the different values of provinces limit the comprehensive collection of disaggregated data in health care.

Another barrier to the collection of disaggregated data is the potential for harm. For example, while nearly 85% of respondents surveyed in the province of Alberta indicated that they felt comfortable having their ethnicity recorded in hospital charts (Quan, Wong et al. 2006), one study suggests there is concern, particularly among leaders of ethno-cultural communities, about the discrimination against and reinforcement of ethnic divisions in asking individuals to report their ethnicity in health care settings (Varcoe, Browne et al. 2009). Furthermore, the collection of ethnicity data is valuable to the extent that it can be used to improve the health or quality of care received by individuals of different ethnicities. The usefulness and need for disaggregated data to ensure equitable care is not a universally held perspective.

Lastly, because ethnicity is a complex and hard-to-measure concept, its usefulness may be questioned. For example, a majority of Canadians report their ethnicity is "Canadian," often without identification of any other ethnic ancestry. These responses do not provide detail regarding cultural traditions, spoken languages, experiences with discrimination, or inequalities of health care access. Therefore, the lack of granularity and potential lack of usefulness of disaggregated data have been cited as reasons not to collect disaggregated data (Browne, Varcoe et al. 2014).

Are there any examples of how the disaggregated has been used to impact on policies, programmes, and population health outcomes?

The Healthy Foods North intervention is one example of how disaggregated data have been used to develop health promotion programs. The Inuit population in northern Canada has experienced changes in their health status over time, with a rise in cardiovascular risk factors and chronic diseases (Bjerregaard, Young et al. 2004, Statistics Canada 2015). This shift can be partially attributed to a

transition in Inuit nutritional intake from traditional animal and high protein foods (e.g., caribou, hare, berries, fish) to foods more reflective of Canadian culture that are higher in fats and refined carbohydrates (Sharma 2010). Healthy Foods North was a 1-year program, consisting of store and community educational interventions that promoted a traditional Inuit diet and increased physical activity, with the goal of reducing the risk of obesity and cardiovascular risk factors (Sharma, Gittelsohn et al. 2010, Public Health Agency of Canada 2012). The program was considered successful, as intention to eat healthier foods and food related self-efficacy increased. However, post-intervention programs remain unimplemented due to lack of funding (Northern News Services 2010).

Another example of an intervention that has been developed based on disaggregated data is mobile screening for breast cancer in Chinese women. In Canada, over 70% of women between the ages of 50 and 69 years reported having had a mammogram within the past 2 years (Shields and Wilkins 2009, Shields and Wilkins 2015). However, there are significant disparities where over half of immigrant women (compared to 26% of modern Canadian born women) did not receive mammograms (Shields and Wilkins 2009). Of the immigrant women, those born in Asian countries had the highest rates of non-use. To address this health disparity, a partnership formed between the Chinese Community Service Association, Alberta Cancer Board, former Calgary Health Region, and Canadian Cancer Society developed culturally sensitive educational materials in Chinese to educate Chinese women living in Calgary, Alberta, about the importance of breast cancer screening. In addition, mobile mammography was offered via a van sent to Calgary's Chinatown to increase breast cancer screening in this community (CBC News 2005). These examples underscore the importance of disaggregated data in understanding how health care access and outcomes differ by ethnicity. Moreover, targeted interventions that tackle health disparities may be most effective where they address the social determinants of health (e.g., culture); such interventions are not possible without disaggregated data.

While it is difficult to measure the policy impacts of disaggregated data sources, Census and NHS data are routinely used to inform federal and provincial health decision-making. The Public Health Agency of Canada (PHAC), for instance, uses data from the census to calculate rates of disease among populations and also to design interventions for populations most vulnerable to certain diseases (Statistics Canada 2015). One specific example of how disaggregated data may come to influence health policy is reflected in the recent announcement by the Government of Canada to increase funding of community-based mental health and addictions programming on-reserve and in territories by $69 million dollars (Government of Canada 2016). This announcement, made in June 2016 by Prime Minister Justin Trudeau, came after repeated reports published by PHAC that found higher rates of depression, suicide, alcoholism, and other mental health outcomes for Aboriginal population groups (e.g., The Human Face of Mental Health and Mental Illness in Canada 2006) (Government of Canada 2006). PHAC currently promotes a number of culturally appropriate interventions for Aboriginals that may tackle these issues, such as the Makimautiksat Youth Wellness and Empowerment Camp in

Nunuvat, where children who attended the 10-day camp showed improved self-esteem, wellness, pride in Inuit culture, enhanced peer support, and coping abilities (Public Health Agency of Canada 2016).

Are there ongoing developments to improve current systems of classifications?

The re-instatement of the long-form census in 2016 marks a significant event for data collection in Canada. The long-form census, which includes questions regarding ethnicity (see table 2), was re-instated through the actions of engaged citizens, scientists, organizations, and the recently-elected Liberal government responsible for overturning the decision of the former Conservative government to eliminate the mandatory long-form census (Campion-Smith 2015, Canadian Library Association 2015, Canadian Sociological Association 2015, Liberal Party of Canada 2016). Additionally, in their 2015-2016 Report on Plans and Priorities, Statistics Canada included developments in the socio-economic statistics program, which is used to help guide decision-making in Canada (Statistics Canada 2015). The 2013 General Social Survey was administered with the theme of Social Identity, and collected data on Canadians' sense of belonging to national, ethnic, geographic, and cultural groups to "help build national measures and support policies on the inclusion and diversity of people living in Canada" (Statistics Canada 2015). The results from this survey are expected to be released mid-2016.

References

Bell, C., Henderson, W (2006) "Rights of Indigenous Peoples."

Bjerregaard, P., T. K. Young, E. Dewailly and S. O. Ebbesson (2004). "Indigenous health in the Arctic: an overview of the circumpolar Inuit population." Scand J Public Health **32**(5): 390-395.

Browne, A. J., C. M. Varcoe, S. T. Wong, V. L. Smye and K. B. Khan (2014). "Can ethnicity data collected at an organizational level be useful in addressing health and healthcare inequities?" Ethn Health **19**(2): 240-254.

Burnet J, D. L. (2011). "Multiculturalism." Historic Canada Retrieved February 14, 2016, from http://www.thecanadianencyclopedia.ca/en/article/multiculturalism/.

Campion-Smith, B. (2015) "Canada's long-form census is back for 2016." The Toronto Star.

Canadian Library Association. (2015). "CLA continues to advocate with the federal government regarding the Long Form Census and funding for StatCan." Retrieved February 19, 2016, from http://cla.ca/cla-at-work/advocacy/long-form-census-and-statistics/.

Canadian Sociological Association. (2015). "Long form census." Retrieved February 19, 2016, from http://www.csa-scs.ca/blog/long-from-census.

CBC News. (2005). "Chinese immigrants offered breast cancer screening info in mother tongue." Retrieved January 19, 2016, from http://www.cbc.ca/news/technology/chinese-immigrants-offered-breast-cancer-screening-info-in-mother-tongue-1.566764.

Government of Canada (2006). The Human Face of Mental Health and Mental Illness in Canada 2006. Ottawa, ON.

Government of Canada (2016). Government of Canada to provide immediate support for indigenous mental wellness. Ottawa, ON.

Grenke, A. (1991). The German Community in Winnipeg 1872 to 1919. New York, AMS Press.

Jedwab, J. (2008). "The rise of the unmeltable Canadians? Ethnic and national belonging in Canada's second generation." Canadian Diversité **6**(2): 25-34.

Liberal Party of Canada. (2016). "Bring back the mandatory long-form census." Retrieved February 19, 2016, from https://www.liberal.ca/petitions/bring-back-the-mandatory-long-form-census/.

Library and Archives Canada. (2015). "About the Federal Census of 1871." Retrieved 2016, February 14, from http://www.bac-lac.gc.ca/eng/census/1871-on/Pages/about-census.aspx.

Mackey, E. (1998). House of difference: Cultural politics and national identity in Canada. . Florence, KY, Routledge.

Miller, J. (2000). Skyscrapers hide the heavens: A history of Indian-White relations in Canada. Toronto, University of Toronto Press.

Northern News Services. (2010). "Healthy Foods North shuts down." Retrieved January 19, 2016, from http://www.nnsl.com/frames/newspapers/2010-09/sep20_10hf.html.

Public Health Agency of Canada. (2012). "Healthy Foods North – A culturally appropriate and community-based program to promote healthy eating and lifestyle." Retrieved January 19, 2016, from http://www.phac-aspc.gc.ca/publicat/2009/be-eb/nunavut-eng.php.

Public Health Agency of Canada. (2016). "Makimautiksat Youth Wellness and Empowerment Camp, Nunuvat." Canadian Best Practices Portal, Aboriginal Ways Tried and True Retrieved August 4, 2016, from http://cbpp-pcpe.phac-aspc.gc.ca/aboriginalwtt/makimautiksat-youth-wellness-and-empowerment-camp-nunavut/.

Quan, H., A. Wong, D. Johnson and W. A. Ghali (2006). "The public endorses collection of ethnicity information in hospital: implications for routine data capture in Canadian health systems." Healthc Policy **1**(3): 55-64.

Sharma, S. (2010). "Assessing diet and lifestyle in the Canadian Arctic Inuit and Inuvialuit to inform a nutrition and physical activity intervention programme." J Hum Nutr Diet **23 Suppl 1**: 5-17.

Sharma, S., J. Gittelsohn, R. Rosol and L. Beck (2010). "Addressing the public health burden caused by the nutrition transition through the Healthy Foods North nutrition and lifestyle intervention programme." J Hum Nutr Diet **23 Suppl 1**: 120-127.

Shields, M. and K. Wilkins (2009). "An update on mammography use in Canada." Health Rep **20**(3): 7-19.

Shields, M. and K. Wilkins. (2015). "An update on mammography use in Canada." Retrieved January 19, 2016, from http://www.statcan.gc.ca/pub/82-003-x/2009003/article/10873-eng.htm.

Statistics Canada (2003). Ethnic Diversity Survey: Portrait of a multicultural society. Ottawa**:** 26.

Statistics Canada. (2007). "The Ethnic Diversity Survey (EDS)." Retrieved August 4, 2016, from http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4508.

Statistics Canada. (2009). "Canadian statistics in 1867." Retrieved February 14, 2016, from http://www65.statcan.gc.ca/acyb07/acyb07_0002-eng.htm.

Statistics Canada (2010) "Study: Projections of the diversity of the Canadian population." The Daily.

Statistics Canada (2011). "Ethnic origin reference guide, 2006 Census."

Statistics Canada. (2011). "National Household Survey (NHS)." Retrieved January 16, 2016, from http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5178.

Statistics Canada. (2011, May 8). "National Household Survey: Immigration, place of birth, citizenship, ethnic origin, visible minorities, language and religion." The Daily, from http://www.statcan.gc.ca/daily-quotidien/130508/dq130508b-eng.htm.

Statistics Canada. (2014). "Canadian Health Measures Survey (CHMS)." Retrieved January 10, 2016, from http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=148760.

Statistics Canada. (2015). "2015-2016 Report on Plans and Priorities. Program 2: Socio-economic statistics." Retrieved February 19, 2016, from http://www.statcan.gc.ca/eng/about/rpp/2015-2016/s02p01-2.

Statistics Canada. (2015). "2016 Census of Population questions, long form (National Household Survey." Retrieved February 14, 2016, from https://www12.statcan.gc.ca/nhs-enm/2016/ref/questionnaires/questions-eng.cfm.

Statistics Canada. (2015). "Canadian Community Health Survey (CCHS) - 2015." Retrieved January 10, 2016, from http://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&Item_Id=238890&TET=1.

Statistics Canada. (2015). "The General Social Survey: An overview." Retrieved February 19, 2016, from http://www.statcan.gc.ca/pub/89f0115x/89f0115x2013001-eng.htm#a4.

Statistics Canada. (2015). "History of the Census of Canada." Retrieved February 14, 2016, from https://www12.statcan.gc.ca/census-recensement/2011/ref/about-apropos/history-histoire-eng.cfm.

Statistics Canada. (2015). "Immigration and Ethnocultural Diversity in Canada." Retrieved August 3, 2016, from https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-010-x/99-010-x2011001-eng.cfm#a4.

Statistics Canada. (2015). "Permanent Resident Landing File." Retrieved January 10, 2016, from http://www.statcan.gc.ca/eng/rdc/data/prlf.

Statistics Canada. (2015). "Research on Census Program data uses and requirements." Retrieved August 4, 2016, from https://www12.statcan.gc.ca/census-recensement/fc-rf/reports-rapports/R1_Part3-Partie3-eng.cfm#A1.

Statistics Canada. (2015 [1876]). "Aboriginal peoples." Retrieved February 14, 2016, from http://www.statcan.gc.ca/pub/98-187-x/4151278-eng.htm.

Statistics Canada. (No date.). "Ethnic origin reference guide, National Household Survey, 2011." Retrieved February 14, 2016, from https://www12.statcan.gc.ca/nhs-enm/2011/ref/guides/99-010-x/99-010-x2011006-eng.cfm.

Stevens G, I. H., Grbic D (2015). "Measuring race and ethnicity in the censuses of Australia, Canada, and the United States: Parallels and paradoxes." Canadian Studies in Population **42**(1-2): 13-34.

Varcoe, C., A. J. Browne, S. Wong and V. L. Smye (2009). "Harms and benefits: collecting ethnicity data in a clinical context." Soc Sci Med **68**(9): 1659-1666.

Verbeeten, D. (2007). "The past and future of immigration to Canada." International Migration & Integration **8**: 1-10.

Chapter 6: Country of birth classification in Denmark

Liv Stubbe Østergaard, Allan Krasnik

Abstract

Since 1850 Statistics Denmark has been the central authority on Danish statistics including demographic data. Statistics Denmark does not register ethnicity directly, but does identify origin based on country of birth and ancestry (parents' country of birth and citizenship). The definitions and classifications of immigrants and descendants are solely Danish definitions and the data used in these statistics derives from the Central Person Registry (CPR Registry). Statistics Denmark's definition of origin divides the population into three groups: Persons of Danish origin (country of birth does not matter, but at least one parent holds a Danish citizenship and is born in Denmark), Immigrants (born abroad, none of the parents hold Danish citizenship and are born in Denmark), Descendants (born in Denmark, none of the parents hold Danish citizenship and are born in Denmark). Furthermore, country of birth is divided into Western/Non-Western countries. National registers on disease and healthcare usage in Denmark do not routinely include data on ethnicity or country of birth. However, these registers can be linked to other registers in Statistics Denmark by using the personal ID number, the CPR-number. Linking between population registers and registers on disease and healthcare provides relatively good opportunities for studying relations between migration, ethnicity (defined as country of birth or parents' country of birth) and health. Research on ethnic minorities and health in Denmark largely use Statistics Denmark's definitions of persons of Danish origin, immigrants and descendants. In sub-sequent categorizations and variables, the focus on country of birth, ancestry and citizenship varies, but country of birth (own or parents') grouped in broad categories is most often used as the main variable.

Measures such as country of birth will always be rough proxies of complex mechanisms; hence, in combination with other data such as language and socio-economic status, it may present a more valid measure of ethnicity. However, country of birth is mainly used because it is easily accessible and easy to make operational, whereas routinely collected data on ethnic affiliation is considered difficult, costly and time consuming, and like religious affiliation is considered as sensitive data which is not included in national registers because of the risk of abuse, discrimination and violation of privacy.

Introduction: Demographic background of the country

Since 1850 Statistics Denmark has been the central authority on Danish statistics including demographic data. Statistics Denmark does not collect data on ethnicity, but about origin, based on country of birth and ancestry (parents' country of birth and citizenship). Based on this information the Danish population is divided into three groups: Persons of Danish origin, Immigrants and Descendants. Furthermore, country of birth is divided into Western/Non-Western countries.

Traditionally, Denmark is regarded as an ethnically homogenous country with a strong sense of national identity and approximately 90% of the population being ethnic Danes. Since the late 1960s migration to Denmark has increased, led by immigrants from non-Western countries. Thus, Denmark's position as a net-receiver country with an increasingly ethnically mixed composition is relatively new (Migration Policy Institute 2006, Aarhus University 2016).

Before 1900, foreigners who migrated to Denmark mainly came from other Nordic and Western countries with cultures and languages close to the Danish[1]. During the first half the of the 19th century refugees came as an effect of persecution and the two world wars, mainly Eastern European Jews and Germans, but also Polish agricultural workers. In a short period from 1960s until 1973 guest workers were invited to the country from Turkey, Pakistan, Yugoslavia and Morocco. Eventually, they were allowed residence, and family reunifications came to account for a large part of the immigration in the 1980s and 1990s. With Denmark's membership of the European Economic Community in 1973 (later the EU) a substantial number of work migrants, especially from Poland and the Balkans, came to the country. Since the 1980s there has been an increase in refugees from Middle Eastern countries such as Iran, Iraq and Afghanistan as well as African countries such as Somalia and Ethiopia (Statistics Denmark 2015).

In January 2015 the total population in Denmark was 5.7 million people. Immigrants and descendants comprised 11.6 % of the total Danish population (657,471 persons) – about 8.9% were immigrants and 2.8 % descendants (see definition later). Statistics Denmark identifies 423,260 people as originating from non-Western countries (see definition later), which amounts to 64.4% (see table 6.1) of the immigrant population. The largest group of immigrants and descendants came from Turkey, namely 61,634 persons or 9.4 % of all immigrants and their descendants. Poland and Germany follow with respectively 39,465 and 31,962 immigrants and descendants in Denmark (see figure 6.1).

[1] Historically, the Kingdom of Denmark has included Sweden, Norway, the Faroe Islands, Greenland, Iceland, the Shetlands and the Orkney Islands under the Union of Kalmar. Adding to this, Denmark has reigned over the Duchies Schleswig, Holstein and Lauenburg close to the German border until 1864.

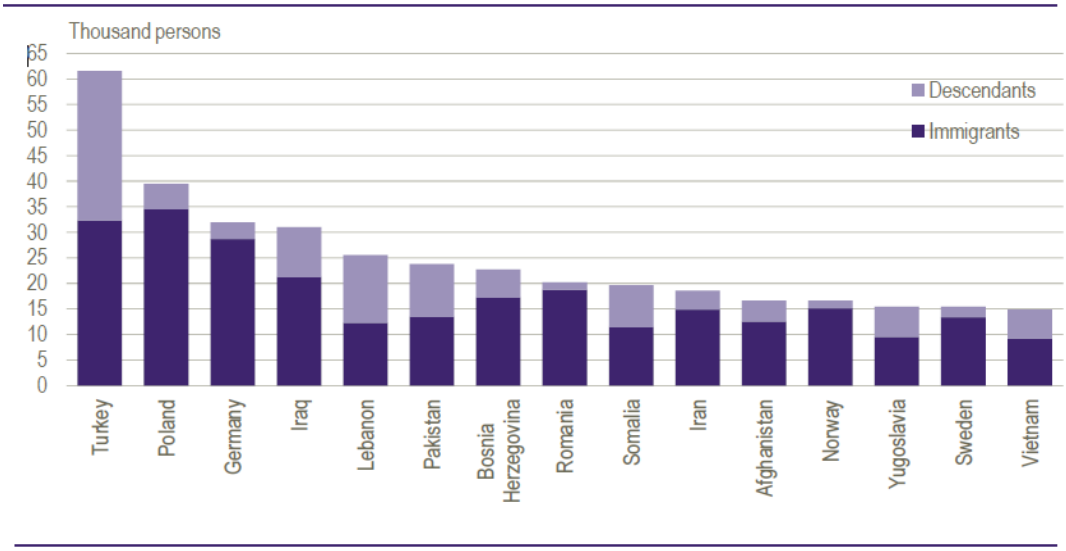
The only formally recognized *national minority group* in Denmark is the German minority group living in North Schleswig close to the German border (app. 15,000), who holds Danish citizenship but self-identify as Germans and use German as their language (The Border Union for a Danish Openness (Grænseforeningen for en Åben Danskhed) 2015). Given their status as a minority group they are guaranteed certain cultural and linguistic rights. Greenlanders and people from the Faroes are not officially recognized as minority groups in Denmark, which has spurred criticism as especially Greenlanders face linguistic, cultural and social challenges and barriers for equal treatment in the Danish society in line with other minority groups (Danish Institute for Human Rights 2015). The Inuit people in *Greenland* are the only *recognized indigenous* people[2] (Committee on the Elimination of Racial Discrimination 2015), and their rights are closely linked to the introduction of a self-government system in Greenland and thereby linked to conditions defined by the self-government for Inuit living in Greenland (Danish Institute for Human Rights 2015).

Table 6.1: Immigrants and descendants from Western and non-Western countries 1. January 2015

	Immigrant	Descendant	Total
Western countries	210 724	23 489	234 213
Non-western countries	290 333	132 927	423 260

www.statbank.dk/folk1

Figure 6.1: Immigrants and descendants by country of origin 1. January 2015



www.statbank.dk/folk1

[2] Committee on the Elimination of Racial Discrimination. 2015, Section 21. The rights of national and ethnic minorities are defined in 'UN Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities' (1992) and EU 'Framework Convention for the Protection of National Minorities' (1995).

Sources of data and their heterogeneity/granularity?

Since 1850 Statistics Denmark has been the central authority on Danish statistics including demographic data (Statistics Denmark 2000). Statistics Denmark does not register ethnicity directly, but does identify origin based on country of birth and ancestry (parents' country of birth and citizenship). The definitions and classifications of immigrants and descendants are solely Danish definitions (Statistics Denmark 2016), and the data used in these statistics derives from the Central Person Registry (CPR Registry[3]). The data in the CPR registry derives from different authorities and the Registry contains information on name, address, civil status, place of birth, citizenship, ancestry and whether you belong to the Danish Folk Church. Data on country of birth is derived from signing up to the civil registry in the municipalities. All persons are assigned a personal ID number, consisting of ten numbers displaying birth date and sex.

Statistics Denmark's definition of origin divides the population into three groups (Statistics Denmark 2016):

- *Persons of Danish origin* (country of birth does not matter, but at least one parent holds a Danish citizenship and is born in Denmark)
- *Immigrants* (born abroad, none of the parents hold Danish citizenship and are born in Denmark)
- *Descendants* (born in Denmark, none of the parents hold Danish citizenship and are born in Denmark)

Children of descendants (3rd generation) were added as a separate group in 2007 and are defined as children born in Denmark where at least one parent is a descendant and none of the parents are of Danish origin(Statistics Denmark 2016).
Children of descendants were introduced as a separate category in 2007 because of political interest and demand for statistics on this group. The group is comprised of parts of the 'persons of Danish origin' and 'descendants' groups (depending on whether one of their parents holds Danish Citizenship)(Danish Ministry of Refugees Migrants and Integration 2009).

Ancestry and ethnic background are not synonyms, since the former solely concerns the parents' country of origin and citizenship, and therefore is based on geographical and legal criteria, while the latter rather is a question of cultural background and identity. There may well be people who in Statistics Denmark are listed as persons of Danish origin, but are perceived and perceive themselves as people with a different ethnic background than Danish. Furthermore, people in the descendant group can have a very

[3] The CPR Registry contains basic information about all people living in Denmark with a personal ID number (CPR number) such as country of birth, citizenship and parents' country of birth.

remote attachment to their ancestral national and cultural affiliation. The main advantage of the classification by parents' place of birth lies in the simplicity of its collection and, even more decisively, the fact that by staying in the 'migration paradigm', this categorization avoids explicit ethnic enumeration. However, the system might be unsatisfactory when it comes to monitoring ethnic discrimination against the so-called third and fourth generations and even more so if race or religion actively shapes prejudices and stereotypes (Simon 2012).

The statistics provide information on numbers, age, sex, citizenship, country of origin and geographical distribution (Statistics Denmark 2016). Countries of origin are listed separately and can be grouped by continents[4] and Western/non-Western countries, where Western countries are comprised of EU28 countries, Andorra , Australia, Canada , Iceland, Liechtenstein, Monaco, New Zealand , Norway, San Marino , Switzerland, USA and Vatican City and non-Western countries comprise the rest.

People from Greenland or the Faroe Islands are not depicted as an individual group in the statistics, as they hold Danish citizenship and are thus included as Danes. However, since 2008 Statistics Denmark has registered the number of people living in Denmark and born in Greenland or the Faroe Islands. This means that people born in Denmark by e.g. parents from Greenland are not included in the numbers of Greenlanders in Denmark (even though they have different features than Danes and probably different cultural background). Different reports of Greenlanders in Denmark use different definitions of what constitutes a Greenlander. Typical criteria are that a persons' place of birth must be Greenland combined with one or both parents and grandparents also being born in Greenland (Danish Institute for Human Rights 2015).
Since 1989 Statistics Denmark has also published statistics on number of asylum seekers, with information on citizenship and type of application, and residence permits granted to refugees and family reunification immigrants based on data from the Danish Immigration Service (The National Information Services (Statens Informationstjeneste) 1991).

National registers on disease and healthcare usage in Denmark do not routinely include data on ethnicity or country of birth. However, these registers can be linked to other registers in Statistics Denmark by using the personal ID number, the CPR-number(Thygesen, Daasnes et al. 2011). Linking between population registers and registers on disease and healthcare provides relatively good opportunities for studying relations between migration, ethnicity (defined as country of birth or parents' country of birth) and health(Norredam, Kastrup et al. 2011).

[4] Continents are divided into EU-28, Europe outside EU-28, Africa, North America, South and Central America, Asia, Oceania, Stateless.

Lessons to be learned from outside the health field

Until 2008 students with an immigrant background were identified based on linguistic criteria in Danish education statistics(OECD 2010). However, Statistic Denmark's classification system is the main system used in reports on ethnicity, work, criminality, education etc.[5] These analyses often depict only Western versus non-Western immigrants and descendants, but for specific purposes more refined groupings are used based on country of origin or groupings of countries into larger regions.

The development of heterogeneity/granularity in classifications

Denmark and the Nordic countries have a long history of collecting information on births, deaths, immigration and emigration, disease incidence etc.[6] Information from the historical censuses shows the development in variables included in population registers where e.g. data on religion (belonging to a religious community), which today is considered sensitive information, was included from 1855 to 1921 and again in 1950. Place of birth was included from 1845 and citizenship was included from 1925(The Danish National Archives (Rigsarkivet) 2016). The last census was conducted in 1970 as the CPR System was introduced in 1968 (The Danish National Archives (Rigsarkivet) 2016).

Since the mid-1970s Statistics Denmark has published statistics on number of immigrants in Denmark based on citizenship. The current classification system of origin used by Statistics Denmark was established in 1991, and records using this type of categorizations are listed from 1980. The system introduced 'immigrants' and 'second generation immigrants' (later changed to 'descendants') as new groups in the national statistics and it became possible to make sub-divisions based on country of origin. Countries of origin were divided into 'less developed' and 'developed' countries following UNs definition from 1988(The National Information Services (Statens Informationstjeneste) 1991). This was changed into Western/non-Western countries in 2003(Statistics Denmark 2016).

Why has the heterogeneity/granularity in these classifications developed in these countries in terms of their social, historical and political context?

Records of immigrants in Denmark based on citizenship were useful as long as it concerned temporary guest workers and smaller groups of refugees. However, with a rising number of immigrants gaining Danish citizenship, these records became insufficient in order to plan for special measures directed towards specific groups of newcomers(Togeby 2004, Statistics Denmark 2016). Adding to this, a new

[5] See e.g. the following reports: Justitsministeriet 2005, UNI•C Statistik & Analyse 2009, Teknologisk Institut, Arbejdsliv 2000.

[6] In Denmark information from church records and censuses has been the foundation for population statistics until modern times. Records on numbers of born and deceased go back to 1735 and stem from the church records. In 1787 the first actual census was conducted and since 1840 censuses were done regularly. However, actual population registers were not established before 1924.

Aliens Law was passed in 1983 that granted better conditions for asylum seekers. This caused heated debates in the Danish media and Parliament where different categories of migrants were conflated. On this background demands were made for clearer and improved definitions and statistics on migrants (The National Information Services (Statens Informationstjeneste) 1991). A working group was established by the Danish Minister of Finance in 1991, in order to improve the current statistics on refugees and immigrants[7].

The working group emphasized that the classification system should contain proxies, which enabled statistics regarding integration and life opportunities such as levels of education, employment, housing, health and criminality (The National Information Services (Statens Informationstjeneste) 1991)[8]. Furthermore, the definitions used in the statistics should be as objective as possible; statistical information about the different groups should be available e.g. from the CPR Registry; and definitions in the national registers should be as close to 'the prevalent opinion' as possible (The National Information Services (Statens Informationstjeneste) 1991). The working group also stated that it was important to be able to differentiate between immigrants from countries similar to and different from Denmark (Western/non-Western), as the conditions for these two groups differed[9].

An over-arching reason explaining why ethnicity is not included in national registers is the Danish Act on Processing of Personal Data (Persondataloven) from 2000[10], which implements EU Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. The Act states that: "*Data on racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership and data concerning health or sex life may not be processed"* (Retsinformation.dk 2000). Such data has historically been used to exclude, discriminate, and eliminate groups and individuals deemed as undesirable minorities[11]. Thus, ethnicity is considered as 'sensitive data', which is not included in national registers because of the risk of abuse, discrimination and violation of privacy(Statistics Denmark 2016)[12].

[7] The working group was established under the presidency of Statistics Denmark and included representatives from the Ministry of Justice, the directorate for Foreigners, the Ministry of Interior, the Ministry of Social Affairs, and the interest group Local Government Denmark (Kommunernes Landsforening).

[8] Statens Informationstjeneste 1991, p. 20. The working group notes that information on life opportunities for immigrants should also include surveys on how immigrants view their own situation and their contact to: "Danes and country fellows". Such surveys should be based on interviews and is not dealt with further in the report.

[9] Non-Western countries are defined in section 2 and cover regions as diverse as Africa, Latin America and Asia. Thus, this is a very crude grouping.

[10] The Act on Processing of Personal Data substitutes The Public Authorities' Registers Act and The Private Registers Act from 1978 (See: http://www.datatilsynet.dk/english/the-danish-data-protection-agency/introduction-to-the-danish-data-protection-agency/).

[11] In the aftermath of the Second World War, two statements on race were promoted by UNESCO in 1950 and 1951 to posit the dismissal of race as a scientific fallacy, and a political danger. These statements are still highly influential in Europe for categorizations in statistics by providing a justification against any importation of ethnicity or race into official statistical apparatus (se ref in endnote 26, p. 1377).

[12]For more information on prohibition of registration of ethnicity in Western Europe see Simon 2012. This article

In an interview in January 2016 Statistics Denmark also explain that the system of country of birth still makes sense in relation to the levels of integration in Denmark compared to e.g. the UK and the USA, which has much longer histories of immigration than Denmark(Statistics Denmark 2016).

Why disaggregated data are not being collected, analysed, or reported more often

Sub-classifications of ethnic groups are often difficult because of the relatively limited number of migrants in Denmark within the sub-groups. Hence, heterogeneous groups of migrants are often collated although they present very different cultural backgrounds. Research on ethnic minorities and health in Denmark largely use Statistics Denmark's definitions of persons of Danish origin, immigrants and descendants. In sub-sequent categorizations and variables the focus on country of birth, ancestry and citizenship varies(National Institute of Public Health (Statens Institut for Folkesundhed) 2007), but country of birth (own or parents') grouped in broad categories is most often used as the main variable.

Measures such as country of birth will always be rough proxies of complex mechanisms. A major problem of validity with country of birth is that people born in the same country may have different ethnic background; hence, in combination with other data such as language and socio-economic status, it may present a more valid measure of ethnicity (Norredam, Kastrup et al. 2011). However, country of birth is mainly used because it is easily accessible and easy to make operational, whereas routinely collected data on ethnic affiliation is considered difficult, costly and time consuming, and like religious affiliation this would probably require special permission by the Danish Data Protection Agency.

Are there any examples in these countries of how the disaggregated data has been used to impact on policies, programmes, and population health outcomes?

Reports on education, employment, crime rates etc. based on crude categorizations of immigrants and descendants are often subject to policy debates and to some extent influential on initiatives regarding specific programs in these fields. In the health field, however, national policies and programs rarely reflect variations in health related data based on the generally used categorizations of migrants and descendants, whereas local initiatives are in some instances more sensitive to specific evidence on i.e. screening uptake, participation in preventive programs etc. thereby inspiring special, local interventions.

---

also suggests that another reason for not including statistical data on ethnicity when describing populations and analysing social processes is that this challenges the representation of ethnically homogeneous societies, see p. 1376.

Developments to improve current systems of classification

According to Statistics Denmark there are currently no major movements towards alterations of the current classification system and no indications that this system will be changed in the near future(Statistics Denmark 2016). However, there has been and still is on-going debate on whether ethnicity should be registered in national databases and registries.

The Danish interest group Local Government (Kommunernes Landsforening) has argued that the Act on Processing of Personal Data should be revised, making it possible to register ethnicity in labor market databases related to applicants. This would make it possible to identify the number of ethnic minority applicants and whether there is systematic discrimination taking place based on ethnicity(Berlingske.dk 2000).

Furthermore, representatives from the UN Committee on the Elimination of Racial Discrimination have recommended that Denmark should note ethnicity in the CPR Registry. The current data collection system makes it difficult for the State to measure results of its strategies as well as assess the enjoyment of economic, social and cultural rights of vulnerable groups protected by the Convention. The committee suggests that the collection of data on ethnic groups in the country may be obtained through surveys, censuses or other appropriate methods based on principles of confidentiality, informed consent and self-identification (Committee on the Elimination of Racial Discrimination 2015).
As part of the enforcement of anti-discrimination policies, European human rights institutions are also urging a reconsideration of not including ethnicity in registers (Simon 2012).

However, Statistics Denmark finds it most unlikely that methods such as censuses will be reintroduced in Denmark because of the related expenses and implementation challenges. Adding to this, the Act on Processing of Personal Data must be changed before it will be possible to register ethnicity in national registers (Statistics Denmark 2016).

References

Aarhus University. (2016). "Indvandring til Danmark efter 1945." from http://danmarkshistorien.dk/leksikon-og-kilder/vis/materiale/indvandring-til-danmark-efter-1945/.

Berlingske.dk (2000). Kommuner vil registrere ansattes etniske baggrund. Berlingske.dk.

Committee on the Elimination of Racial Discrimination (2015). Concluding observations on the twentieth and twenty-first periodic reports of Denmark**:** Section 21.

Danish Institute for Human Rights (2015). Ligebehandling af Grønlændere i Danmark.

Danish Ministry of Refugees Migrants and Integration (2009). Rapport om marginaliserede nydanske børn og unge.

Migration Policy Institute. (2006). "Denmark: Integrating Immigrants into a Homogeneous Welfare State ", from http://www.migrationpolicy.org/article/denmark-integrating-immigrants-homogeneous-welfare-state. .

National Institute of Public Health (Statens Institut for Folkesundhed) (2007). Folkesundhedsrapporten, Danmark, Etniske minoriteter.

Norredam, M., M. Kastrup and K. Helweg-Larsen (2011). "Register-based studies on migration, ethnicity, and health." Scandinavian Journal of Public Health **39**(7): 201-205.

OECD (2010). OECD Reviews Of Migrant Education Denmark.

Retsinformation.dk. (2000). "Lov om behandling af personoplysninger." from https://www.retsinformation.dk/Forms/R0710.aspx?id=828.

Simon, P. (2012). "Collecting ethnic statistics in Europe: a review." Ethnic and Racial Studies **35**(8): 1366-1391.

Statistics Denmark (2000). Befolkningen i 150 år.

Statistics Denmark (2015). Statistical Yearbook 2015.

Statistics Denmark. (2016). "Befolkning og Valg." from http://www.statbank.dk/statbank5a/SelectTable/Omrade0.asp?SubjectCode=02&PLanguage=0.

Statistics Denmark. (2016). "Comparability." from http://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/immigrants-and-descendants-1-january/comparability. .

Statistics Denmark. (2016). "Immigrants and Descendants 1 January." from http://www.danmarksstatistik.dk/en/Statistik/dokumentation/documentationofstatistics/immigrants-and-descendants-1-january.

Statistics Denmark. (2016). "Indvandrere og Efterkommere." from http://www.danmarksstatistik.dk/da/Statistik/emner/indvandrere-og-efterkommere/indvandrere-og-efterkommere.

Statistics Denmark (2016). Interview Statistics Denmark. L. Østergaard and A. Krasnik.

The Border Union for a Danish Openness (Grænseforeningen for en Åben Danskhed). (2015). "Det tyske mindretal." from http://www.graenseforeningen.dk/det-tyske-mindretal.html-0.

The Danish National Archives (Rigsarkivet). (2016). "Folketællinger - kom godt i gang! ." from https://www.sa.dk/brug-arkivet/laer/intro-folketaellinger.

The Danish National Archives (Rigsarkivet). (2016). "Hvilke oplysninger finder jeg i folketællingerne." from https://www.sa.dk/wp-content/uploads/2014/10/Hvilke-oplysninger-finder-jeg-i-folketaellingerne.pdf. .

The National Information Services (Statens Informationstjeneste) (1991). Bedre statistik om flygtninge og indvandrere. Betænkning nr. 1214.

Thygesen, L. C., C. Daasnes, I. Thaulow and H. Brønnum-Hansen (2011). "Introduction to Danish (nationwide) registers on health and social issues: Structure, access, legislation, and archiving." Scandinavian Journal of Public Health **39**(7): 12-16.

Togeby, L. (2004). Fra fremmedarbejdere til etniske minoriteter (Magtudredningen), Aarhus Universitetsforlag.

Chapter 7: Ethnic group classification in Great Britain

Peter Aspinall

Abstract

In the second half of the twentieth century migrant flows were largely related to Britain's colonial past. The marked increase in immigration since the early 1990s, for reasons of asylum-seeking, education and work, and family migration and from an increasing number of countries, has transformed the country's ethnic diversity and ushered in an era of *superdiversity*, challenging census ethnic group categorisation as never before. This diversity has been captured in three decennial censuses which have listed subgroups in the Asian and Black pan-ethnicities and, since 2001, in the White and Mixed groups. However, the major contribution of the decennial census to granularity has been through the analysis of free-text responses, the extensive release of detailed country of birth data, and the use of cross-tabulation in the cultural question set. The use of granular ethnicity categories in health datasets is more limited, comprising the NHS Personal Demographics Service Birth Notification Data Set, the Family Origin Question in antenatal settings, and Medical Read and SNOMED CT ethnic origin codes in general practice. Most of the 40 or so routine health datasets still use the 2001 Census ethnic group classification, to the exclusion of the new groups added in 2011. The main set of granular ethnicity categories outside the health field is the Department for Education's 'Extended Categories' list, containing around 100 ethnic categories.

The main barriers to the introduction and use of granular ethnicity categories in official health datasets are organisational, involving complex bureaucratic processes and substantial costs. Further, in Britain there has been no strong advocacy or leverage for greater granularity from professional bodies in medicine, the NHS, and Public Health England. There are competing data priorities for the NHS, including the demands of the public sector's Equality Duty under the Equality Act 2010. Finally, there are technical issues with granular data itself, starting with the selection of these categories and the complex process of cross-mapping fine-grained categories back to census ethnic group categories for reporting. Ongoing developments across government to obtain greater granularity in ethnicity classifications are limited. They include continuing review of the case to add the new 2011 Census categories to routine health datasets and consideration of new ethnic categories for the upcoming 2021 Census.

Introduction

The USA has pioneered the development of 'granular ethnicity categories' based mainly on responses to the census free-text ancestry question. These 'code sets' rather than classifications, now numbering seven or more, are much more fine-grained than the census race and ethnicity options and contain between 143 and 993 categories, some inflated by American Indian and Alaska Native tribes (Ulmer, McFadden et al. 2009). They are intended as comprehensive national standard lists, including even small geographically isolated populations, which are regularly updated. Their intended purpose is to serve as a 'pick list' for voluntary use by local entities (jurisdictions) with the intention of granular ethnicity counts being retained in data systems.

With no history of census ancestry questions and consequently no ancestry code lists, the British interest in granularity has been more limited. It has focused on extended ethnicity classifications rather than exhaustive national 'code lists'. The main source of granularity is the 1991, 2001, and 2011 census ethnic group questions, including predesignated (closed) categories, the analysis of free-text (open response) options, and the cross-tabulation of variables in the census cultural question set. There are also a few mainly health-related extended classifications, such as the NHS Personal Demographics Service Birth Notification Dataset, the NHS Family Origin Questionnaire used in antenatal booking settings, controlled vocabularies/nomenclatures – notably Medical Read Codes and SNOMED CT – used mainly in primary care, the Department for Education 'Extended Ethnic Categories', and ethnicity derived from distinctive names, including Nam Pehchan, Sangra, and Onomap subgroups. Finally, ONS maintains a comprehensive ethnic group code list akin to the US 'code sets'.

Health analysts in Britain have access to the decennial census ethnicity categorisation in routine NHS and health and social care datasets, parents' country of birth in birth registration statistics, and country of birth in mortality statistics and some health protection data collections, and limited sets of optional granular ethnicity categories in local collections. The next sections describe Britain's changing ethnic group and migrant population and how census classifications have evolved to capture this population.

The changing ethnic group and migrant composition of Britain's population

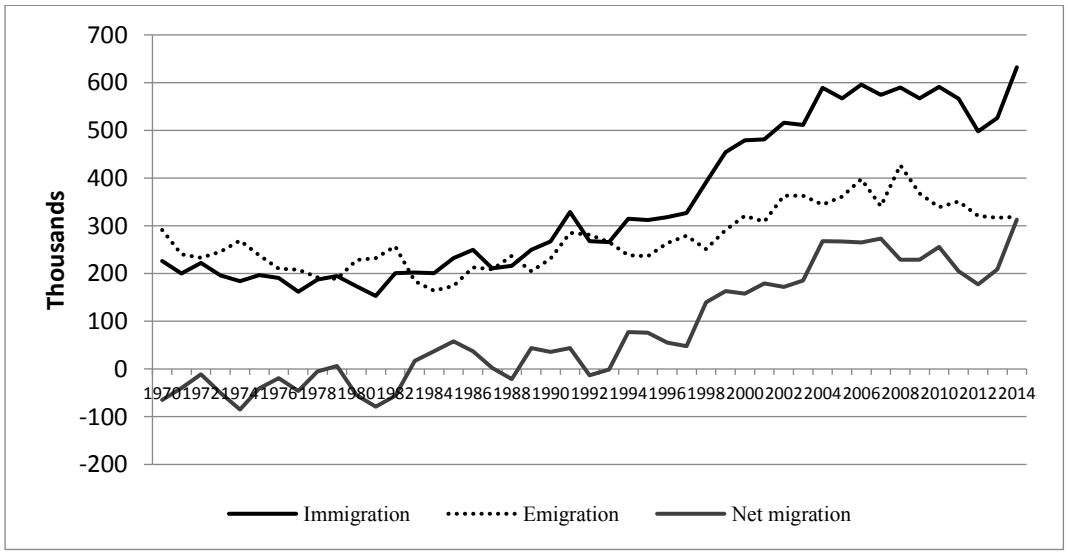
In the 2011 Census Britain had significant minority ethnic populations. In England and Wales 19.5% of the 56.1 million population belonged to ethnic groups other than White British and in Scotland the proportion was 16.0% of its 5.3 million population. The largest minority ethnic groups in England and Wales were Other White (4.4%), Indian (2.5%), Pakistani (2.0%), African (1.8%), and Other Asian (1.5%) and, in Scotland, Other White (1.9%), Polish (1.2%), Irish (1.0%), and Pakistani (0.9%) groups.

These minority ethnic group populations have grown through processes of net migration (immigration minus emigration) and natural change (births minus deaths). In 2011, 13% (7.5 million) of the resident

population in England and Wales were born outside the UK, substantially higher than the 4.3% (1.9 million) in 1951(ONS 2013). While the total resident population of England and Wales increased by 28% (from 43.7 million to 56.1 million) between 1951-2011, the non-UK born population almost quadrupled. Thus, migration has contributed to just under half (45%) of the total population change over the last 60 years and has been a key driver of diversity. For example, in 2011 the top ten non-UK countries of birth accounted for 45% (3.4 million) of the total foreign born population (7.5 million), up from 60% (1.1 million) of the total foreign born population (1.9 million) in 1951.

Historic censuses (showing the growth in the non-UK born populations) and 2011 Census data on year of arrival of migrants who were usual residents reveal the complexity of migrant flows. Migrants have come to the country in substantial numbers at different times. The Irish-born were the largest non-UK country of birth group at each England and Wales census from 1951 to 2001, reflected in the 38% of Irish-born residents in England and Wales in 2011 arriving before 1961. Indians have been the second largest foreign-born group in five censuses (1961-2001, almost doubling between 1961 and 1971) and the largest in 2011. Similarly, the Pakistani-born population saw a noticeable increase between 1961 and 1971 and to a lesser extent between 1971 and 1981. Bangladesh did not enter the top ten non-UK countries of birth till 1991. 35.8% of Indian migrants in 2011 had arrived before 1981 compared with 28.5% of Pakistanis and 19.2% of Bangladeshis. The sub-Saharan African-born and Caribbean-born have had markedly different migration experiences. The Jamaican born population saw substantial rises between 1951 and 1971, peaking in the latter year, while the Nigerian country of birth group, the largest in sub-Saharan Africa, only entered the top ten non-UK countries of birth in 2011. The most dramatic changes have come about in the migrant Other White group, 86% of Polish-born residents in 2011 arriving in 2004 or later. A marked increase in immigration since the early 1990s has transformed the country's ethnic diversity (see figure. 7.1) and ushered in an era of superdiversity (Vertovec 2007) challenging census ethnic group categorisation as never before.

Fig. 7.1. Long-term International Migration, UK, 1970-2014.



The capture of Britain's cultural diversity by the decennial census

In capturing the country's ethnic diversity, the decennial census in Britain first included a self-assigned ethnic group question in 1991, following a series of field trials covering the years 1975-1989 (see appendix 7.1). In 2001 and 2011 Scotland asked its own ethnic group question. Though the 1991 Great Britain Census contained an unsubdivided *White* option, the 2001 Censuses offered granularity: *British, Irish*, and an open *Any Other* in England and Wales and *Scottish, Other British, Irish*, and an open *Any Other* in Scotland. In the 2011 Census *Gypsy or Irish Traveller* (but not *Roma*) was added in England and Wales and *Gypsy/Traveller* and *Polish* in Scotland. Thus, the White group has evolved from one to a maximum of six categories across three censuses and is now the most granular of the pan-ethnicities. National identity (added in 2011) provides a poor proxy for 'home country' ethnic groups as the set contains '*British*'. However, in 2008, the Scottish Government promoted the use of 'Scotland's New Official

Ethnicity Classification' for Scottish Official Statistics, new tick boxes in the 'White' group for English, Welsh, Northern Irish, and British making 9 categories in all (Scotland. 2008) (This was not, however, introduced in 2011 census). Amongst categories incorporating minority ethnic groups, there was no categorisation for 'Mixed' in the 1991 Great Britain Census. However, in the 2001 and 2011 England and Wales Censuses three 'exact combination' categories were added (White and Black Caribbean, White and Black African, and White and Asian) and an open response Any other Mixed option, under the pan-ethnic labels of Mixed and Mixed/multiple, respectively. The Scotland Censuses offered an open response Any Mixed option as numbers were small (<20,000 in 2011). All three decennial censuses have captured Asian/Asian British categories of Indian, Pakistani, and Bangladeshi. In the 2001 and 2011 Censuses an open response Other Asian category was added. Chinese was moved into this set in 2001 in Scotland and 2011 in England and Wales. Similarly, categorisation has been included for the Black African, Black Caribbean, and Other Black groups, the last through open response, in all three decennial censuses, though in 2011 Scotland offered a separate section for African encompassing a tick box and dedicated African other free text option and dropping the 'descriptor' of Black. In 2011 both the England and Wales and Scotland censuses added an Arab category.

However, the major contribution of the decennial census to granularity has been through the analysis of free-text responses, the extensive release of detailed country of birth data, and the use of cross-tabulation in the cultural question set. Only 2 write-in options were offered in the 1991 Great Britain Census. However, the 2001 and 2011 England and Wales Censuses brought distributed residual open response categories, one for each of the five pan-ethnicities (six in Scotland in 2011). Their use has substantially increased over the last two decades, from 740,257 (1.3% of the population) in the 1991 Great Britain Census to 2.11 million (4.0%) of the England and Wales population in 2001 and 4.23 million (7.5%) in 2011. They are an important source of information on more granular ethnic categories for respondents who eschew the predesignated options. However, they do not yield an accurate count

of particular categories as some will tick predesignated options and their analysis also makes demands on scarce information resources as some groups use multiple open-response options.They have greater utility for those groups who fall between categories, such as Sri Lankans in the Asian/Asian British pan-ethnicity and Somalis in the Black pan-ethnicity, but provide important indicative evidence for local jurisdictions in deciding their fine-grained ethnic categories. In the England and Wales 2011 Census, ONS released counts at local authority level for all coded write-ins in the five pan-ethnic Other categories. For example, 57 coded responses were issued for Any other White background, including groups as diverse as Albanian, Australian/New Zealander, Baltic States, Chilean, Cypriot, Iranian, Israeli, Kurdish, Mexican, North African, Polish, Turkish, and White African (see table 7.2). Open response options are also useful to monitor the efficacy over time of the pre-specified categories and as a source of new categories.

This was accompanied in the England and Wales 2011 Census by the release of 'Small Population' tables(ONS 2011), by age group for 13 ethnic groups based on the write-ins (Afghan, Filipino, Greek, Greek Cypriot, Kurdish, Latin Central South American, Nepalese, Polish, Somali, Sri Lankan, Tamil, Turkish, and Turkish Cypriot) and 17 country of birth groups (Bangladesh, Bulgaria, Cyprus EU, France, Ghana, India, Ireland, Jamaica, Nigeria, Pakistan, Philippines, Poland, Romania, Somalia, Sri Lanka, South Africa, and Turkey) and 17 tables for each of Kashmiri, Nepalese, and Sikh (ethnic groups), Cornish (national identity), and Ravidassia (religion). Indeed, there has also been much more extensive release of granular country of birth data in both univariate and cross-tabular data: counts are now available, for example, for 227 country of birth groups cross-tabulated by the 18 ethnic group categories (a matrix of over 4,000 migrant ethnic categories) and period of arrival (before 1981, 1981-2000, 2001-2006, and 2006-2011). Additionally, ethnic group data can now be cross-tabulated by religion (2001, 2011), main language (2011), proficiency in English (2011), and passports held (2011). For example, counts are now available for African languages where they are respondents' main language in the Black African group, important for the provision of linguistically appropriate care: Amharic, Tigrinya, Somali, Krio, Akan, Yoruba, Igbo, Swahili/Kiswahili, Luganda, Lingala, Shona, Afrikaans, Any other Nigerian language, West African language (all other), and African language (all other). This represents a huge increase in the availability of granular ethnic and migrant group data, the latter needed by health analysts as denominators for the calculation of rates and ratios (births and deaths data recording only country of birth, though Scotland now ethnically codes deaths) (Christie 2012).

Table 7.2. Coded responses (n=57), 'Any other White background, write in' (Local Authority level)

Afghan	Greek	Other Middle East
Albanian	Greek Cypriot	Other Western Europe
Anglo-Indian	Iranian	Peruvian
Argentinian	Israeli	Polish
Australian/New Zealander	Italian	Polynesian/Micronesian/Melanesian
Baltic States	Japanese	Serbian
Bosnian	Kashmiri	Somali
Brazilian	Kosovan	Somalilander
British Asian	Kurdish	Sri Lankan
Burmese	Latin/S/Central American	Tamil
Chilean	Malaysian	Thai
Colombian	Mexican	Turkish
Comm. of (Russian) Indep States	Moroccan	Turkish Cypriot
Croatia	Multi-ethnic islands	Venezuelan
Cuban	Nepalese	Vietnamese
Cypriot	Nigerian	White African
Ecuadorian	N African	White Caribbean
European Mixed	N American	Any other White group
Filipino	Other Eastern Europe	

The development of ethnic group categorisation

The development of ethnic group categorisation towards greater granularity following the 1991 Census has been driven by user needs and ethnic group lobbies. At the start of the 2001 ONS Census Development Programme, a consultation exercise was undertaken with members of Census Advisory Groups (CAGs), mainly users of Census data, which invited responses to suggested changes to the ethnic group question (Aspinall 1995). There was unanimous support for the inclusion of categorisation for the 'Mixed' group which was agreed by ONS, though lengthy discussion on how this group could best be captured. However, the recommendation that 'Irish' be included in the classification met with resistance from some CAG members and was only finally accepted in 1997 (necessitating the addition of 'White British' and write-in 'Any Other White background' additions). This came about following a robust campaign by Irish organisations and the national convenor, based on the evidence of inequalities experienced by the Irish, especially those related to health and health care, and was unrelated to the incoming Labour administration as some have speculated (Nagle 2009). Amongst other candidates for change, the recommendation that South Asian ethnic group members should be able to differentiate their ethnicity by religion was met by ONS's decision to include a question on religion, an initiative driven by the main faith communities late in the census development programme. In the 2011 Census the addition of 'Gypsy or Irish Traveller' and 'Arab' from a set of 22 candidates was decided by a prioritisation tool, on the grounds that 18 options was reaching a limit with respect to respondent and administrative burdens (ONS 2009).

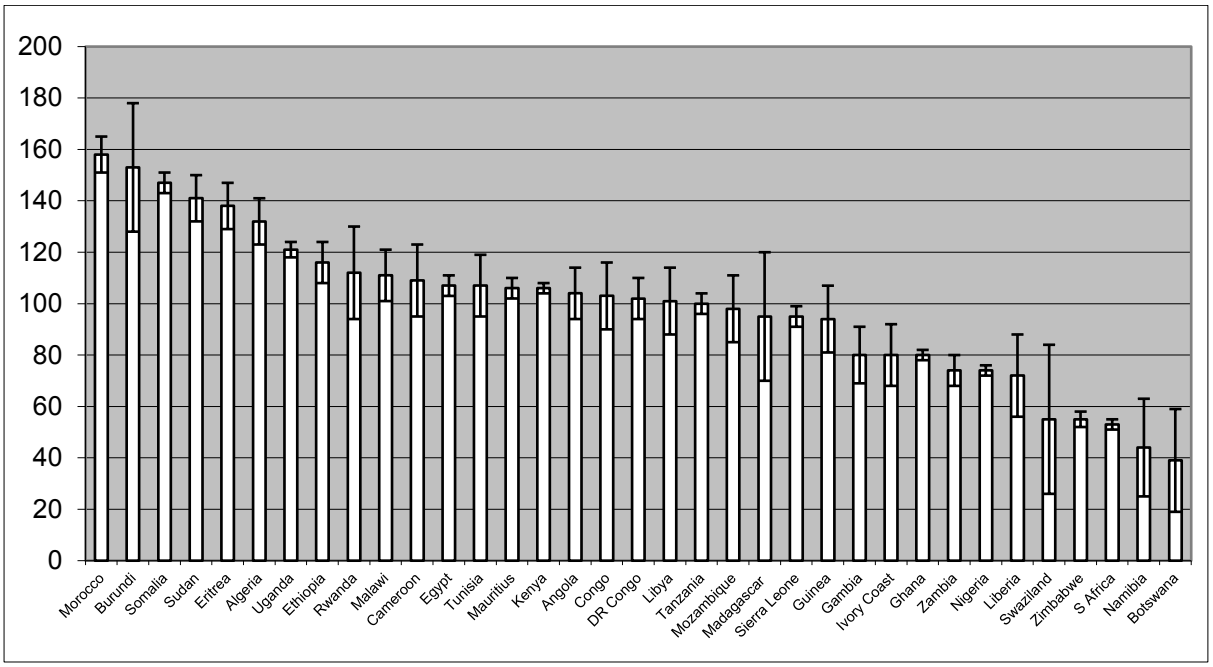
The rationale for granular ethnicity categories in health datasets

In Great Britain the demand for greater granularity in ethnicity classifications has come mainly (but not exclusively) from the public health community. It is influenced by the composition of the population under study, tending to be stronger in those metropolitan areas which have been recipients of high levels of international migration. Such demand is conditioned by two factors: whether the size of subgroups is numerically large enough to report statistically reliable comparisons (the so-called 'small numbers problem' or 'sparse data bias') and whether the subgroup differences identify distinct needs that are masked by data aggregated into broader categories (named the 'fallacy of homogeneity' (Bhopal 2016) or 'concealed heterogeneity') (Aspinall 2011). The main driver is the identification and targeting of inequalities to achieve healthcare quality improvement. By and large the needs of the diversity agenda, with its now multiple strands, are met by census ethnicity data.

The wider public health literature on this concealed heterogeneity has reported numerous examples of differences in health behaviours and outcomes when census ethnic categories are unpacked. Limiting long-term illness in 2001 showed an almost four-fold difference in age-standardised rates across African countries of birth (see figure 7.2) (John Aspinall and Chinouya 2008). Current cigarette smoking, as revealed in the Integrated Household Survey, shows important differences by ethnic group, sex, and

country of birth. While 24.3% of UK-born Mixed: White and Asian females are smokers, this proportion falls to around 6-8% in UK-born Indian, Pakistani, and Bangladeshi females (see appendix 7.2) (Aspinall and Mitton 2014). Similar variability has recently been reported in US racial groups and subgroups, with the highest smoking rate amongst Asian subgroups during 2010-13 being seen among Korean people (20%) and the lowest amongst Chinese and Asian Indian people (7.6%) (Martell 2016).

Fig. 7.2. Age standardised rates of limiting long-term illness by African countries of birth, London, 2001.



Source: Extracted from: Piggott (2006), *DMAG Briefing 2006-3* [Table A10], citing 2001 Census, Commissioned Table C0116, as the source. Error bars show 95% confidence intervals. (Note: countries of birth may conceal considerable ethnic heterogeneity so caution is required in interpreting these rates, for example, only 37.9% of those born in Zimbabwe identified as 'Black African').

The use of granular ethnicity categories in health datasets

The overall volume of granular ethnicity data that is available to health analysts in Great Britain is largely determined by bodies such as the NHS and Public Health England as they are the custodians of centrally reported official datasets. This top-down structure invokes a bureaucratic process of implementation of new ethnic categories. Following the 2001 Census enumeration the government mandated the use of these ethnic group categories across government. Consequently, most routine health datasets were populated with them, replacing the 1991 categories. However, most of these datasets - such as Hospital Episode Statistics for admitted patient, outpatient, and accident and emergency services - do not collect data on country of birth and other variables like main language and religion, so the granularity available in the census is much reduced. Moreover, the NHS in England has not yet introduced the new ethnic group categories added in the 2011 Census, so routine reporting datasets numbering over forty (see appendix 7.3) currently use the 2001 classification. Only a few social

care datasets include the *Gypsy or Irish Traveller* and '*Arab*' categories. The 2011 ethnicity categories have been adopted in Scotland and also in a number of health surveys in England (notably, the NHS GP Patient Survey and Health Survey for England). Scotland has also benefited from the linkage of its 2001 Census and health records for most of its population (Bhopal, Fischbacher et al. 2010) and a second round of linkage of 2011 Census and health records is under way. However, even the benefits of using the England and Wales 2001 Census 16 ethnic categories may be compromised as some datasets (and even Census tables) only report these at the pan-ethnic or section level (*White, Mixed, Asian, Black, and Other*), e.g. the National Child Measurement Programme and NHS workforce tables. Others omit certain categories in reporting, such as the *Mixed* group in the Health Survey for England.

Relatively few official health datasets have been developed that use categories more granular than the census. There are optional second character ethnicity codes in the NHS Personal Demographics Service Birth Notification Data Set (68 national codes including 51granular (see appendix 7.4), that encompass, for example, 25 White and 10 Asian codes: *Mixed Asian, Punjabi, Kashmiri, East African Asian, Sri Lankan, Tamil, Sinhalese, British Asian, Caribbean Asian, and Other Asian*). These can also be used in the NHS Electronic Staff Record and, since 2015, in the NHS Emergency Care Data Set where they are the primary classification. The Family Origin Question (FOQ) - comprising 22 family origin categories - is now asked in all antenatal settings in the UK and is used specifically for NHS Sickle Cell and Thalassaemia Screening Programmes, though there is evidence of poor capture of mixed origins (Aspinall 2013). In addition, the clinical terminologies used in general practice (the Medical Read and SNOMED CT [Systematized Nomenclature of Medicine Clinical Terminology] Codes) offer over 200 fine-grained ethnicity categories but are scarcely used in national reporting. The RiO care record system, used by many mental health trusts and community health services in London and the South East, also collects information on detailed ethnicity, nationality, and first language spoken.

This parsimony in the official use of extended classifications means that many rapidly growing ethnic groups with increasing second generation members - such as Poles, Sri Lankans, Filipino/a, and Somalis - remain concealed in the write-in categories and are only partially counted in country of birth or free-text responses. Indeed, the Black African group as a whole is now a prime candidate for subdivision, given the huge spectrum of inequality and disadvantage hidden within this broad collectivity that is still captured by the 'Black African' category. In such circumstances some health analysts have had recourse to a further *derivative* method of distinctive first/surnames to identify Asian subgroups through specialised algorithms like Nam Pehchan and Sangra, a practice that has been substantially extended by Mateos (Mateos 2014), to include a wide range of ethnic subgroups, including ten within the African collectivity.

While official (NHS) health dataset development suffers from the dampening effect of infrastructural routines and processes, greater innovation is possible amongst local jurisdictions or entities. GP

practices, for example, use a plurality of ethnic coding systems and frequently also collect data on religion and language. Moreover, in England the Health and Social Care Act 2012 embedded new public health functions in local authorities. One benefit has been that these organisations, accustomed to collecting more fine-grained ethnicity data in education settings, have now extended these practices to public health, in some cases implementing borough- or district-wide extended classifications that are tailored to the composition of their local populations. Such practices confer clear benefits as quality improvement activities and interventions are frequently designed and implemented locally.

Heterogeneity/granularity of ethnic classifications outside the health field and lessons learned

The main set of granular ethnicity categories outside the health field is the Department for Education's Extended Categories list, containing around 100 ethnic categories (84 detailed) and available since 2003 for optional use by local authorities[i]. Following a consultation exercise in 2002 on a key list of extended ethnicity categories, local authorities were given the option of using these extended ethnicity categories in their schools if they felt that the main (predominantly census) ones did not meet local management needs. These categories, as well as pupil information on language spoken at home, are now extensively used by London Boroughs, metropolitan authorities, and wider jurisdictions (e.g. Greater London Authority), to monitor educational attainment, school exclusions, and social inclusion. The fact that local authorities usually choose only a selection of the codes (if any) for local purposes precludes national reporting and the Department for Education does not hold complete data for the ethnic background categories and now makes little use of them. This practice does point to the potential utility of developing a standardised extended ethnic category list to NHS trusts and other health and social care providers which would be of clear benefit to those in London and metropolitan districts with diverse populations. The Extended Categories list is much superior to the PDS Birth Notification Data Set and would be an obvious starting point for the development of a national standard.

Why disaggregated data are not being collected, analysed, or reported more often.

The main barriers to the introduction and use of granular ethnicity categories in official health datasets are organisational. The NHS Data Model and Dictionary set the national standards and changes to them represent substantial bureaucratic challenges. The continuing difficulties encountered in adding the 2011 Census categories to NHS data collections almost five years after the census enumeration

[i] For the latest list (for implementation in the 2015-16 collection), see: Department of Education. Education Data Division – Request for Change Form for CBDS. March 2015. Accessed: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/418148/RFC_784_changes_to_ethnicity_codeset.pdf

illustrates this point. Such changes incur huge administrative burdens, requiring changes to 43 datasets, 6 central returns, 6 messaging schema, and system supplier technologies and would incur significant costs, estimated at £660,000, a significant sum with respect to other priorities in a cost-constrained NHS (May 2014, McCrirrick 2014). A further barrier is the complex organisational framework in the NHS for achieving change to current ethnicity data collections or initiating new ones, as the Health and Social Care Information Centre (HSCIC) no longer has the statutory power to undertake this. This time-consuming process, involving multiple organisations, acts against innovation in ethnic group data collection in official health datasets.

In Britain there has been no strong advocacy or leverage for greater granularity from professional bodies in medicine or other sectors, a role performed in exemplary fashion by the Institute of Medicine in the USA. Moreover, there is no clear bridge between expressions of need from the academic and public health practitioner community and the complicated bureaucratic processes that have to be followed to achieve change. The process even remains protracted when commissioned academic advisors feed their recommendations through bodies like the government's National Inclusion Health Board (Aspinall 2014).

Further, there are competing data priorities for the NHS. The quality of ethnic data is so poor in some routine administrative data sets, when validated against gold standard self-assignment (Saunders, Abel et al. 2013), that the pursuit of more granular data may be seen as a step too far. In Scotland this problem has been resolved through linking Census and health records, while the NHS in England somewhat optimistically relies on validation through patients' online access to their health records. In addition, the NHS has to address ethnicity categorisation anomalies, when data flows between health and social care organisations and across national (home country) borders, and the demands of the public sector's Equality Duty under the Equality Act 2010, which now covers nine protected characteristics.

Finally, there are technical issues with granular data itself, starting with the selection of categories that capture the composition of the local population. The optional and selective use by organisations of granular categories, while contributing to quality improvements and better targeting of interventions locally, preclude fine-grained reporting for larger, geographically-defined populations. Further, the intractable tendency towards aggregating ethnic category data at the reporting stage - driven by such considerations as speed and simplicity, minimisation of error, protection of confidentiality, obtaining robust counts, and use of census denominators - may breach the important principle of self-identification. Collapsing the granular categories into the census classification is frequently undertaken on the basis of probability assignment. Yet the process of mapping fine-grained categories back to census ethnic group categories for reporting (what UK analysts call cross-maps and the US roll-up) may be complicated by the fact that some fine-grained migrant and ethnic groups map back to multiple census categories (Aspinall and Chinouya 2016), reducing the validity of operational procedures.

Ongoing developments

Ongoing developments across government to obtain greater granularity in ethnicity classifications are limited. They include continuing consideration by NHS standards bodies of the case to include the 2011 Census categories in routine NHS health datasets, following representations by the National Inclusion Health Board. Moreover, the 2021 census development programmes are now reviewing the 2011 cultural question sets, public consultation exercises having yielded requests for new response categories, including Sikh, Jewish, Roma and extension of the White-Other category within ethnic group; Cornish within national identity; the use of specific languages, including Welsh proficiency across the UK, British Sign Language and the Cornish language; and requests to collect additional information about religious denominations. A number of minority ethnic communities, notably Kashmiris, are pursuing their own campaigns to get their respective groups on the upcoming Census ethnic group classification, while others are represented by overarching bodies. This route to granularity may become of increasing importance as ONS has introduced a Special Populations (renamed Diversity) Advisory Group alongside Census Community Liaison initiatives.

Clearly, inter-governmental work towards developing a single national standard set of granular ethnicity categories would be desirable. This could be informed by the 2011 Census findings (including the ONS Small Population datasets) and made available for optional use as the granular categories most important for collection will vary according to the specific locales for analysis. Where roll-up to Census ethnic categories is required, an additional question on these should be asked, rather than use made of probabilistic methods of assignment. In the absence of a census ancestry question, a template or cumulative list of granular categories could draw on the Department of Education's Extended category list as a starting point. This might lend itself to a multi-tier structure, level one encompassing the Census pan-ethnicities, including *African*; level two, national origins, such as *Sri Lankan, Filipino/a*, and *Somali*; and level three more fine-grained including locally grounded categories, such as *Kenyan Asian, Kenyan African, Pakistani Kashmiri, Pakistani Mirpuri, and Nigerian Igbo, Nigerian Yoruba*, etc. However, given the current difficulties in even getting the 2011 Census new ethnic categories added to official health datasets and ONS's current work towards an Administrative Data Census post-2021, more modest proposals may have a better chance of success. These might include representations to the current 2021 Census Development Programmes for additional category options and revisions to the NHS Personal Demographics Service extended classification using the evidence base of 2011 Census

findings. Great Britain already has a comprehensive national code list of the US kind in the full ONS ethnic group classification (760 categories listed alphabetically under the five pan-ethnic groups)[ii].

In addition, greater effort is needed to ensure that data captured at the granular level is, where feasible, used at this level rather than collapsed into broader categories, analyses shared, and opportunities to influence routine data collections to include a greater number of ethnic subgroups are fully utilised.

[ii] ONS. 2011 Census User Guide. 2001 Census Variable and Classification Information: Part 6. January 2013. Each category has a unique numeric code. Similar classifications are available for main language, passports held, religion, and country of birth. Accessed at:vcipart6classificationsv03_tcm77-297950.pdf.

References

Aspinall, P. (1995). The development of an ethnic group question for the 2001 Census. London: United Medical and Dental Schools.

Aspinall, P. (2014). "Hidden needs: identifying key vulnerable groups in data collections: vulnerable migrants, gypsies and travellers, homeless people, and sex workers." Centre for Health Services Studies, University of Kent.

Aspinall, P. and L. Mitton (2014). "Smoking prevalence and the changing risk profiles in the UK ethnic and migrant minority populations: implications for stop smoking services." Public Health **128**(3): 297-306.

Aspinall, P. J. (2011). "The utility and validity for public health of ethnicity categorization in the 1991, 2001 and 2011 British Censuses." Public health **125**(10): 680-687.

Aspinall, P. J. (2013). "When is the use of race/ethnicity appropriate in risk assessment tools for preconceptual or antenatal genetic screening and how should it be used?" Sociology **47**(5): 957-975.

Aspinall, P. J. and M. J. Chinouya (2016). African Diaspora Population in Britain, Springer.

Bhopal, R., C. Fischbacher, C. Povey, J. Chalmers, G. Mueller, M. Steiner, H. Brown, D. H. Brewster and N. Bansal (2010). "Cohort profile: Scottish Health and Ethnicity Linkage Study of 4.65 million people exploring ethnic variations in disease in Scotland." International journal of epidemiology: dyq118.

Bhopal, R. S. (2016). Concepts of Epidemiology. Integrating the ideas, theories, principles, and methods of epidemiology, Oxford University Press.

Christie, B. (2012). "Scotland introduces record of ethnicity on death certificates." Bmj **344**: e475.

John Aspinall, P. and M. Chinouya (2008). "Is the standardised term 'Black African'useful in demographic and health research in the United Kingdom?" Ethnicity and health **13**(3): 183-202.

Martell, B. N. (2016). "Disparities in Adult Cigarette Smoking—United States, 2002–2005 and 2010–2013." MMWR. Morbidity and Mortality Weekly Report **65**.

Mateos, P. (2014). "Names, ethnicity and Populations." Advances in Spatial Science.

May, J. (2014). Ethnic Category Coding. . DSCN 11/2008. Leeds, HSCIC.

McCrirrick (2014). Ethnicity Coding Change Review Proposal: Guidance Required. Leeds, HSCIC.

Nagle, J. (2009). "Multiculturalism's Double-Bind: Creating Inclusivity." Cosmopolitanism, and Difference (Farnham, Surrey, England: Ashgate Publishing Ltd, 2009) **5**.

ONS. (2009). " Information Paper. Deciding which tick-boxes to add to the ethnic group question in the 2011 England and Wales Census."

ONS. (2011). " Census Small Population Tables for England and Wales." from http://www.nomisweb.co.uk/census/2011/small_population.

ONS. (2013). "Immigration Patterns of Non-UK Born Populations in England and Wales in 2011."

Saunders, C. L., G. A. Abel, A. El Turabi, F. Ahmed and G. Lyratzopoulos (2013). "Accuracy of routinely recorded ethnic group information compared with self-reported ethnicity: evidence from the English Cancer Patient Experience survey." BMJ open **3**(6): e002882.

Scotland., S. G. a. t. G. R. O. f. (2008). Scotland's New Official Ethnicity Classification for Scottish Official Statistics and Recommended for Scotland's 2011 Census. Edinburgh, Scottish Government.

Ulmer, C., B. McFadden and D. R. Nerenz (2009). Race, ethnicity, and language data: standardization for health care quality improvement, National Academies Press.

Vertovec, S. (2007). "Super-diversity and its implications." Ethnic and racial studies **30**(6): 1024-1054.

Appendix 7

Appendix 7.1 Ethnic category content of the 1991, 2001 and 2011 Census in Britain

	1991 Great Britain	2001 England and Wales	2001 Scotland	2011 England & Wales	2011 Scotland
WHITE	□ White	□ British	□ Scottish	□ English/Welsh/Scottish/ Northern Irish/British	□ Scottish
		□ Irish	□ Other British	□ Irish	□ Other British
		□ Any other White background (FT)	□ Irish	□ Gypsy or Irish Traveller	□ Irish
			□ Any other White background (FT)	□ Any other White background (FT)	□ Gypsy/Traveller
					□ Polish
					□ Other White ethnic group (FT)
MIXED	-	□ White and Black Caribbean	□ Any mixed background (FT)	□ White and Black Caribbean	□ Mixed or multiple ethnic groups (FT)
		□ White and Black African		□ White and Black African	
		□ White and Asian		□ White and Asian	
		□ Any other Mixed background (FT)		□ Any other Mixed/multiple ethnic background (FT)	
ASIAN	□ Indian	□ Indian	□ Indian	□ Indian	□ Pakistani, Pakistani Scottish or Pakistani British
	□ Pakistani	□ Pakistani	□ Pakistani	□ Pakistani	□ Indian, Indian Scottish or Indian British
	□ Bangladeshi	□ Bangladeshi	□ Bangladeshi	□ Bangladeshi	□ Bangladeshi, Bangladeshi Scottish or Bangladeshi British
		□ Any other Asian background (FT)	□ Chinese	□ Chinese	□ Chinese, Chinese Scottish or Chinese British
			□ Any other Asian background (FT)	□ Any other Asian background (FT)	□ Other Asian, Asian Scottish or Asian British (FT)
BLACK	□ Black-Caribbean	□ Caribbean	□ Caribbean	□ African	□ African, African Scottish or African British
	□ Black-African	□ African	□ African	□ Caribbean	□ Other African (FT)
	□ Black-Other (FT)	□ Any other Black background (FT)	□ Any other Black background (FT)	□ Any other Black/African/Caribbean background (FT)	□ Caribbean, Caribbean Scottish or Caribbean British

					□ Black, Black Scottish or Black British
					□ Other Caribbean or Black (FT)
OTHER	□ Chinese	□ Chinese	□ Any other background (FT)	□ Arab	□ Arab, Arab Scottish or Arab British
	□ Any other ethnic group (FT)	□ Any other (FT)		□ Any other ethnic group (FT)	□ Other ethnic group (FT)

Notes: □ = tick boxes on the form; FT = free text ('please write in'). The ordering in the 1991 Census was: White, Black-Caribbean, Black-African, Black-Other (FT), Indian, Pakistani, Bangladeshi, Chinese, Any other ethnic group (FT). The categories relate to the tick boxes rather than the overarching labels. The overarching labels used in the 2001 census were: England and Wales – 'White', 'Mixed', 'Asian or Asian British', 'Black or Black British', and 'Chinese or other ethnic group'; Scotland – 'White', 'Mixed', 'Asian, Asian Scottish or Asian British', 'Black, Black Scottish or Black British', & 'Other ethnic background'; 2011 Census: England and Wales – 'White', 'Mixed/multiple ethnic groups', 'Asian/Asian British', 'Black/African/Caribbean/Black British', & 'Other ethnic group'; Scotland: 'White', 'Mixed or multiple ethnic groups', 'Asian, Asian Scottish or Asian British', 'African', 'Caribbean or Black', & 'Other ethnic group'.

Appendix 7.2 Current cigarette smoking status by ethnic group, gender, and UK or non-UK born, England and Wales, Integrated Household Survey, 2009/10 to 2011/12.

Ethnic group	*Sex*	*Country of birth*	*current cigarette smoker*
White British	Male	Non-UK	20.7%
		UK	22.0%
	Female	Non-UK	17.6%
		UK	19.8%
Other White	Male	Non-UK	30.5%
		UK	24.3%
	Female	Non-UK	20.9%
		UK	18.5%
White and Black Caribbean	Male	Non-UK	15.9%
		UK	33.0%
	Female	Non-UK	8.6%
		UK	37.5%
White and Black African	Male	Non-UK	31.9%
		UK	36.0%
	Female	Non-UK	17.6%
		UK	33.2%
White and Asian	Male	Non-UK	17.3%
		UK	23.3%
	Female	Non-UK	14.1%
		UK	24.3%
Indian	Male	Non-UK	12.7%
		UK	17.3%
	Female	Non-UK	2.0%
		UK	5.8%
Pakistani	Male	Non-UK	23.2%
		UK	25.7%
	Female	Non-UK	3.0%
		UK	6.7%
Bangladeshi	Male	Non-UK	31.4%
		UK	22.4%
	Female	Non-UK	3.6%
		UK	7.8%
Black Caribbean	Male	Non-UK	21.0%
		UK	29.6%
	Female	Non-UK	7.2%
		UK	23.7%
Black African	Male	Non-UK	14.3%
		UK	18.5%
	Female	Non-UK	4.4%
		UK	11.1%
Chinese	Male	Non-UK	21.4%
		UK	16.9%
	Female	Non-UK	5.8%
		UK	15.3%

Source: Integrated Household Survey, pooled date.

Appendix 7.3 Main datasets in the English NHS using the 2001 Census ethnic group categorisation

Cancer Outcomes and Services Data Sets
Accident and Emergency Commissioning Data Set (CDS)
Outpatient CDS
Future Outpatient CDS
Admitted Patient Care
Finished Birth Episodes CDS
Finished General Episode CDS
Finished Delivery Episode CDS
Other Birth Event CDS
Other Delivery Event CDS
Detained &/or Long Term Psychiatric Census
Unfinished Birth Episode CDS
Unfinished General Episode CDS
Unfinished Delivery Episode CDS
Child and Adolescent Mental Health Services Secondary Uses Data Sets
Children and Young People's Health Services Data Set
Chlamydia Testing Activity Data Set
Choose and Book Messaging Service
Community Information Data Set
Diagnostic Imaging Data Set
Genitourinary Medicine Clinic Activity Data Set
HIV and AIDS Reporting Data Set
Improving Access to Psychological Therapies Data Set
Maternity Services Secondary Uses Data Set
Mental Health and Learning Disabilities Data Set
National Neonatal Data Set
National Renal Data Set - Demographics
National Workforce Data Set (and Electronic Staff Record)
NHS Health Checks Data Set
Personal Demographics Service Birth Notification Data Set
Personal Demographic Service Create Initial Record Request Sata Set
Sexual and Reproductive Health Activity Data Set
Stop Smoking Services Quarterly Data Set
Systemic Anti-Cancer Therapy Data Set

Appendix 7.4 Sets of Granular Ethnicity Categories: PDS Ethnic Category Code List, NHS Data Dictionary, and Department of Education 'Extended Category' List

PDS Ethnic Category Code List

Permitted National Codes:

A	British, Mixed British
B	Irish
C	Any other White background
C2	Northern Irish
C3	Other white, white unspecified
CA	English
CB	Scottish
CC	Welsh
CD	Cornish
CE	Cypriot (part not stated)
CF	Greek
CG	Greek Cypriot
CH	Turkish
CJ	Turkish Cypriot
CK	Italian
CL	Irish Traveller
CM	Traveller
CN	Gypsy/Romany
CP	Polish
CQ	All republics which made up the former USSR
CR	Kosovan
CS	Albanian
CT	Bosnian
CU	Croatian
CV	Serbian
CW	Other republics which made up the former Yugoslavia
CX	Mixed white
CY	Other white European, European unspecified, European mixed
D	White and Black Caribbean
E	White and Black African
F	White and Asian
G	Any other mixed background
GA	Black and Asian
GB	Black and Chinese
GC	Black and White
GD	Chinese and White
GE	Asian and Chinese
GF	Other Mixed, Mixed Unspecified
H	Indian or British Indian
J	Pakistani or British Pakistani

K	Bangladeshi or British Bangladeshi
L	Any other Asian background
LA	Mixed Asian
LB	Punjabi
LC	Kashmiri
LD	East African Asian
LE	Sri Lanka
LF	Tamil
LG	Sinhalese
LH	British Asian
LJ	Caribbean Asian
LK	Other Asian, Asian unspecified
M	Caribbean
N	African
P	Any other Black background
PA	Somali
PB	Mixed Black
PC	Nigerian
PD	Black British
PE	Other Black, Black unspecified
R	Chinese
S	Any other ethnic group
SA	Vietnamese
SB	Japanese
SC	Filipino
SD	Malaysian
SE	Any Other Group
Z	Not stated

Department for Education 'Extended Category' List

This extended category list offers substantially greater granularity than the *PDS Ethnic Category Code List.* However, it has not been aligned with the England and Wales 2011 Census. For example, SE Asian categories are listed under 'Any other ethnic group'.

WRIB	**White-British**
WCOR	White-Cornish
WENG	White-English
WSCO	White-Scottish
WWEL	White-Welsh
WOWB	Other White British
WIRI	**White-Irish**
WIRT	**Traveller of Irish heritage**
WOTH	**Any other White background**
WALB	Albanian (excl. Kosovan)
WBOS	Bosnian-Herzegovinian
WCRO	Croatian

WGRE	Greek/Greek Cypriot
WGRK	Greek
WKRC	Greek Cypriot
WITA	Italian
WKOS	Kosovan
WPOR	Portuguese
WSER	Serbian
WTUR	Turkish/Turkish Cypriot
WTUK	Turkish
WTUC	Turkish Cypriot
WEUR	White European
WEEU	White East European
WWEU	White Western European
WOTW	White other
WROM	**Gypsy/Roma**
WROG	Gypsy
WROR	Roma
WROO	Other Gypsy/Roma
MWBC	**White and Black Caribbean**
MWBA	**White and Black African**
MWAS	**White and Asian**
MWAP	White and Pakistani
MWAI	White and Indian
MWAO	White and any other Asian background
MOTH	**Any other mixed background**
MAOE	Asian and any other ethnic group
MABL	Asian and Black
MACH	Asian and Chinese
MBOE	Black and any other ethnic group
MBCH	Black and Chinese
MCOE	Chinese and any other ethnic group
MWOE	White and any other ethnic group
MWCH	White and Chinese
MOTM	Other mixed background
AIND	**Indian**
APKN	**Pakistani**
AMPK	Mirpuri Pakistani
AKPA	Kashmiri Pakistani
AOPK	Other Pakistani
ABAN	**Bangladeshi**
AOTH	**Any other Asian background**
AAFR	African Asian
AKAO	Kashmiri other
ANEP	Nepali
ASNL	Sri Lankan Sinhalese
ASLT	Sri Lankan Tamil

ASRO	Sri Lankan other
AOTA	Other Asian
BCRB	**Black Caribbean**
BAFR	**Black African**
BANN	Black-Angolan
BCON	Black-Congolese
BGHA	Black-Ghanaian
BNGN	Black-Nigerian
BSLN	Black-Sierra Leonean
BSOM	Black-Somali
BSUD	Black-Sudanese
BAOF	Other Black African
BOTH	**Any other Black background**
BEUR	Black European
BNAM	Black North American
BOTB	Other Black
CHNE	**Chinese**
CHKC	Hong Kong Chinese
CMAL	Malaysian Chinese
CSNG	Singaporean Chinese
CTWN	Taiwanese
COCH	Other Chinese
OOTH	**Any other ethnic group**
OAFG	Afghan
OARA	Arab other
OEGY	Egyptian
OFIL	Filipino
OIRN	Iranian
OIRQ	Iraqi
OJPN	Japanese
OKOR	Korean
OKRD	Kurdish
OLAM	Latin/South/Central America
OLEB	Lebanese
OLIB	Libyan
OMAL	Malay
OMRC	Moroccan
OPOL	Polynesian
OTHA	Thai
OVIE	Vietnamese
OYEM	Yemeni
OOEG	Other ethnic group
REFU	**Refused**
NOBT	**Information not yet obtained**

Source: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/494334/2016_alternative_provision_census_guide.pdf

Chapter 8: Ethnic group classification in Hungary

Inez Zsófia Koller

Abstract

Hungary has had a long history of migration, particularly marked by state border changes after WWII. Nowadays Hungary has 13 official nationalities, entitled to some rights such as to form their ethnic minority self-governments, plus four migrant groups which are enumerated in censuses. This country collects ethnic data through two official registration procedures, decennial censuses and, since 2006 registration lists for ethnic minority self-governments.

In Hungary only the ethnic group of Gypsy/Roma people can be distinguished by their health conditions from the majority of the society. For example, life expectancy of Gypsy people is ten years shorter than the average of the society.

Since the change of regime in Hungary, determining ethnic background by physical traits has not been an accepted way. It is important in Hungary that counting ethnic background cannot be the basis of discrimination. This is why ethnic data have not been registered in health care institutions since then to protect the Roma people. Since 2001, answering questions on ethnicity is not compulsory and is placed in a specific block of questions in the census. Moreover, the Hungarian Central Statistical Office has started a wide-scale co-operation platform involving civil organisations in the process of forming the census questionnaire. Granularity is provided in the census questionnaire in two ways: questions which enable the expression of voluntary motivations and then balance them with an objective focus on language.

In Hungary, the lack of database information on ethnicity and the health condition of the population derives from the new ideological frames of the change of regime since the very beginning of the 1990s. Due to two grounding arguments: On the one hand, questioning the worthiness of research on health conditions of Gypsy/Roma people, as health conditions of Gypsy/Roma people did not differ from those of the poor people in general. The second argument supposes that there might be some who would derive political consequences from statistical results, as a way to blame Gypsy/Roma people for spreading epidemics and other diseases which is very dangerous, however, the aptitude for these diseases depends on family genetics not on ethnic differences.

Granular data is not being collected more often due to the general demographic data collection methods. These methods have changed radically after the change of regime, mirroring sensitiveness to antidiscrimination, which makes it more difficult to gather exact information on the relations of ethnicity to health conditions.

Introduction: Hungarian approaches to concepts on ethnicity

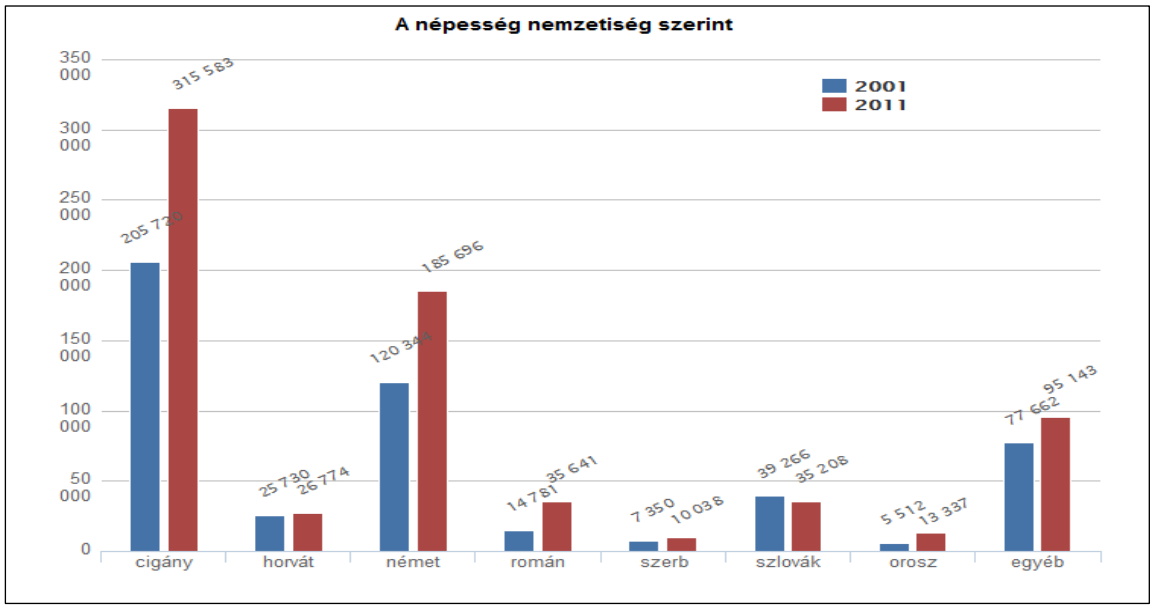
The Hungarian definition on ethnicity differs in specifications from the general definition of the study. According to the Hungarian version a person belongs to a certain ethnic group – which is called now "nationality" (1) as a general term – if he himself or she herself identifies with it, so it is a self-determined action. Furthermore, identification focuses on culture, language and historical traditions in the territory of Hungary, which limits the number of the accepted nationalities in the country. Finally, it lacks physical features and third party determination also for historical reasons and avoids the usage of the term race which is considered as a discriminating and humiliating word in political and cultural contexts. Ethnic minority groups are all accepted nationalities except from the Hungarians according to law, but common usage calls them minorities just as all other ethnic groups living in the country who are also called migrant communities. Temporary residents who live mostly in reception centres are called migrants. (2) Hungarian terminology tries to avoid ethnocentrism. Due to this aim Hungarians are also called a nationality, including those Hungarians who live in neighbouring countries, and for example census questionnaires are available in 18 languages. In Hungary there is also development in gaining more granularity in data collection on ethnicity.

What is the demographic background of Hungary in terms of the development of its ethnic composition and its history of migration?

Historically, Hungary used to be a multi-ethnic society but today it is a nation state with thirteen officially accepted nationalities. The Armenian, Bulgarian, Croatian, German, Greek, Gypsy, Polish, Romanian, Ruthenian, Serbian, Slovakian, Slovene and Ukrainian nationalities are accepted officially and entitled for some special rights in Hungary such as to form their ethnic minority self-governments. (3) In a population of 9897541 (2015) Hungarian inhabitants constitute the majority with 85.6% of the population. (4) There are only two nationalities above one percent of the whole population, Gypsy (3.2%) and German (1.9%). In comparison to census data of 2001, according to Tóth and Vékás (5) newly introduced methods in the latest census had generally positive effects on the growing number of declared nationalities, as their proportion has grown 45.6%, although not in every case. The group of Gypsy nationality growth is the most spectacular, they rose from 205 thousand to 315 thousand. However, according to social researchers their real number is still much more than that, as the estimated number of Gypsy people living in Hungary is around 700 to 800 thousand. On the other hand, Greeks, Slovakians and Slovenians lost between 10-40% of their members. Not just these small ethnic groups show losses. The entire country population has also decreased from 10.2 million to 9.94 million while the number of Hungarians has shrunk from 9.4 million to 8.3 million according to Index analysis of the released census data in 2013. (6) Further interesting change is that there is information gained from

census questionnaires on immigrants as well. As the Hungarian Central Statistical Office explains, general demographic loss is due to decreasing birth rate and the increasing emigration rate as hundreds of thousands of Hungarian economic migrants have left the country since the early 2000s.

Figure 8.1 Diagram on changes in the number of ethnic nationalities between 2001 and 2011



Source:Index.hu

Beside nationalities there are also other ethnic groups living in Hungary. There is a meaningful sized community of Jewish people, approximately one hundred thousand, living mostly in Budapest. They did not claim official state recognition on ethnic grounds as they declared themselves a religious community. Other ethnic minority groups do not fulfil all the official criteria for recognition, such as Russians who constitute the largest migrant community in Hungary (13337), Arabians, Chinese people, Kurdish people and some communities from African countries. (7)

The history of migration is more than a thousand years old in Hungary. The *Avar* indigenous people lived in the Carpathian Basin before the mostly Magyar (Hungarian) tribes arrived to the Carpathian basin in the 9th century and founded Hungary. Although the state itself is more than a thousand years old, the basis of modern ethnic composition of the country has been shaped significantly after the Ottoman occupation. At this time, Slavic peoples came to re-inhabit the Northern and Southern parts while other regions were resettled by Germans due to state organised order of the Habsburg Empire during the 17th and 18th centuries. In the period of nation state creation, the country's non-Hungarian ethnic groups constituted more than a half of the total population, nevertheless, Hungarian political elite claimed independence from the Habsburg Empire on ethnic national basis to reach matching national and state borders and fought for it in its independence war. After their defeat, the Austro-Hungarian Kingdom was formed on the territory of Hungary and the ratio of Hungarians increased up to around

80 percent. Due to state border changes after the World War I, the ratio of Hungary's ethnic minorities varied from 10 to 20% while 33% of Hungarians became citizens of neighbouring countries. Under the socialist regime handling the question of ethnicity was twofold. On one hand, ethnic differences were seen as grounds for the division of society and were not compatible with socialist ideology which proclaimed for social equality. On the other hand, revealing the situation of ethnic minorities was perceived as a source of international tensions as many Hungarians lived in neighbouring countries which also belonged to the Soviet area. Taking all this into consideration, it is not surprising that the proportion of ethnic minorities decreased radically to 1-2% in census records in these decades. After the change of regime of 1990, their proportion rose again to approximately 10% altogether. The Minority Act of 1993 recognised them as constituent components of the state.

What sources of data are available about ethnic composition of the population?

There are two official registration procedures collecting data on ethnic diversity in Hungary. These are the decennial censuses, which enable international comparison, and since 2006 registration lists for ethnic minority self-governments. There have been elections held for nationalities to form minority self-governments as local authorities and to be channelled into national level minority self-governments every four years since 1994. For these elections there are exclusive lists for members of nationalities introduced in 2006 which are signed in voluntarily and checked by prominent members of these communities. After the elections the exclusive lists are stamped out. These techniques of observation show some disadvantages. The most problematic element in them is that they enable voluntary self-professing which leads to the distortion of reality as many members of nationalities, especially, older Germans or Slovakians still remember the forced emigration and population changes and mistrust these data gatherings. Also, many Gypsy who face discrimination, as they can be identified easily by their different physical features, tend to declare themselves as Hungarians. Lastly, and in most cases, not confessing ethnic minority background is the result of growing disinterest in ethnicity in general in the country.

There are some estimations of ethnicity data as well. Most of the sources are civil organisations of nationalities. Unfortunately, these estimations vary in great amplitudes, for example, the number of Gypsy is estimated from 400 thousand to 800 thousand. Also there are some representative surveys undertaken but with small samples and only on population with specific demographic peculiarities.

Fitting to the general antidiscriminative approach of minority affairs in Hungary, ethnic background is not stated in official identification papers. There are no birth certificates which include ethnic nationality, mother tongue or religion although this is a widespread practice in Eastern Europe. However, there are increasingly institutions carrying out research on ethnicity. The Minority Research

Institute of the Hungarian Academy of Sciences publishes research projects, monographs and co-authored volumes mostly in the fields of history, anthropology, sociology, law or political science. Interestingly, research is more about ethnic Hungarians in minority status living in other, mostly neighbouring, countries and less about the nationalities living in Hungary. Besides that, there are several research groups in the academic area dealing with different fields of minority affairs but, as stated before, they have great challenges in data gathering as they have to undertake surveys for themselves.

What sources of data are especially inside the health field? A distinguished ethnic group – the Gypsy/Roma people and their health conditions

In Hungary only the ethnic group of Gypsy/Roma (8) people can be distinguished by their health conditions from the majority of the society, although, there are academic debates about the relevance of this distinction. All other nationalities show only different cultural patterns but have almost the same living conditions as the rest of the society. According to researchers in social sciences (9) and official surveys, before the change of regime, life expectancy of Gypsy people was ten years shorter than the average of the society. Root causes go back 10 to 15 years as mortality data correspond with these periods of time. According to the Hungarian Central Statistical Office the size of the Gypsy population decreases radically in people aged above forty years, when life expectancy is generally 75 years in Hungary. However, exact statistical data do not provide information about the causes of this disparity. Babusik (2002) cites a survey from 1978 conducted in Szabolcs-Szatmár county which revealed not just age partitions in death of Gypsy population in this county but characteristic death causes as well which differed from the rest of the population of the county. According to this investigation the main reasons for early mortality are cardiovascular diseases (cerebral haemorrhage, heart attack) and harmful nutrition habits, and inheritable features played key roles in the evolution of these disease. As it will be explained, it is more difficult today to reveal such information as this proper investigation from 1978. The only relevant recent survey is the Health Status Data on Hungary from 2014 which reported not just affirmation of the ten years shorter life expectancy of Gypsy men in comparison with the majority male population, but also a greater gap for women which is 18 years. The Health Status Data in Hungary collects series of comparative data on different fields of health status of Gypsy and non-Gypsy populations in Hungary, such as infant mortality rate, infectious diseases, hepatitis A and B, TB infection, injecting drug use, illicit drug use, smoking, iron deficiency, cardiovascular diseases, hypertension, diabetes, depression, asthma, stomachic ulcer, cancer, as well as hospitalisation and discrimination from healthcare personnel.

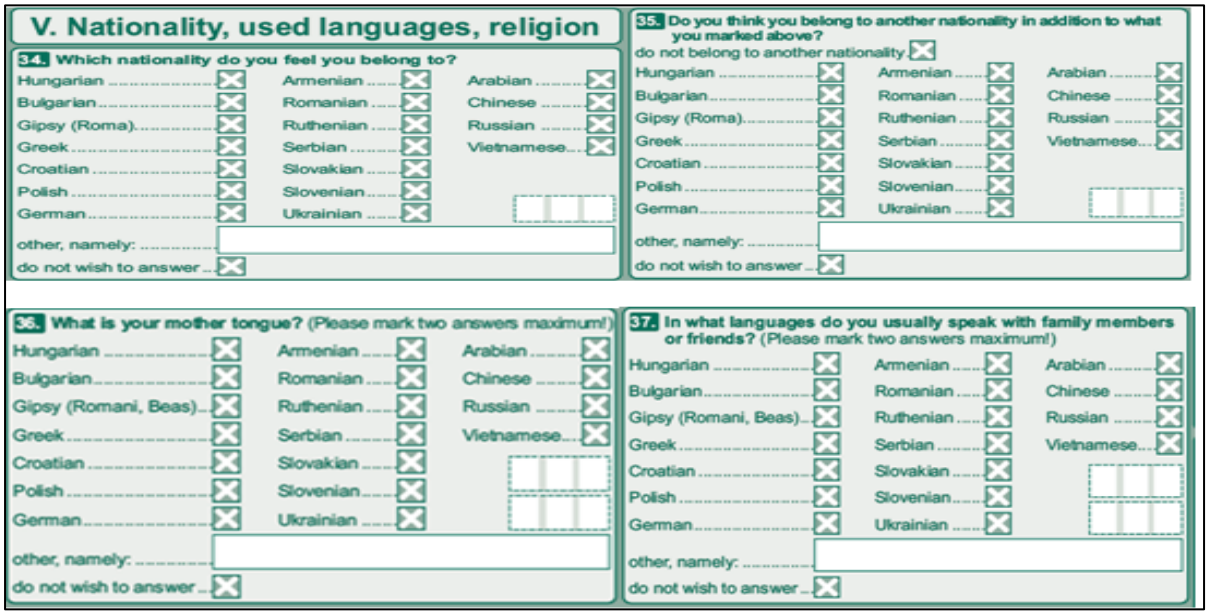
What level of heterogeneity/granularity is used in collection, analysis and reporting?

There are several ways for official state certificates on ethnicity to gain a higher level of heterogeneity and granularity. It is useful to include more questions on the use of mother tongue, special education, cultural claims and perspectives. Although, from the other side, the confession on ethnic background is

both a form of collective action and a form of self-identification (10). There are some who are fearful to indicate their ethnic difference as they want to avoid collateral disadvantages, social exclusion, and various other forms of discrimination. This is why census questionnaires are revised quite often, especially in Hungary. In the Hungarian census questionnaire the topic of ethnicity is put in a separate block of Voluntary questions in the last section, and it is separated from the question on citizenship which is placed in the very first part of the questionnaire. Respondents are free to confess their ethnic ties, they do not have to prove it, and they will not be controlled later on.

As direct questions are criticized for mirroring subjective motivations, especially in Hungary, the questionnaire uses more indirect questions. The question on mother tongue is perceived as an objective criterion not a subjective standing point, since it determines ancestry and the environment in which the language was learnt. However, in cases where bi- or multilingualism is common, the determination of mother tongue becomes challenging. Yet, it is a very useful categorising tool as many minority rights are associated with the use of mother tongue (11) (for example, in minority language education). The census questionnaire contains questions on not just one ethnic relation of a respondent, not just one mother tongue, and about not just one language used with family members or friends in order to map a more realistic picture on ethnicity (12). This results in huge differences between data on mother tongue and identity. For example, in the census of 1990 48072 respondents responded they speak Gypsy as their first language while 142683, nearly three times more stated their Gypsy identity. All questions on ethnicity are closed with 18 possible choices (Hungarian, 13 officially accepted nationalities and plus the four migrant communities mentioned earlier) and the *Other* and *Do not wish to answer* categories. (see figure 8.2).

Figure 8.2 Hungarian census questionnaire, topic of ethnicity



Source: Hungarian Central Statistical Office

Finally, it uses all languages in the questionnaire blanket which are officially recognised in Hungary, which means you can fill your copy in 18 different languages. It is a way to show respect towards nationalities, however, the coding methodology needs some further development (it took nearly two years for the Hungarian Central Statistical Office to analyse and reveal the census data of 2011).
Granularity is provided in the census questionnaire (13) in two ways. It lists beside Hungarians all officially recognised nationalities of the country and involves the four largest migrant groups and beside that in section of "Nationality, used Language, Religion" the questionnaire contains four questions to tone the complexity of ethnicity. The first two are about ethnic identity: "Which nationality do you think you belong to?" and "Do you think you belong to another nationality in addition to what you marked above?" These questions enable people to express voluntary motivations, while the other two questions balance it with the objective focus on language: "What is your mother tongue?" and "In what language do you usually speak with family members or friends?" Granularity and the lack of granularity also appear in the given answers. In the question on identity 'Gypsy/Roma' is not divided into subgroups, while the largest migrant groups 'Chinese', 'Russian' and 'Vietnamese' follow citizenships, however, 'Arabian' is a more complex denotation. Of course, there are some developments towards more granularity as well: in the language questions Romani and Beas (they omitted to list Lovari as well) are marked as subgroups for Gypsy, although respondents cannot tick them, and also in the language questions respondents are allowed to choose two different languages at the same time.
(see figure 8.2).

Are there any lessons to be learned from outside the health field relating to the heterogeneity/granularity of ethnic classifications?

Political and economic tensions during the transitional period in the nineties induced uncertainty in citizens and reduced willingness to provide data. Since the change of regime, determining ethnic background by physical traits has not been an excepted way in Hungary. It is important in Hungary, that the counting of ethnic background cannot be used as the basis of discrimination. This is why ethnic data have not been registered in health care institutions since then to protect the Roma people. It is not possible to renew former practise after the Data Protection Act and The Minority Act were passed. Just after the change of regime, and as a consequence of the special problems of Gypsy/Roma people emerging, there was an elementary need for statistical data on ethnicity. Since the census of 1990 data providing on ethnicity is not compulsory. Those persons are conceived as members of certain nationalities who provide voluntarily their ethnic background. According to this census only 1.4% of the whole population belonged to the Gypsy/Roma ethnic minority group. Changes in census data and divergent results of estimations using different sources induced a complex survey by the Hungarian Central Statistical Office in 1993. This survey used a special research method to reveal the number, and

living conditions, of Gypsy/Roma people. Among classification systems voluntary data supply is the most unreliable (14). This is why researchers employed special data collectors with personal local experiences to conduct a non-self-professed survey on household lifestyle classification. As a result of this research 3.9% of the whole population was classified as having a Gypsy lifestyle (15). This research method led to a more complex overview on Gypsy/Roma people, at the same time, it led to the strengthening of discrimination against Gypsy/Roma people as during the inquiry local non Gypsy/Roma population not just identified them as Gypsy but separated their lifestyles from the Gypsy lifestyle.

Disadvantages in the case of Gypsy/Roma people grew in access to jobs and education as well. Judgements on them were worsened by demographic prognoses from decades-long data. For example, according to the comparison of this 3.9% Gypsy/Roma people in the whole population, and the 3% from the census of 1971 means that the proportion of Gypsy/Roma citizens grew in time in the total population as the live-birth number is three times higher than the average (16). So the prediction has estimated 7% Roma population for our days.

On the other hand, the disappearance of ethnic data also had some positive effects. Criminals with Gypsy/Roma ethnic background were recorded in distinct lists in police departments and juridical courts before the change of regime, inducing the thought on a generic connection of Roma ethnicity and criminal willingness in the public opinion. These statistic gathering ways were also prohibited.

How has the heterogeneity/granularity in these classifications developed historically?

As the Hungarian Central Statistical Office history lists (17), in the Habsburg Empire there was an effort taken to categorise ethnic background of the population as it had a heterogeneous ethnic composition. Dating back to the ages of the Austro-Hungarian Monarchy, it has been a multi-ethnic state. 1869 was the first year when a complete enumeration of the population took place extended to the whole country. Since the second census (1880) Hungary has had a decennial national statistics gathering harmonised with other European countries.

Following World War I Hungary became a nation-state which means Hungarians predominantly have been in majority in the ethnic map of Hungary by having over 80% share in the country's population. The aim of the census held in 1920 was to survey the precise losses in population, however, it was not completed as Yugoslavian forces controlled the Southern part of Hungary till the end of 1921.

As many ethnic Hungarians (approximately 3.3 million) became citizens of neighbouring countries, there was an emerging consciousness of the importance of observation, which resulted in the census of

1941 introducing native language questionnaires for those who were not Hungarian-speaking citizens in the country. As a lesson learned after the second world war, all the losses and population migration (forced emigration of Germans and population exchange between Hungary and Czechoslovakia making more than a half million people leave their homes) drove census statisticians to give a sacramental confidence to gain trust from the citizens.

During the last decade of Socialism in Hungary a new passport was introduced which enabled Hungarian citizens to travel freely to other countries. This passport symbolised the finish of isolation, the ease of state party control and gave way to the strengthening of traditional cultural values and growing ethnic identity among others.

While the main goal of the 1980 census was to provide general but basic information, the program of the 1990 census mirrored peculiarities of this transitional period with all its social claims. Former administration register systems, such as the state demography register and the united employment register, were still suitable for further data records, but shortening financial frames and resistance from different parts of the society enabled only the operation of the state demography register to be maintained.

In the 2001 census a new question category appeared in the otherwise compulsory data providing process. There were some sensitive questions to answer voluntarily which needed special treatment for assuring personality rights – mother tongue, nationality, national identity, home language, religion, disability. Moreover, all the data collection went anonymously. These new measurement ways met a positive response from society. Finally, the last census in 2011 also introduced some new techniques towards gaining more satisfaction from the citizens such as the possibility to provide personal data via the internet but this did not affect the willingness to provide data on ethnicity. Two thirds of data were recorded by census conductors, while a great number of filled census questionnaires arrived online and via post. This multiple-method for filling the census facilitated the recording process, although it made the data process more difficult as summing up paper-based answers took more than a year.

At the same time, the Hungarian Central Statistical Office has started the Civilian Partner Program, which is a wide scale co-operation platform involving civil organisations (organisation leaders and representatives of disabled people, nationalities, churches and local authorities) in the process of forming the census questionnaire. Moreover, during the census campaign several civil organisations, among nationality organisations, encouraged citizens to provide their specific identities and ties to bring census results closer to reality. Not considerably, however this program resulted in the average rise in the number of those who declared themselves as belonging to a certain ethnic minority group in Hungary.

Why has the heterogeneity/granularity in these classifications developed in terms of their social, historical and political context?

The reason why Hungary lacks exact databases of information on ethnicity and the health condition of the population derives from new ideological frameworks from the change of regime in the very beginning of the 1990s. Official views which are criticised by many social researchers (19) have two grounding arguments. The first one questions the worthiness of research on health conditions of Gypsy/Roma people as, according to the survey results of the Institute of Nation and Health Care from 1989, health conditions of Gypsy/Roma people did not differ from those of the poor people in general. While the second argument supposes that there might be some who would derive political consequences from statistical results as a way to blame Gypsy/Roma people for spreading epidemics and other diseases which is very dangerous, however, the aptitude for these diseases depends on family genetics not on ethnic differences. Challenging views of social researchers built a more complex interpretation on this question. According to Puporka and Zádori (1998) health conditions of Gypsy/Roma people are special and differ, although slightly, from those of the poor people in general. At the same time, survey results of Delphoi Consulting Research Group (Babusik, 2002) on the general health status of the Roma population in Borsod-Abaúj-Zemplén county discovered some special relationships between poverty, deprived living circumstances, and certain illnesses. Regarding the second official argument Puporka and Zádori argue, through exact databases on the relation of ethnicity and health conditions that targeted prevention and intervention programs could be introduced to improve general health conditions. However, today data gathering on ethnicity meets challenges in Hungary as many sources are unreferenced and varying, and categorise indicators differently and scantly.

Why are disaggregated data are not being collected, analysed, or reported more often

The general demographic data collection methods have changed radically after the change of regime, mirroring sensitiveness to antidiscrimination which makes it more difficult to gather exact information on the relations of ethnicity and health conditions. Since then, the main aim has been to provide general and basic data for national territorial economic planning for the future and to give proper information on demographic stratification. National level data gathering on ethnicity has been used for describing majority and minority relations of society.

How have disaggregated data been used to impact on policies, programmes, and population health outcomes.

There are some additional research studies which focus not exactly on the relation of ethnicity and health conditions, however, these could serve as good examples for concentrated investigations on ethnicity and health conditions. Babusik (19) used a 1500 sample of Roma households concentrating

on young adults between 19 and 34 to reveal possible employment and health conditions of Roma people. In their study they mixed their own database with county level ethnicity estimates and discovered that this social group is a target of significant discrimination. The EC EUROPA EU Report of 2003 (20) also revealed studies on the health conditions and health behaviours of the adult Roma population and their relationship with services in the health care system. They stated that all minority groups whose members can be stigmatised by their different ethnic traits, are in fact discriminated against at every association with representatives of institutions of the society. Forray in her general overview on the health conditions of Roma people in 2013 (21) provides a complex list of research related to ethnicity and health conditions of Roma people from the last decades. In her list she cites two adequate surveys on the relation of living circumstances and health conditions in counties where poor inhabitants are concentrated (2003, 2006). The first one studies connections between health damaging characteristics and low education particularities of the Roma and non Roma population, while the second one concentrates on poverty and not on ethnicity, but reveals that most Roma people belong to the deprived part of the population.

Among others these studies and reports contributed to the National Social Closing Up Strategy of 2011 (22) that aims to roll back pauperisation, which is most characteristic to Roma people in Hungary and to introduce special antidiscrimination programs for enhancing social integration for Roma people. This strategy involves governmental implementation plans in the fields of child welfare, education, employment, healthcare, housing and antidiscrimination through shaping attitudes and involving all parties concerned.

Are these ongoing developments to improve its current system of classification?

Today's policy developments both alter and improve the current system of ethnicity classification in Hungary. Census data gathering is revised quite often in order to map ethnic complexity of the population and to ensure more cultural prosperity of nationalities. However, a core principle of antidiscrimination hinders the applicability of data on ethnicity in other fields, such as in the field of health care.

Notes

(1) The Hungarian Minority Act of 1993 stated "All groups of people who have lived in the territory of the Republic of Hungary for at least one century, who represent a numerical minority in the country's population, whose members are Hungarian citizens, who are distinguished from the rest of the population by their own languages, cultures, and traditions, who demonstrate a sense of belonging together that is aimed at preserving all of these and expressing and protecting the interests of their historical communities". In Hungarian alphabetic order the Bulgarian, Roma, Greek, Croatian, Polish, German, Armenian, Romanian, Ruthenian, Serbian, Slovakian, Slovene and Ukrainian communities were defined as national and ethnic minorities in Hungary. In the phrase 'national and ethnic minority' 'national' referred to those minorities who had a mother country and living ties with its people while 'ethnic' referred to a minority which lacked any of these ties. This way Hungary had only one ethnic minority group, the Roma while twelve national minorities, however, Ruthenians do not have their own country. Since 2012, this phrase has been replaced by the term 'nationality' (nemzetiség) which comes from the Hungarian word for nation (nemzet) and means ethnic group or ethnic community as Act CLXXIX of 2011 on the Rights of Nationalities came into force on the first of January, 2012.

(2) According to the International Organisation for Migration and to the Project site of Research on Migrant Organisations in Hungary the country is on latest migration routes to Europe (on the Eastern Mediterranean Route) but the country itself is rather a transition than a destination country. This is why residing migrants live mostly in reception centres of Hungary for a short period of time. Migration Issues in Hungary. IOM. http://www.iom.hu/migration-issues-hungary (access 12/07/2016) and http://www.kisebbsegkutato.tk.mta.hu/migrans-szervezetek-magyarorszagon-1 (access 12/07/2016)

(3) Act LXXVII of 1993 on the Rights of National and Ethnic Minorities, Chapter 1, Section 1, Subsection (2), then Act CLXXIX of 2011 on the Rights of Nationalities, Chapter 1. Section 1, Subsections 1-3. NetJogtár. http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=A1100179.TV (access 12/01/2016) and Website of the Minority Research Institute of the Hungarian Academy of Sciences. http://www.kisebbsegkutato.tk.mta.hu/ (access 12/01/2016)

(4) The Hungarian Central Statistical Office. Központi Statisztikai Hivatal. KSH. www.ksh.hu

(5) Tóth, Ágnes and Vékás, János (2013): A magyarországi nemzetiségek létszámváltozása 2001 és 2011 között. KSH Statisztikai Szemle. 2013/12. http://www.ksh.hu/statszemle_archive/2013/2013_12/2013_12_1256.pdf (access 21/02/2016)

(6) Kevesebb a vallásos, több a cigány. (2013) Index. http://index.hu/belfold/2013/03/28/kevesebb_a_vallasos_tobb_a_cigany/ (14/06/2016)

(7) same reference

(8) How Gypsy people are called and how they call themselves is a topic of disputes in various scientific fields. Zsombor Grétsy linguist traced back the historical rootes of the most common denominations Gypsy and Roma which have different kinds of interpretations in Hungarian. Today people of this community prefer to call themselves Roma instead of Cigány which is quite the same as Gypsy although has a different lingual basis and many times is commonly used by non Roma people in a pejorative way. Zsombor Grétsy (2011):Cigány? Roma? Dzsipszi? Melyik a píszí? Nyelv és Tudomány. http://www.nyest.hu/hirek/cigany-roma-dzsipszi-melyik-a-piszi (access 10/05/2016)

(9) In the 1990s there were five surveys conducted on determining the number of Roma people in the country with useful documentations exactly with the reason of finding relations between ethnic background of Roma people and their health conditions. Moreover, during the census of 1990 which unfolded demographic, social, educational, employment and wealth circumstances of the country's population, gathering data on the living conditions of Roma people could also serve as a useful basis for researches on this question. Puporka and Zádori (1998), Babusik (2002), Dombayné et all (2003), Forray (2013), the Health Status Data on Roma in Hungary (2014) and Gyukits (2015).

(10) Koller, Inez Zsófia (2014): Ethnic Minorities and Censuses. In: István Horváth, Ibolya Székely, Tünde Székely, Tonk, M. (ed.) Minority representation and minority language rights. Kolozsvár: Scientia Publisher, pp. 320.

(11) Such as in the UN International Covenant on Civil and Political Rights (1976), int he UN Declaration ont he Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities (1992), in the European Charter for Regional and Minority Languages (1992) and the Framework Convention for the Protection of National Minorities (1998) adopted by the Council of Europe or int he Copenhagen Document (1990) of OSCE.

(12) Koller, Inez Zsófia (2014): Ethnic Minorities and Censuses. In: István Horváth, Ibolya Székely, Tünde Székely, Tonk, M. (ed.) Minority representation and minority language rights. Kolozsvár: Scientia Publisher, pp. 324-330.

(13) The Hungarian Census Questionnaire can be found ont he website of the United Nations Statistics Division Demographic and Social Statistics. http://unstats.un.org/unsd/demographic/sources/census/censusquest.htm

(14) According to Puporka and Zádori (1998) Gypsy/Roma people in Hungary live their lives evidently differently from non Gypsy/Roma population in their surroundings, and because of it, non Gypsy/Roma people reckon them as Gypsy/Roma while they do not confess themselves as Gypsy/Roma for different reasons.

(15) Forray (2013) writes, it was recognised as a sociological statement in the 1970s that Gypsy inhabitants of Hungary differ in their lifestyle from the majority of the society and that in the

forming and differing of these groups the scale and the character of inclusion into social division of labour is determining.

(16) Puporka and Zádori, 1998.

(17) The History of Censuses in Hungary. Hungarian Central Statistical Office. https://www.ksh.hu/nepszamlalas/magyarorszagi_nepszamlalasok_tortenete (02/02/2016)

(18) Babusik, Ferenc and Papp, Géza: A cigányság egészségi állapota. Szociális, gazdasági és egészségügyi helyzet Borsod-Abaúj-Zemplén megyében. Delphoi Kutatók Magyarországi Roma népességre vonatkozó kutatásai. Esély, 2002/3 http://www.delphoi.hu/romanep.html (access 10/12/2015)

(19) The EC EUROPA EU Report was prepared by a working team in 2001 an 2002 (Dombainé Arany Vera, Solymosy József, Kanyik Csaba, Daróczi Gábor) ordered by the Ministry of Health Affairs. The Report was part of an international program with contributors from Bulgaria, the Czech Republic, Greece, Hungary, Italy, Portugal, Romania, Slovakia and Spain.

(20) Forray R. Katalin. Cigány egészség, cigány betegség. 2013 www.hier.iif.hu/hu/letoltes.php?fid=tartalomsor/2244 (access 10/12/2015)

(21) Nemzeti Társadalmi Felzárkóztatási Stratégia. 2011. http://romagov.kormany.hu/nemzeti-tarsadalmi-felzarkozasi-strategia (access 24/07/2016)

References

Babusik, Ferenc and Papp, Géza: A cigányság egészségi állapota. Szociális, gazdasági és egészségügyi helyzet Borsod-Abaúj-Zemplén megyében. Delphoi Kutatók Magyarországi Roma népességre vonatkozó kutatásai. Esély, 2002/3 http://www.delphoi.hu/romanep.html (access 10/12/2015)

Dombainé Arany Vera, Solymosy József, Kanyik Csaba, Daróczi Gábor: Egyenlőség, egészség és a roma/cigány közösség. 2003 http://ec.europa.eu/health/ph_projects/2004/action3/docs/2004_3_01_manuals_hu.pdf (10/12/2015)

Egészségügy, kommunikáció, cigányság. Soros Alapítvány. http://www.kka.hu/_soros/kiadvany.nsf/44cfa372d3c5a279c1256e9600682640/6f90012f0359381fc1256e900045df2e?OpenDocument (10/12/2015)

Forray R. Katalin. Cigány egészség, cigány betegség. 2013 www.hier.iif.hu/hu/letoltes.php?fid=tartalomsor/2244 (access 10/12/2015)

Grétsy, Zsombor (2011):Cigány? Roma? Dzsipszi? Melyik a pízí? Nyelv és Tudomány. http://www.nyest.hu/hirek/cigany-roma-dzsipszi-melyik-a-piszi (access 10/05/2016)

Gyukits, György: Az egészségügy vesztesei. http://beszelo.c3.hu/cikkek/az-egeszsegugy-vesztesei (10/12/2015)

Hoóz, István: Population census and nationality. Minorities Research 3. http://www.hhrf.org/kisebbsegkutatas/mr_03/cikk.php?id=1229 (access 12/06/2012)

Kemény István (ed.) (2000): A magyarországi romák. Változó Világ 31., Budapest

Kertzer, David, Arel, Dominique (ed.) (2004): Census and Identity. The Politics of Race, Ethnicity and Language in National Censuses. Cambridge. New Perspectives on Anthropological and Social Demography.

Koller, Inez Zsófia (2014): Ethnic Minorities and Censuses. In: István Horváth, Ibolya Székely, Tünde Székely, Tonk, M. (ed.) Minority representation and minority language rights. Kolozsvár: Scientia Publisher, pp. 319-333.

Majtényi, Balázs: The Autonomy of Minorities and Registration. Minorities Research 5. http://www.hhrf.org/kisebbsegkutatas/mr_05/cikk.php?id=1208 (access 12/06/2012)

Puporka, Lajos and Zádori, Zsolt (1998): A magyarországi romák egészségi állapota. http://siteresources.worldbank.org/EXTROMA/Resources/health_status_hun.htm (10/12/2015)

Roma Health Report. Health status of the Roma population. Data collection in the Member States of the European Union (2014) Health and Customers. European Union. 99-102. pp. http://ec.europa.eu/health/social_determinants/docs/2014_roma_health_report_en.pdf (02/06/2016)

Tóth, Ágnes and Vékás, János (2013): A magyarországi nemzetiségek létszámváltozása 2001 és 2011 között. KSH Statisztikai Szemle. 2013/12.

Appendix 8

Appendix 8.1

Table on changes in the number of ethnic nationalities between 2001 and 2011

	2001	2011
Total	10 198 315	9 937 628
Roma	205720	315583
German	120344	185696
Romanian	14781	35641
Slovakian	39266	35208
Croatian	25730	26774
Ukrainian	7393	7396
Serbian	7350	10038
Greeks	6619	4642
Polish people	5144	7001
Slovenians	4832	2820
Bulgarians	2316	6272
Ruthenians	2079	3882
Armenians	1165	3571

Source: Hungarian Central Statistical Office

Chapter 9: Ethnic group classification in Malaysia

Shyamala Nagaraj and Chiu-Wan Ng

Abstract

Malaysia is a multi-ethnic society. Historically, the country is home to a multitude of indigenous tribal groups. The country's geographical position in the middle of maritime trade routes between the east and the west, as well as British colonial policies of bringing in migrant workers from countries in the region to work in rubber plantations and tin mines, helped set the scene for increasing ethnic diversity over the past two centuries. Malaysia has also seen an increasing presence of migrant workers in agriculture, construction and services mostly from Indonesia, but also from Nepal, Bangladesh and the Philippines, often through inter-governmental arrangements. Different from earlier British policy, these migrants are required to return home after a fixed period. Economic opportunities have also made Malaysia a magnet for illegal economic migrants from neighbouring countries with which it shares borders.

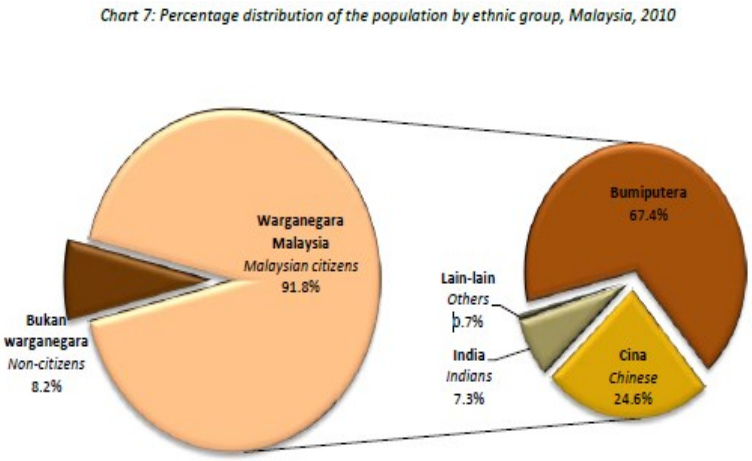
Public agencies, in particular the Department of Statistics Malaysia, take the lead in efforts to accurately measure ethnic diversity for purposes of policy formulation and evaluation. Ethnicity is essentially self-reported and only one ethnic category is recorded per person. The granularity and identification of ethnic categories have changed and improved over time in line with changes in size of a group or its importance to public policy. Though data capture is often granular, information on ethnicity is mainly reported by only a few broad ethnic groups: Bumiputera, Chinese, Indian and others. Malaysia provides a unique example of the impact of public policy and concerns on ethnicity classification. The ethnic category Bumiputera (translation: princes of the soil) is a result of the New Economic Policy (NEP) first introduced in 1971 that provides special benefits to Malays, the largest ethnic sub-group in the Bumiputera category and to selected indigenous groups.

Malaysia has a welfare based health system. Health policies have been aimed at reducing health disparities between sub-populations which may or may not coincide with ethnic classification. The poor health status of rural communities has been a policy focus in the past but policy attention is shifting towards health needs of the urban communities. Health data by ethnicity captured by public agencies, in particular the Ministry of Health Malaysia, are often quite granular. Studies consider ethnicity a socio-political construct that can be used essentially as a social determinant of health.

Introduction: Demographic background

Malaysia is a multi-ethnic society. Information on ethnicity is often reported by a few broad ethnic groups and country of birth or citizenship: Bumiputera, Indians, Chinese, Others and Non-citizens, as for example in a table from the 2010 Census (see Figure 9.1). The Bumiputera category, a politically defined ethnic group, comprises mostly Malays. Data collected, however, is quite granular as seen in the 110 ethnic codes of the 2010 Population Census (see Figure 9.2). Ethnicity is essentially self-reported and only one ethnic category is recorded per person. Ethnicity is officially documented in the national registration documents, Identity Card (MyKad), birth certificate and death certificate. MyKad information is, however, accessible only via appropriate card-readers and its use limited by legislation.

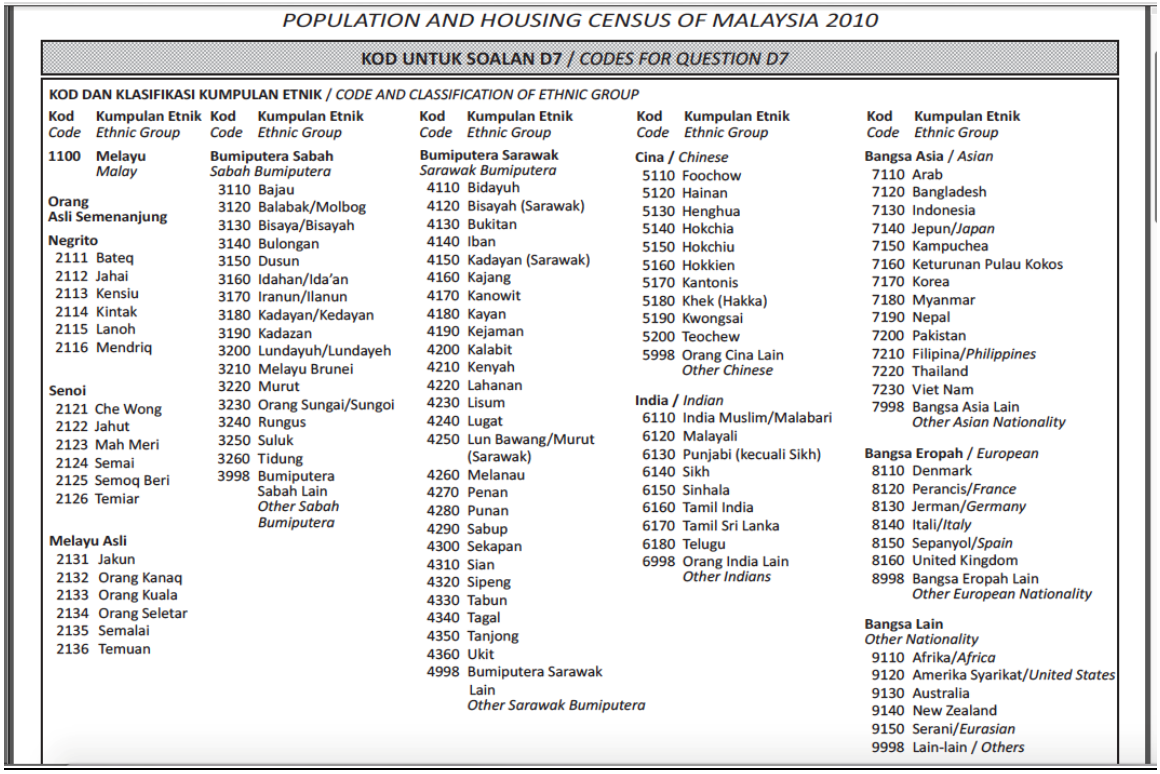
Figure 9.1: Percentage Distribution of the Population by Ethnic group, 2010.



Source: Chart 7, Population Distribution and Basic Demographic Statistics 2010, Department of Statistics, Malaysia, 2011

The historical circumstances for the great variety of ethnic, racial and linguistic groups in Malaysia lie in its geographical location (see Figure 9.3). The region that is now Malaysia comprises Peninsular Malaysia on the Asian continent, and Sabah and Sarawak on the island of Borneo. Peninsular Malaysia lies at the crossroads of maritime trade between the West (India, Arabia) and the East (China). The seas between Sabah and the Sulu islands are trading routes between Australia and China. Even ancient maritime empires from India or China governed lands in the region. There have thus long been movements of peoples between the West and the East and within Southeast Asia itself (Andaya and Andaya 1982).

Figure 9.2 Codes for Ethnicity, 2010 Census, Malaysia



POPULATION AND HOUSING CENSUS OF MALAYSIA 2010

KOD UNTUK SOALAN D7 / *CODES FOR QUESTION D7*

KOD DAN KLASIFIKASI KUMPULAN ETNIK / *CODE AND CLASSIFICATION OF ETHNIC GROUP*

Kod *Code*	Kumpulan Etnik *Ethnic Group*
1100	**Melayu** *Malay*
	Orang Asli Semenanjung
	Negrito
2111	Bateq
2112	Jahai
2113	Kensiu
2114	Kintak
2115	Lanoh
2116	Mendriq
	Senoi
2121	Che Wong
2122	Jahut
2123	Mah Meri
2124	Semai
2125	Semoq Beri
2126	Temiar
	Melayu Asli
2131	Jakun
2132	Orang Kanaq
2133	Orang Kuala
2134	Orang Seletar
2135	Semalai
2136	Temuan

Kod *Code*	Kumpulan Etnik *Ethnic Group*
	Bumiputera Sabah *Sabah Bumiputera*
3110	Bajau
3120	Balabak/Molbog
3130	Bisaya/Bisayah
3140	Bulongan
3150	Dusun
3160	Idahan/Ida'an
3170	Iranun/Ilanun
3180	Kadayan/Kedayan
3190	Kadazan
3200	Lundayuh/Lundayeh
3210	Melayu Brunei
3220	Murut
3230	Orang Sungai/Sungoi
3240	Rungus
3250	Suluk
3260	Tidung
3998	Bumiputera Sabah Lain *Other Sabah Bumiputera*

Kod *Code*	Kumpulan Etnik *Ethnic Group*
	Bumiputera Sarawak *Sarawak Bumiputera*
4110	Bidayuh
4120	Bisayah (Sarawak)
4130	Bukitan
4140	Iban
4150	Kadayan (Sarawak)
4160	Kajang
4170	Kanowit
4180	Kayan
4190	Kejaman
4200	Kalabit
4210	Kenyah
4220	Lahanan
4230	Lisum
4240	Lugat
4250	Lun Bawang/Murut (Sarawak)
4260	Melanau
4270	Penan
4280	Punan
4290	Sabup
4300	Sekapan
4310	Sian
4320	Sipeng
4330	Tabun
4340	Tagal
4350	Tanjong
4360	Ukit
4998	Bumiputera Sarawak Lain *Other Sarawak Bumiputera*

Kod *Code*	Kumpulan Etnik *Ethnic Group*
	Cina / *Chinese*
5110	Foochow
5120	Hainan
5130	Henghua
5140	Hokchia
5150	Hokchiu
5160	Hokkien
5170	Kantonis
5180	Khek (Hakka)
5190	Kwongsai
5200	Teochew
5998	Orang Cina Lain *Other Chinese*
	India / *Indian*
6110	India Muslim/Malabari
6120	Malayali
6130	Punjabi (kecuali Sikh)
6140	Sikh
6150	Sinhala
6160	Tamil India
6170	Tamil Sri Lanka
6180	Telugu
6998	Orang India Lain *Other Indians*

Kod *Code*	Kumpulan Etnik *Ethnic Group*
	Bangsa Asia / *Asian*
7110	Arab
7120	Bangladesh
7130	Indonesia
7140	Jepun/*Japan*
7150	Kampuchea
7160	Keturunan Pulau Kokos
7170	Korea
7180	Myanmar
7190	Nepal
7200	Pakistan
7210	Filipina/*Philippines*
7220	Thailand
7230	Viet Nam
7998	Bangsa Asia Lain *Other Asian Nationality*
	Bangsa Eropah / *European*
8110	Denmark
8120	Perancis/*France*
8130	Jerman/*Germany*
8140	Itali/*Italy*
8150	Sepanyol/*Spain*
8160	United Kingdom
8998	Bangsa Eropah Lain *Other European Nationality*
	Bangsa Lain *Other Nationality*
9110	Afrika/*Africa*
9120	Amerika Syarikat/*United States*
9130	Australia
9140	New Zealand
9150	Serani/*Eurasian*
9998	Lain-lain / *Others*

Source: 2010 Census, Department of Statistics, Malaysia.

The inflow of immigrant workers in the somewhat large numbers during the last two centuries has helped solidify the ethnic fabric of the country. The British brought in migrant labour to work on rubber estates (from India) and tin mines (from China) in the nineteenth and early twentieth centuries. Since the 1970s, Malaysia has seen an increasing presence of migrant workers in agriculture, construction and services mostly from Indonesia, but also from Nepal, Bangladesh and the Philippines, often through inter-governmental arrangements. Different from earlier British policy, these migrants are required to return home after a fixed period. Economic opportunities have also made Malaysia a magnet for illegal economic migrants from neighbouring countries with which it shares borders.

Figure 9.3 Geographical Location of Malaysia in Asia

Malaysia provides a unique example of the impact of public policy and concerns on ethnicity classification (Nagaraj, Lee et al. 2009). We highlight two groups. The first is Bumiputera (translation: prince of the soil), an ethnic classification resulting from the New Economic Policy (NEP) first introduced in 1971. The NEP provides special benefits to Malays, the largest ethnic sub-group in the Bumiputera category, and to selected indigenous groups defined or identified constitutionally. In particular, Malays are defined by religion (Islam), language (Malay) and 'Malay customs'. Thus, the Bumiputera ethnic group is less about common customs (although a majority do share common customs) and more about meeting official policy targets. The second group is Non-citizens, a classification often cited alongside information on ethnicity, although it is really about citizenship.

The presence of large numbers of migrant workers (so large it competes in size with Indian Malaysians) has meant that this group, even if heterogeneous, must be identified as a group in Malaysian society. We note in passing that since many of these foreign workers are Muslim, one can expect that people who once identified as non-Malaysians may at a later identify as Malay or *Bumiputera*, and not just from inter-marriages, but also from assimilation into Muslim communities.

Sources of official and unofficial data

The National Registration Department (NRD) maintains vital statistics data. All births must be registered, and ethnicity and religion, among others, are noted on the birth certificate. For deaths, the informant would also need to provide information on the deceased's ethnic group, religion, and identity card to the approved authorities to receive a burial permit. With self-reporting, it is expected that details on the deceased's ethnicity and religion are counterchecked with that in the MyKad. The NRD does not produce routine reports although information on births and deaths are made available to the Department of Statistics Malaysia (DOSM) for reporting.

The counting of Malaysia's ethnic groups is an important function of the population censuses run by the DOSM. Granularity in the censuses is broadly similar to birth and death data but there may be more or less disaggregation for specific sub-groups. For example, the 2010 Census had 110 categories for ethnicity, whereas the NRD has 149 categories (see table 9.1). Specifically, the Census had just one Malay category whereas the NRD has six. In national surveys, the number of categories is usually much fewer than the census as some minority groups are few in number. Information on languages and/ or dialects spoken is no specifically collected, although some dialect groups have been incorporated into ethnic sub-groups listed in the coding. The census also captures religion, birthplace and citizenship, but cross-tabulations with ethnicity are limited in public documents.

Information on ethnicity may be recorded by an enumerator or by the respondent on a form, or even digitally as for the recent 2010 Census. Only one ethnicity is captured. Ethnic categories are pre-coded. Since only one ethnicity is recorded, mixed parentage is not captured. Children of mixed parentage

often report their father's ethnicity, but may also report the ethnicity they identify with. There is mostly no public interest in the statistical capture of such information, although society acknowledges the presence of such groups through the common use of terms like "Chindian" (Chinese-Indian mix). Data on mixed marriages are also not published, although such information can be extracted from household information in censuses. We use the term marriage in a broad sense, since there is no information obtained on the formal nature of a couple's relationship in the census. Such studies are usually carried out intermittently by researchers, or other agencies like the National Population and Family Development Board (NPFDB), through small scale surveys. Thus, the rich tapestry that is Malaysia's ethnicity is, despite deep granularity in recording, not completely documented.

Table 9.1 Granularity of Ethnicity, National Registration Department (NRD) and 2010 Census, Malaysia

Number of Categories	**NRD**	**2010 Census**
Malay	6	1
Chinese	10	11
Indians	13	9
Bumiputera, Sabah	33	17
Bumiputera, Sarawak	58	27
Indigenous, Peninsular Malaysia	8	18
Foreigners	21	27
Total	149	110

Source: 2010 Census, NRD.

Unlike many countries in the Asia Pacific region, the Malaysian health system is not based on a single payer public insurance system and thus does not have a centralised comprehensive health information system common in all such systems. The Malaysian public can freely choose to seek care from low priced taxation-based public sector providers or from the significantly more expensive privately funded private sector providers. The Ministry of Health (MoH) Malaysia is the main agency responsible for the governance of the health sector in the country as well as the provision of health care services. As such, the MoH has the mandate to collect data on inpatient and outpatient care episodes from public facilities and increasingly from private facilities as well. Patient-level clinical data, including information on ethnicity, are thus captured in diverse health provider localities. In addition, since

patients are freely able to choose their providers, there may be duplication of clinical data collected for individual patients.

The MoH or its agencies maintain several disease specific registries like the cancer and HIV/AIDS registries. HIV/AIDS is a notifiable disease under the Prevention and Control of Infectious Diseases Act 1988. The information obtained is used for various policies including the decision to fund antiretroviral therapy for HIV patients and the estimation of the resources needed to implement this policy. Distribution of HIV cases by ethnic groups is also routinely shared among various governmental and international agencies such as the National Antidrug Agency of Malaysia and the Prisons Department (Ministry of Health Malaysia 2010).

The MoH conducts regular national household health surveys to collect data (including ethnicity) on health status and health seeking behaviours to aid development of health policies. The most comprehensive of these surveys are the National Health and Morbidity Surveys (NHMS) which started in 1986. The NHMS 2011-2014 included student health and adult nutrition. The on-going NHMS 2015-2018 focuses on non-communicable diseases and selected infectious diseases such as tuberculosis and dengue. Since the late 1960s, the NPFDB has collected data (including ethnicity) on fertility, family planning and contraceptive use.

The number of ethnic categories is usually far fewer in a survey than that in the census. For example, the 1996 and 2006 NHMS surveys provided for 46 groups, the 2011 NHMS survey had 18 groups. However, even if the sizes of some of the ethnic groups are small at certain levels and hence not identified, these groups may be identified in a regional sub-population of interest where they are a significant group. For example, patients at MoH clinics in Peninsular Malaysia who are aborigines may be classified as 'Others' but in Sabah and Sarawak, the actual aboriginal group may be noted. The MoH routinely publishes reports on morbidity based on its NHMS surveys and includes findings by broad ethnic groupings.

In reporting, ethnicity data is usually a priority and may even be released along with other essential demographic data well before the general report on a census (Chander 1972). Regardless of how granular the categories of ethnicity collected, data for the nation may be released, or findings may be reported, usually only for the broad ethnic groups, that is, Malays, Other Bumiputera, Chinese, Indians and Others. There are exceptions when information of a local nature is required, as for example, with ethnic composition in a parliamentary constituency. The granularity in release is also constrained by the size of the ethnic group in the population.

Ethnicity is an important factor in empirical research, including public health. However, while information on ethnicity at a national level is collected and maintained by public producers of data, release of such information beyond the primary tables, either as cross-tabulations of interest or raw data samples, is limited. In particular, it is rarely available to the public for in-depth study, and researchers need to apply for permission of use; confidentiality is seen as a rein on ethnic sensitivities. Primary data collected by researchers to study distributions of health conditions and disease risk factors often include

ethnicity as a category in addition to other demographic information like age, gender and education. For these, we expect granularity to follow the scope and objectives of the specific study.

Lessons outside the health field

The most important lesson is actually a question, what does ethnicity really mean when formulating health policies that affect a particular ethnic group? The identification of a specific ethnic group can be only as good as its measurement, and that too so long as the group is reasonably homogenous. The issue is especially pertinent in Malaysia for the politically defined ethnic group, Bumiputera. The reality is that the Bumiputera group, especially Malays, are an increasingly heterogeneous group. In using Bumiputera to identify a sub-group of the population, we are really using a socio-political construct and a grouping that has a degree of impermanence. Interpreting data by ethnicity for health policy requires therefore careful assessment and evaluation.

It may be interesting to also note that over time Malaysia's health professionals may encounter medical problems common elsewhere in the world. For example, Malaysia is increasingly attracting Muslims from elsewhere in the world to its educational facilities, and students who stay in the country for extended periods of time have families. There may come a time when birth complications arising from maternal female genital mutilation (highly prevalent among some African countries) become more commonplace in local hospitals. The level of disaggregation by ethnicity presently available in data collection will not capture the extent of such problems.

How has the heterogeneity/granularity in these classifications developed historically in these countries?

Malaysia has long been concerned with the measurement of its many ethnic groups, whether for political, economic or social reasons. The recording of ethnicity by the NRD for the purpose of identity cards is based on self-reporting, and discussions of the coding are not a matter of public record. The deliberation of both the terms used to capture ethnicity and the actual categories have probably been most important in the decennial population censuses, which seek to enumerate and document the diversity in the nation's population (see appendix 9).

Over the years, the specific form of the question measuring ethnicity in the population census has been modified to capture ethnic/ dialect groups, and the term used has changed from 'nationality' to race to ethnicity/ community/ dialect(Nagaraj, Lee et al. 2009). Today, the information on ethnicity captured in the census is a mix of one ethnic group with no sub-groups (Malays), many indigenous groups (grouped under Other Bumiputera*)*, dialects (Chinese) and origins (Indians). The classification, one could argue, is the result of more than a century of experience in measurement as the ethnic fabric of society has evolved, and reflects the careful efforts of the various Superintendents of Census to define

a diverse population. Most public collectors of demographic data for national surveys take the lead from the census in definition and granularity.

Why has the heterogeneity/granularity in these classifications developed in these countries in terms of their social, historical and political context?

The experience of measurement of ethnicity in Malaysia shows that not only does measurement of ethnic data support policy but policy also drives ethnic measurement in data. The measurement of ethnicity reflects attempts to capture a conceptualization of an ethnic group as one that shares common interests such as language, religion and customs (Hirschman 1987, Nagaraj, Lee et al. 2009). The granularity and identification of ethnic categories have changed and improved across time, in line with changes in size of a group or its importance to public policy. Statisticians have demonstrated their determination in collecting census data from people of "many tongues", even against the odds of collecting data in the remotest parts of Sabah and Sarawak, doing so on a relatively regular interval. This has made possible the fairly detailed ethnic classification now used in censuses.

Since information on ethnicity is obtained by self-identification, it essentially measures identity or the group the respondent perceives himself or herself to be from. This is particularly true today for the Bumiputera category. The somewhat loose constitutional definition of a Bumiputera has entered the social realm to the extent that the Bumiputera community is seen as an ethnic group. It can be argued that this meets Sawyer's (Sawyer 1998) criteria for establishing an ethnic category for statistical purposes: consistency and comparability of data over time as well a category that is widely understood, so that meaningful comparisons can be made to evaluate social progress.

Nevertheless, it cannot be denied that despite all these years of experience in counting, there can still be confusion in society about concepts such as race (example, Chinese), dialect group (example, Hokkien or Cantonese), language group (example, Tamil, Telegu), nationality (Indian vs Sri Lankan) or even the term ethnicity.

Why disaggregated data are not being collected, analysed, or reported more often if the field generally agrees that this is critical to understanding disparities?

Reporting and analysis are usually presented by broad ethnic categories, in part recognition of the small size of certain groups in relation to especially health outcomes, and in part recognition of the politically appropriate and acceptable groupings in society. We have noted that Malays is quite a heterogeneous category; however for political reasons, further disaggregation may not be acceptable.

Health concerns of the MoH, however, are generally recognition of needs and the provision of care to specific communities rather than to specific ethnic groups. Moreover, identifying health concerns by ethnic groups does not always mean that their healthcare needs will be recognized or addressed. The

groups need to be recognized as being important enough or the healthcare concerns great enough for intervention.

For example, the rapid urbanisation over the last few decades has led to equity concerns about access to affordable primary care for urban folk since the vast network of public clinics exist mainly in the rural areas. On the other hand, the health of the rural orang asli living in remote areas of the Malaysian Peninsular and North Borneo remains a concern. Some tribal groups are open to assimilation to modern society whilst others retain traditional nomadic lifestyles. In general, the higher health needs of the orang asli have not been comprehensively evaluated, officially acknowledged nor provided for.

Another area of health disparity is that of migrant health. There are large numbers of both legal and illegal migrants. Legal migrants are able to access public care but need to pay higher fees than citizens. However, illegal migrants risk detention and deportation if they seek care in public institutions. Since policy-wise this vulnerable group is not entitled to public care, their health needs have also not been studied comprehensively.

Are there any examples of how the disaggregated data have been used to impact on policies, programmes, and population health outcomes?

The Malaysian government has made known its intention to reduce health disparities across all communities and not across specific ethnic groups. In many instances, however, it is not immediately obvious how disaggregated health morbidity and mortality data have influenced public policies in Malaysia. Though such information is known to be generated and shared between public agencies and ministries, the policy making processes in such public agencies are generally not openly discussed in the public domain. Sometimes, addressing these disparities across communities can have outcomes that impact certain ethnic groups. For example, the main concern at the time of Independence in 1957 was to reduce rural/urban disparities in recognition of the much lower health status of rural communities. The Rural Health Services (RHS), which were implemented in stages from the early 1960s, have been acknowledged both locally and internationally, to have successfully allowed rural communities access to near-free primary care close to home with referrals to higher levels of care in larger towns when needed. The programme led to the dramatic improvement within a decade in the health status of Malays who at the time made up the majority of the rural communities outside of plantations (Tan, Kwok et al. 1988).

In certain instances, survey findings of higher incidence of disease/risk factors for a particular ethnic group have contributed to heightened public discourse and may have led to policies targeting that ethnic group. An example is the finding of high prevalence of tobacco consumption among Malay males (Institute for Public Health 2012). In addition to an anti-smoking public health campaign aimed at the general public, the MoH conducted a campaign to encourage Muslims to stop smoking during the fasting month of Ramadhan. It is not always clear whether and how information has influenced health

reforms. For example, ethnic differences have been observed in the pattern of health seeking behaviours. Surveys have shown that Malays have a preference for public care and Chinese a preference for private care, but it is uncertain if this has informed current debate.

Information on the mortality and morbidity patterns by ethnic groups has also influenced clinical decisions. Such information is usually obtained from studies that seek answers to clinical questions. The country's multi-ethnic population presents a unique opportunity to study a given disease that may present differently in different ethnic groups. Reports or publications often present findings by ethnic groups that are then discussed in the context of the disease and clinical management.

Importantly, even in the area of clinical research one should ask the question, How relevant is ethnicity in health policies in Malaysia? Ethnicity as a biological determinant of disease is likely to become less relevant over time since many Malaysians are able to trace their ancestries from more than one ethnic group. Gene pools are getting more mixed. On the other hand, ethnicity as a social determinant of disease is likely to become increasingly more relevant. Ethnicity in the Malaysian context is a social construct with the rights and privileges accorded to one ethnic group helping to distinguish it from other groups. This may have a bearing on the way different ethnic groups live their lives and thus affect the distributions of disease.

Are there ongoing to improve current systems of classification?

There has no public movement demanding improvements in classification as such, although the political impact of ethnic classification on social integration has seen much debate (Cheong, Nagaraj et al. 2009) with some voices asking for the removal of such identification altogether. One possible improvement is the ability to select more than one ethnic group in collecting information on ethnicity. However such a move would require at the very least transparent guidelines on how federal agencies should tabulate, publish, and use the data.

References

Andaya, B. W. and L. Y. Andaya (1982). A history of Malaysia. New York, St. Martin's Press.

Chander, R. (1972). 1970 Population and Housing Census: Community Groups. Kuala Lumpur, Department of Statistics, Malaysia.

Cheong, K.-C., S. Nagaraj and K.-H. Lee (2009). "Counting ethnicity: The national economic policy and social integration." Malaysian Journal of Economic Studies **46**(1): 33-52.

Fell, H. (1960). 1957 Population Census of the Federation of Malaya. Report No. 14. Kuala Lumpur, Department of Statistics, Malaysia.

Hare, G. T. (1902). Federated Malay States: Census of the Population. Kuala Lumpur, Government Printer.

Hirschman, C. (1987). "The meaning and measurement of ethnicity in Malaysia: An analysis of census classifications." The Journal of Asian Studies **46**(03): 555-582.

Institute for Public Health (2012). Report of the Global Adult Tobacco Survey Malaysia 2011. Kuala Lumpur, Institute for Public Health.

Jones, L. W. (1953). A Report on the Census of Population Held on 4th June, 1951. London, Government of the Colony of North Borneo.

Khoo, T. H. (1983). General Report of the Population Census of Malaysia, 1980. Kuala Lumpur, Department of Statistics, Malaysia.

Ministry of Health Malaysia (2010). Malaysia: 2010 UNGASS Country Progress Report. A. S. S. Disease Control Division. Putrajaya, Ministry of Health Malaysia,.

Nagaraj, S., K.-H. Lee, N.-P. Tey, C.-W. Ng and J. Pala (2009). "Counting Ethnicity in Malaysia: the Complexity of Measuring Diversity*." Malaysian Journal of Economic Studies **46**(1 & 2): 5-32.

Noakes, J. L. (1948). A Report on the 1947 Population Census. Kucing, Sarawak, Government Printers.

Sawyer, T. C. (1998). "Measuring race and ethnicity: Meeting public policy goals." The American Statistician **52**(1): 34-35.

Tan, P.-C., K.-K. Kwok, B.-A. Tan, S. Nagaraj, N.-P. Tey and S. Norazah (1988). "Socio-economic development and mortality patterns and trends in Malaysia." Asia Pacific Population Journal **2**(1): 3-20.

Appendix 9

Appendix 9.1 Terms and Categories for Ethnicity in Malaysia: A Historical Perspective

Malaysia has much experience with the measurement of its many ethnic groups, whether for political, economic or social reasons. The deliberation of both the terms used to capture ethnicity and the actual categories have probably been most important in the decennial censuses, which seek to enumerate and document the diversity in the nation's population. Hirschman (Hirschman 1987) and Nagaraj et al. (Cheong, Nagaraj et al. 2009) have explored the meaning and measurement of ethnicity in Malaysia's population census.

The first modern census was carried out in 1871 for the Straits Settlements (Penang, Malacca and Singapore) that were parts of what is now Peninsular Malaysia, then under British rule. Carried out every ten years or so subsequently, the census was slowly extended and eventually covered in 1921 the Straits Settlements, Federated Malay States and the Unfederated Malay States. North Borneo conducted its first census in 1891, while the first census for Sarawak was done carried out in 1947. With the formation in 1963 of Malaysia and the subsequent secession of Singapore in 1965, the decennial censuses since 1970 have covered Peninsular Malaysia, Sabah and Sarawak. Thus, regular censuses (other than the war years) have been carried out despite the difficulties of taking a census in a population "with so many races speaking different" (Hare 1902)(p4) or the need to have census questionnaires prepared in several languages as well as enumerators who can speak the language of the respondents.

The specific form of the question measuring ethnicity in the population census has, not surprisingly, seen change. The early years used the term 'nationality', but there were obvious difficulties in using this term to capture the various groups in the population. G. T Hare, the Superintendent of the 1901 Census of the Federated Malay States preferred the word 'race' as it is "a wider and more exhaustive expression than 'nationality' and gives rise to no such ambiguous question in classifying people" (as cited in (Hirschman 1987) p561). By 1911 the term had been changed to 'race' for the Straits Settlements as well, but 'nationality' continued to be used in North Borneo up till the 1931 census. L. W Jones, the Superintendent of the 1951 Census of North Borneo reported that the term 'nationality' was dropped, as enumerators could not distinguish between nationality and race (Jones 1953). This issue did not arise in Sarawak as the first census in 1947 used the term 'race'. J. L. Noakes, Superintendent of Census of Sarawak and Brunei 1948 recognized that there were many indigenous groups that regarded "Sarawak as their homeland" and who were "regarded as natives by their fellowmen" (Noakes 1948)(p29).

H. Fell, the Superintendent of the 1957 Census of the Federation of Malaya, instructed enumerators to use the term 'race' as "understood by the man in the street and not physical features as used by ethnologists"(Fell 1960)(p12). Nevertheless, there was still dissatisfaction with the measurement. The 1970 Population Census for Malaysia used the term 'community', a term used in the 1947 Census as well. Chander (Chander 1972)(p22) justifies the return to the practice of earlier Malayan censuses, noting that "the term race has not been used as it attempts to cover a complex set of ideas which in a

strict and scientific sense represent only a small element of what the Census taker is attempting to define." The term 'community' was used to identify a group "bound by a common language/ dialect, religion and customs."

There were further refinements and from the 1980 census, the term 'ethnic / dialectic/ community group' has been used, although its description is the same as that used for 'community' (Khoo 1983). Furthermore, although the term 'dialect' was introduced formally only in 1980, enumerators have long been instructed to note the dialect when enumerating the Chinese community. Hare (Hare 1902)(p6) recommended that in the following census language be added in a separate column as "if a person now writes 'Chinese' it is hard to say to which race of Chinese he belongs."

A major criterion for the inclusion of a group as a category should be its size in the population. Tom Harrison, in assisting in determining the categories for the Census, observes that (Noakes 1948)(p271), "classification should be as scientifically accurate as possible, the groups must be reasonably balanced in size, and it should be in sufficient detail to provide a sound basis for future scientific investigations."

In the first census, the categories for ethnicity appear to be a collection of identities either known to authorities or recognised by society. There were 46 categories in the first 1871 census for the Straits Settlements, 18 of which were sub-groups of 'European and Americans'. Subsequent censuses saw not only an increase in the number of ethnic categories but also a revision in the categories themselves reflecting the recognition of the sizes of groups in society as identified from the experiences of previous censuses. In contrast, Sarawak's census began in 1947 with 129 categories, reflecting the attempt – with the aid of Tom Harrison, Curator of the Sarawak Museum and Government Ethnologist - to document the many indigenous groups in its society, and then reduce the number when group size was ascertained.

The categorization of ethnic groups has also changed to accommodate changes in society. Categories have been updated as required, for example, with adjustments to new political entities or names being revised as necessary. Sometimes an ethnic group has appeared in one census but not in another in line with group size and policy needs. Policy has also created or modified ethnic categories. The somewhat loose constitutional definition of a *Bumiputera* has entered the social realm so that we now consider the '*Bumiputera'* community as an ethnic group. Politics has also influenced the categorization of the Kadazan-Dusun, an Other *Bumiputera* grouping, in Sabah. The seemingly easy shifts between 'Malays', 'Other *Bumiputera*' and 'Other Malaysians' reflect in part the commonalities in origin of a considerable part of the populace from the neighbouring regions that are politically different, that is, Indonesia, Philippines and Thailand. The movement of such peoples across the region in search of economic prosperity is not new, and continues to occur.

The granularity and identification of ethnic categories have thus changed and improved over time. The central data collection agencies have over the years been quite creative in defining and redefining ethnicity as Malaysian society and needs evolve. Each subsequent census has seen changes in line with size of group or its importance to public policy. A reading of the various census reports indicates experiences from censuses were shared historically between the Straits Settlements, North Borneo and

Sarawak, especially with regard to the measurement of ethnicity. The Malaysian experience with the population census reflects attempts to capture a conceptualization of an ethnic group as one that shares common interests such as language, religion and customs. Statisticians have demonstrated their ability in collecting census data from people of "many tongues", even against the odds of collecting data in the remotest parts of Sabah and Sarawak. This has made possible the fairly detailed ethnic classification now used in censuses, and which shows the great diversity across the regions in the country.

Chapter 10: Overall discussion of findings and recommendations

Background to the HGEC project

Many countries worldwide collect official demographic data to gain an accurate and reliable description of their population. The principal source of such data for most countries in this report is a census. Some countries use (or supplement census data with) alternative official sources such as population registers, health surveys, household surveys and health insurance registers.

As populations become increasingly heterogeneous worldwide, official demographic agencies are being challenged to devise valid data collection methods to classify people according to their ethnicity. Obtaining high-quality ethnicity data is important as it has a major impact on the formulation and monitoring of policies/programmes designed to identify and eliminate disparities (Swee-Hock and Kesavapany 2006). Therefore, practical principles for, and approaches to, the collection of granular ethnic/racial data are being sought.

To this end, the RWJF in the US developed an initiative to explore the disaggregation of ethnic/racial groups into more granular categorisations within health data. Alongside five literature reviews focused on the US, the RWJF invited the HGEC project to carry out an analysis of ethnicity data disaggregation across the EU and in four countries outside Europe, aiming to answer the following questions:

1. How do some major surveys collect, analyse, or report data for ethnic groups that go beyond the five categories that form the backbone of reporting in the US?

2. Are there any lessons that can be learned from outside the health field?

3. Why are disaggregated data not being collected, analysed, or reported more often if the field generally agrees that this is critical to understanding disparities?

4. Ideally, how should data be analysed and reported given the health outcomes that the Foundation is interested in?

Principal findings according to the RWJF questions

How do some major surveys collect, analyse, or report data for ethnic/racial groups that go beyond those five US aggregated categories?

Our findings demonstrate a diversity of ethnic group classifications which follow a complex pattern. Twenty-one of the thirty-five countries we explored employed granular ethnic classification, stipulated as extending beyond six categories, some also including the freedom of a write-in option. The rest of the countries collected proxy measures for ethnicity such as country of birth and/or nationality.

Within Europe, the UK, Hungary, Poland, Slovakia, Latvia, Romania and the Czech Republic, appear to have the most disaggregated ethnic group classification. Outside Europe, Malaysia, Canada, New Zealand and Bolivia also have disaggregated approaches to classification.

Throughout these classifications, the focus of disaggregation varies; for example, concentrating on heterogeneity within Asian- and White-origin populations in the UK, and within indigenous populations in Bolivia. From the selected countries, New Zealand appears to have the most opportunity for granularity, utilising eight ethnic group tick boxes, which allow for multiple (up to six) responses, and also a free text option. To analyse these data they have developed a four tier hierarchical output system, which goes from aggregated broad groupings to more granular codes and thereby can be analysed to an extremely detailed level of the classification (level four), that accommodates over 230 separate categories. This system can therefore be adapted for use at varying levels depending on the setting and purpose of data collection (see chapter 3).

We found that wide variation in ethnic classification internationally relates to each country having its own context, which shapes the collection and categorization of ethnic group data. The multi-faceted nature of the concept of ethnicity also means that the factors considered to constitute ethnicity differ across countries. We observed variations utilised throughout the EU, reflecting cultural, linguistic and geographical influences. For example, in some countries ethnic group categories are based mainly on nationality (e.g. Estonia, Lithuania) or ethnic affiliation (e.g. Hungary); in others they include a combination of nationality, ethnic minority group, language and religion (e.g. Poland); whilst in others we observed that they encompass ethno-religious groups (e.g. Cyprus). There are also countries that do not directly collect ethnicity data *per se*, but use proxy measures to infer ethnic background (e.g. Country of birth in Denmark, and The Netherlands). Proxy data may encompass a long list of variables including origin, country of birth, identity, language, nationality, religion and tribe (Costanzo 2016). These measures aim to convey an account of origins or ancestry (Hollinger 1998). From our selected countries outside Europe we observed a strong indigenous perspective (e.g. Bolivia, New Zealand, Canada) and, in the case of Malaysia, ethnic classification has been politically influenced (e.g. Bumiputera) and groups have sometimes been combined for political benefit (e.g. Kadazan-Dusun group (Nagaraj, Nai-Peng et al. 2015).

Complexity was also observed in the terminology used in the questions and how these were articulated. In countries selected for this project there is no mention of race but rather questions based on terms of ethnicity (e.g. ethnic affiliation, ethnic group, ethnic origin). This reflects a historical shift, outside of the US, since the 1970s, partly in response to increased migration, varying migration patterns (Allen and Macey 1990), and also a conscious move away from the biological concept of race (Bhopal 2013).

Likewise, the way in which the ethnicity related question is posed differs from country to country. For example, in the UK the question is phrased "*What is your ethnic group?"* In other countries a

combination of nationality and ethnicity was used in a single question, e.g. “What is your nationality understood as the national or ethnic affiliation?” in Poland. In other countries the question refers to a sense of belonging such as in Romania e.g. “What ethnic group does the person consider he/she belongs to?”; Hungary e.g. “Which nationality do you feel you belong to?’; New Zealand e.g. “Which ethnic group do you belong to?”; and also in Bolivia e.g. “As Bolivian do you belong to any nation or indigenous farming peoples or Afro-Bolivian origin?”.

The response options provided in these official sources also vary, and take three principal formats: an open-ended option (free text) (e.g Romania); a closed-ended response with a list of categories (tick-boxes) (e.g. Cyprus); or a closed-ended, multiple response option (e.g. New Zealand) also sometimes including a free text option (e.g. Poland, Canada). In some countries, questions relating to ethnicity were not compulsory to answer, as this is considered sensitive information (e.g Bulgaria, Croatia, Slovenia and Hungary), whereas in other countries it is compulsory (e.g. UK) and, on the whole, this information is safeguarded by well governed confidentiality and data protection laws.

Overall, we found multiple variations in the way in which ethnic classification is undertaken including differences in the underlying concept of ethnicity, the number of categories used, the way in which questions are phrased, the format of responses permitted, to what level responses are analysed, and whether the questions are compulsory. The practices appear to be contingent on contextual factors unique to each country, including the country’s social, political, economic, historical and geographical circumstances. These factors are therefore important to bear in mind when identifying practices and considering if these are generalizable to another context, such as within a US setting. We discuss these contextual considerations next.

Contextual factors influencing the development of granular ethnic categorization

This project provided us with insight into multidimensional factors which require consideration when creating granular ethnic group categories (see figure 2). We observed that ethnic group classifications are influenced socially taking into account the pivotal role played by political, historical and economic factors (Petersen 1969). We reflect on these contexts further drawing on our European overview, the country reports and the international investigators’ meeting.

Historical context

It is apparent that growth in international migration over time has provided impetus for ethnic group enumeration. For countries receiving migrants from diverse population groups, the need for classification has arisen and they have responded to these demographic changes in their population by developing different methods of categorization within their official sources (Castles 1995). Migration flows have influenced the development of particular categorisations, some of these patterns of migration being politically and geographically influenced and others are due to historical and economic links between countries. After the Second World War (WWII), migration in Europe was driven by the need

for labour, leading to an increase in the diversity of the population of most countries. Migrants initially comprised mainly of "guest workers" from surrounding countries, plus migrants from former colonies into the UK and France (Turton and González 2000). Over the ensuing years ethnic diversity has continue to grow, as result of increasing globalization, both with continued migration and the mixing of resident populations of different backgrounds within countries. Currently we are living in an era of *super diversity* (see chapter 7) for which more granular classifications are required in order to understand the structure of our complex populations.

The influence of particular historical events has also massively influenced methods of ethnic enumeration within countries, in particular history of conflict. From the countries studied in Europe, the majority collect country of birth and not ethnicity, as a historical legacy of the Nazi period and the abuses of such microdata during this era. Similarly, the dissolution of Yugoslavia in the 1990s was reflected in major changes in data collection where previously granular classifications were replaced with only a few categories as a result of the process of ethnic cleansing during this war (e.g. Croatia) (Hayden 1996).

Geographical

As mentioned before, migration flows have been influenced by a country's geographical location with principal migration in many cases coming from neighbouring countries. Formation of ethnic group categorizations therefore have had to encompass, and cater for, these complex patterns of migration. Some countries have also been strongly influenced by continual shifting and reshaping of their geographical boundaries, which has challenged categorisation and influenced population responses to ethnic classification. For example, in border areas of Hungary, where people are often bilingual and bicultural, they may have consciously ascribed themselves to a different ethnic category depending on where they were and where they wished to reside (Kocsis and Hodosi 2001). These geographical changes are also historical processes and are deeply rooted in political motivations, which shows that these contextual influences are not isolated but interrelated and overlapping.

Social /political/economic context

Social, political and economic dimensions have also influenced ethnic categorization. Some societies embrace their multicultural identity and are freely permissive of expression of diversity (e.g. UK, Canada); conversely, there are countries where a unified national identity and assimilation are promoted (e.g. Denmark). The social acceptability of collecting and utilising data on ethnicity vary accordingly. In the case of Denmark, whilst it has been traditionally considered a liberal and tolerant society, it has relied on the cultivation and preservation of a homogenous national identity giving rise to cultural racism and immigration being perceived a threat (Wren 2001). This is reflected in their population register, as the population registers collect country of birth and not ethnicity (see chapter 6).

Outside Europe, we observed that in countries such as Canada, Bolivia and Aotearoa New Zealand, although political influences are very strong, the action of specific communities within society have also been a key factor in mobilizing recognition of specific communities, for example indigenous rights movements (see chapters 3,4,5) (Kukutai and Didham 2012).

Some governments have responded with proposals to restore and support indigenous communities (Kukutai and Didham 2012). For instance, the introduction of the official ethnicity standard in 1993 in Aotearoa New Zealand (see chapter 3). These measures have influenced ethnic categorization through the collection of ethnic data in official sources and have helped address economic disadvantages for indigenous population groups (Callister 2007). Furthermore, in the Bolivian census, people were given the chance to self-identify as belonging to indigenous groups, which led to recognition in their constitution (see chapter 4) (Canessa 2007).

There have also been social movements which have inhibited, rather than promoted, granularity in ethnic classification. In Aotearoa New Zealand and Canada there has been promotion of national naming within ethnic classification as 'New Zealander' or 'Canadian'. These campaigns were led by the media sources prior to the census and did result in boosted responses from some sectors of the population using a national identifier (Kukutai and Didham 2012).

Lastly, religion is another societal factor which some scholars have suggested as a key factor in ethnic identity formation (Oppong 2013). For example, in Poland, religious affiliation and national identity are closely linked (e.g. "Polak-katolik" stereotype) (Zubrzycki 2001).

Are there any lessons that can be learned from outside the health field?

Although lessons garnered in this report are mostly from within the health field, there are some countries where examples were identified of granular data being collected outside this field. For example, in the UK, granular ethnic categories are found within the Department for Education's 'Extended Categories' list (see chapter 7). Such extended classifications have their greatest utility in allowing local jurisdictions to choose categories relevant to the composition of their populations in their ethnic monitoring initiatives. In Denmark, the main focus outside the health field is also in education, where immigrant students are identified based on linguistic criteria (see chapter 6).

Other countries collect granular ethnic data outside the health field to understand socio-economic factors (e.g. of the Gypsy/Roma people in Hungary) and to understand how people report their ethnicity (e.g. Canada). We also found examples of granular categories being used to guarantee the rights of indigenous people, for example through the use of indigenous language at schools (e.g. Bolivia) and for the purposes of policy formulation and evaluation (e.g. Malaysia). In Aotearoa New Zealand, granular data is also collected as part of routine survey programmes, and in administrative collections across the social sector including schools, the police and courts (see chapter 3).

Although we have been able to identify examples from outwith the health field (e.g. education, and policy), there is again great diversity in the classifications used and no particularly exemplary practice found which unquestionably warrants adoption within the health arena. One finding from both within and outside the health field, however, is that the greater the granularity within the classification, the more adaptable it is to different settings and to being analysed for differing purposes (see chapter 3).

Why are disaggregated data not being collected, analysed, or reported more often if the field generally agrees that this is critical to understanding disparities?

We found a broad range of explanations as to why granular data is not being collected, analysed and reported at a more granular level. These included organisational factors; for example, the logistics and cost of designing and implementing new categories (e.g. UK, Denmark) (see chapter 6, 7). There were also methodological reasons; a lack of advocacy for greater granularity; fear of stigma for particular ethnic minority groups (e.g. Hungary); political reasons; geographical barriers for the actual collection of data from isolated populations (e.g. Malaysia); administrative barriers; potential for harm (e.g. Canada); and in countries where data is actually collected in a granular manor, the tendency to still aggregate data at the point of analysis (e.g. Aotearoa New Zealand)

Ideally, how should data be analysed and reported given the health outcomes that the Foundation is interested in?

Central to the Robert Wood Johnson Foundation's (RWJF) vision of building a *culture of health* are the outcomes of improved population health, well-being, and equity. RWJF is committed to helping everyone in the US to have an equal opportunity to pursue a healthier life (RWJF website). Obtaining and utilising high-quality ethnicity data is pivotal to such a vision, to identify and address inequities in health which exist between population groups and have historically been masked by the use of large aggregate ethnic group categories.

However, it is problematic to stipulate an ideal way that data should be collected, analysed and reported, as we found that developing appropriate classifications is contingent on recognising and understanding the contextual factors of each country. We can identify good practices such as the, hierarchical four-tier classification system of analysis in Aotearoa New Zealand, but most likely these practices will need to be adapted in another context. In place of recommending specific practices, a set of considerations and principles for developing classifications for use within the health field have been highlighted during this work and this can be found in the overall conclusions.

Strengths and weakness

Strengths of this work are that our findings are from official demographic sources and are relevant to policies and strategies within multi-ethnic societies. This project was strengthened by its collaborative approach in working with seven international experts and their teams to provide in depth information from their country, who had access to the most pertinent literature and also to other instrumental people

in this field in their country. The project also endeavoured to dig deeper than the RWJF main questions in order to explore the contexts of each country and help to understand important factors when planning the development of classification systems. This is the first study to our knowledge that attempts to contextualize the development and use of granular ethnic data in Europe and in selected countries outside Europe.

A weakness is that the European overview was rarely able to examine data sources beyond population registers and censuses. This was due to the large number of countries involved and the difficulties in reliably accessing these additional sources. It may be that there are examples of more granular uses of data in literature that we were not able to obtain. We acknowledge that we also only examined seven selected countries internationally as examples. These were chosen after considering which examples may have the most lessons to offer in terms of approaches to granular classifications; however, there may be other countries that it would be useful to also take into consideration in generating principles for consideration in the US setting.

Overall conclusion

Aggregating data under large ethnic group denominations could mask diversity. There is a need to understand country's contextual factors to reconceptualise the categories of ethnicity, nationality, CoB and other indicators and thereby maximise the granularity of approaches internationally. In an attempt to inform this process we have developed, from this work, a set of main principles and recommendations. The adoption of such principles may assist in the development of country-specific ethnic systems of classification and also could support comparisons of data being made over time and across countries.

Main principles and recommendations

1. Ethnicity is predominantly a social construct, there is no global consensus about the concept and definition of ethnicity, and it varies across contexts – it should be made explicit what underlying concepts are being applied when developing a classification.

2. Operationalising ethnicity as a concept to classify population groups needs to take into account country ***contexts*** which greatly influence the feasibility of implementing granular ethnic enumeration.
3. Clarifying the 'objective' of collecting ethnicity data is paramount to determining what should be collected. It should be made explicit what the data are to be used for, how granularity can operate in classifications given the outcomes in mind and what the societal objectives are (e.g. health equity).
4. We need to consider how ethnicity data are ***collected*** in relation to the outcomes in mind. (In the health field, for example, self-assigned ethnicity is the standard as it correlates with health

behaviours and risks. However, there are instances where socially assigned ethnicity has been demonstrated to correlate with quality of healthcare received and inequitable outcomes – (see appendix 10)

5. The number and order of ethnic group categories needs to be considered to determine the usability and equity of a classification. (e.g. using a long list of categories might be confusing and particular ordering of groups may favour dominant groups).

6. Allowing free text responses as well as multiple response may assist in achieving greater granularity and accommodating the increasing population of people identifying themselves as *mixed-origin* (although this may present challenges in data processing and analysis).

7. Developing a flexible hierarchical categorisation which can be expanded or collapsed and enables a very high degree of granularity when appropriate may be the most advantageous approach to pursue for analysing data from free text responses (as seen in New Zealand)
8. Developing protocols to guide the analysis of granular ethnic group may assist in appropriately utilising available data and motivating the collection of granular data
9. Processes to enable community involvement in the development of ethnic group categories and in decision-making on analysis of and inclusion in statistical surveys should be considered.

References

Allen, S. and M. Macey (1990). "Race and ethnicity in the European context." The British journal of sociology **41**(3): 375-393.

Bhopal, R. S. (2013). Migration, ethnicity, race, and health in multicultural societies, OUP Oxford.

Callister, P. (2007). "Special measures to reduce ethnic disadvantage in New Zealand." Wellington: Institute of Policy Studies, Victoria University.

Canessa, A. (2007). "Who is indigenous? Self-identification, indigeneity, and claims to justice in contemporary Bolivia." Urban Anthropology and Studies of Cultural Systems and World Economic Development: 195-237.

Castles, S. (1995). "How nation-states respond to immigration and ethnic diversity." Journal of Ethnic and Migration Studies **21**(3): 293-308.

Costanzo, J. M. (2016). "Census Data and Ethnicity." The Wiley Blackwell Encyclopedia of Race, Ethnicity, and Nationalism.

Hayden, R. M. (1996). "imagined communities and real victims: self-determination and ethnic cleansing in Yugoslavia." American Ethnologist **23**(4): 783-801.

Hollinger, D. A. (1998). "National culture and communities of descent." Reviews in American History **26**(1): 312-328.

Kocsis, K. and E. K. Hodosi (2001). Ethnic geography of the Hungarian minorities in the Carpathian Basin, Simon Publications LLC.

Kukutai, T. and R. Didham (2012). "Re-making the majority? Ethnic New Zealanders in the 2006 census." Ethnic and Racial Studies **35**(8): 1427-1446.

Nagaraj, S., T. Nai-Peng, N. Chiu-Wan, L. Kiong-Hock and J. Pala (2015). Counting Ethnicity in Malaysia: The Complexity of Measuring Diversity. Social Statistics and Ethnic Diversity, Springer**:** 143-173.

Oppong, S. H. (2013). "Religion and Identity." American International Journal of Contemporary Research **3**(6): 10-16.

Petersen, W. (1969). "The classification of subnations in Hawaii: An essay in the sociology of knowledge." American Sociological Review: 863-877.

Swee-Hock, S. and K. Kesavapany (2006). Malaysia: recent trends and challenges, Institute of Southeast Asian Studies.

Turton, D. and J. González (2000). Ethnic diversity in Europe: challenges to the nation state, Universidad de Deusto.

Wren, K. (2001). "Cultural racism: something rotten in the state of Denmark?" Social & Cultural Geography **2**(2): 141-162.

Zubrzycki, G. (2001). "“We, the Polish Nation”: Ethnic and civic visions of nationhood in Post-Communist constitutional debates." Theory and Society **30**(5): 629-668.

Appendix 10

Notes for the HGEC Meeting (main points)

Date: 2nd - 3rd May, 2016 - CPHS, Teviot Place, Edinburgh.

Chair: Prof. Raj Bhopal

Present:

Raj Bhopal (RB)

Emma Davidson (ED)

Peter Aspinall (PA)

Liv Stubbe (LS)

Allan Krasnik (AK) (videoconference)

Inez Koller (IK)

Hude Quan (HQ)

Donna Cormack (DC)

Pamela Pereyra-Zamora (PP)

Shyamala Nagaraj (SN) (videoconference)

Chiu Wan (CW) (videoconference)

Tina Kauh (TK) (teleconference)

Nazmy Villarroel (NV)

Notes May 2nd, 2016:

1. RB welcomed the HGEC group and thanked everyone for attending and participating in meeting. All members of the group introduced themselves and briefly outlined their background

2. RB went through the agenda and explained how and why the Robert Wood Johnson Foundation (RWJF) funded this project. RB referred to his book on Migration, ethnicity, race and health in multicultural societies and the diverse approaches internationally to defining ethnicity and developing ethnic classifications. He mentioned the rationale behind exploring and developing more granular classifications.

3. RB started the meeting. The first presentation was the EU overview by NV, which provided information about ethnicity data collection across the EU-28 countries. The questions posed by this presentation were:
   - Why some countries use a combination of the different aspects of ethnicity? Is this a helpful way to identify ethnic groups?
   - Are standard classifications within official sources in EU countries meeting user's needs? Especially when taking into account current social processes (international migration, super-diversity- hybridization among others)
   - How standard classifications could be comparable across EU countries? Giving the different ethnic categorization in EU countries.
   - If ethnic group categorization might be influenced by certain ideologies such us ethnocentrism or historical conflicts. Is it possible to move towards a more granular classification?

   After the presentation there was discussion about whether it is possible to classify ethnicity when there is not yet global consensus about the concept and definition of ethnicity, and also because it is a social concept and varies across contexts. It was proposed that clarifying the objective of collecting ethnicity data was also paramount to determine what should be collected. A recommendation was put forward that when developing classifications it should be make explicit what the data is to be used for, what the societal objectives are and, therefore, what underlying concepts are being applied. PA posed some questions regarding harmonising operational criteria (ethnicity) across EU countries and how this was extremely difficult – it could be possible to agree on terminology to be used, but that it is probably not possible to compare groups across different settings e.g. what it is to be of Nigerian–origin in the UK is different to being Nigerian in another country in Europe where there is not a colonial history. The group concluded that it may be possible to agree on a set of principles and concepts internationally, but not one classification.

Ethnicity is a social concept and the relationship of ethnicity to health is only meaningful within its social setting or context and varies across contexts. Perhaps we could agree on what the concept of ethnicity means globally, what question wording is appropriate, what categories are acceptable to use?

4. RB then introduced the country reports (CR) from Denmark (presented by LS), the UK (presented by PA) and Hungary (presented by IK). LS outlined the focus in Denmark of origin based on country of birth and ancestry, and described the distinction between immigrants and descendants, and between people from 'Western' and 'non-Western' countries; she also discussed the difference in rights between minority ethnic groups and indigenous groups e.g. Inuit groups living in Denmark versus Greenland. The questions raised were:
    - Is it possible to develop more fine grained ethnic categorisation systems that can be comparable across countries?
    - Is it meaningful to include 2nd and 3rd generation in data on health. Proxy for ethnicity, but is it actually useful?

    There was discussion of the use of the term immigrant and how for 2nd and 3rd generation populations they are not immigrant but ethnic minority populations (less stigmatising). The division between Western and non-Western countries was also considered and how this relates to people's experiences of migration. How public opinion contributes to categorisation was also raised (in Denmark, their definitions are said to be close to 'prevailing opinion) and both the pros and cons of this process, including the effects of lobbying - as DC explained regarding the inclusion of a 'New Zealander' group in the NZ census. PA contrasted the US and UK situation, then described sources of granular data in the UK and the rationale for increased granularity of categories, mentioning the African diaspora population in Britain as an example. He posed the questions:

    - Which are most useful – comprehensive ethnicity code sets of US-type or more modest extended classifications?
    - What is the appropriate level of resolution for granularity? E.g. 1 - level 1: Black African, level 2: Nigerian (country origins); level 3: Igbo, Yoruba, Hausa ('local' ethnic); E.g. 2 – level 1: Asian; level 2: Pakistani; level 3 – Mirpuri, Kashmiri
    - Should these granular classifications be voluntary & selected (pick-lists) or mandatory national standards
    - How should cross-mapping be achieved: probabilistic assignment or by direct question on census categories?

    The group discussed these levels and the methods of mapping and how one level 3 category could map to several different higher categories, which then loses 'self-determination' by the process

of assignment. The order in which categories are arranged my also influence peoples choices as they may not read all the way to the bottom of a long classification and approaches vary to how the list is arranged e.g. with some countries ordering the categories alphabetically, and some by population size. Increasing granularity risks that people may be less likely to consider the whole list if it is longer. Perhaps there should be two questions? - One higher level and one more granular was suggested. The question was posed as to whether there is a bias towards the White population in the UK question, and also why colour based classifications are acceptable in the UK and US and not elsewhere e.g. NZ and Canada. The census development programme in the UK was outlined by PA and the influence of lobbying and community groups. The overarching experience is, although there is a feeling that ethnicity is a sensitive question, that people don't mind being asked and in the UK more people were lobbying for inclusion of their category than could be represented.

IK outlined the situation in Hungary and the data sources. She mentioned the Roma people in Hungary and the various Roma groups that have different history, language and living standards; and also mentioned the mix groups such as the Serbian catholic and the Croatian Orthodox. She posed the questions:

- Do you think ethnic granularity system needs a definitional ground? E.g. criteria for accepting groups for inclusion in an ethnic category
- Which is better a compulsory or voluntary system of collecting ethnicity data? – considering the usefulness of the data versus the potential for discrimination

It was felt that compulsory completion provides the most useful data, but in some countries e.g. Denmark legislation means that this is not possible. PA suggested that it would be interesting to carry out a survey to see how many people you would lose data for if the question was voluntary – create an evidence base to help inform policy. However, RB mentioned that there are also issues that, even when compulsory, people don't tell the truth e.g. write Jedi! In the UK it is compulsory to complete the census. However the culture in the UK is adapting so that the question is becoming more wide-spread and more widely acceptable.

5. Following lunch, two further CR were presented for Bolivia (presented by PP) and for Malaysia (presented by SN and CW by videoconference). PP outlined the complex distribution of ethnic groups within Bolivia and the focus of ethnicity data collection on the many indigenous groups, with a very high level of granularity. She explained the importance of the recognition of thesegroups to their political, economic and social rights; for example, determining the number of seats in parliament, the right to speak their own language, to manage their land, to apply their own justice. PP posed the questions:
    - How to formulate an ethnic category that captures the non-native population? What

lessons can we learn from the other countries?

- How to use the census information to design, monitor and evaluate public policies when there is a lack of other sources of administrative data? How to deal with the evidence gaps?
- Is it necessary to categorise groups that were seen as mestizo in colonial times?

The point was made that collecting data for all groups, rather than focusing on one group, helps people to understand why data is being collected. Sometimes we don't collect for those who are doing well/privileged but it is necessary to measure everyone to assess equity. Collecting data is strongly influenced by politics globally as to which groups are recognised and why. The group discussed cross tabulation as a means of understanding the diversity of populations and PA highlighted the UK websites where cross tabulation can be done on line and tables also requested (requested tables are then made available online). The idea of keeping indigenous group categories in the classification, as opposed to having this as a separate question, if the classification broadens out was put forward. The New Zealand census was given as a good example where you can identify as being Maori, but there are also separate questions about ancestry and iwi affiliation which relate to political and land rights etc.

SN and CW presented the Malaysian report, focusing firstly on the collection of data and then the relationship of ethnicity and health. There has been a long history of collecting ethnicity data and the question is widely acceptable. The aim is that 'classification should be as scientifically accurate as possible, the groups must be reasonably balanced in size, and it should be in sufficient detail to provide a sound basis for future scientific investigations'. Definitions have changed and influences have included policy, changes in society over time, and politics e.g. Particularly the Bumiputera group. In terms of health policy, ethnicity has importance as a social determinant of disease and is used to examine equity. RB asked more about the Bumiputera group. SN talked about the historical development of this category in Malaysia as an economic and political group. SN explained that being part of this group is constitutional and the identification of the ethnic groups is based on the National Registration Identity Card issued by the National Registration Department; the Bumiputera "label" is a political construct and is a growing category. In the census, people tick their ethnic census category and when the census data is compiled they will be classified as Bumiputera, Non-Bumiputera and Non-citizens. HQ, asked about the Ministry of health focus in Malaysia. CW responded that in the healthcare system there is the principle of equity and healthcare distribution for all. However, there are equity issues which are related to the migrant/non-citizen population. Legal migrants may access healthcare but at higher costs, which may not be affordable, and non-legal migrants would only receive lifesaving treatment and are then deported.

6. RB then introduced TK by teleconference to present the RWJF perspective and the US project. Tina spoke about the RWJF model of a Culture of Health and the framework supporting this vision. She talked about hidden disparities within aggregate ethnic classifications and the rationale behind seeking greater granularity. She then outlined the US project methods and process. RB, asked about the timelines of the US project and suggested to TK to contact Professor Jay Kaufman for the White population review for which they were still looking for a lead investigator. The US project aims to have the reviews completed by the end of the summer 2016. RB told TK that we

   aim to have a solid draft of our report ready within the next 2 months and the finished version by September 2016. HQ asked TK the action there will be following this project. TK mentioned that this is still an exploratory area. The goal is to get experts together for a series of convenings at the end of the project and to develop recommendations; the RWJF will then decide how to take these further. RB mentioned to TK that, as PA explained, there is no better potential for examining statistics by a granular approach than in the US. However, the data that's available needs to be accessed and appropriately used. TK mentioned that the lack of granularity in data is at the core of many issues that the RWJF is trying to address. HQ asked TK about immigration status (illegal and legal migrants). TK mentioned that this is part of the individual review processes about variables related to race/ethnic status that are also strong predictors of health outcomes. However, there is a high level of sensitivity around asking immigration status in the US. TK mentioned that they want to ask questions in a way that feels safe for the respondents. In the convenings they will make sure that social justice is represented at the table for this project. TK asked RB about the historical context of the ethnicity classification in other countries. RB mentioned that the US actually was one of the first countries to take the lead on granularity. He also commented on ethnicity data in the UK where the White group contains at least 4 categories, and mentioned the case of Bolivia where there are more than 30 indigenous groups collected in the census. He described our discussions around the possibility of harmonising principles and theory, although perhaps not categories themselves. RB mentioned to TK the 2018 World Congress on Culture, Ethnicity and Health and the hope that the RWJF would be involved with this event.

Notes May 3rd, 2016:

1 RB went through the agenda for the second day and introduced presentations from New Zealand (presented by DC) and Canada (presented by HQ).

2 DC presented the report on Aotearoa New Zealand and the changing approaches to ethnic enumeration. The Statistics New Zealand definition of ethnicity and the characteristics of an ethnic group were outlined. Key sources of official ethnicity data and data for health were examined and the granularity of classifications described – including their hierarchical four-tier classification system and the possibility to select multiple (up to 6) ethnic groups. DC posed the following

questions:

- How does granular data support a deeper understanding of ethnic health priorities and inequities?
- What are the ethnicity data that we need to better understand how ethnicity is associated with experiences and outcomes in health?
    - What are the processes that make ethnicity meaningful in relation to a particular outcome or experience (Garner 2010)?
    - When is self-identified "expressed" (Roth 2010) ethnicity the most appropriate measure? When is socially-assigned ("reflected" or "observed") more relevant?

RB commented that NZ appear to be leading the field both in terms of being the earliest to collect ethnicity data on death certification and health data sets, and in terms of their granular approach. HQ agreed that the levels of classification is a very good model and commented that NZ has a very comprehensible approach to granular classification. HQ asked if the all the health service collected ethnicity. DC stated that the minimum required for the health sector is level II which includes 22 categories. RB asked if the question was mandatory or a recommendation. DC explained that there is not a penalty but it is under the control of the Ministry of Health and part of the primary care contract, so it is in the interest of services to collect ethnicity data as it impacts on their funding formula. DC also informed the group that the census is moving to 70% of the collection being online. Their hierarchical classification system goes from aggregated broad groupings to more granular codes, with level IV containing up to 230 categories, but most data is still reported at level I. Statistics NZ is proposing a more dynamic classification. They are also developing a big integrated data infrastructure including data from health, accident and injury, education among others. All this data is reported at a National level. DC mentioned that having granular data will help us understand health priories and inequities. Also, it is important to try to understand socially-assigned ethnicity because, although self-determined ethnicity relates more to health behaviours and risks, socially-assigned ethnicity correlates more with the quality of healthcare received and this can translate into disparities in health outcomes. IK commented that communities' identification of people's ethnicity is also important to consider. DC referred to research in NZ examining the relationship of self-assigned and socially-assigned ethnicity, and similar research in the US by Camara Jones. It was emphasised that this relates again to the importance of knowing what the data will be used for and that can determine what question is asked or what concept of ethnicity is sought. The group discussed the sensitivity of asking ethnicity questions and DC raised the importance of having a 'refused' category in healthcare services so that those who do not want to respond do not get repeatedly asked the question upon healthcare contact.

HQ presented the final CR from Canada. He outlined the history and background to collection and the concept of 'visible minorities' which is unique to Canada. This population is rapidly growing in Canada. He outlined the changing definitions over time and the granularity of the categories available, and then examined the relationship of ethnicity and health. HQ put forward the questions that:

- Ethnicity is complex and hard to measure, and thus may not be useful
- What do Canadians perceive as ethnicity? - What is "Canadian"?

The influence of politics and lobbying was again discussed, including the influence of senior statisticians. RB asked if Canada operates the same way in every province. HQ pointed out that in the health sector ethnicity collection is fairly similar. HQ mentioned that in the hospital service facilities they capture ethnic groups but when data is coded it is recorded just as aboriginal or non-aboriginal. HQ also mentioned two methodologies for identifying ethnicity - linking immigration data (where are you from), but this will not capture second generation populations, and also identification by surname. PA raised a question of data quality. RB mentioned that the problem was often related to incompleteness and the numerator (outcome, health status) and denominator (census or population register) mismatch/bias. DC informed the group that New Zealand have done a lot of work on quality and have developed both adjusters to apply to data sets and also audit tools to examine quality – she can share this work with the group. PA also mentioned that in England there are challenges coding and analysing mixed ethnic groups.

3 A general discussion was undertaken and RB commenced by asking the group about what we have learned from the presentations. LS asked what are we trying to capture with this kind of data at a national level and what are the different objectives, purposes or goals between countries of the data collection. HQ added that it's important to specify which data bases we're referring to/using (census, civil registers systems) as they are designed for different purposes. AK suggested it would be very interesting to know to what extend the health dimension is including in the agenda (e. g. in the census activities) as in many countries this is not a reason why the information is being collected. LS suggested that it would be helpful to share the presentations from our colleagues to be able to refer to the information from different settings. SN commented on the purpose of collecting data and the categories and reminded the group that the number of observations or the number of persons in some groups will determine what categories will make sense to be collected. So, some groups may be possible at a census or national level, but at a health survey level might not be possible (e.g. reaching indigenous populations) as the numbers are too small to

analyse. PP agreed that use of granular data is important, but may be dictated sometimes by the statistical methodology (small numbers). However, the smaller (absolute) numbers may be useful for social purposes too e.g. planning services. RB commented on the combination of groups for statistical purposes and how this should be done depending on the health outcome of your study (e.g. tobacco vs. cholesterol) – in some cases this is appropriate, but in others the health practices vary too much to combine groups. AK suggested that the main message in this regard is when the objective for collecting the data is not clear and the census and register are used in many different purposes, then it is best to collect with as much granularity as possible and then those using the data will have to judge whether it should be combined or not for the purposes they are using it for. This returned to the principle discussed earlier of the importance of knowing what the data will be used for and ED reminded the group of the RWJF question which was: how should data be analysed giving the health outcomes that the RWJF is interested in. So our primary focus should be on considering how granularity can operate in classifications with their outcomes in mind. HQ asked about the format and tables of the CR. ED and DC suggested having a general table which compares the basic principles across our selected countries. RB suggested having a chapter summarizing the main points of the CR together and the main lessons arising. PA commented on how difficult it is to implement granularity and contrasted the different needs and processes at local and national levels. RB also commented on the potentially huge administrative costs - so there is a cost to increasing granularity and heterogeneity.

4 RB reviewed the timelines for the final CR and when it will be possible to have a revised version of all the CR. HQ commented on the need of a guidance for this. RB suggested to set up a system of informal peer reviewing the CR; for example, two people giving feedback on each CR. For this to happen it will be appropriate to send the CR to the peer reviewers before the end of May and then at the end of June give the feedback to the CR. SN suggested to create a list of questions for the peer reviewer in a way that is useful for the final report. RB remarked that the CR writers will finish their job at the end of June and then read the full report by the end of September. The main conclusions and recommendations were discussed. SN highlighted that ethnicity is a social construct, and RB remarked that the more granular you become the more social the construct becomes. There are also issues of disclosure with increasing granularity. AK raised issues of the data collection, who are responsible for these procedures and providing the data and access, and also the cost of data collection procedures. These issues should also be mentioned in the discussion. PP mentioned that the census is the gold standard and that a question should be developed for the census which could then be used in health surveys etc. There was then discussion around the census coverage, particularly for undocumented migrants and homeless people. RB commented that in the UK the census is based on households so it includes undocumented migrants. PA pointed out that in the UK there are special enumerators whose job is to find people who are difficult to reach (homeless, nomadic populations). SN mentioned that

the Malaysian statistical office will similarly capture the illegal workers who have a place of residence, although people may be hesitant to complete it. RB noted that the census is the most protected of all data sources in any country. DC remarked that in New Zealand there is the development of more integrated data infrastructure (including justice and welfare data sets) and whether the perceived risk that data could be used to profile individuals or targeting some groups (e.g. immigrants or indigenous people) may adversely affect people completing ethnicity data items, particularly with a history in New Zealand and Australia of the forced removal of children. There should be risk modelling about the use of data and Tahu Kukutai is currently involved in consultations of a data sovereignty group in New Zealand, addressing issues as to who owns data and how it should and shouldn't be used. AK reflected on undocumented migrants in the data and how this varies across countries. In some EU countries they can register in some municipalities and make use of health services for example, whereas in others this is not possible. PA mentioned a group called "No recourse to Public Funds Destitution Network" administrated by local authorities. They have data on ethnic groups and reasons for destitution. RB proposed that for our full report we should try and develop a set of principles and that the focus should be on what data is needed to improve population health and particularly to improve equality/equity. This could include the needs for consistency of classification, quality of data, different levels of data, stipulating the use of the data (e.g. promote equity in health) and the more granular the data the more it can help groups with most need, where disparities are often otherwise hidden. It was also suggested that we could perhaps produce a schematic for the principles.

5 The structure of the report was discussed and LS suggested to a short introduction for the CR and why these countries were chosen and the methods for this. It would also be good to include the methods, including of the peer review of CR. PA suggested that we should have a general conclusion chapter and a recommendation chapter, and that there would be the potential to influence the next round of census consultation in the US. IK asked about the potential for publications. ED commented on the US timelines of the RWJF and that the RWJF may be keen that publications are not published before we submit our report to them. RB reiterated that the full report will go to the RWJF first at the end of September and after that we are free to produce our work as we wish. The CR belongs to the CR team and they can involve whomever they want or not from the wider team, and just have to make the appropriated acknowledgements of other people. RB suggested to have at least one publication with all the group which may outline the general principles. AK mentioned that there is a lot of cross cutting issues and this might be a good idea for one paper which compares across the CR – there should be a lot of interest in this. Everyone will give some thought about potential ideas for publications. PA suggested that there are dedicated journals on special issues and it was proposed that there may be the option to write pieces for a mini symposium in the Public Health journal. SN supported PA idea on submitting

articles to a special issue. RB may explore this with the editor informally. The revision of CR was mentioned and the format (particularly the number of words) and RB suggested that it's possible to have another page if people need the space and then any additional information can be added to an appendix at the end. We will aim to publish papers after we have submitted the full report to the funder (end of September).

Actions:

- NV to contact the Investigators to share their PowerPoint presentations
- NV/ED will create a list of questions to peer review the CR
- NV will circulate the list of questions close to the end of May to the investigators group
- The HGEC team to think on potential topics for publication and potential journals