Population data

Barbara Edgar, Cathy Day, and Steven McEachern

July 2019

To a large extent, population studies are driven by the availability of data, often used by population specialists to look at differentials or change. Differentials refer to variations among different subgroups of the population. For example, mortality differentials between high and low-income subgroups. Change can include change over time and change with age (Plewis 1985:4) such as human growth from birth to adulthood. Data are used to understand human behaviour and to test theories.

Data are needed to see if problems exist, to measure their extent, and to test whether interventions can overcome the problem. The subfield of *demographics* or *applied demography,* described in Chapter One, is the application of demographic science to practical problems (Weeks 2012:538).

Demographers can use data that is already available (secondary data), or they can collect their own (primary data). The main sources of population data are:

1.   Population censuses
2.   Sample surveys
3.   Registration systems:
    - Vital registration of *vital events* such as births, deaths, marriages
    - Population registers

4.   Administrative statistics such as:
    - Service statistics (from education and health services or family planning programs)
    - Border statistics
    - Special registers (e.g. notification of diseases)

6.   Other quantitative data sources
7.   Qualitative data.

Methods of data collection may be combined. For example, a census may include a *built-in sample* where additional questions are asked of selected populations or from a sample of households. The quality of a census may be checked by a post-enumeration survey (Economic Commission for Africa 1974:13). A census of a small community may precede the collection of qualitative data (Randall 1988).

The characteristics, uses, and limitations of each data source are discussed in the sections that follow. Each type of data has its strengths and weaknesses, which will be discussed in turn.

POPULATION CENSUSES

The United Nations (1998:33) has described a population census as:

> The total process of collecting, compiling, evaluating, analysing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country.

Censuses provide aggregate or macro-data and have been used for centuries to describe populations at the national, sub-national and small area level. Demographers prefer large numbers as they often subdivide the population into age-sex and other categories. Censuses have sufficiently large numbers for analysis of relatively small sub-groups, such as the Myanmar-born in Australia, which may not be adequately represented in a survey. When census data are used with economic data (such as Gross Domestic Product) as the numerator, important indicators of economic growth (such as GDP per capita), can be calculated and macro-level change can be measured. Timely and reliable census data are important for planners in economics, education, health, and other fields.

Early census counts or *enumerations* were often related to taxation and the word 'census' comes from the Latin *censere* which means to value or tax. Other reasons for census-taking were military service and food supply. The ancient Greeks counted the adult males in times of war, and the general population when food was in short supply. The first national censuses of Britain, which commenced in 1801, were partly intended to provide an estimate of how many soldiers could be provided in the event of a war, and partly intended to determine how many people needed to be fed (Thorvaldsen 2017).

A significant advance in census-taking was made in the United States in the 1797 Constitution which required the government to hold a census every ten years. This feature of a modern census - that it is taken at regular intervals, usually every ten or five years - makes comparisons over time much easier. Other important features of a modern census are that it covers everybody *(universality)* and that it refers to a well-defined moment or period of time *(simultaneity).*

In the past some countries practiced *census by assembly* (gathering together all the residents of an area and then counting them) or only collected group data from households. This meant that the data analysis was simple but limited. *Individual enumeration* is another feature of a modern census (Shyrock and Taeuber 1993:3-17). With individual enumeration, each house is contacted and information is recorded for each individual separately.

Most European countries began to hold modern-type censuses in the 19th century. In Asia, a census covering the Indian sub-continent and Ceylon was completed in 1872, followed by another in 1881. Since then India has had regular censuses every ten years (Davis 1951:4).

Many non-European countries only began census-taking in the 20th century. Between 1955 and 1964 an estimated 68% of the world's population was covered by censuses. The coverage was almost complete in Europe (97%), but was 62% in Africa and 53% in Asia (Shryock and Siegel 1973:17). Between 1975 and 1984, 191 of 213 countries carried out census enumerations, covering 95% of the world's population (Laroche 1993). Countries which held their first census in the 1980s included Ethiopia (1984) and Laos (1985).

However before the technological advances of the 20th century, data analysis was very tedious and limited in scope. In the case of India, little use was made of the data (Bose 1973:18). By the 1960s, in countries such as Ghana, the government was keen for researchers to illuminate the findings of the 1960 census (Caldwell 1969:1).

Census topics

The basic population topics recommended by the United Nations for the 1980 round of censuses are shown in Table 2.1. A number of housing topics were also recommended. All of the 18 selected Economic and Social Commission for Asia and the Pacific (ESCAP) countries covered in Table 2.1 asked about Relationship to Head of Household (or other reference member), Age, Sex, Marital Status, Literacy, Activity Status, Occupation and Industry.

TABLE 2.1 Population topics recommended for 1980-round censuses and the number of ESCAP countries using them

	Number of countries
Geographic and migration characteristics	
Place found at time of census	12
Usual residence	5
Place of birth	15
Duration of residence	15
Previous residence	8
Household	
Relationship to head of household or other reference member	18
Demographic and social characteristics	
Age	18
Sex	18
Marital status	18
Citizenship	10

Religion	13
Language spoken	9
Fertility and mortality	
Children born alive	16
Children living	13
Education characteristics	
Educational attainment	17
Literacy	18
School attendance	13
Economic characteristics	
Activity status	18
Occupation	18
Industry	18
Status in employment	14

Source: Cho and Hearn 1984: xvii-xix.

The United Nations (2008) has expanded the recommended core topics for population censuses to include, from 2010:

- International migration characteristics: Citizenship; Year or period of arrival
- Fertility and mortality: Date of birth of last child born alive; Household deaths in the past 12 months
- Disability characteristics: Disability status.

The Australian Census

The first population counts of Australia, conducted as early as 1788 and known as musters, involved all members of the community assembling at specified locations to be counted. Census-taking in Australia began in New South Wales in 1828, with the individual colonies holding their own censuses until 1886. The first census to produce population figures for all the Australian colonies on the same day was held in 1881, as part of the first simultaneous census of the British Empire. After Federation, the 1905 *Census and Statistics Act* provided that Australia would take a census in 1911, and then every tenth year. From 1961 the census has been held every five years (Australian Bureau of Statistics 2011:3-4).

The Australian Bureau of Statistics describes the purpose of the census:

> The Census provides the characteristics of the population and its housing to support the planning, administration and policy development activities of governments, businesses, communities, researchers and other users. While some of this information is available from other sources, only a Census can provide the information for the country as a whole and for small geographic areas and small population groups (Australian Bureau of Statistics 2016a).

Because Australian censuses are held every five years, users had faced the problem that their analysis of one census was still in progress when the next census was held. This situation has now changed as data are released much more quickly (e.g. the first releases of the 2016 data were in 2017) and powerful online software such as Tablebuilder has been developed. See:http://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder.

Table 2.2 shows the questions asked in the 2016 Australian Census. Unlike some of the countries in Table 2.1 where the oldest male in the household is assumed to be the head of the household, Australia uses a reference person who may be any male or female adult. Australia has a large percentage of its population born overseas, hence the importance of questions 11 to 18. The Aboriginal and Torres Strait Islander population has different demographic characteristics to the rest of the population, so Question 7 enables this group to be looked at separately. Australia included only one fertility and mortality topic (Question 32), relying instead on its comprehensive birth and death registration systems. A list of all topics included in every census from 1911 to 2011 has been produced by the Australian Bureau of Statistics (2011: Appendix 4).

TABLE 2.2 2016 Australian Census of Housing and Population Household Form

1. Address
2. Name of each person
3. Male or female?
4. Age?
5. Relationship to Person 1/2?
6. Marital status?
7. Aboriginal or Torres Strait Islander descent?
8. Usual residence?
9. Usual residence 1 year ago?
10. Usual residence 5 years ago?
11. Australian citizen (Yes/No)
12. Country of birth?
13. Year of arrival in Australia?
14. Fathers country of birth?
15. Mothers country of birth?
16. Language spoken at home
17. How well is English spoken?
18. Ancestry?
19. Religion (optional)
20. Needs help with self-care activities?

21. Needs help with body movement activities?
22. Needs help with communication activities?
23. Reasons for needing help with activities
24. Attending school or other educational institution?
25. Type of educational institution? completed
26. *Continue only if person aged 15 or more*
27. Highest level of schooling
28. Any educational qualification?
29. Highest qualification completed?
30. Main field of study of highest qualification
31. Qualification completed before 1998? (Yes/No)
32. For females: number of babies ever given birth to?
33. Total of person's income
34. Job last week?

35.	Working for employer or own business?	49.	Provided unpaid care?
36.	Own business: incorporated or unincorporated?	50.	Looked after a child without pay?
37.	Number of employees in own business?	51.	Did voluntary work?
38.	Occupation?	52.	Any usual residents absent?
39.	Main tasks performed?	53.	Sex, age, relationship of absent persons
40.	Employers business name?		*Dwelling questions*
41.	Workplace address?	54.	Number of motor vehicles?
42.	Industry of employer?	55.	Number of bedrooms in dwelling?
43.	Goods or services produced?	56.	Dwelling owned or rented?
44.	Hours worked in all jobs?	57.	Name of landlord
45.	Method of travel to work?	58.	Rent/mortgage payments
46.	Actively looked for work last four weeks?	59.	Access to internet?
47.	Could have started work last week?	60.	Permission to release personal information after 99 years?
48.	Did unpaid domestic work?		

Source: Australian Bureau of Statistics, 2016b.

Sources of population census data

Much census data and other details are now available online. National statistical offices or population departments may publish census data on their websites, for example:

- United States Census Bureau
  http://www.census.gov/

- Australian Bureau of Statistics
  http://www.abs.gov.au/

- Myanmar Department of Population
  http://www.dop.gov.mm/

- Statistics New Zealand
  http://www.stats.govt.nz/

- Statistics South Africa
  http://www.statssa.gov.za/.

Various international agencies support national census-taking and publish results:

- The United Nations Population Fund (UNFPA) provides technical support during all stages of a national census and helps ensure the data are widely disseminated. For example, UNFPA supported the 2014 Myanmar census. See: http://countryoffice.unfpa.org/myanmar/census/

- The United Nations Statistics Division publishes demographic and social statistics and has a website featuring World Population and Housing Censuses. See: http://unstats.un.org/unsd/demographic/

Integrated comparative projects provide a useful resource for demographic research, particularly Integrated Public Use Microdata Samples, International. Working from one to ten percent samples of individual records from national censuses including Thailand, Bangladesh, Vietnam, Cambodia and China, key variables in the file have been harmonised to enable comparability between countries. See: https://international.ipums.org/international/.

One advantage of using censuses for demographic study is that results can be presented for small geographical areas, unlike much survey data. Thus, the census is a major source of *geodemographics,* which are demographic data for specific geographic regions (Weeks 2012:139). However, a disadvantage of censuses is that, compared to surveys, results may be published years after the data collection (National Research Council 1993:3-5).

S AMPLE SURVEYS

A sample survey is cheaper and quicker than a census because it involves the selection of people who represent the whole population, or a particular section of it. Sampling saves human and financial resources. Thus, sampling textbooks emphasize the benefits of obtaining precise estimates at minimum cost. A survey can obtain information not available from other sources, and on topics, thus providing more scope for analysis (Fowler 1988:11-12).

The theory of survey sampling was established in the 1950s, allowing unbiased and precise estimates of populations or variables to be made if scientific probability samples are used (Bethlehem 2009:16). Thus, samples can provide detailed information on, for example, a women's reproductive health or on a person's life course.

Warwick and Lininger (1975: Chapter 3) consider that surveys will have one or more of these purposes:

- Exploration or clarification of a problem
- Description of groups, individuals, events or phenomena
- Causal explanation
- Hypotheses testing
- Evaluation

- Predicting or forecasting future events
- Developing social indicators.

The process of selecting survey respondents creates *sampling errors,* which do not occur when the whole population is covered. The main problem with sample surveys, therefore, is whether the propositions established from the sample can be generalised to the larger population from which the sample was drawn. Also, because a national sample contains relatively few people, it is often impossible to make judgements about small areas of the country or about small ethnic and other subgroups.

If the sizes of subgroups in a sample are large enough, surveys are very useful in studying differentials. Change can be calculated if two or more surveys have estimated the same indicator for the same geographic area at different times. A single survey can also provide measures of change by asking retrospective questions (such as 'Where did you live five years ago?') or by obtaining a respondent's complete migration, birth or employment history.

The World Fertility Survey (WFS) program that began in 1972 and ran until 1987 deserves a special mention. The WFS was described as the largest single social science research project ever attempted (*People* 1978:30). Its purpose was 'to assist a large number of interested countries, particularly the developing countries, in carrying out nationally representative, internationally comparable, and scientifically designed and conducted surveys of human fertility behaviour' (World Fertility Survey 1977:7). The implementation began in mid-1974 after a survey in Fiji of almost 5,000 women earlier that year. In all, 62 countries comprising about 40 per cent of the world's population participated (Gille 1987:18-20). Although these included 20 developed countries (17 in Europe plus Israel, Japan and the United States), Australia did not participate. The WFS did not provide the means to evaluate national family planning programs so a new series of surveys, the Contraceptive Prevalence Survey (CPS), commenced in the late 1970s (Morris et al. 1981).

The Demographic and Health Surveys (DHS) program, initiated in 1984, was designed to update and expand the data from the WFS and CPS. The objectives of the DHS included the provision of adequate data for policymakers, planners and researchers (Fisher and Way 1988:15). The DHS has led to a substantial expansion of the international health and population database. During the first phase, which ended in 1989, 34 surveys were carried out in 29 developing countries. The second phase ran from 1988 to 1993 and the third phase from 1992 to 1999. In 1997 DHS began incorporating traditional DHS features with expanded content on maternal and child health, and became known as DHS+.

By 2015 the DHS program had collected and disseminated data on population, health, HIV, and nutrition through more than 300 surveys in over

8

90 countries, including 173 surveys in Sub-Saharan Africa, 58 in Latin America and the Caribbean and 57 in South, Southeast and Central Asia. See: http://www.dhsprogram.com/. Countries that have held only one DHS include Afghanistan in 2010, Mexico in 1987, Samoa in 2009 and Thailand in 1987. DHS data are available through the World Bank Microdata Catalogue. See: http://microdata.worldbank.org/index.php/catalog/dhs.

Other sources of multi-country survey data include:

- Multiple Indicator Cluster Surveys (MICS), available from UNICEF http://www.childinfo.org/mics2_datasets.html

- World Bank Microdata catalogue http://microdata.worldbank.org/index.php/catalog

- World Health Organization microdata catalogue http://apps.who.int/healthinfo/systems/surveydata/index.php/catalog

- Demographic and Health Surveys (USAID/RAND) http://www.measuredhs.org

- INDEPTH: Demographic surveillance system http://www.indepth-network.org.

Some Australian surveys

The Australian Bureau of Statistics (ABS) conducts several surveys on different subject areas.

The Labour Force Survey (LFS) has been conducted on a monthly basis since February 1978. One of the key statistics derived from this survey is the unemployment rate. When redesigning the questionnaire in 2000, the ABS was careful to maintain the continuity of time series for key labour force variables (that is, employed, unemployed and not in the labour force). (Australian Bureau of Statistics 2001:1).

The National Health Survey (NHS) collects information about the health status of Australians. The ABS conducted the NHS every 5 years from 1995 until 2001, and every 3 years after that. The 2001 survey included approximately 30,000 people from all states and territories and across all age groups. A supplementary survey of Indigenous people was undertaken which increased the number of Aboriginal and Torres Strait Islander adults and children surveyed to 3,681. The aim was to support national estimates for the Indigenous population and enable comparisons between Indigenous and non-Indigenous populations for selected health characteristics. (Australian Bureau of Statistics 2003).

During 2011-13 the ABS conducted the most comprehensive study of the health of Australians ever undertaken, the Australian Health Survey (AHS),

which was an expansion of existing surveys including the NHS. Over 48,000 people participated, comprising around 35,000 from the general population and 13,250 from the Aboriginal and Torres Strait Islander population. The AHS collected information on physical activity, nutrition behaviours, and biomedical measures of nutrition status and chronic disease risk in the population. See: http://www.health.gov.au/nutritionmonitoring.

The National Aboriginal and Torres Strait Islander Social Survey (NATSISS) began in 2002 and has been conducted every six years. It provides self-reported information across key areas of social interest for Aboriginal and Torres Strait Islander people. Over 11,000 Aboriginal and Torres Strait Islander people took part in the latest survey in 2014-2015.

The above surveys generate cross-sectional data from different, but representative, samples of the population at specific points in time. In contrast, the Household, Income and Labour Dynamics in Australia (HILDA) survey has a longitudinal design, involving repeated observations of the same individuals over many years.

HILDA collects information about economic and subjective well-being, labour market dynamics and family dynamics. It is an important panel study of a nationally representative sample of Australian households which has led to research publication in many areas. See: http://www.melbourneinstitute.com/hilda/biblio/. Data relate to both individuals aged 15 and over, and households (Heard and Arunachalam 2015:241). Of the 13,969 individuals interviewed in the first wave in 2001, 8,543 were re-interviewed in wave 12. In wave 11 a new, top-up, sample was introduced, 92 per cent of whom were re-interviewed in wave 12 (Melbourne Institute 2012).

The Longitudinal Survey of Immigrants to Australia (LSIA) has been another important study whose outcomes have been used to assess government policies and programs, and in academic research. Australia's Department of Immigration and Multicultural Affairs stated:

> The LSIA seeks to provide government and other agencies with reliable data to monitor and improve immigration and settlement policies, programs and services... The benefit of using a longitudinal approach is that it provides a better picture of settlement progress than would be captured from a point-in-time survey.

> (Department of Immigration and Multicultural Affairs, 2001:1).

The Department of Social Services funded and managed a longitudinal survey of humanitarian migrants settling in Australia, called Building a New

Life in Australia. The survey, which ran from 2013 to 2018, 'aims to increase understanding of the well-being and experiences of humanitarian migrants settling in Australia. It focuses on their participation in society, their general health and happiness and investigating factors that may facilitate positive outcomes.' (Department of Social Services 2019).

The Personality & Total Health (PATH) Through Life project is an ongoing, population-based, longitudinal cohort study comprising approximately 7,500 participants ranging from early to late adulthood. It is administered by the Centre for Research on Ageing, Health & Wellbeing at the Australian National University. See: https://rsph.anu.edu.au/research/projects/personality-total-health-path-through-life. It assesses participants across the full adult lifespan.

The 45 and Up Study comprises over 267,000 residents of NSW who were aged 45 and older when they joined the study in 2006-2009. Participants completed an initial baseline questionnaire, and then follow-up questionnaires periodically. They have agreed to have their questionnaires linked with their Medicare, hospital and other health records, including their death records. See: https://www.saxinstitute.org.au/our-work/45-up-study/

VITAL REGISTRATION

Vital registration is concerned with the recording of key life cycle or *vital events*. According to the United Nations (2003) these are 'A live birth, death, foetal death, marriage, divorce, adoption, legitimation of birth, recognition of parenthood, annulment of marriage, or legal separation'.

Unlike censuses that describe the state of the population at a fixed point in time, vital statistics are collected on a *continuous* basis, and are thus important for the study of population change. Today, for legal purposes, various *registers* are kept on deaths, births, and marriages. In many countries birth certificates are a requirement for school enrolment.

In some countries religious records were kept before there were national systems, such as parish registers in Europe or temple records in Japan. Local or parish registers were kept by some churches in Europe from the 14th century onwards but civil or state registration systems did not develop until the 19th and 20th centuries. Both the parish registers and national records of births, deaths and marriages are recorded in separate registers, with no cross-referencing.

Registers and administrative statistics are key sources of data in historical studies, especially if they can be linked with other data. In the 1950s historical demographers used French parish registers to develop the technique of *family reconstitution.* This technique often begins with a record of a marriage, and then links this with other records showing the dates of birth and death of the

spouses and their children. Sophisticated computer-based techniques of record linkage are now being used in Europe and North America. The work of Rallu (1990) in French Polynesia and Vanuatu shows that family reconstitution can be possible in developing countries as well. Sources of Australian historical demography are discussed later in the Historical Demography chapter which follows.

A good registration system should be continuous, complete, permanent and compulsory. If registration is effective and compulsory, as in Australia and similar developed countries, the aggregate data can be used to calculate demographic rates and changes over time. For example, the registered numbers of births and deaths, used together with the most recent census and migration statistics, enable the calculation of birth and death rates and up-to-date estimates of the population.

Registration provides flow data for demographic analysis, so the data relate to *incidence* (rate of occurrence of new cases) rather than *prevalence* (proportion of cases in the population at a given time). Because public health administrators are putting considerable emphasis on preventing disease and reducing mortality, the analysis of certain items of registration data (such as causes of death, age at death, and the dead person's occupation) is increasingly needed.

In England and Wales, an 1837 Act of Parliament provided the basis for a national registration system, with a General Registry Office headed by the Registrar General and with a provision to record the cause of death (Devis and Rooney 1999). In Australia, each State and Territory has its own registers. In Victoria, for example, there is a Registry of Births, Deaths and Marriages within the Department of Justice. See: http://www.justice.vic.gov.au.

A number of developing countries in Central and South America have complete vital registration (of births, marriages, and deaths), and countries such as Mauritius and Fiji have a fairly complete recording. Sri Lanka is one of the few developing countries in Asia with a compulsory and comprehensive registration system. However, the difficulties and costs of establishing a complete registration system are so enormous that registration is unlikely to provide reliable demographic data for most developing countries within the next few decades.

Only a few characteristics, including age and sex, are common to both the census and vital registration. The United Nations recommendations for first priority topics for registration include the date and place of occurrence of the event, the date of registration, and place of residence. In addition there are various priority items for the different types of registration:

- For *birth registration*, multiple or single birth, attendant at birth or delivery, legitimacy, mother's age, parity and duration of marriage

- For *death registration,* cause of death, certifier, and the age, sex and marital status of the deceased

- For *marriages,* type of ceremony, age and previous marital status of the bride and groom

- For *divorces,* age, number of dependent children and duration of marriage.

It is often difficult and time-consuming to access data from registration systems which are jurisdiction specific. For example, researchers seeking to access Australian mortality data need access to data from eight different jurisdictions, each with their own registrar and coroner. The Australian Institute of Health and Welfare manages the National Mortality Database and the National Death Index, which provide a centralised source for many variables contained in state death registries. See: https://www.aihw.gov.au/about-our-data/our-data-collections/national-death-index/about-national-death-index

Population registers

As discussed above, a vital registration system provides continuous data on vital events but each register is kept separately so it is difficult to get a complete picture of the individuals in the population. A *population register* also provides a continuous record of vital events, but is an integrated and more comprehensive system because a personal record showing vital events and migration is kept for everyone in the population.

Population registers were first kept in ancient China and were later adopted by the Japanese. Today universal population registers which cover the whole population are less common than censuses or vital statistics. Keeping a universal population register often means that everyone has to carry an identity card, and in some countries, this is thought to infringe on the freedom of the individual. In the 1970s only 16 countries had registers with almost complete coverage. Of those countries only three were outside Europe: Taiwan, Israel, and Japan (Shryock and Siegel 1973:34). In Europe, Denmark, the Netherlands and Sweden have chosen not to take censuses deeming that sufficient information is available from registers and surveys (Laroche 1993:154). The last traditional censuses were taken in 1970 in Denmark, in 1971 in the Netherlands and in 1990 in Sweden.

Demographic Surveillance Systems

In the absence of an adequate national registration system, *Demographic Surveillance Systems* which monitor births, deaths, causes of death, migration, and other indicators within a defined population over time may be used. In Asia, for example, the Matlab field research station has recorded births, deaths and migration in rural Bangladesh since 1966. Matlab was established in Dhaka by the Cholera Research Laboratory, Bangladesh, and succeeds the

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B). In 2002, ICDDR,B stated that the fundamental mission of the Centre is to develop and disseminate solutions to major health and population problems facing the world, with emphasis on simple and cost-effective methods of prevention and management. See https://www.icddrb.org/

To achieve these aims, the ICDDR,B has undertaken a vast number of research projects, many of which are described on its website. Much of this research is based on Matlab field studies involving 225,000 people under the Health and Demographic Surveillance System and 110,000 people under the Maternal and Child Health-Family Planning Project.

INDEPTH is an international network of field sites with continuous demographic evaluation of populations and their health in developing countries. In 2015 there are 52 INDEPTH Demographic Surveillance Sites, with 39 sites in Africa, 11 sites in Asia and 2 sites in Oceania. Details can be found at http://www.indepth-network.org/

ADMINISTRATIVE STATISTICS

Administrative statistics are collected by Government Departments and other bodies for primarily administrative purposes. The data produced are a by-product of administrative processes within the government or bureaucracy and include service statistics, border statistics, and special registers of diseases. Currently the push for open data programs is expanding access to administrative statistics, but often not at suitable levels for research. Like vital registration data, administrative data are generally provided as aggregates and are not available for individuals. See, for example:

- The U.S. Governments open data website: http://www.data.gov/

- The Australian Governments public data website: http://data.gov.au.

Service statistics

Service statistics are administrative statistics collected by service providers. For example, schools keep records of students enrolled, and hospitals and clinics keep records of patients and clients who attend. In countries such as Indonesia that have government sponsored family programs, these service statistics provide important indicators of program success. Indonesia has better family planning service statistics than Australia where family planning clinic statistics cover only a small part of the population.

Border statistics

Border statistics are important for studying international migration and the movement of people across international boundaries. Arrivals and departures are normally recorded either electronically or by completion of a form. One

exception, where such statistics are not collected, relates to travel between European countries that have signed the Schengen Agreement.

Many nations such as Australia and New Zealand have forms for passengers arriving from another country. In Australia, arrivals complete an Incoming Passenger Card, while people leaving complete an Outgoing Passenger Card. In New Zealand, the corresponding forms are called the Arrival and Departure Cards. The information is required for the administration of Customs, Health, Immigration, and other laws and it can also be used for data matching between agencies administering and entitled to receive it under relevant laws. For example, in Australia international movements of people on social security benefits are provided to the relevant government agency.

The information supplied by a passenger arriving in Australia includes name, passport number, nationality, date of birth, occupation, flight number or name of ship, country of embarkation, contact details in Australia, emergency contact details, and intention (or not) to live in Australia for the next 12 months.

Passengers also have to indicate if they are:
- migrating permanently to Australia or
- residents returning to Australia or
- visitors or temporary entrants.

Those who are not Australian residents are asked if they suffer from tuberculosis and if they have criminal convictions. Similar information is obtained on the Outgoing Passenger Card, which also records country of residence of visitors/temporary entrants.

The data enable the Department of Home Affairs to produce statistics on:
- settlers
- temporary entrants
- permanent departures.

Special registers

Some diseases are notifiable, whereby medical staff must inform the health authorities when cases are diagnosed. Registers can be based on such notifications. Some special registers on chronic conditions can be used to supplement death registration data.

The National Diabetes Register is an example of a voluntary register. The data are compiled by the Australian Institute of Health and Welfare and record people in Australia with diabetes who use insulin and who have consented to be included in the register. Information collected includes the person's name, address, country of birth, date of diagnosis and whether they identify as an Aboriginal or Torres Strait Islander person. Thus, the register shows the number of new cases of diabetes diagnosed each year and their geographical distribution. It facilitates estimates of the services and products needed by

people with diabetes, and helps researchers find better ways to prevent and treat diabetes. See: http://www.aihw.gov.au/national-diabetes-register/.

An example of a non-voluntary registry is the Cancer Registry in each state and territory of Australia. Data on individuals with cancer is collected from a wide variety of sources including pathology laboratories, hospital records, aged care facilities and state or territory death registries. Data can be linked with other data sources at the individual level to develop a comprehensive picture of cancer risk, incidence, treatment and outcome. See https://www.aihw.gov.au/about-our-data/our-data-collections/aacr

OTHER QUANTITATIVE DATA SOURCES

An increase in the use of technology in the everyday lives of people in both developed and developing countries has opened up some new data sources which can be used to answer novel questions in demography. One such new source is mobile phone data. In 2017, 91% of adults in developed countries and 90% of adults in developing countries owned or had ready access to a mobile phone (Deloitte 2018). By tracking the pings that mobile phones make to cell towers, researchers have been able to estimate the population density of cell areas and to track population movement after natural disasters (Hughes et al 2016). Mobile phone data has been used to track internal migration in Rwanda (Blumenstock 2012). If the movement data can be linked with other demographic features such as age and employment status, mobile phone data could become a useful tool in future studies.

Another non-traditional data source is social media, such as Twitter, Foursquare, Flickr, Facebook and LinkedIn. In theory, social media could be used to track migration and employment, particularly when communication is geotagged (Hughes et al 2016). One study used data on the number of users who were classified by Facebook as 'expats' to estimate the number of expatriates in 17 European Union countries. They had mixed results, finding that they accurately estimated increase in Venezuelan migration to Spain, but significantly over-estimated the number of expatriates from the Philippines in Spain and Italy (Spyratos et al 2018). Whilst research on using social media for demographic study is still in its infancy, it shows some promise.

Other quantitative data that can be used for demography research include numbers of calls to Quitline stop-smoking services to estimate the prevalence of smoking; traffic monitoring systems to estimate short-term migration patterns of workers; infra red heat sensing to estimate the size of a population in an open or loosely covered area such as open rural land and refugee camps; and sewerage outflow monitoring to estimate population size.

QUALITATIVE DATA

Qualitative methods can be defined by default, that is, they use non-quantitative data, namely words rather than numbers; tend to be micro- rather than macro-level; and seek to understand processes at the individual rather than the aggregate level (Obermeyer 1997:814).

Qualitative methods include:
- life histories and genealogies
- participant observation in groups
- interviewing key informants.

Demographic reference books in the past have often failed to mention qualitative methods. Stycos (1981:450) concluded that qualitative methods had not flourished in the field of population studies; rather the survey had become the standard instrument for basic data collection. During the 1980s qualitative methods gained greater attention (Caldwell and Hull 1988). Demographers are now more appreciative of the methods of data collection used by anthropologists, and mixed methods are useful.

However, Susan Greenhalgh, an anthropologist, says that what appears to be the same qualitative method can in fact mean different things to demographers and anthropologists. For this reason, she argues against the use of terms such as anthropological methods by demographers (Greenhalgh 1997:820, 823).

Qualitative data can also be collected from focus groups, consisting of a small number of participants of roughly equal status. With this method the researcher organises a small group to discuss a particular topic and records the interaction of the group (Morgan 1988, Knodel 1997). For example, Knodel et al (1984) held focus group sessions in rural Thailand to discuss family size. In each village there were four separate groups: young men, young women, older men, and older women. Fricke (1997:830) sees advantages in small-scale and intensive community studies that combine survey, focus groups and general ethnographic enquiry.

Scrimshaw (1991:240) has identified the following advantages of qualitative methods:
- Cross-checking (triangulation) can be used
- Real, as well as ideal behaviour, can be identified
- Sensitive topics can be explored in context
- Attitudes can be revealed
- Observation is possible
- Open ended, so that any factors can be explored.

The disadvantages, particularly when compared with survey methods, include:
- The lack of random sampling and statistical testing of data

- With small samples, it is difficult to generalise
- Data collector bias
- It is time-consuming and labour intensive (for example qualitative data often involves the use of transcripts of in-depth interviews and focus groups).

Qualitative data collection involves repeated contact with the field (Obermeyer 1997:816). This is in sharp contrast with armchair demography where researchers can analyse quantitative data from a country or locality without ever having been there.

QUALITY AND COMPARABILITY

Evaluation of data is very important: 'The population researcher must be ingenious not only in seeing opportunities for original research using data from various sources, but also equally ingenious in evaluating the adequacy of any existing or proposed data source.' (Baum 1993:3.1)

Factors for evaluation include:
- Adequacy of subject matter content
- Adequacy of geographical, population group and temporal coverage
- Response accuracy and sensitivity
- Cost considerations
- Organisational considerations
- Timeliness of results.

Of these, the second and third points are most important when two or more datasets are compared. In particular, problems inevitably arise if measuring change involves the use of two or more datasets or if different countries are compared. One of the most difficult problems faced in planning a comparative survey is that of ensuring equivalence in concepts and procedures across the various cultural units studied (Warwick and Lininger 1975:44). Language is one potential source of difficulty, often ignored by demographers. Not all terms can be easily translated from a world language such as English into a local language. For example, terms such as domestic violence or underemployment may not be easily translated.

There are international recommendations for the conduct and analysis of censuses (United Nations 2008:109-110). Some programs such as the Demographic and Health Survey have standard procedures which help to ensure that data are comparable across countries. In addition, UNICEF assists countries in collecting and analyzing data in order to fill data gaps for monitoring the situation of children and women through its international household survey initiative the Multiple Indicator Cluster Surveys (MICS).

Coverage

National censuses are intended to cover the whole population at a particular moment in time (the census moment). For sample surveys the target population needs to be specified. For example, the Australian Family Project focused on detailed life histories of women, but for reasons of economy, rural areas with fewer than 0.07 dwellings per square kilometre were excluded, as were non-private dwellings (Bracher 1987:109).

Sample surveys are not intended to cover the whole population. Instead they collect information on the target population which consists of the elements that the researcher would like to study. In some studies, listing is undertaken (e.g. of households or individuals) to identify those in the target population. For example, in a study of returned migrants, listing may identify those households that contain return migrants and identify the migrants. With some fertility surveys, 'never married' women have been excluded, despite pre-marital sexual activity increasing. For example, 'never married' women were excluded from fertility surveys in the USA prior to the 1980s and in China in the 1990s (Li and Newcomer 1996).

Non-coverage refers to failure to include elements that ought to be in the sample. It may be easier to obtain good coverage for a single city or local government area than for the whole country (Kish 1965:527-535). One example of non-coverage arises when boundaries are unclear, and in such cases it may be better to choose a sample design based on compact clusters such as villages. Non-coverage can be reduced by tighter supervision, and by good listing and mapping. Changes in geographic boundaries can create substantial problems for researchers measuring demographic change over time.

Non-response refers to failure to obtain observations on some of the selected elements. It includes 'movers' in panel studies and 'not at home', 'refusals' and 'not found' in mail surveys. Evaluating and comparing surveys should look at the extent of total non-response and also non-response to particular questions, which is called missing data.

Problems of comparability of demographic data can also arise as definitions change over time or if different countries use different definitions. For example, several early censuses showed very low female labour force participation because women's work in agriculture was not recognised as work (Ware 1981:210-15). All countries do not use the same definition of gestational age and this makes it impossible to perform international comparisons that aim to identify cases needing a better pregnancy assistance (Loghi et al 2008:89).

Definitions of 'household' and of 'urban area' also vary considerably between countries. Some censuses have only four categories of marital status:

single, married, widowed, and divorced. Others have additional categories of separated and consensually married. With vital registration, marriage and divorce are defined for statistical purposes according to the laws of individual countries.

National definitions of urbanisation vary considerably across countries (Alkema and Jones 2013:291). About half of the countries use administrative criteria, while the remainder use population size and density or socio-economic criteria, or give no definition at all (Alkema and Jones 2013:292). Institutions such as the United Nations (2012) regularly produce wall charts for the world. In such instances, attention must be paid to the notes relating to the data.

Measuring ethnicity is another example of where care is needed. The United Nations in its 1998 Principles and Recommendations for Population and Housing says that no internationally relevant criteria can be recommended. A study of the way that countries classified their populations found that 54 out of the 141 countries planning a census for 2000 did not include a question on ethnicity. This was sometimes because of ideas of national unity. Of the remaining 87 countries that enumerated ethnicity, only 45 used 'ethnicity' as a primary term. Other countries measured 'race', 'nationality', or 'colour', and for some countries the question was optional. The definitions of 'ethnicity' and 'nationality' varied widely between countries (Morning 2008).

THE FUTURE OF DATA COLLECTION

Statistical standards have been steadily improving throughout the world while technological advances are enabling the provision of more timely and accessible data to researchers and government officials. Statistical data integration in which data from different sources are linked is becoming more widespread. There are many benefits from statistical data integration:

> Analysis of integrated datasets offers valuable opportunities to investigate more complex and expanded policy and research questions than would be possible using only separate, unlinked data sources. Integration can produce new official statistics (such as those based on analysis of longitudinal and small area data) to inform society.

> Data integration can reduce the need for costly collections by better leveraging existing data to meet current and emerging information requirements. Maximising the use of existing data, rather than establishing new collections, avoids additional load on respondents, helps to ensure cost-effectiveness and can improve timeliness.

> (Australian Bureau of Statistics 2013)

Two important linked datasets created by the Australian Bureau of Statistics are the Australian Census Longitudinal Dataset (ACLD) and the Australian Census and Migrants Integrated Dataset (ACMID). ACLD brings together data from the 2006 and 2011 censuses to create a research tool for exploring how Australian society is changing over time. ACMID enables the characteristics of individual migrants collected in the 2011 census to be cross classified with administrative visa data from the Department of Home Affairs.

Large-scale linked data is increasingly used in health research. For example, a project at the National Centre for Epidemiology and Population Health uses Australian census records linked to death records where the cause of death is related to heart disease, to develop a very powerful tool to examine social, economic, geographic and other factors and their association with heart disease. Other projects link questionnaire data from the 45 and Up Study with records of hospital admission, death registration, cancer registry, pharmaceutical prescriptions and medical treatment both within and outside hospital.

Some countries, such as the United Kingdom, now have such a body of data that much remains unanalysed, partly because of a shortage of skilled researchers. There also remain serious deficiencies in data in some areas, such as forced migration, that will be discussed in later chapters. Wars and other conflicts have meant that data collection has been interrupted in many regions. Many governments, including the Australian Government, are considering lower cost alternatives to the census. According to Martin (2015:1), the Australian Bureau of Statistics suggested abandoning the 2016 census in order to pay for an upgrade of its ageing computer systems. When both were funded, the Government was congratulated on retaining the census, with commentators noting that if we had lost it, we would have been at risk of making uninformed decisions about social services and urban planning (Martin 2015:2).

Once data are collected it is important that they are retained, documented and easily accessible. In Australia, two data archives are:

- Australian Data Archive: http://www.ada.edu.au
- Inter-University Consortium for Political and Social Research (ICPSR): http://www.icpsr.umich.edu

Citizens around the world are increasingly interacting with digital technologies. The resulting digital trails provide new sources of data including data from satellites, mobile phones, search engine queries and social media. This data can be mined relatively cheaply for interesting patterns. These have the potential to provide more timely information on emerging trends such as epidemics and economic upturns/downturns. On the other hand, these new data sources are typically characterised by statistical bias and measurement error, which limit the validity of statistical inferences that can be drawn. The

magnitude of these types of error cannot be reduced by increasing the size of the data set, unlike errors due to sampling. The challenge is to find effective and valid ways of utilising these new and emerging data sources (Australian Bureau of Statistics 2015).

Demographers may make use of non-random opinion data, such as phone or online surveys on subjects such as abortion, same sex marriage and euthanasia. Bethlehem (2009:24) has expressed concern about non-probability samples, particularly web surveys, where problems include under-coverage and self-selection of respondents.

In 2013 a United Nations High Level Panel called for a data revolution to improve the quality of data available to citizens, noting that data is an effective means to achieve greater accountability of government, corporation, and civil society (IUSSP 2014a). In 2014 the International Union for the Scientific Study of Population (IUSSP) convened a meeting of 22 prominent demographers to discuss how demographers and population scientists could contribute to this data revolution. The demographers present identified two crucial issues: data quality; and interoperability principles that facilitate data linking, including addressing any biases of big data (IUSSP 2014b).

As the world becomes more technologically focussed, the opportunities to identify and exploit new data sources for demographic research can only increase. The problem will not be collecting the data, but managing and analysing it. This is a challenge that has already begun.

REFERENCES

Alkema, Leontine and Jones, Gavin, 2013. Levels of urbanization in the world's countries: testing consistency of estimates based on national definitions. *Journal of Population Research* 30(3):291-304.

Australian Bureau of Statistics, 2001. *Implementing the Revised Labour Force Survey Questionnaire, Australian Bureau of Statistics Information Paper 6295.0.*

——————————, 2003. *National Health Survey: Users Guide, 2001*, Australian Bureau of Statistics, Canberra.

——————————, 2011. *How Australia Takes a Census, Catalogue No. 2903.*0, Australian Bureau of Statistics, Canberra.

——————————, 2013. *Key Facts About Statistical Data Integration.* Available at: http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Key+facts+about+statistical+data+integration

——————————, 2015. *Research Paper: Big Data, Statistical Inference and Official Statistics, Catalogue No. 1351.0.55.054,* Australian Bureau of Statistics, Canberra.

——————————, 2016a. *Using Census Data.* Available at: https://www.abs.gov.au/websitedbs/censushome.nsf/home/datause?opendocument&navpos=510

——————————, 2016b. *Household Census Form.* Available at: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/2901.0Main%20Features802016/$FILE/2016%20Census%20Sample%20Household%20Form.pdf

Baum, Samuel, 1993. Editors introduction, in D. Bogue et al. (eds), *Readings in Population Methodology,* Vol 1., Social Development Centre, Chicago, 3.1-3.2.

Bethlehem, Jelke, 2009. *The Rise of Survey Sampling,* Statistics Netherlands, The Hague.

Blumenstock, Joshua, 2012. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. *Information Technology for Development,* 18(2):107–125

Bose, Ashish, 1973. The Census of India: a historical perspective, *Demography India,* 2(1):18-33.

Bracher, Michael, 1987. The Australian Family Project, *Journal of the Australian Population Association,* 4(2):106-122.

Caldwell, John, 1969. *African Rural-Urban Migration,* Australian National University Press, Canberra.

Caldwell, John and Hull, Valerie (eds), 1988. *Micro-Approaches to Demographic Research,* Kegan Paul, London.

Cho, Lee-Jay and Hearn, Robert L., 1984. *Censuses of Asia and the Pacific: 1980 round,* East-West Center Population Institute, Honolulu.

Davis, Kingsley, 1951. *The Population of India and Pakistan.* Princeton, Princeton University Press.

Deloitte, 2018. *Global mobile consumer trends,* 2nd edition. Available at: https://www2.deloitte.com/global/en/pages/technology-media-and-telecommunications/articles/gx-global-mobile-consumer-trends.html

Department of Immigration and Multicultural Affairs, 2001. *Introduction to the LSIA*, Canberra. http://www.immi.gov.au/research/lsia/lsia01.htm

Department of Social Services, 2019. *Building a New Life in Australia* (BNLA): The Longitudinal Study of Humanitarian Migrants. Available at: https://www.dss.gov.au/our-responsibilities/families-and-children/programmes-services/building-a-new-life-in-australia-bnla-the-longitudinal-study-of-humanitarian-migrants

Devis, Tim and Rooney, Cleo, 1999. Death certification and the epidemiologist, *Health Statistics Quarterly,* Spring:21-33.

Economic Commission for Africa (ECA), 1974. *Manual on Demographic Sample Surveys in Africa,* ECA, Addis Ababa.

Fisher, Andrew and Way, Ann, 1988. The Demographic and Health Surveys Program: an overview, *International Family Planning Perspectives*, 14(1):15-19.

Fowler, Floyd, 1988. *Survey Research Methods,* Sage, Newbury Park.

Fricke, Tom, 1997. The uses of culture in demographic research; a continuing place for community studies, *Population and Development Review,* 23(4):847-853.

Gille, Halvor, 1987. Origin and nature of the World Fertility Survey, *The World Fertility Survey,* Oxford, Oxford University Press, 7-28.

Greenhalgh, Susan, 1997. Methods and meanings: reflections on disciplinary difference. *Population and Development Review,* 23(4):825-832.

Heard, Genevieve and Arunachalam, Dharmalingam (eds), 2015. *Family Formation in 21st Century Australia*, Springer, Netherlands.

Hughes, Christina, Zagheni, Emilio, Abel, Guy J.,Wiśniowski, Arkadiusz, Sorichetta, Alessandro, Weber, Ingmar and Tatem, Andrew, 2016. *Inferring Migrations: Traditional Methods and New Approaches based on Mobile Phone, Social Media, and other Big Data*, European Commission project #VT/2014/093, Brussels.

International Union for the Scientific Study of Population, 2014a. Data Revolution Background Document. Available at: http://iussp.org/sites/default/files/Data%20Revolution%20Meeting%20Background%20Document.pdf

——————, 2014b. *Defining and successfully accomplishing the Data Revolution: The perspective of Demographers*. Available at: http://iussp.org/sites/default/files/Data_Revolution_Demographers_IUSSP.pdf

Kish, Leslie, 1965. *Survey Sampling*, Wiley, New York.

Knodel, John, 1997. A case for non-anthropological qualitative methods for demographers, *Population and Development Review,* 23(4):847-853.

Knodel, John, Havanan, Napaporn and Pramualratana, Antony, 1984. Fertility transition in Thailand: a qualitative analysis, *Population and Development Review,* 10(2):297-328.

Laroche, Benoit, 1993. The future of population census, in *International Population Conference, Montreal, Vol.* 3, International Union for the Scientific Study of Population, Liege:151-5.

Li, Rose, and Susan Newcomer, 1996. The exclusion of never-married women from Chinese fertility surveys, *Studies in Family Planning* 27(3);148-154.

Loghi, Marzia, Giorgi, Piero and Pierannunzuio, Daniela, 2008. Negative pregnancy outcomes by gestational age in Italy, *Genus* 64(3-4):83-9.

Martin, Peter, 2015. Census saved, $250m cash boost for ABS. *Canberra Times*, 7 May 2015:1-2.

Melbourne Institute, 2012. The Household, Income and Labour Dynamics in Australia (HILDA) survey. Available at: http://www.melbourneinstitute.com/hilda/

Morgan, David, 1988. *Focus Groups as Qualitative Research,* Sage, London.

Morning, Ann, 2008. Ethnic classification in global perspective: a cross-national survey of the 2000 round, *Population Research and Policy Review* 27(2):239-272.

Morris, Leo, Lewis, Gary, Powell, Dorian, Anderson, John, Way, Ann, Cushing, Jeanne and Lawless, Gregory, 1981. Contraceptive prevalence surveys: a new source of family planning data, *Population Reports,* Series M, 5:M161-1M198.

National Research Council, Panel on Data Collection, 1993. Data collection (for fertility and mortality), in D. Bogue et al. (eds), *Readings in Population Methodology,* Vol 1., Social Development Centre, Chicago, 3.2-3.8.

Obermeyer, Carla, 1997. Qualitative methods: a key to better understanding of demographic behavior? *Population and Development Review,* 23(4):813-818.

People, 1978. Verdict on world survey, *People,* 5(2):30.

Plewis, Ian, 1985. *Analysing Change,* Wiley, New York.

Rallu, Jean-Louis, 1990. *Les Populations Oceaniennes aux XIXe and XXe siecles,* Institute National Demographiques, Presses Universitaires de France, Paris.

Randall, Sara, 1988. Use of ethnological knowledge in the collection of Demographic data, in *African Population Conference, Dakar,* Vol. 1:1.2.1-1.2.12, International Union for the Scientific Study of Population, Liege.

Scrimshaw, Susan, 1991. Combining quantitative and qualitative methods in the study of intra household resource allocation, in J. Cleland and A. Hill (eds), *The Health Transition: Methods and Measures,* Health Transition Centre, The Australian National University, Canberra:237-250.

Shryock, Henry, and Siegel, Jacob, 1973. *The Methods and Materials of Demography,* Part I, U.S. Department of Commerce, Washington, DC.

Shryock, Henry and Taeuber, Conrad, 1993. Characteristics of an adequate census program, in D. Bogue et al. (eds), *Readings in Population Methodology,* Vol. 1, Social Development Centre, Chicago, 3.16-3.22.

Spyratos, Spyridon, Vespe, Michele, Natale, Fabtizio, Ingmar, Werner, Zagheni, Emilio and Rango, Marzia, 2018. *Migration Data using Social Media,* Publications Office of the European Union, Luxembourg

Stycos, J.Mayone, 1981. A critique of focus and survey research: the machismo case, *Studies on Family Planning,* 12(2):450-6.

Thorvaldsen, Gunnar, 2017. *Censuses and Census Takers*. Routledge, London.

United Nations, 1998. *Principles and Recommendations for Population and Housing Censuses,* United Nations, New York.

—————————, 2003. MBS On-line data dictionary, *Monthly Bulletin of Statistics On-line*, January. http://unstats.un.org/unsd/mbs/ddgrid.asp

——————, 2008. Principles and Recommendations for Population and Housing Censuses Revision 2, *Statistical papers Series M No. 67/Rev.2*. http://unstats.un.org/unsd/publication/SeriesM/Seriesm_67rev2e.pdf

——————, 2012. *World Population 2012 Wallchart*, Population Division, New York. Available at: http://www.un.org/en/development/desa/population/publications/

Ware, Helen, 1981. *Women, Demography and Development*, Development Studies Centre, ANU, Canberra.

Warwick, Donald, and Charles Lininger, 1975. *The Sample Survey: Theory and Practice*, McGraw-Hill, New York.

Weeks, John, 2012. *Population,* Wadsworth, Belmont, California.

World Fertility Survey, 1977. *The World Fertility Survey: January 1976 -December 1976,* International Statistical Institute, The Hague.