Data Analysis and Report Writing for Civil Registration based Vital Statistics

SUPPORTED BY THE BRISBANE ACCORD GROUP (BAG)

Compiled by: SPC (Karen Carter, Renee Sorchik) and Brisbane Accord Group Partners (Nicola Richards-PHIN/ UQ /WHO) from sources as referenced. Additional worksheets to be used with this course provided by UNSW (Prof. Richard Taylor).

For course use ONLY

CONTENTS

PREFACE

Accurate data on births, deaths and cause-of-death are indispensable for accurate monitoring of population health, identifying health priorities and evaluating health program impacts. They also provide valuable context for a broad range of social development investment pertaining to education, social security, and child protection among others. This is particularly important in the Pacific due to the need to continue to monitor and report progress against the sustainable development goals (SDGs); which heavily rely on population data, and the need to respond to the impact of non-communicable diseases (NCDs) on the population. In addition, the rise of the NCD epidemic has meant that many of the traditional methods for calculating life expectancy and adult mortality from census data are less reliable as the models used in many of these calculations do not sufficiently account for this higher proportion of premature adult deaths. Civil registration is therefore an essential source of population data. The United Nations recognises it to be the most reliable source of vital statistics [UN Statistics Division; 2001].

Improved vital statistics data (focusing on births, deaths and causes of death) is a priority area under the Ten Year Pacific Statistics Strategy (TYPSS) phase II (2015-2017). Additionally, in July 2013, both the Pacific Ministers of Health Meeting in Apia and the Heads of Planning and Statistics Meeting in Noumea confirmed that improving vital statistics and cause of death data, along with other health information, is a key priority for the Pacific Islands. Both meetings called upon Pacific governments to undertake an assessment of their systems (if not already done) and develop national plans for CRVS improvement with clear targets. Ministers of Health also highlighted a need for additional capacity development in analysis and reporting of health data. This was reiterated at the 2015 Pacific Ministers of Health meeting in Fiji.

THE PACIFIC VITAL STATISTICS ACTION PLAN (PVSAP) AND THE BRISBANE ACCORD GROUP (BAG)

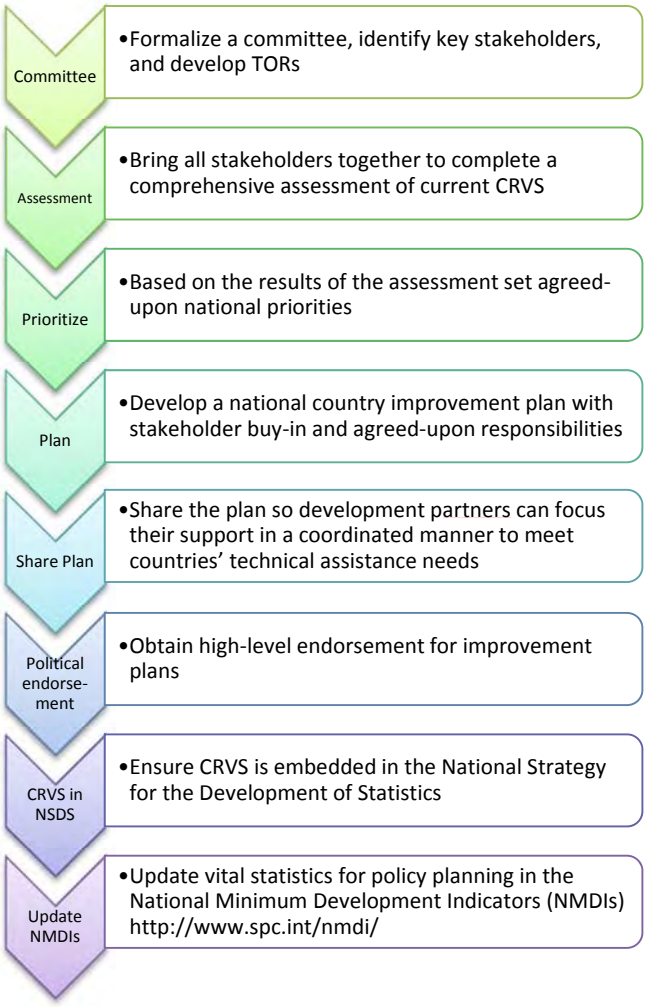
Significant progress has been made in improving reporting processes and data quality since the inception on TYPSS, although much remains to be done. Implementation of work related to CRVS is undertaken as a collaborative effort through multiple agencies under the Brisbane Accord Group (BAG), coordinated by SPC.

The Brisbane Accord Group was established in 2010 with the aim of supporting countries to improve their vital statistics and to improve coordination between development partners. Members of BAG include SPC, UQ, UNFPA, WHO, UNICEF, the Pacific Health Information Network (PHIN), the Australian Bureau of Statistics (ABS), Queensland University of Technology (QUT), University of New South Wales (UNSW), and Fiji National University (FNU). Through this group the Pacific Vital Statistics Action Plan (PVSAP) was developed to assist Pacific Island countries to understand the critical importance of vital statistics on births, deaths and causes of deaths and to improve their availability, accuracy and use. The plan sits under, and supports the Pacific Ten Year Statistics Strategy, which has been endorsed by all SPC member countries.

The basic premise of the Pacific Vital Statistics Action Plan is to work with countries to undertake an assessment of their collection and reporting systems for births, deaths and causes of death. Assessment findings are used by countries to develop a country-specific Vital Statistics Improvement Plan. Partner agencies are then able to focus their support in a coordinated manner to meet the technical assistance needs identified in the country plans. This country driven approach recognizes the importance of local context. It acknowledges that while there are similarities in the issues that countries face in strengthening

their civil registration and vital statistics systems, for technical support to be appropriate and for its desired impact to become sustainable, a good understanding of existing processes and structures is required.

Steps in developing a national CRVS improvement plan



The plan aims to implement activities within four categories of countries for which development priorities and needs are likely to be similar. Country groups are based loosely on geography, cultural affiliations, size, and known level of system development.

The PVSAP complements a range of global initiatives and partnerships including the Commission for Information and Accountability for Women and Children's Health. BAG agencies have agreed to coordinate work on vital statistics/civil registration improvements through the BAG structure and in accordance with the PVSAP to ensure coordination across these strategies. The PVSAP has also been recognized in the UN ESCAP regional CRVS plan and Regional Action Framework, with agreement that support and activities under this plan will be coordinated in the Pacific through the existing BAG structure.

INTRODUCTION TO THE COURSE

As countries work towards improving their CRVS systems, a number of countries are now in a position where their routine administrative processes either are able to, or will soon be able to generate reasonably reliable data. Ensuring that system improvements are reflected in an actual improvement in the availability and reliability of data is a key regional challenge. It is critical that this data is analysed, reported, and made available to decision makers. This may be a significant change for many NSO's who have not previously had a clear role in the analysis or reporting of this data. This course aims to build capacity in these areas.

The objectives of the workshop are as follows:

- Assist countries to complete a vital statistics report that illustrates current levels and trends of births, deaths, and cause of death over time; and that can be used for planning and policy review purposes (including providing the basis for monitoring related SDGs)
- Assist countries to meet the agreed-upon vital statistics regional guidelines for reporting
- Assist participants to develop key skills to critically appraise their data collection systems and sources
- Assist participants to understand the purpose and structure of the International Classification of Diseases, and how cause of death data is coded and tabulated
- Assist participants to build proficiency in key analytical skills required to meet reporting guidelines including:
    - Data aggregation
    - Quality assessment and re-distribution of unknown values
    - Basic rates (including CBR, CDR, IMR, U5M)
    - Understanding fertility calculations
    - Understanding life table calculations (including adult mortality)
    - Age-standardisation for mortality
    - Cause-specific mortality rates and interpretation.
    - Confidence intervals for key measures
    - Interpretation of results (plausibility, stability, comparison against other sources etc.)
- Assist participant to build proficiency in key data presentation skills including tabulations, graphical representation of data and interpretation
- Assist participants to practice and develop key report writing skills to make data easily accessible and understood by readers.

The course draws upon material from a range of primary sources as indicated throughout the text, in particular tools and materials developed by SPC, WHO, UQ, UNSW, ABS, the US Census Bureau, and Open courseware from John Hopkins University.

PART 1: Background

1 IMPORTANCE OF CRVS

Statistics for health and development

Accurate data on births, deaths, and cause-of-death are essential for:

- monitoring the health of a population;
- identifying health priorities and evaluating health and program impacts;
- providing access to real-time information on population size and structure; and
- providing the necessary data for the calculation of all population-based development indicators used to track development progress.

While there are multiple sources of birth and death data, well-functioning civil registration systems are recognised by the United Nations as being the most reliable sources of vital statistics [UN Statistics Division; 2001]. The latter is based on recognition of a number of advantages posed by the systems, over other data sources, primarily due to their continuous and universal character which also make them cost effective to implement.

Civil registration systems provide a bulk of the data needed for monitoring of social and economic development in countries. Notably, at least 24 of the 169 targets of the recently endorsed sustainable development goals (SDGs) will rely on civil registration data for their monitoring (SPC, 2016). In addition, the systems will also provide the denominator for measurement of all the population based targets and indicators of the SDGs. The following is a list of health related targets of the SDGs whose monitoring will directly depend on the availability of timely birth and death data; and preferably data from civil registration systems.

- 3.1 By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births
- 3.2 By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under 5 mortality to at least as low as 25 per 1,000 live births.
- 3.3 By 2030, end the epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water-borne diseases and other communicable diseases
- 3.4 By 2030, reduce by one third premature mortality from non-communicable diseases through prevention and treatment and promote mental health and well-being.
- 3.6 By 2020, halve the number of global deaths and injuries from road traffic accidents.
- 16.1 Significantly reduce all forms of violence and related death rates everywhere

Figure 1.1: Key uses of CRVS records and data in the Pacific

Real data for real decisions

While estimates of fertility and mortality serve an important function in highlighting the potential scale of a problem or calling attention to issues which would otherwise go unnoticed, they are not a substitute for real data. Estimated data cannot adequately monitor changes over time or account for differences in local conditions that may not be built into models. There is a real risk that policy decisions based on incorrect estimates could lead to poorer health and development outcomes, and that policy makers may not understand the uncertainty that is associated with estimated statistics. The only way to reduce the uncertainty of health and development statistics is to improve the collection, analysis, and dissemination of empirical data, sourced from a robust civil registration and vital statistics system (CRVS).

CRVS is cost effective

In the long term, investing in CRVS reduces costs and inefficiencies by lessening dependence on very costly surveys (such as the DHS or MICS), and censuses. Because it is routinely collected, CRVS provides more timely data, supporting more efficient and effective planning and service provision.

Registration protects rights

Civil registration also has an important legal function in providing a **legal identity,** and setting the foundation for the realisation of other rights and privileges that are associated with proof of legal identity.

"The child shall be registered immediately after birth and has the right to a name and nationality..."

International Covenant on Civil and Political Rights, Convention on the Rights of the Child, Article 7 – Signed by all countries of the Pacific

Civil Registration:

- provides proof of nationality by birth or descent
- facilitates access to social services such as education, health and social protection
- enables individuals to claim other critical rights and privileges the are dependent on the possession of a legal identity such as rights to participate in governance processes e.g. voting, rights to claim inheritance benefits from one's family etc.
- offers protection under law.

A birth certificate may be needed to obtain other forms of identity documents such as a national id card, a passport, a marriage or driver's license; and to provide proof of identity and family relationships while opening a bank account, seeking formal employment or claiming rights to inherit property.

Children without a legal identity are at greater risk of trafficking, statelessness, and arguably violence. Registration supports social and government accountability, and unregistered children cannot access the same social protection benefits as registered children.

Other children's rights associated with the right to birth registration include:

- The right of a child to be supported to live past their first, and fifth birthdays
- The right of a child to be supported to live into adulthood, and for young adults to be supported not to die prematurely, but to live through to old age
- The right to the protection afforded by a system that investigates deaths properly in order to understand and respond to deaths that did not occur due to natural causes.

At a societal level, **communities have the right to know why they are losing key community members, parents, workers, and leaders to premature mortality so that they can respond.** We will examine data users and their needs more closely later in the workbook.

1.1 REVIEW

- Accurate data on births, deaths and cause-of-death are important for setting priorities and monitoring the effectiveness of programs across a broad range of sectors including health, education, social security, etc.
- Population data is central to planning for government services and underpins development.
- Vital statistics are also needed for providing reliable, timely denominator data for the calculation of all population based development indicators – such as for education, welfare, gender, economy etc.
- Health issues in the Pacific such as the non-communicable disease epidemic mean that it is critical that countries can generate and make available measures of births, deaths and causes of death.
- Countries have a range of international reporting obligations that require vital statistics data, including the SDGs and the National Minimum Development Indicators.
- Improving the reporting of vital statistics was identified as a priority by both the Pacific Ministers of Health and the Heads of Planning and Statistics.
- The Pacific Vital Statistics Action Plan provides a regional framework to assist countries to improve their civil registration and vital statistics data, and is collectively supported by a group of regional and international agencies known as the Brisbane Accord Group.

2 SOURCES OF DATA, THEIR ADVANTAGES AND LIMITATIONS

Extract adapted from: Carter K. 2013. Mortality and Causes of Death in the Pacific. PhD Thesis. UQ, Brisbane.

Additional material in italics sourced (adapted) from Manual for "Training in the Use of Existing Health Datasets" [Adair, T. 2012]

Vital statistics may be collected through censuses, surveys, or routine vital event collections such as civil registration or registries operated through the MoH [Mahapatra P, et. al. 2007]. The three collection methods should be seen as complementary, rather than alternative options [UN Statistics Division; 2001]. Additional administrative data collections including records of births and deaths may also exist, and be used for analysis of levels of fertility, mortality and CoD distribution.

*These data sources can best be classified as **institution-based** and **population-based**.*

Institution-based data sources include those collected routinely from administrative and operational activities, including Health Information Systems (HIS) and hospital discharge data, police records for attended deaths, and social security records. It can also include surveys, such health facility surveys, where data are collected in an institution.

Population-based data sources are those that are representative of the whole population. These sources include population surveys, censuses and civil registration.

2.1 POPULATION-BASED DATA SOURCES

Census collections

The primary source of mortality data in PICTs has historically been through periodic censuses. *A population census is a compulsory, universal and simultaneous enumeration of the national population, conducted on a periodic basis. They provide population counts by age and sex as denominators that can be used in calculating rates where the numerator data are obtained from other data sources* [UN Statistics Division; 2008]. *They also collect socio-economic data.*

In the absence of reliable death registration at a population level, census data may be used to provide measures for all-cause mortality by age group, i.e. adult mortality, life expectancy at birth, and life expectancy at 40.

Information on births and deaths may be derived either directly or indirectly in a census.

Direct measures ask about births and deaths in the household over a nominated reference period – usually within the last 12 months.

Indirect measures collect some information on births and deaths then approximate when these events would have occurred.

Childhood mortality can be derived using the children ever born/children surviving technique (CEBCS), where women of child bearing age are asked how many children they have ever had and how many are still alive. Deaths are then distributed in time according to national fertility patterns and the age of the

respondent [United Nations, 1983]. Adult mortality can be estimated using sibling survival (how many brothers and sisters the respondent has had, how many are still alive, and age of respondent), widowhood (has the respondent ever been married, is that spouse alive, and age of respondent), and/or orphanhood techniques (is the respondent's mother or father alive, and age of respondent) with deaths distributed according to the age distribution of the respondents [United Nations, 1983]. The UN manual X [United Nations, 1983] provides a good overview of calculations associated with these methods.

These measures can then be used as inputs for model life tables that generate estimates of age-specific measures of mortality and LE. Models that use a single input parameter based on childhood mortality have been shown to have a tendency to under-estimate mortality in PICTs where low infant mortality does not necessarily imply low adult mortality [Carter K et. al. 2011]. Models that use both childhood and adult mortality generally perform better [Carter K et. al. 2011]. These will be discussed later in the workshop.

Indirect methods, based on the age distribution of respondents, are generally preferred by statistical agencies over direct questions related to the number of deaths in the previous twelve months which have "not proved to be very satisfactory" [United Nations, 2001] due to serious concerns around recall bias and subsequent under-enumeration of deaths [Byass P, 2007].

Estimates of adult mortality may also be derived from changes in population (by age and sex) based on survival and migration recorded between multiple census rounds. The technique of inter-censal survival is particularly reliant on accurate migration data, which is frequently unavailable at the level of disaggregation required in PICTs [Tomas P, 2009].

Enumeration of deaths through a census is noted to be more difficult in developing countries as there tends to be a wider family structure (less nuclear families) [United Nations, 2001], increasing the level of uncertainty in data captured through questions about family (such as how many siblings do you have) or households (such as have there been any deaths in the household in the previous 12 months). Some Pacific cultures also have social conventions that mean death cannot be freely spoken about, which may introduce response bias. Despite these concerns, censuses have remained the main source of mortality data for the region over the past 60 years due to the absence of reliable CRVS systems and concerns about data quality [Haberkorn G, 1998].

Periodic Surveys

Surveys are increasingly used in many countries to collect information on: the health status of a population, health facilities, and socio-economic status. Historically, surveys in PICTs have been more concerned with infant and child mortality, or deaths from specific causes such as maternal deaths, rather than all-age all-cause mortality. Globally, estimates of causes of death often utilise cause-specific mortality rates from epidemiological studies and surveys. Examples of direct cause-specific mortality studies include diarrhoea, acute respiratory infection, malaria, HIV/AIDS, TB and war/conflict. There are limitations in using epidemiological studies that focus on a certain disease(s). They can lead to overestimation of mortality due to multi-causality of some conditions and may lead to unawareness of its magnitude. Critical judgment therefore must be applied when analysing data from epidemiological studies.

Surveys are not ongoing data collections, but are collected in a specific period of time. Unlike the census that collects information on the whole population, surveys collect data for a proportion of the population considered to be representative of the broader population of interest. As with censuses, surveys may

collect data to undertake direct estimation of mortality from reported deaths, or indirect estimation based on the age distribution methods described previously. Repeated household surveys may also be used to estimate mortality based on survival and migration.

Surveys fill an important role in the provision of mortality estimates in the absence of routine data collection. The reliability of estimates from survey data is driven by how well the **sample selection** reflects the broader population of interest [Rowland DT, 2003]. Surveys are also subject to **recall bias** as they collect information on events that may have happened several years prior, and **response bias** (where people that complete the survey are different in some way to those who opt out or cannot be contacted) [Rowland DT, 2003; Lilienfeld AM, 1983]. As with censuses, further **response bias** may be introduced by cultural issues around discussing deaths. The use of surveys to collect data for monitoring health status and evaluating health programs can also prove costly [Haberkorn G, 1998].

Two important survey types for mortality estimation in PICTs are Demographic and Health Surveys (DHS) and the UNICEF Multi-Indicator Cluster surveys (MICS).

DHS surveys have been implemented across the region since the early 2000's. These are retrospective cohort studies that use a birth history (all births with dates and outcomes) from respondents to collect inputs for estimation of infant and child mortality. Estimates often become less plausible for earlier periods (5-10 and 10-15 years prior to the survey) [Stanton C, et. al. 2000; Bairagi R, et. al. 1997] due to recall bias [Obermeyer Z, et. al. 2006]. To date, the only MICS in PICTs was conducted in Vanuatu in 2007 [Vanuatu MoH, 2008], although more are planned for the region over the next several years. These collect data for estimation of infant mortality using the CEBCS techniques as employed in the censuses. Both surveys can be used to derive life tables from the estimates of infant and child mortality using model life tables as for censuses; however this is routinely done only for DHS's.

Bias is a systematic rather than stochastic (or random) effect on a statistic

Using survey data

Before we do any statistical analysis using survey data we must carefully examine the concepts and definitions used to collect the data. Questions to ask include:

- *What was the target population?*
- *What was the survey population?*
- *What was the statistical unit?*
- *How were the observations made?*
- *What standards or classifications, if any, were used? What were the coding decisions made?*
- *What edits or consistency checks, if any, were applied?*

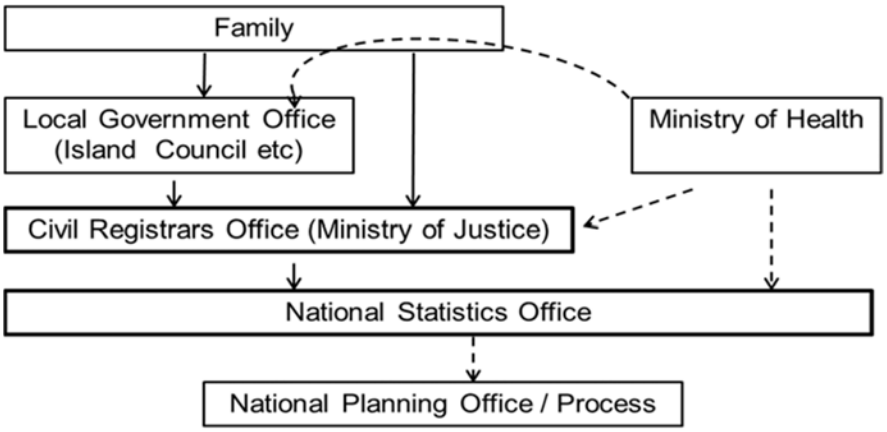
Other points to consider include whether there are enough observations for analysis (for example, in some sample surveys detailed statistical analysis of variables might not be possible). Also look at the level of non-response to the question [Ryan C, et. al. 2011].

Routine vital registration collections

Civil registration provides a legal basis for the recording of vital events such as live births, deaths, foetal deaths, marriages and divorces. An efficient routine CRVS system, with medical certification of CoD, provides ongoing and relatively low cost data collection and therefore timely mortality data for decision making [United Nations, 2001; Mahapatra P, et. al. 2007].

Routine data collections for births and deaths are usually managed through civil registration or health system reporting systems, but may also rely on lay reporting by a local chief or leader. Though this varies by country, civil registration is in many PICTs managed by the Ministry of Justice (or local equivalent) with local functions often delegated to local government (such as Town or Island Councils). The process comprises of two major components, recording the event in the official records and therefore establishing the civil status of the baby or deceased person, and the issuance of a legal document, the birth or death certificate, certifying the event for legal purposes; such as insurance and land inheritance in the case of a death or enrolment in school and identification for births. A death certificate may also be required prior to burial [United Nations, 2001].

Figure 2.1: Diagram of the reporting and registration processes for deaths



Official registration of death usually requires the family to provide proof of the death (often in the form of a medical certificate completed by a qualified doctor), attend the local civil registration office to complete the required paperwork, and pay a nominated fee. Although the civil registration process may record CoD, this is not the primary purpose of the system (other than to rule out unnatural causes that may require criminal investigation [AbouZahr C, et. al. 2007], and may vary in usefulness according to the source of the information. Systems that require a medical certificate are likely to be more reliable than systems that record CoD as reported by family members. As the legal record, civil registration is frequently legislated as the official source for mortality data [Mahapatra P, et. al. 2007] despite not always being the most reliable data source in country. An effective civil registration system requires coordination across Ministries and government agencies.

The ***medical certificate*** is considered the reference standard for CoD information, in the absence of an autopsy, as a qualified practitioner is required to assess the case and make an informed decision concerning the sequence of events that led to the death [WHO, 2011; United Nations, 2001]. This is particularly attractive in PICTs where small population sizes make complete routine registration feasible, despite the challenges of dispersed populations and limited infrastructure. Where universal certification of death is not possible, verbal autopsy should be considered for areas not covered by medical certification in order to

develop population based estimates of proportional mortality by cause (this will be discussed further later in the workbook).

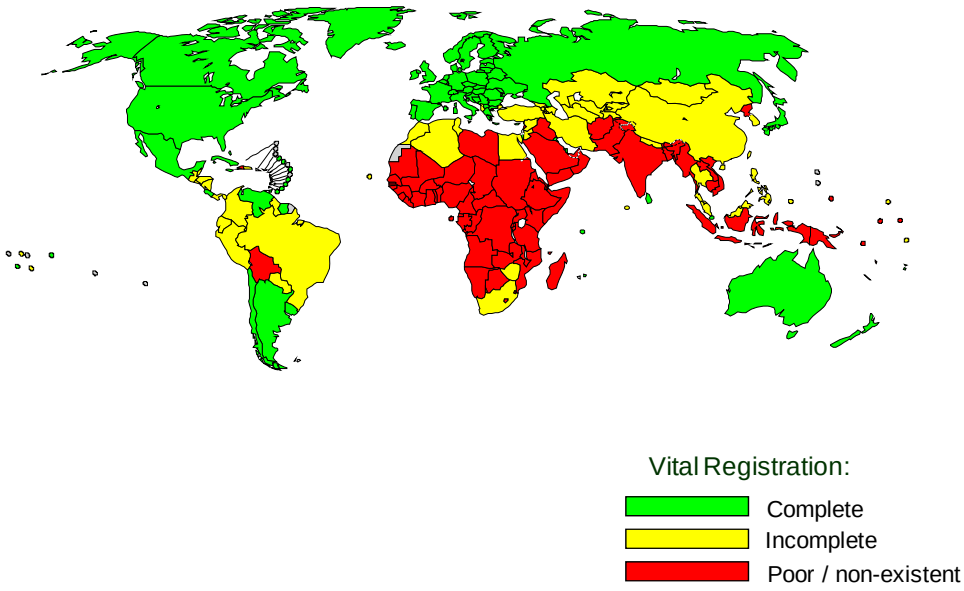
While civil registration is frequently upheld as the "gold standard" [AbouZahr C, et. al. 2007; Lopez AD, et. al. 2007 ; Setel PW, et. al. 2007 ; Mahapatra P, et. al. 2007] of mortality data (when complete) due to the relatively low cost of data collection, the ongoing nature of collection (and therefore timeliness) and the ability to combine this function with medical certification to obtain CoD [United Nations, 2001]; only seven of the 23 countries of the WHO Western Pacific Region were assessed by Mathers et. al. to have complete data collection systems in 2003 [Mathers CD et. al. 2005]. The 2004 assessment of mortality data in the region by Taylor et. al. [Taylor R, 2005] also found mortality from civil registration was implausible and demonstrated significant under-recording.

Figure 2.2 on the following page shows that vital registration is incomplete, poor or non-existent in most countries throughout the world. These include the poorest countries such in South Asia and sub-Saharan Africa, where mortality levels are high and accurate health information to inform policy is most required.

Registration of vital events within the health system

Vital events such as births and deaths are also often recorded through routine data collections within the health system. In comparison to the legal focus of the civil registration system, health data collections are primarily to inform operational decisions, and CoD is central to this purpose. Health systems for reporting deaths may include a vital registration system (a record of all deaths in both the health facilities and the community) based on medical certificates or community nursing reports, or facility based data (discussed in the following section).

Figure 2.2: Global availability of vital registration data



Demographic surveillance sites

A demographic surveillance system captures all vital events in a specified area. Often it is combined with disease detection (sentinel surveillance), especially for epidemiological studies of specific diseases. Demographic surveillance systems can be useful where there is incomplete vital registration, as it can provide interim mortality statistics, as well as serve as a basis for complete registration in the future.

This technique has been used in PNG [Riley, 2009] but it is unlikely to be a suitable solution in other PICTs where small populations mean rates are often quite unstable and already need to be averaged over several years to be interpretable [Taylor, 2001]. A further disadvantage of DSSs is that focussing enumeration in these areas over a number of years could result in a shift in community knowledge and attitude to health interventions and reporting that may not be reflected in the broader population. These sites may therefore become less representative of the broader population over time [Baiden et. al. 2007].

Surveillance systems

Surveillance systems utilise a range of data sources, both institution-based and population-based, to detect and report certain diseases, and their outcomes. They utilise a number of data sources, including individual health records and disease registries. Sentinel surveillance systems are used in health facilities or a defined population subset to monitor disease trends intensively. Surveillance systems also use household surveys to track risk factors, e.g., behavioural risk factors.

Summary of population based data collection approaches

As outlined in the preceding sections, and summarized in Table 5.1, while all of the data collection approaches have strengths and weaknesses, routine vital registration, done well, should be the preferred approach for mortality data in PICTs due to the continuous nature of collection, ability to generate data on both mortality level and CoD distributions, and the small population sizes which result in a great deal of uncertainty when dealing with rare events such as deaths, especially when disaggregated by age and sex as required to generate meaningful data for policy [UN Statistics 2001; AbouZahr C, et. al. 2007]. Although censuses remain the primary data collection in the region, they are not able to capture CoD, and in most instances rely on extrapolation of adult mortality and LE from childhood mortality based on model life tables, rather than measuring actual events. DHS and censuses will continue to be an important source of data in the absence of reliable routine CRVS systems, and for generating data to assess their performance.

Direct methods of estimating mortality from routine reporting have the advantage of timeliness, can better recognise differentials, and are better suited for providing estimates for infant and child mortality [United Nations, 1983]; although indirect methods are less expensive.

Table 2.1: Summary of Population based Data Collection Approaches for Mortality Data

Data Source		Periodicity	Sample frame	Period of interest	Data collection (mortality level data)	CoD Data Collected?
Census		Periodic – 5-10 years	Whole Population	Retrospective	Direct - (deaths in the household) Indirect – partial birth history (CEB/CS) & orphanhood data	No
Survey	DHS	Periodic – ~ 5 years	Selected sample – representative of whole population	Retrospective	Direct – complete birth history	No
	MICS	Periodic – ~ 5-10 years	2 stage clustered sample – representative of whole population	Retrospective	Indirect - partial birth history (CEB/CS)	No
	Other household based surveys	Usually once-off	Varies	Retrospective	Varies	Possible - using verbal autopsy
Routine vital registration	Civil Registration	Continuous	Whole population (depending on coverage)	Current	Direct reporting of event	Yes
	Health vital registration	Continuous	Whole population (depending on coverage)	Current	Direct reporting of event	Yes
	Hospital discharge records	Continuous	Hospital cases only	Current	Direct reporting of event	Yes
Other routine databases	Various	Continuous	Varies – usually targets sub-population of specific interest.	Current	Direct reporting of event	Usually limited
Demographic surveillance Sites		Continuous	Selected areas – usually not representative of whole population over time.	Current	Direct reporting of event	Yes

2.2 INSTITUTION-BASED DATA SOURCES

Institution-based data sources in the health sector primarily include those based at hospitals and health centres and cover:

- Data available on preventive services, acute curative services, follow-up of chronic disease, inpatient care, laboratory or radiographic examinations, referrals, immunisations etc.
- Utilisation of case records and disease records (consultation and discharge) to analyse mortality and morbidity of individuals utilising these services.
- Delivery of services, e.g., growth monitoring of children, antenatal care.
- Cancer registries, pregnancy registers: these are specialised collections of data. Cancer registries contain complete information of patient history, diagnosis and treatment of cancer patients. Pregnancy registers contain information about antenatal visits, birth delivery, birth weight, and neonatal survival.

Health resource data focus on the quality, availability and logistics of health service inputs.
The types of data include:

- *Human resources – number/type of staff, where they work, qualifications/experience*
- *Budgets and expenditure*
- *Health facilities - type and size of facilities, location/distribution*
- *Drugs/commodities*
- *Key services*.

In the hospital setting, unit record data on deaths and CoD may be collected through ***hospital separation data*** (hospital records that indicate whether the patient was discharged home from hospital, transferred to another facility or died). In these collections CoD is based on the principal diagnosis at the time of death (and hence discharge from hospital) as recorded by the physician on the front sheet of the medical record. While these data are based on cause as nominated by a trained physician, the reason for treatment is not always the underlying cause that led to death and therefore there will be some small degree of error in using these data in place of a medical certificate [UN Statistics 2001].

Countries may also collect data on deaths through primary or community health care nursing programs, with deaths recorded on a separate form to be sent to the local area nursing manager and/or recorded on monthly reports. Details collected about each death vary between country and system, but usually include name, place of death (or of report), age at death, and may include CoD as reported by the local health worker nurse or family members. These systems may focus on specific categories of deaths such as maternal or child deaths. There has been limited analysis (at least in the published literature) of CoD based on this type of data collection, and the usefulness of these data may be limited, or restricted to very broad categories of cause. Similarly lay reporting by community leaders, such as utilised in Tonga [Carter K, et. al. 2012], can provide accurate collection of mortality level but is limited in scope for collecting anything more than broad CoD. CoD is collected through these systems essentially to rule out whether a death is suspicious and requires investigation.

Registries

Cancer registries can be another source of outcome data, although they tend to be more useful for measures of morbidity rather than mortality (as they are generally a secondary data source in this case). When registries are used for mortality indicators, it is important to note that most registries record anyone who dies WITH the disease rather than FROM the disease (as the designated underlying cause) and therefore the data is not directly comparable to that obtained from civil registration.

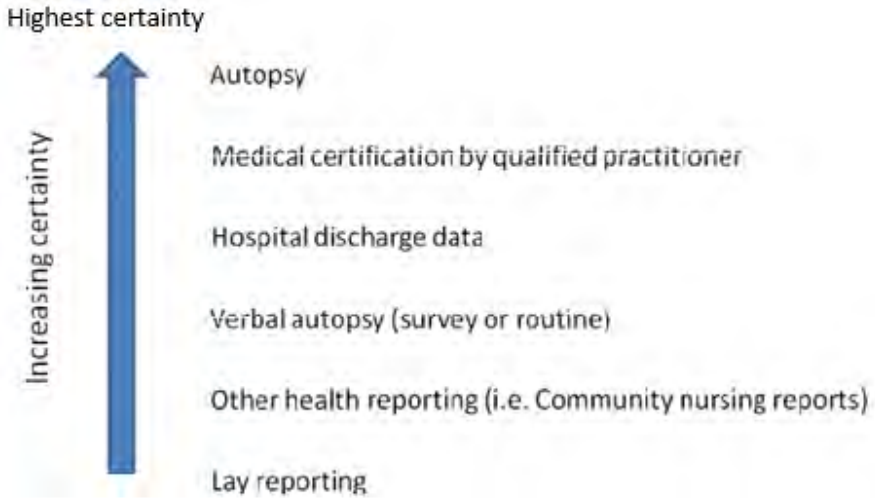
Police data

Police records may include information on road traffic accidents, homicides, suicides and unintentional injuries that can be used for analysis. This is often filed by case rather than being searchable by event, and is therefore generally of limited use at present.

2.3 DATA SOURCES TO ASCERTAIN CAUSES OF DEATH

Cause of death data can be obtained through medical certification, hospital discharge records, verbal autopsy, nursing reports or lay reports from family and other community representatives. As doctors are trained in diagnosis of disease, sources that collect data based on direct observation by the doctor (provided the doctor is aware of the patient's history) are obviously likely to be more accurate than reporting by nurses or family members who do not have the same level of training. While both medical certificates and diagnosis for hospital discharge records are completed by doctors, the medical certificate is specifically structured to capture underlying CoD and there is a clear set of coding rules in order to identify this (ICD 10). In comparison the hospital discharge record captures the "primary diagnosis" which may not necessarily reflect the cause that initiated the sequence of conditions that ultimately resulted in the death occurring.

Figure 2.3: Level of certainty of underlying cause of death by data source



(Carter K, 2013)

Although medical certification is therefore accepted as the preferred source of CoD data, the quality of this information varies significantly, with numerous studies identifying discrepancies between the medical certificate and information from medical records.

Verbal Autopsy

The shortcoming of using CoD information from medical certificates is that in many of the PICTs, a large proportion of deaths occur outside the hospital system without a medical attendant. This means either CoD is not provided, or is based on the determination of someone without a health background (such as the police or local registrar). The verbal autopsy (VA) arose out of work in "population laboratories" in India and Bangladesh in the 50's to 70's to address this issue. It is now recognised internationally as a viable alternative for collection of accurate CoD statistics, at a population level, where universal certification is not practiced.

A standardised questionnaire on symptoms displayed by the deceased prior to death is completed by a trained interviewer who visits family members, with CoD established by analysis of the questionnaire (either by a physician, or more recently by automated analysis). In 2012, WHO also released a short version of the standard VA tool designed for use in conjunction with routine reporting. This however is yet to be tested in the field. Standardised approaches with trained interviewers are a significant improvement on earlier systems of lay reporting of CoD through community leaders which had limited quality control around recording or standardisation of collection. Deaths should be reported separately from certified deaths.

NOTE: Verbal autopsy should only be used where doctors are not present and cannot be accessed by the general public.

2.4 USE OF MULTIPLE DATA SOURCES

When computing indicators, we can utilize multiple data sources in different ways. A common method of using multiple data sources is when using population data from a census for our denominator and another data source as our numerator.

Where using multiple data sources, care must be taken ensuring that the areas of enumeration (e.g., the region/province) are consistent, so that residents of different areas are not included in one data set and not the other.

Finally, it is good practice when reporting vital statistics to examine the plausibility of your results and provide context for the information by comparing your data to other sources.

2.5 USING INCOMPLETE DATA

Many datasets in the real world, particularly those from administrative data sets for CRVS, may not be a complete record for all events, or may be missing data. As a general rule, as long as we know how incomplete the data set is, and at least 70-80% of the events are captured, we can use the data collected to generate population based indicators such as fertility and mortality by adjusting the final results upwards

by the completeness of our records. However we must be careful, as this assumes that the under-reporting of events is general and not limited to particular sub-groups within the population.

If our data set is very biased or very incomplete, while we cannot directly use the data to generate population indicators, we may be able to make an assessment of what is happening in more complete sub-groups of the population. For example, if our cause of death data is not available at a population level, we may be able to focus on deaths in one hospital; or if births are better reported in one geographical area, we may be able to look at fertility rates in that area. It is important not to generalize too much from these results, but looking at sub-groups may help provide part of the picture of what is happening when we otherwise would have no direct information.

We will discuss assessing how complete a data set is and using incomplete data in more detail later in this workbook. If your data set is not either complete or nearly complete, you will need to refer to this material before calculating many of the indicators in the following chapters. The following chapters assume that your data set is complete or nearly complete and no adjustments for under-counting are included.

Writing exercise 1: Data sources and scope

This first writing exercise is to start to document the strengths and weakness of the data you will be using in the analysis this week, and what are data sources are available that will be used for comparison. Don't be too worried about getting it "just right" at this stage. The most important bit for now is getting the key points on paper so that you can refine it later. Think about starting each paragraph with a topic sentence and go from there.

Start with a paragraph that describes the data source. Include details such as which department collects the data, how is it collected and collated, and why the data is collected. Is there legislation or policy that requires births and deaths to be reported? You may want to think about adding a simple diagram of what the flow of data looks like (from the event such as the birth or death through to when it is in a national data set). What years are the data available for?

In your second paragraph, start to think about the strengths and weakness of your data set. Does the data collection process have any routine data quality checks built in, do all events in the country get included, what events might be missed out from your data. How do you know these things? Think about what events are counted – i.e. residents only, all events on island, and what happens with overseas events – such as deaths of people who are referred overseas for medical treatment.

Without delving into long descriptions, it is important to accurately note key strengths and weaknesses of your data set. This will help people to be able to interpret the analysis you are presenting and add credibility to your findings. You may need to later revise this paragraph against the findings from your national CRVS assessment.

Finally, in a separate paragraph, make a note of the other data sources in your country which you will use to compare to your results or which may provide additional context information. This would generally include sources such as the census and health related surveys.

Now go back and re-read your paragraphs. Do they make sense? Would someone who knows nothing about your country understand the key points of how the data is collected?

Swap your work with a colleague to read. Take a few minutes to highlight the key points they have made, then discuss what you have read. Make a note of anything that was missing or seemed confusing to them to review for their report.

2.6 REVIEW

- Bias is a systematic rather than stochastic (or random) effect on a statistic. Common types of bias include:
    - sample bias - where the population surveyed does not reflect the broader community (this is a common problem when using facility based data to infer information about the general population),
    - recall bias - where we are more likely to recall some events than others, and some key events in our lives will feel more recent than they really were,
    - non-responder bias – where people who refuse to answer specific questions in a survey may be different in some way (often having had an experience that is in some way related to the question being asked) than those who do not.
- Data sources may be population-based or facility based.
- When complete, vital registration data in conjunction with routine medical certification of death is considered the “gold standard” data source for vital statistics as it is continuous and representative of the whole population.
- It is good practice when reporting vital statistics to consider the effect that the data collection method may have on the results. In this regard, it is important to compare and contrast your findings to those from other data sources, in order to assess the plausibility of your findings.

PART 2: Data and Analysis

3 DATA QUALITY AND TABULATION

In this course (and as outlined throughout this workbook) we will be using routinely collected data from the CRVS system, and other routine vital registration collections (such as held by the Ministry of Health).

3.1 INDIVIDUAL RECORDS: VARIABLE FORMATS, DUPLICATES, AND MISSING DATA

Before we can use our data, we need to make sure that it is "clean" and is the most complete data set available.

In order to obtain the most complete data set on births and deaths for your country, it may be necessary to merge to different data sources such as civil registration and health data. If you are using a single data source, make sure that it is the most complete data as based on the data collection processes you described in the initial writing exercises. For example, if data is primarily collected through health registration and a notification from the health department is required to legally register a birth or a death in your system, it is likely that the health records may be more complete than the civil registration data.

Once we are assured that we are working with the best data set available, we need to clean the data before we can prepare our tables and begin our analysis. This involves checking the raw unit-record data to make sure that all:

- required data fields have been carried over into our working spreadsheet
- records have been carried over into our working spreadsheet
- duplicate records have been removed
- inappropriate records have been excluded (for example, still births have been removed from your sheet on live births)
- records use variables which are consistent and can therefore be readily aggregated (for example – if we are using M/F for sex, then all records should have one of these values in the field, rather than some having recorded as male, Male, or 1 etc.)
- missing values are minimized wherever possible (for example, if there is no age recorded for the mother in a birth record, but we have her date of birth, an age could be calculated and inputted into the records).

Many of these problems can be addressed at the data collection stage by setting up our database to force data entry to conform to a set of rules or standards. This is certainly what we would do in setting up a new survey. However, as we are analysing data from systems that already exist and may not be set up the way we would wish them to be if their sole purpose was the collection of data for statistical analysis, we must accept that there is likely to be a significant amount of data cleaning to be done before our analysis can

commence. This does not mean however that we should not work with our systems and departments to build these checks into our existing systems where possible.

Variables or fields required

For each record, the following fields should be available (as a minimum):

Births	**Deaths**
First name	First name
Surname	Surname
Date of Birth	Date of Birth
Sex	Sex
Place of Birth (Hospital, Village, Island)	Date of death
Place of Residence (Village, Province, Island)	Age (use separate fields for days, months, and years)
Mother's first name	Place of Death (Hospital, Village, Province, Island)
Mother's surname	Place of Residence (Village, Province, Island)
Mother's age	Spouse's first name
Live or still birth (or all live births)	Spouse's surname
Birth weight (optional)	Causes of death (by line of death certificate – 1 variable per line)
Length (optional)	Underlying cause of death (more on this in chapter 16)
Period of gestation in weeks (optional)	External cause (if applicable)
Ethnicity (optional)	Occupation (optional)
	Ethnicity (optional)

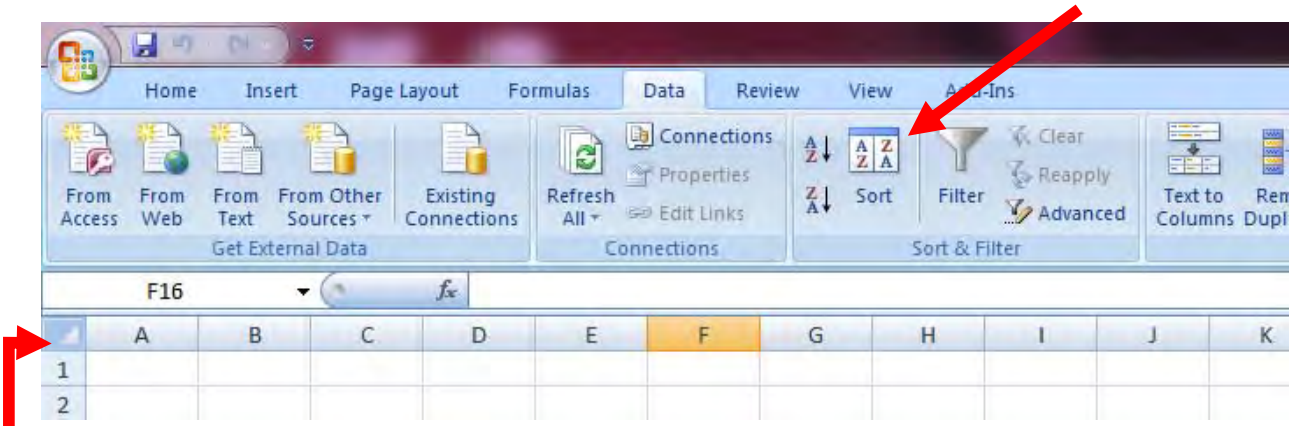
[United Nations, 2001]

3.2 INTRODUCTION TO EXCEL: DATA SORTING

For larger surveys and censuses, data cleaning would generally be done in the collection software such as CSPro. However, for the relatively small number of records we are working with, excel can be used to do a reasonable job of cleaning our data.

Sort function

The most important tool when using excel to clean the data is the **sort function** which appears under the **data tab** in Excel 2007 and Excel 2013. By clicking on the button marked, you can sort highlighted text by any of the fields in your data set.



Important points to remember when sorting data:

1/ Always set your data up with **one record per row and one field per column**, otherwise data will become separated and will not be able to be analysed.

2/ When selecting data to sort – it is good practice to always **select ALL data** by clicking on the arrow in the left upper corner (between the A and 1). Otherwise some columns may be sorted while data in others will stay in the same order, creating chaos in your records.

3/ Make sure there are **no blank columns or rows** which may interrupt the sort function

4/ **Label your fields** in the top row and make sure these are not repeated later in the data set. You can then sort by headings by ticking on the box that says "my data has header rows" in the sort dialogue box.

5/ **Save your data before** you sort it, so that you can come back if you make a mistake.

3.3 DERIVED VARIABLES

Once our individual records are clean, we can begin to set up our tables for analysis. However before we do this, we will also need to derive some additional variables to make our analysis easier.

Firstly, if you have not already done so, you will want to combine your data so that all years are in the same sheet. Make sure that before doing this you add a column for year.

To add a column in excel

Right click on the column after where you want this column to go and select "Insert column". Label this new column before adding any data.

Year can be derived from the date data already included in your spreadsheet by using the formula **=YEAR( )**

Insert the cell letter and number that corresponds to the date containing the year. Then copy this down to all the other records.

To copy cells

Place the cursor on the cell you wish to copy near the bottom right corner. When the open cross changes to a black cross, click to select the cell, and drag this down to the cells you want to copy the formula to before lifting your finger from the mouse key.

Secondly, you will need to insert an additional column for age groups to make your analysis easier.

Sort your data by age to make this easier. Then using 5 year age groupings for mother's age for your birth data (<15 years, 15-19, 20-24 etc.) and age groupings as below for deaths, fill this in by entering the required group at the start of each section and copying it down until you reach a new age group. If you find that excel wants to change the numbers as you copy the age group down, change the **cell type** under the general tab, from either "general" or "number" to **"text"**.

Age groups for collation of deaths

Neonatal deaths (under 28 days)
28 days to <1 year
1 to 4 years
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85+
(and unknown)

Data quality lab 1:

The first data quality lab will focus on tidying up our data set so that we can use it for further analysis. There is no point in tabulating our data and calculating indicators unless we are using the correct data. If your data is well managed, there should not be too much work in cleaning the data for analysis. If you do find lots of problems at this step, you should document what these are for discussion with your national CRVS committee when you go home.

Please make sure that you are NOT using the original copy of your data for this exercise. Open a new folder on your computer or flashdrive, label it with "Cleaned data" and the date, and copy each of the raw files on births and deaths that you will need into this folder. From here on, we will only work on these copies of the files.

For each of your data sets, go through the following steps:

- Make sure that there is one record per line and one line per record. If each record spreads out over more than 1 line you will need to re-format your data. There should be one field per column and one record per row.
- Similarly, if you have multiple fields in one column you will need to split these out to separate columns.
- If you have multiple data sheets or files for different years, you need to check that each file has the variables or fields in the same order so that you can later put these together.
- Within each year of reporting, check for duplicates by sorting the data, make sure all the data appears in the record that you will keep, and delete the duplicate. Make sure that the record is truly a duplicate (i.e. be careful of twins when cleaning birth records). Use the sort function in excel to sort by surname, first name, and date of event, then repeat this twice more with the sort fields in a different order (i.e. do it again with sort by first name, then date, then surname). It is important to sort using several criteria to check for duplicates, as data sets will frequently have two records for the same person with slightly different information. This frequently occurs for example when a child is initially registered as "baby of….." and is later re-recorded with their given first name.
- If you think there are duplicate records, but the information is not exactly the same in each record, create a set of rules as discussed in this section and document these before applying them to your data set. In very large data sets, this process would need to be automated and should be carried out in conjunction with someone who is trained in data processing and can help develop and run these programs. For our purposes, we will however be doing this manually.
- Make sure you have removed all records of stillbirths from your data set on deaths and births. Move these to a separate labelled spreadsheet. Terms to look out for in your data include "foetal death", "FDIU", "stillbirth", "macerated" amongst others. Review carefully the deaths for children aged less than one day.
- Check that date formats are consistent across all of your files (and within each file)
- If your data set does not have the YEAR as a field on its own, you will need to insert a column for this. It is important that this is done before you combine data across years where you are less likely to be able determine the correct year if this information is missing.

- Once you have completed these steps, if your data for each year is in separate files you will need to copy these into one single file. At the completion of this exercise, for each data source (or reconciled data set) you should have one file for births and one file for deaths. Make sure these files are clearly labelled (we suggest something along the lines of births_years_cleaned_datecleaned) as you will need to refer back to these files for all future analysis.
- Sort the records to determine how many of the records are missing core data fields, and document this information. Determine whether the missing data can be reliably obtained by extrapolating from other fields. For example, if the sex of the person is missing – is this readily identifiable from the first or given name? If the age is missing, can this be extrapolated from the date of birth? Complete the missing data fields where possible. It is suggested that you use a different colour text so that the corrected data is readily identifiable in your files.

Notes:

If you have software in the system from which you have extracted the data that allows identification and merging of duplicate records (and you have done this before extracting the data) you may be able to skip the initial couple of steps to identify potential duplicates. This will not be possible if your data was subsequently merged with another source.

If your CoD data is separate to the overall records of deaths, you will need to ensure that the file for CoD is cleaned in the same manner. Keep as much detail as you can, and remember to ensure there is a field for each line of the death certificate (if your data is coded, use separate fields for text and coded causes of death). We will come back to causes of death later in the course.

Finally, ensure that you record in your notes the steps that you took so that you can include these in your final report.

3.4 BASIC TABULATIONS

We can now go ahead and think about the basic tables that we will need for the remainder of our analysis.

For births these are:

- Total Number of births by year
- Total Births by year by sex
- Number of births by year, by age of mother (5 year age groups)
- Number of births by geographic sub-region (where relevant) (and potentially by sex and age of mother if there is sufficient data)
- Number of births by major ethnicity (where relevant)

For deaths this will be

- Total number of deaths
- Number of deaths by sex and age groups
- Number of deaths by sex, major ethnicity (where relevant), and age group
- Number of deaths by geographic region (where relevant) by sex and age group
- Number of neonatal deaths (deaths in infants aged 28 days or less)
- Number of deaths by age (for ages <1, 1-4, 5-9, 10-14.....65-59, 70-74, 75+ years), sex, and underlying cause of death (according to the ICDv10 103 cause – General Mortality list 1).

These will be generated as we work through the analytical chapters, by using pivot tables in excel.

PIVOT TABLES IN EXCEL

To create a pivot table, highlight the data sheet by clicking in the top left corner (between A and 1) and go to the insert tab. On the far left is a button that will allow you to create a pivot table from your selected data. **Always choose the option of creating the table on a new sheet so as not to confuse your source data.** You can then specify which variables to use as columns and rows to tabulate your data by moving them into the appropriate place. Use the count of function, and a variable which has no blanks to populate your table.

Once a table is set up the way you want, copy it and paste it into a new worksheet, as pivot tables cannot be locked. Make sure you label the worksheet (right click on the tab to re-name it) and table so that you know what they refer to when you come back to them later.

3.5 AGGREGATING DATA OVER TIME

Due to the small population size and level of disaggregation required to report events for specific age groups or causes, single year data for all indicators is likely to be unstable and is not recommended. Minimum aggregations should be over three years, although 5 years may be more suitable in smaller countries. As such annual reports are likely to be less useful than investing in reporting at a more detailed level every 3 or so years.

Depending on the data available, countries will need to determine whether they will use 3 or 5 year aggregation. It is preferable that the same time periods are used for births and for deaths for consistency.

3.6 RE-DISTRIBUTING DEATHS BY AGE

Where the data includes a substantial amount of deaths without details of age recorded, we will need to estimate at what age these deaths occurred. To do this, we will need to use the age distribution of those deaths for which age at death was distributed to estimate how many of our unknown age deaths should end up in each age group. As age patterns are different for males and females, the re-distribution of these deaths should be done separately by sex. Whether the re-distribution is done by year, or over an aggregated period will depend on the overall number of deaths, and the proportion for which no age is reported.

Start by setting up a table of deaths by age group by sex for the year(s) you are interested in as shown below. Calculate the percentage of deaths for which age is known that fall in each age group, and then multiply this percentage to the total number of deaths (including deaths of unknown age) to get the revised number of deaths by age. Round your results to the nearest whole person (after all – we don't get part of a person dying!) as shown.

Table 3.1: Re-distributing deaths of unknown ages.

Age	Total deaths		Percentage of total excluding unknown ages (%)		Re-distributed deaths by age	
	M	**F**	**M**	**F**	**M**	**F**
<1 year	14	12	3.8	3.9	15	13
1-4	6	4	1.6	1.3	7	4
5-9	2	1	0.5	0.3	2	1
10-14	1	4	0.3	1.3	1	4
15-19	5	6	1.4	2.0	5	6
20-24	9	13	2.5	4.3	10	14
25-29	16	12	4.4	3.9	17	13
30-34	23	12	6.3	3.9	25	13
35-39	25	14	6.8	4.6	27	15
40-44	22	15	6.0	4.9	24	16
45-49	26	22	7.1	7.2	28	24
50-54	35	26	9.6	8.5	38	28
55-59	38	28	10.4	9.2	41	30
60-64	48	32	13.1	10.5	52	35
65-69	58	44	15.8	14.4	63	47
70-74	36	36	9.8	11.8	39	39
75+	2	24	0.5	7.9	2	26
Unknown	33	24				
TOTAL	399	329	100.0	100.0	399	329

To get the highlighted result for males less than 1 year for example, we would do the following:

- Calculate the percent of deaths of known age that occur in this age group
  $= \frac{14 \text{ deaths}}{(399\text{-}33) \text{ deaths}} \times 100 = \frac{14 \text{ deaths}}{366 \text{ deaths}} \times 100 = 3.8\ \%$
- Apply this percentage to the total number of deaths
  $= \frac{3.8}{100} \times 399 = 15.16$
- Round the figure to the nearest whole person
  = 15 deaths

It is suggested that age groups lower than one year are not used in this general re-distribution due to the small numbers of events.

This method can also be used to re-distribute births by age of mother.

3.7 DATA QUALITY OF TABULATED DATA

Data quality is something that we should consider at every step of our tabulation and analysis. Once our unit record data is as good as it can be, and we have started to generate tabulated data for further analysis, we can start to make judgments about the quality and reliability of our tabulated data.

In general you should aim to identify the most important sources of error and provide quantitative measures where possible or qualitative descriptions otherwise. The result should be a balanced discussion which addresses itself to specific sources of error or bias and is therefore informative to readers.

Some **core concepts of data quality** which we should consider as we look at our data include:

Table 3.2: Core data quality concepts

Term	Meaning
Consistency	This is a description of the data over time, and whether it follows a similar pattern from year to year (month to month) etc.; or whether there are significant gaps or peaks in our data set. Best assessed from tabulations or graphs of the total number of events by year or month. It is good practice to do these for sub-regions as well to identify any reporting problems in the data. Also known as ***Comparability over time:*** it may be appropriate to discuss comparability with the results of the same activity for a previous reference period, especially if there has been a change in methodology, concepts or definitions. ***Benchmarking and revisions****:* the effects of benchmarking or revisions on comparability over time should be described (for example, with cause of death data, a change between ICD 9 and ICD 10 would be expected to have an impact on the data consistency when tabulated by cause). Guidance on the possible impact of future benchmarking should be given based on past experience. The very small populations and subsequently the small numbers of births and deaths can result in poor consistency of data over time due purely to **stochastic or random effects**. Data should be **aggregated** over several years before calculating measures of fertility or mortality to account for this.

Term	Meaning
Coverage (or scope)	This describes the area or population that the data set includes, noting any groups of events that may not be able to be recorded in the current system. For example coverage of a CRVS system may be national, or it may in practice exclude remote and rural areas, births or deaths overseas, or events related foreign nationals.
Representativeness	Related to coverage, this is how similar the population for which you are collected data reflects the broader population for which you want to use the results. This is very important in the use of demographic surveillance sites and for survey design. For example, if you interviewed only school teachers on their nutrition and eating habits, this well-educated, employed group of people would not necessarily be very representative of nutrition and eating habits at a national level. In a survey situation, we can examine this concept via ***Sampling error:*** if the survey is based on a random sample then estimates of the sampling error of tabulated data based on the sample should be provided, together with an explanation of how these standard error figures should be used to interpret the data.
Completeness	Completeness is used to assess what proportion of the events within our area of coverage that we intended to capture and that we actually managed to collect data for. For example, if our area of coverage is births in the national hospital, what proportion of the births in the hospital were recorded in our data set? It is general convention that data that is more than 80% complete[1] for CRVS may be able to be used for analysis without adjustment if it is also representative (although the completeness should be reported for context and each situation should be considered carefully). For surveys, completeness is generally reported as a ***Response rate.*** Any known differences in the characteristics of respondents and non-respondents should also be described (responder bias), but there should also be a careful consideration about whether the data is sufficiently complete to analyse based on what you are trying to find out. **Completeness** may also be used to refer to the completeness of key fields within the data set. For example, what proportion of births had the age of the mother recorded? The effect of ***editing and imputation*** on the quality of data should be assessed and described.

[1] Note that less complete data (say70-80% complete), may still be usable, but will need to be adjusted to account for under-reporting.

Term	Meaning
Validity	The plausibility of both our raw data (in terms of number of events) and of calculated measurements. For example, it is generally implausible (in most settings where there is no systematic processes to cause this) that the infant mortality rate would be substantially different for males and females. One way of assessing validity is through ***Comparability with other data sources:*** if similar data from other sources exist they should be identified. Where appropriate, a reconciliation should be attempted describing how the data sets differ and the reasons for these differences.
Reliability	Whether the system is able to produce results of a similar quality over time.
Bias	A systematic effect on a statistic or measurement rather than a stochastic or random one. Generally related to some aspect of the data collection which results in us being more likely to see particular answers in the data set over others.

(including material adapted from New Zealand Statistics – Data Quality Framework)

Each death should be registered once and once only, thus enabling universal coverage. Additional quality criteria are the amount of detail on the decedent including age, sex, date and place of death, cause of death (medically certified according to International Classification of Diseases (ICD) standard).

Off-island events

Deaths that occur off island (or in the case of Guam, New Caledonia, and possibly Fiji – deaths in patients that have been referred for treatment from other islands) can have a significant impact on distribution of causes of death and overall mortality rates, and subsequently bias results if not considered. For example, cancer patients may be referred overseas for treatment. If they subsequently die and are not counted in their home countries records (as a death certificate would be issued in the country where the death occurred) proportional mortality for cancer may be under-represented.

3.8 COMPLETENESS AND COVERAGE

What is completeness and why is it important?

Completeness assesses what proportion of the events (births and deaths) we intended to capture in our coverage area, we actually managed to collect data for. For example, if our area of coverage is births in the national hospital, what proportion of the births in the hospital were recorded in our data set?

We must take careful consideration about whether the data is sufficiently complete to analyse based on what you are trying to find out. Many datasets in the real world, particularly those from administrative data sets for civil registration, may not be a complete record for all events, or may be missing data. As a general rule, as long as we know how incomplete the data set is, and at least 70-80% of the events are captured, we can use the data collected to generate population based indicators such as fertility and mortality by adjusting the final results upwards by the completeness of our records. CRVS data that is more than 90% complete can generally be used for analysis without adjustment (although the completeness should be reported for context). However we must be careful, as this assumes that the under-reporting of events is general and not limited to particular sub-groups within the population.

If our data is not adjusted for completeness, we may make assumptions about fertility and mortality rates that are not true. For example, if we are not capturing all deaths, we may think mortality is lower than it is, which would result in a higher life expectancy. This kind of indicator can have impacts not just in the health arena, but for other sectors such as for social services (retirement funds). If our birth registration is not complete, the resulting fertility rates will be too low and we cannot accurately know how many babies are being born. This could impact the Department of Education's planning for schools, the Ministry of Health's planning for immunizations, and a variety of other social service departments that rely on this data.

Coverage vs. completeness

While they are closely related, it's important to not confuse coverage with completeness. Coverage (scope) describes the geographic area or population that the data set includes, noting any groups of events that may not be able to be recorded in the current system. For example, coverage of a CRVS system may be national, or it may in practice exclude remote and rural areas, births or deaths overseas, or events related foreign nationals.

If coverage is very low on remote outer islands or very rural areas, this means the dataset is not nationally representative, and is essentially not 100% complete and should be adjusted accordingly.

Assessing completeness

All data should be assessed for completeness before being used for analysis. Below are some methods to assess completeness of registration for both births and deaths, adopted from the UN ESCAP CRVS monitoring guidelines.

Completeness for birth registration

Measuring completeness of birth registration can be done in several ways. A basic approach involves dividing the number of births registered within the year(s) of occurrence in the country by the total estimated number of births for the same period, and then multiplying by 100 to give a percentage figure:

$$\frac{\text{Number of registered births}}{\text{Estimated number of live births}} \times 100$$

The number of registered births (numerator) would be sourced from the civil registration authority. National estimates on the number of births (denominator) can be developed based on estimates from the ministry of health, population census data or sample surveys. In the absence of national estimates, birth rates produced by the United Nations Population Division can be used for the denominator, as follows:

$$\frac{\text{Number of registered births}}{\text{Crude birth rates as estimated by your NSO or the UN (per 1000) x Total pop size (in '000s)}} \times 100$$

Crude birth rates will be discussed in more detail later in this workbook.

Completeness for death registration

Method of estimation/calculation

Measuring completeness of death registration can be done in several ways. A basic approach involves dividing the number of registered deaths in the country for a given year by the total estimated number of deaths for the same period, and then multiplying by 100 to give a percentage figure.

$$\frac{\text{Number of registered deaths}}{\text{Estimated number of deaths}} \times 100$$

The number of registered deaths (numerator) would be sourced from the registration authority. National estimates on the number of deaths (denominator) can be developed based on population census data or sample surveys.

In the absence of national estimates, death rates produced by your National Statistical Office or the United Nations Population Division can be used, as follows:

$$\frac{\text{Number of registered deaths}}{\text{Crude death rates (per 1000) x Total population size (in '000s)}} \times 100$$

There are various indirect demographic techniques for estimating the completeness of death registration; for example, the Bennett–Horiuchi, Sekar–Deming and Brass growth balance methods. These methods are not described here.

Example

Assume there are 450 registered deaths in Country X in 2014.

The census from 2013 estimated that in a population of 120,000 there was a crude death rate of 5 per 1000.

… which means $\frac{5 * 120{,}000}{1{,}000}$ = 600 deaths per year

The estimated completeness is therefore = 450/600

= 0.75
= 75 %

Issues and considerations

Disaggregation

> It is important to determine completeness by geographic area (outer islands vs main islands) and by sex. National averages can hide major variations in coverage. A main island/outer islands or urban/rural split should be the minimum level of sub-national disaggregation but further breakdowns may provide more meaningful information.

Completeness of death registration should be disaggregated by:

a) Sex – any differences in the tendency to register men more than women may point to gender issues that require targeted education or services.
b) Location – geographic location to the level of outer vs main islands. If further information is available, disaggregation can occur down to the island level.
c) Age – Are some age groups less likely to be registered than others? This is particularly important for young children as it will affect infant and under 5 mortality rates.

Checking completeness of death data by age using age-specific mortality rates

Example:

Country Z has vital statistics data from 2011-2013. They had a census in 2010 that provided age-specific mortality rates by age and sex. They also have population projections for 2011-2013. To assess the completeness of their vital registration data, they will use the age-specific mortality rates from the census and apply them to the population projections to determine the estimated number of deaths by age and sex.

Age specific mortality rates				Population Projections						Estimated Number of deaths					
Age group	ASMR 2010			2011		2012		2013		2011		2012		2013	
	Male	Female	Total	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-4	14.5	8.2	11.4	2601	2217	2697	2325	2760	2421	37.8	18.2	39.1	19.1	40.1	19.9
5-9	0.6	1.3	1.0	1713	1755	1812	1827	1953	1911	1.1	2.3	1.1	2.4	1.2	2.5
10-14	0.6	0.0	0.3	1635	1512	1644	1554	1638	1596	1.0	0.0	1.0	0.0	1.0	0.0
15-19	2.0	0.7	1.4	1431	1320	1449	1323	1491	1353	2.8	1.0	2.9	1.0	2.9	1.0
20-24	0.7	0.0	0.4	1548	1545	1524	1503	1485	1446	1.1	0.0	1.1	0.0	1.0	0.0
25-29	1.5	3.1	2.3	1470	1482	1497	1521	1518	1545	2.2	4.6	2.3	4.8	2.3	4.8
30-34	3.9	2.1	3.0	1245	1125	1290	1191	1338	1266	4.8	2.4	5.0	2.5	5.2	2.7
35-39	7.3	2.7	5.1	921	915	978	951	1035	984	6.7	2.4	7.2	2.5	7.6	2.6
40-44	8.2	18.0	12.8	744	732	753	741	777	771	6.1	13.1	6.2	13.3	6.4	13.8
45-49	21.4	9.6	14.9	714	774	720	768	714	750	15.3	7.4	15.4	7.3	15.3	7.2
50-54	35.5	19.0	26.7	528	624	558	657	591	690	18.7	11.9	19.8	12.5	21.0	13.1
55-59	32.5	33.7	33.1	399	441	420	471	435	501	13.0	14.9	13.6	15.9	14.1	16.9
60-64	81.5	12.1	47.5	198	267	228	297	264	324	16.1	3.2	18.6	3.6	21.5	3.9
65-69	102.4	58.7	80.6	69	99	84	123	99	150	7.1	5.8	8.6	7.2	10.1	8.8
70-74	298.4	116.6	192.5	51	54	36	48	33	51	15.2	6.3	10.7	5.6	9.8	5.9
75+	114.5	36.6	74.8	48	75	60	78	60	81	5.5	2.7	6.9	2.9	6.9	3.0
				15315	14937	15750	15378	16191	15840	154.5	96.2	159.5	100.5	166.5	106.1

For example, for males aged 0-4, the estimated number of deaths for 2011 is: $14.5 * \frac{2601}{1000} = 37.8$

For 2012 it's: $14.5 * \frac{2697}{1000} = 39.1$

For 2013 it's: $14.5 * \frac{2760}{1000} = 40.1$

And so on and so forth for each age group by sex. This exercise can easily be completed in excel.

When you perform this exercise, compare the results to your data. Are there any large differences? If you see large variation in one or more age groups between the two data sources, talk to your facilitators to determine the most appropriate course of action.

Checking completeness of death data by age using projected deaths by age

Sometimes a census will provide projected number of deaths by age and sex of the decedent or a percent distribution of deaths by age and sex. If this data is available to you, you should check to see how your registration data compare to the distribution of the census. Remember to average your data over your 3-5 year time period before making comparisons to the census data.

If you see large variation in one or more age groups between the two data sources, talk to your facilitators to determine the most appropriate course of action.

Data sources

National statistics offices for data on:

- total population and population by census enumeration area
- national mid-year estimates by age and sex
- estimates of deaths for the given year, if available
- crude death rates

Civil registration authorities for data on:

- deaths registered by location and demographic characteristics of the deceased (e.g. age, sex, place of occurrence, ethnicity, place of usual residence)

United Nations Population Division for:

- **crude death rates (in the absence of national estimates of deaths) – see** www.un.org/en/development/desa/population/

Adjusting incomplete data

Adjusting for incomplete birth registration data

In the example above for County X, the estimated completeness for birth registration was 83% at the national level. This exercise was to give us an idea of how complete the civil registration data was. If completeness is under 90%, the data will need to be adjusted. As you will be aggregating over 3-5 years, the adjustment of your data will involve a few more steps.

Adjusting for incomplete data using projected number of births

Yearly birth projections are often performed following a census. If your country has performed these estimations and they are believed to be reliable, you can use these to adjust your data for completeness. Example:

Let's say Country Y is analysing data from 2010-2012. They had 3,000 registered male births and 2,000 registered female births over this period. However, their national birth projections estimate there should have been 3,100 male births and 2,900 female births for a total of 6,000 births. This results in a completeness rate of 97% for males but only 69% for females, indicating there is a problem with female birth registration.

As the number of male births are very close to 100% complete, they probably do not warrant adjustment. However, Country Y will want to adjust the female births.

- The projected female births for 2010 was 960.
- The projected female births for 2011 was 965.
- The projected female births for 2012 was 975.

The total number of female births for 2010-2012 is now **2,900**.

2,900-2,000 = 900. 900 births will not have information on mothers' age and Country Y will need to address how to re-distribute their data.

This example is also applicable to birth projections for outer vs main islands or other geographic breakdowns.

Adjusting for incomplete data using the crude birth rate

If you do not have any information for projected number of births by sex or by area for each year your data covers, or you do not have the crude birth rate by area or sex, you will use the national crude birth rate to

adjust your data. Multiply each years' population estimate by the crude birth rate, divide by 1,000, and sum up the total number of births. This is your new total. Note that for the additional births you have added, you will not have information for them such as sex of the baby or age of the mother. You will need to use the methods discussed in the section above to perform a redistribution of your data for these important attributes.

Example:

Let's say Country X has data covering 2012-2014. Assume there are 8,900 registered births in Country X in from 2012-2014. However, we know from the previous exercise that the data is only about 83% complete. The only other birth information available for Country X is the crude birth rate from the 2013 census, which was a CBR of 30 per 1,000.

The population estimate for 2012 was 117,000

The 2013 census estimated a population of 120,000.
The population projection for 2014 is 125,000.

To adjust the data, we need to apply the crude birth rate to each year.

For 2012 we have: $\frac{30 * 117{,}000}{1{,}000}$ = 3,510 births for 2012

For 2013 we have: $\frac{30 * 117{,}000}{1{,}000}$ = 3,600 births for 2013

For 2014 we have: $\frac{30 * 125{,}000}{1{,}000}$ = 3,750 births for 2014

We sum up all the births for 2012-2014 and we have 3,500+3,600+3,750 = **10,860**.

This is our new number of births. This means 10,860 – 8,900 = 1,960 or 1,960 births will need to be assigned sex and mother's age using a redistribution process.

Note that if your census reported the crude birth rate by sex or by area, you would perform these calculations for each sex, or for each area. For sex, you would apply the CBR for males to each yearly population projection, and then the CRB for females to each yearly population projection your data covers. You would then add up the births for males and females for a new total.

Adjusting for incomplete death registration data

In the example above for County X, the estimated completeness for death registration was 75% at the national level. This exercise gave us an idea of how complete the civil registration data was. If completeness is under 90%, the data will need to be adjusted. As you will be aggregating over 3-5 years, the adjustment of your data will involve a few more steps.

Adjusting for incomplete data using projected number of deaths

Yearly death projections are often performed following a census. If your country has performed these estimations and they are believed to be reliable, you can use these to adjust your data for completeness.

Example:

Let's say Country Y is analysing data from 2010-2012. They had 1,000 registered male deaths and 500 registered female deaths over this period. However, their national death projections estimate there should have been 1,300 male deaths and 700 female deaths for a total of 2,000 deaths. This results in a completeness rate of 77% for males and 71% for females.

As the number of male and female deaths are below 90%, they will need to be adjusted. You can use the projected totals for each year to get the new number of male and female deaths: **1,300 male deaths and 700 female deaths for a total of 2,000 deaths**

2,000-1,500 = 500. 500 deaths will not have information on age or cause of death and Country Y will need to address how to re-distribute their data accordingly.

This example is also applicable to death projections for outer vs main islands or other geographic breakdowns.

Adjusting for incomplete data using the crude death rate

If you do not have any information for projected number of deaths by sex or by area for each year your data covers, or you do not have the crude death rate by area or sex, you will use the national crude death rate to adjust your data. Multiply each years' population estimate by the crude death rate, divide by 1,000, and sum up the total number of deaths. This is your new total. Note that for the additional deaths you have added, you will not have information for them such as sex of the baby or age of the mother. You will need to use the methods discussed in section [ ] to perform a redistribution of your data for these important attributes.

Example:

Let's say Country X has data covering 2012-2014. Assume there are 8,900 registered deaths in Country X in from 2012-2014. However, we know from the previous exercise that the data is only about 83% complete. The only other death information available for Country X is the crude death rate from the 2013 census, which was a CBR of 30 per 1,000.
The population estimate for 2012 was 117,000
The 2013 census estimated a population of 120,000.
The population projection for 2014 is 125,000.

To adjust the data, we need to apply the crude death rate to each year.

For 2012 we have: $\frac{30 * 117{,}000}{1{,}000}$ = 3,510 deaths for 2012

For 2013 we have: $\frac{30 * 117{,}000}{1{,}000}$ = 3,600 deaths for 2013

For 2014 we have: $\frac{30 * 125{,}000}{1{,}000}$ = 3,750 deaths for 2014

We sum up all the deaths for 2012-2014 and we have 3,500+3,600+3,750 = **10,860**.

This is our new number of deaths. This means 10,860 – 8,900 = 1,960 or 1,960 deaths will need to be assigned sex and mother's age using a redistribution process.

Note that if your census reported the crude death rate by sex or by area, you would perform these calculations for each sex, or for each area. For sex, you would apply the CBR for males to each yearly population projection, and then the CRB for females to each yearly population projection your data covers. You would then add up the deaths for males and females for a new total.

Other approaches to assessing completeness

More sophisticated approaches for measuring completeness of birth and death registration, such as 'capture-recapture' and the Brass method, are described in the *United Nations Principles and Recommendations for a Vital Statistics System*. However, these methods are complex to apply so would only be done on an occasional basis with the involvement of trained demographers.

Extract from: Carter K. 2013. Mortality and Causes of Death in the Pacific. PhD Thesis. UQ, Brisbane.

Methods to assess the completeness (or consistency between the death and population data) of the reported deaths (through vital registration, census or survey) include those developed by Brass, Preston, Bennet and Horiuchi (131, 199, 222-225). The Brass "Growth Balance Method" assumes population stability, so that growth rate is constant with age. It is based on the true growth of a population aged x and over is equal to partial birth rate (people who reach age x in a given year divided by the person years lived above age x in that year) minus the partial death rate (deaths in a year in persons aged x at last birthday and above divided by person years lived above age x in that year). Adapting this equation so the recorded number of deaths equals the completeness (C) of the data set multiplied by the true number of deaths provides a linear equation where the Partial Birth rate = growth (r) + 1/C * (partial death rate). The slope of the line formed by this equation is 1/C and can be used to both assess the data and correct for under-reporting. This method is very sensitive to error in the data which changes the slope (and therefore the estimation of completeness of the data source) and to changes in the assumption of stability (225, 226). The selection of age groups to "Crop" in order to fit the line may also have a significant impact on the estimate of completeness (124).

Modifications include the Preston-Coale method, which uses growth rates instead of the age distribution of deaths used by Brass, and the Bennet and Horiuchi method which discards the assumption of stability (225). These methods require inter-censal data which is often not available (227).

A second approach for assessing completeness is direct comparison with another source or "capture-recapture" techniques (127). These methods, adapted originally from population biology, compare the overlap between the different sources to estimate the number of deaths missed by all sources (228-230) (see Figure 7). Two source, or dual record methods such as described by Sekar and Deming (231), also known as the "Peterson method" (232), assume independence between the two sources (228). That is, reporting the death to one source should not affect the probability that the death will be reported to the other source, a scenario unlikely to hold true in most settings. Log-linear methods are also possible for three or more sources. Although the three source model is more robust (233), it is uncommon (especially in the PICTs) to have access to multiple data sets.

Figure 3.1: The two source capture-recapture model

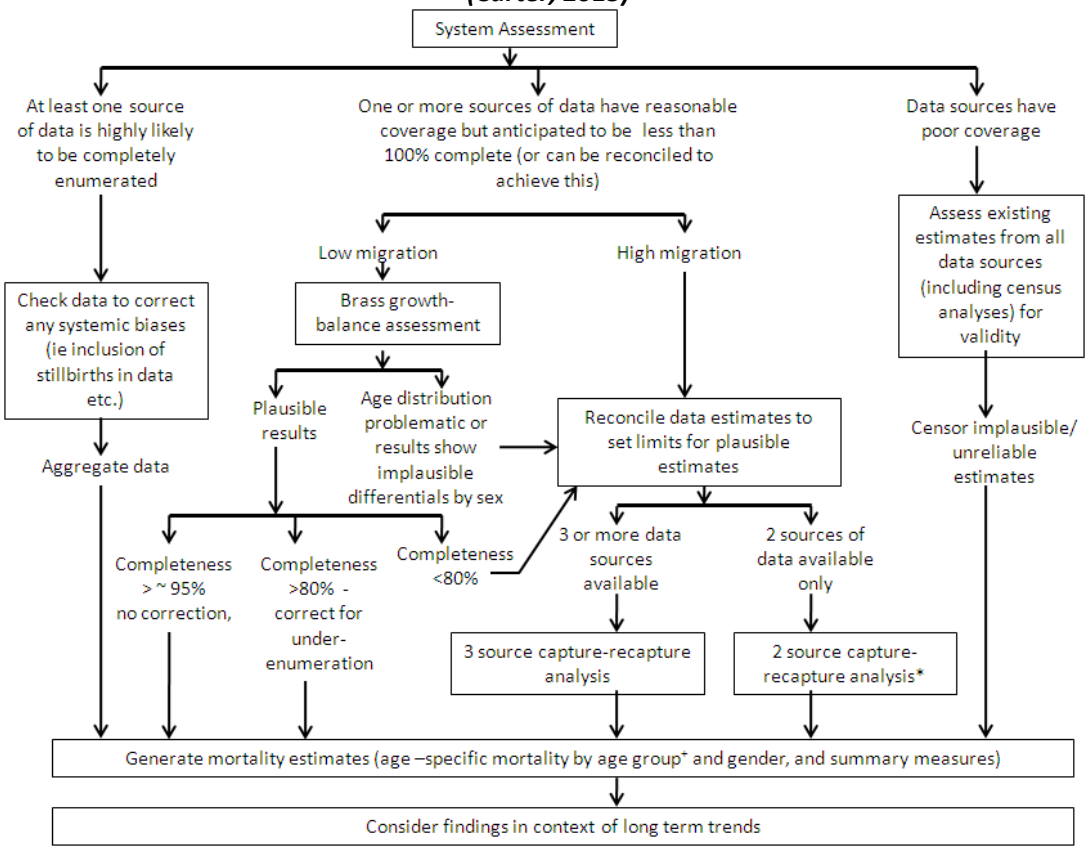
		Source 1	
		YES	NO
Source 2	YES	a	b
	NO	c	X

- The missing number X = bc/a
- Total = a+b+c+X

(229, 230).

Other assumptions of the capture-recapture method include a closed population and equal "catchability" of deaths. That is, every death has an equal probability of being reported. While this is unlikely across an entire population given the incentives for reporting a death for an adult (who has property etc.) are likely to be greater than for a child, the advantage of the capture-recapture method is that analysis can be done by sub-group (such as by age group) where these assumptions are more likely to hold.

Figure 3.2 : Recommended analysis approach for assessment of mortality level in the Pacific Islands (Carter, 2013)



3.9 THE ABS DATA QUALITY FRAMEWORK

The ABS data quality framework is a useful method of examining outputs from a CRVS system and how these affect the data quality, and subsequently the ability to use the collected data for policy and planning decisions.

Figure 3.3: The ABS data quality framework

https://www.nss.gov.au/dataquality/aboutqualityframework.jsp

Data quality lab 2

Before we get into the analysis in more detail, we will conduct some basic data quality checks as outlined below.

Create pivot tables of the number of events by year. Look at the results to check that the number of events is not too widely different year to year, and that no years are missing from your data set. If one year has a very low number of events for example, you may be missing data and will need to check with the original source. Have a look to see if the data is essentially consistent over time. Create a new worksheet for each pivot table (as outlined above) so that you can refer back to you workings.

Repeat this process by creating a pivot table by region (or other geographical measure such as hospital/ province/ island etc.) by year. Look for consistency in the data over time. Are there any years that look very different to those either side? Are there any locations where there were no events reported in a given year? Is this plausible? Consider your data and if you think there might be data missing (or mis-recorded), go back to the original data source or contact your data collection unit to discuss this further.

These checks will help identify any broad errors or gaps in the data set. However we will continue to review data quality as we work through the basic tabulations for births, deaths and causes of death in the following sections.

Writing exercise 2: Describe the data quality

Repeat the process you followed in the first writing exercise to write a couple of short paragraphs on the following:

1/ The process that was used to review the data quality and subsequently clean the data

2/ The data quality, such as the proportion of missing values by data set, the number of duplicates etc.

3/ The major corrections that were made, such as re-distributing data with unknown values, removal of duplicates, data matching etc.

You will come back and review this section of your draft report once the analysis is completed.

Again, share your draft with a colleague.

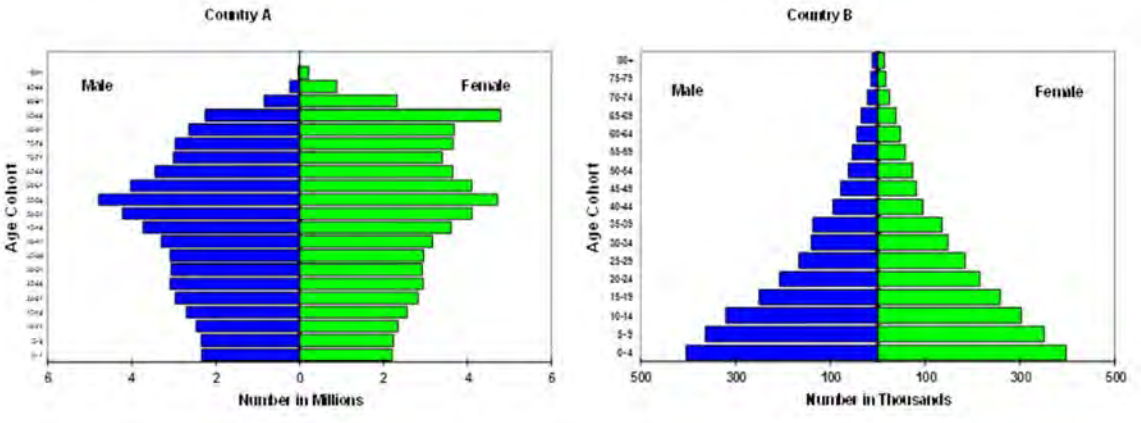
4 POPULATION DATA

Population data and vital statistics

Population data is a critical component in calculating vital statistics. As you will see in the following sections, population data is the denominator for calculating many rates such as the crude death rate (CDR) and crude birth rate (CBR). These rates are a starting point in examining the health of a population.

Analysing vital statistics may not always be intuitive. For example, a large number of reported deaths may not necessarily indicate poor health in a population, but rather that the population is large. China is likely to have many more deaths in a year then Burundi, but that does not mean that China's population is worse off mortality-wise than that of Burundi. The large number of deaths in China come from China's large population numbers. By dividing the number of deaths (aggregated over several years in our case) by the population (using the population midpoint), we can arrive at a better understanding of mortality in a given population. This calculation provides us with the crude death rate, which we will discuss in more detail in a later chapter. However, while the crude death rate may give you a general understanding of the risk of dying in a given population, comparisons of CDRs may be misleading. It is possible for two countries, where mortality is the same among all age groups, to have two different CDRs. This is because the CDR is heavily influenced by the age structure of a population. Assuming the same mortality at each age group, if Country A has an older population compared to Country B, Country A's CDR will be higher. The same concept is true for births. If a large proportion of the population is women of childbearing ages, crude birth rates can be increased when all other factors are held constant. This is why we use age-standardized rates for comparison purposes. These will be discussed in more detail in a later chapter.

Figure 4.1 Comparing populations using Population Pyramids



Creating a population pyramid

It is useful for policy planners to know the age and sex distribution of a population in order to plan for the future. For example, they might want to know the number of children entering school in 5 years' time, the number of young men that could be conscripted into the military next year, or the number of middle aged people who will likely be retiring in 10 years. A population pyramid is a very helpful way of graphically

representing the age and sex distribution of a population. See the graphics above for examples of a population pyramid for Country A and Country B. When population pyramids look like pyramids, as is the case in Country B, this indicates that the population is very young. A large proportion of the population is concentrated in the base of the pyramid. When we see a more top-heavy pyramid, as is the case for Country A, this indicates that the population is aging and that a larger proportion of people are in the middle and older age groups.

Population pyramids are usually constructed by examining the male and female populations by 5-year age groups. The data used in population pyramids can be the actual counts of males and females in these age groups (as is the case in the example above), or percent distributions. Percent distributions can either be the percent of the total population those in that sex and age group represent, or the percent of the male (or female) population that that age group represents. For example, if we examine males aged 15-19, they may represent 6 percent of the total population or 11 percent of the male population.

We can use population pyramids to get an understanding of how are population is distributed by age and sex and how this might affect any resulting vital statics rates.

Population midpoints

As discussed in the previous chapter, due to small population sizes and the relatively small number of occurrences, vital events will need to be aggregated over 3-5 year periods. In order to accurately calculate rates with aggregated data, the midpoint population should be used in the denominator. For example, if you have data on the number of deaths for the years 2008-2012, the midpoint would be July 1, 2010. The population on July 1, 2010 would be used in the denominator as this is exactly halfway between 2008-2012.

When your birth and death data cover the full calendar year from January-December, it's important to use to July 1st as the population midpoint. This is due to the fact that when a population is growing, using census data that is later than mid-year will bias the CDR downwards because the denominator will be too large (the population is too large because more births will have occurred since midyear). Using population data earlier in the year will bias the CDR upwards because the denominator will be too small (the population is too small because not enough births have occurred yet). Likewise, it's important to get the population midpoint correct when aggregating data over a number of years in order to avoid bias.

Some rates use population totals (such as the CDR or CBR), but some rates require male and female populations by 5 year age groups. For the purposes of this course, you will need midpoint populations broken down by 5 year age groups and sex that correspond to the midpoints of your aggregated birth and death data. If your national statistical office has already done yearly population estimates for the years you need, you should use this data. Alternatively, we can use excel spreadsheets to interpolate between censuses to arrive at the midpoint populations needed. In order to perform the interpolation exercise, you will need data from 3 censuses broken down by 5 year age group and sex.

Population Lab 1 – Creating a population pyramid

What you will need: Population data by sex and 5-year age group.

Carefully enter your data into PYRAMID.xls or PYRAMID_SPC.xls. To compare populations at different points in time enter your data for the 2 time periods into Pyr2.xls.

1) What does the population pyramid tell you about your data? Is your population young, old etc.?
2) Is it the outcome you expected? Are there any anomalies in your data?
3) What does the pyramid tell you about recent fertility?
4) How will these results influence your population indicators such as crude birth and death rates?
5) What are your results from Pyr2.xls? How does your population change over time?

Population Lab 2 – Interpolating census data to get midpoint populations

What you will need: Population data from 2 censuses broken down by sex and 5-year age groups.

Carefully enter your data into AGEINT.xls. You will notice that there is only 1 column for each census - we will start by interpolating total population data by age and sex.

- Make sure to enter the year of your census data in cell B11 for your earlier census, and C11 for your later census. Enter the Month and Day of the censuses in the cells below.
- Enter the year of your midpoint in cell D11. Use '7' for month and '1' for day for your midpoint.
- Make sure cell C8 ='1' – this is so we can perform an exponential interpolation.
- Enter your data by age group in columns B and C.
- Column D will automatically show you the population at the time of your midpoint.

Now perform this exercise separately for males and females. You will need to save 3 spreadsheets total in the end- one for your total population, one for your male population, and one for your female population.

Writing exercise 3: Population data

As described in this section, in order to generate population-based rates and proportions from your birth and death data, you need to have population data for the mid-point of each of the time periods which you are analysing.

As it is unlikely that you are lucky enough to have census data that exactly matches that moment in time, you will have had to interpolate or extrapolate from existing census data to obtain the population by age and sex that you are interested in.

If you have had to do population interpolations specifically for this analysis, you will need to describe how these were done. Write a paragraph here that describes the source of the data used, census dates, the age groups, and how you interpolated or extrapolated the data to the time period required (US Census Bureau spreadsheet AGEINT.xls).

5 BIRTHS AND FERTILITY DATA

5.1 BIRTHS

As noted in earlier chapters, births are important to us for many reasons - both as data in their own right, and as denominators for later mortality calculations such as the infant mortality rate (IMR).

Before we begin using births to calculate rates, we need to define what we mean by a 'birth.' Below are some definitions to consider when reviewing your data:

Live birth: The WHO defines a live birth as the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of the pregnancy, which, after such separation, breathes or shows any other evidence of life - e.g. beating of the heart, pulsation of the umbilical cord or definite movement of voluntary muscles - whether or not the umbilical cord has been cut or the placenta is attached. Each product of such a birth is considered live born.

In more practical terms, a **live birth** is a birth where a newborn, regardless of the length of his or her gestation, is born and demonstrates any sign of life including a heartbeat, taking a breath, movement of voluntary muscles, or umbilical pulsation.
Still births are not considered live births and should not be included in birth data for calculating vital statistical rates.

A **stillborn** baby is a baby born after the 24th week of pregnancy who does not show any signs of life. If the baby dies in the womb, it is known as an intra-uterine stillbirth. If the baby dies during labour, it is called an intra-partum stillbirth. If the baby dies before 24 weeks, it is known as a **miscarriage**.

Uses and limitations of birth data

Births are expressed as numbers, which are necessary for many applications such as determining absolute population growth and planning for school enrollments and immunizations. However, the number of births occurring in a country over a given period of time does not tell us anything about the fertility of the average woman in that country. As discussed in the previous section, a large number of births may simply reflect a large population of mothers.

5.2 BIRTHS PER YEAR BY SEX

We will start by tabulating total births by year, and also births per year by sex of the baby to fill in the table below.

Table 5.1: Total Number of Births by Sex of the Baby per Year, {Years}

Year	Female	Male	Total

Make sure your unit record data is set up with one record per row as discussed in previous chapters, and that you have sensible column headings that accurately describe your data (such as year the birth occurred, age of the mother at the time of birth, sex of the baby, resident status of the mother etc.).

Pull the field for year of the birth into the row labels and the field for sex of the baby into the column labels and the count field into the values box. If you have date data for year of birth by day/month/year, you will need to make sure this is set up to be in the short date format. You can then group this field by year so that single years are displayed in the rows. You should end up with something that looks like the table below.

Table 5.2: Births per year by sex of the baby

Row Labels	Females	Males	Grand Total
2008	2	2	4
2009	2	1	3
2010	2	3	5
2011	3	1	4
2012	2	2	4
Grand Total	**11**	**9**	**20**

Review your data to make sure that there are no missing years. If your data is missing any values they are often shown as "(blank)" as you can see in the table below. Check that all of your years and counts make sense. In the table below, it appears there is a typo where they year 3010 was entered instead of 2010.

You should also check how your numbers are changing over time. Are the numbers what you would expect? Do they make sense? For example, in Table B we see a huge jump in births in 2011. Can we explain these births or is there a problem with our data? You will need to go back and edit any records that need to be fixed, such as erroneous years.

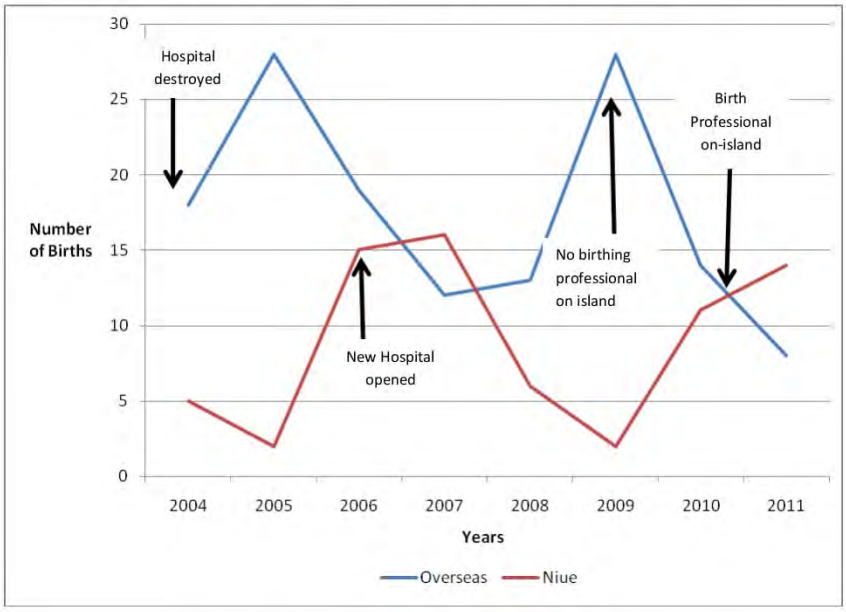
Table 5.3: Births per year by sex of the baby

Row Labels	Females	Males	Grand Total
(blank)		1	1
2008	2	2	4
2009	2		2
2010	1	3	4
2011	22	1	23
2012	2	2	4
3010	1		1
Grand Total	**30**	**9**	**39**

Stochastic variation

When examining vital events in small populations, the numbers of births and deaths are likely to "jump around" a lot over time. An example of this is displayed below for births to Niuean citizens occurring in Niue or overseas for the years 2004-2011 (Figure 1.6 – Niue Vital Statistics Report 1987-2011)[2].

Figure 1.6: Total Births by Place of Birth (Niue or Overseas): 2004 – 2011

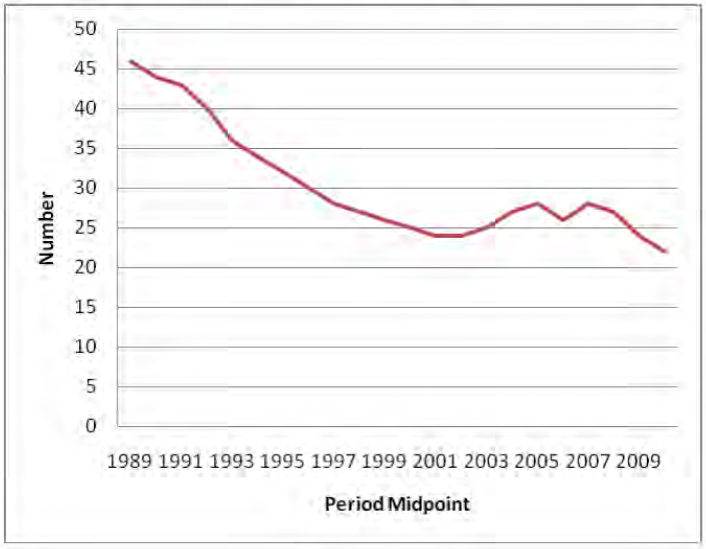


You can see that there are large variations between years. To smooth out some of this noise and to see more clearly what is going on over time, we can calculate the average over a given period, or we can calculate rolling averages.

Below is an example of the rolling average number of births per year in Niue. You can see how this smoothed out the data a bit more, but the story the data was telling in relation to hospital and birthing care available on the island was not lost. We can also see that, in general, the number of Niuean births is decreasing with time.

[2] Available at: http://www.spc.int/prism/

Figure 1.1: 5 Year Rolling Average for Total Number of Births in Niue: 1987 – 2011



Now you will calculate the average number of births over a given time period for your data. What years of data do you have available? What aggregation of time should you use? If possible, we recommend you use 5 year time periods. However, keep in mind you will need to use the same time periods for your death data. If you don't have birth and death data going back in time far enough to use 5 year time periods, aggregate over 4 or 3 year time periods and find the average number of births over this time period.

To calculate the average number of births over 5 years, sum up all the births during that time and divide by 5. For example, if we had 20 birth occurring over 5 years (2008-2012), we divide 20 by 5 to get 4 births per year. We will then add up all the female births and divide by 5, and then do the same for the males births to complete the table below for each of our time periods.

Table 5.4: Average Number of births per year by sex, 2008-2012

Period (years)	Female	Male	Total
2008-2012	2.2	1.8	4.0
2003-2007	2.5	2.2	4.7
1998-2002	2.4	2.6	5.0

Using a rolling average

Using a rolling average (also known as a moving average) can help smooth out some of the noise in our data so that any given point that may be unusually high or low does not distort overall trends over time. We can average data points over 5 years and graph the midpoints of our data to see how things are changing. When we have enough data to support this method, using a rolling average is preferable to having just one average for your 5 year period when dealing with small populations. This method is especially salient when policy makers request annual figures to understand changes in births or deaths over time.

We will again use the Niue Vital Statistics Report (Exert from Appendix 1 below) as an example of how to calculate a rolling average. Niue had birth data for the years 1987- 2011. They started with the average number of births from 1987-1991, which was 46 births. The midpoint of 1987-1991 is 1989 so as you can see in the table below, 46 births were entered for 1989. Next they shifted up one year and calculated the average number of births for the period of 1988-1992. The average was 44 and the midpoint year was 1990 so they entered 44 births for 1990. The next period would be 1989-1993 with a midpoint year of 1991 and so on and so forth. The last period Niue had data for was 2007-2011. The midpoint of that period was 2009, thus that is the last year they were able to report a 5 year rolling average for as you can see in the graph above and the table below. Please note that numbers were rounded down to the nearest whole integer.

Table 1.1: Total Number of Births by Sex per year: 1987 - 2011

YEAR	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
F	25	19	26	23	21	17	20	15	15	14	16	16	11	15	13	16	17	10	14	22	15	7	17	9	14
M	25	25	23	27	16	26	18	18	15	14	17	14	11	10	10	9	9	13	16	12	13	12	13	16	8
TOTAL	50	44	49	50	37	43	38	33	30	28	33	30	22	25	23	25	26	23	30	34	28	19	30	25	22

APPENDIX 1: STATISTICAL TABLES

1.1 BIRTHS

5 Year period of Rolling Average of Number of Births: 1987 – 2011

YR	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
F	20	19	18	16	15	15	14	14	14	14	14	14	14	14	14	14	14	13	13	12	11
M	23	23	22	21	18	18	16	15	14	13	12	10	9	10	11	11	12	13	13	13	12
TOT	46	44	43	40	36	34	32	30	28	27	26	25	24	24	25	27	28	26	28	27	24

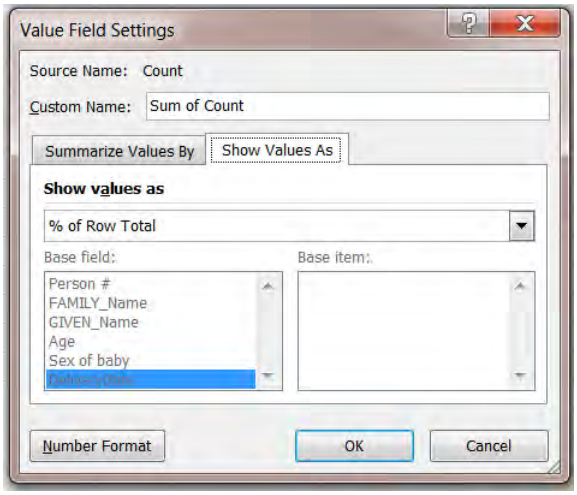
If you have enough data, calculate a 5 year (or a 3 or 4 year) rolling average for your births. Complete a table similar to the one above and then make a graph of your trends over time.

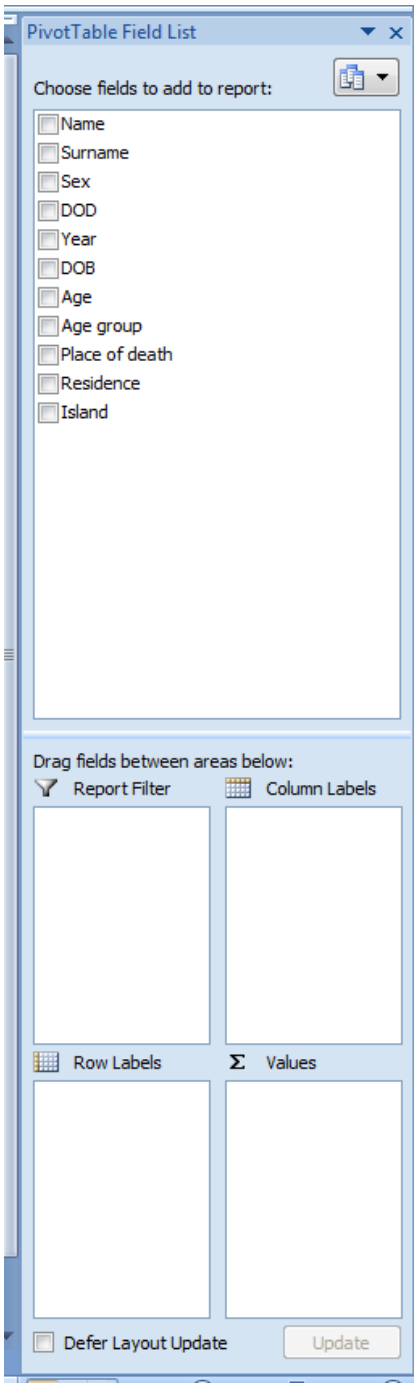
5.3 BIRTHS BY AGE OF MOTHER

Next we will examine the number of births occurring by age of mothers. This statistic is of interest to policy makers because when women are having their children at younger ages, the population grows faster because there is less time between generations. The children that younger women are having now are able to have their own children sooner than the cohort of children born to those mothers when they are older. For example, if a woman has her first child at age 20, that child will be able to have children and contribute to population growth in 15 years (if we use age 15 as the beginning of childbearing years). However, if that mother waits until she is 30 to have her first child, an additional 10 years is added to the time her first child will have children and contribute to population growth. Additionally, very young or older mothers are more likely to have complicated pregnancies and risks associated with the infant's health and may require more health monitoring.

Women of childbearing age are generally considered to be between the ages of 15 to 49. Although less common, births do occur to mothers younger than 15 or older than 49 and it is important we capture these events, especially if greater service provisions are needed for these high risk pregnancies.

Copy the pivot table you just constructed for births to a new tab and remove the sex of the baby field from the column headings. Add the field for age of the mother as the new column label. Group this field into 5 year age groups from ages 15-49, but make sure you also include a column for mothers younger than age 15 or older than 49. In the field settings, click on the 'show values as' tab and select '% of Row Total' from the dropdown list as shown below.





You should be able to create a table similar to the example below. We will examine this table again to calculate age-specific fertility in the next chapter so don't forget to save your work.

Table 5.5: Percent Distribution of births by Age of mother, {Years}

Periods	Age Groups									Total
	<15	15 - 19	20 -24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50+	
2008-2012	0	40	20	20	0	20	0	0	0	100
2003-2007	25	0	25	0	0	25	0	25	0	100
1998-2002	0	50	0	0	25	25	0	0	0	100

5.4 CRUDE BIRTH RATE (CBR)

The crude birth rate (CBR) is the number of births per 1,000 population over a given period of time. Crude birth rates are important because they tell us how much our population is growing assuming mortality and migration are equal to zero. They can also help us plan for the future to determine how many children will be entering school in the coming years, or how many adults will be entering the workforce in the future.

Generally, the crude birth rate is calculated as the number of births occurring in a year divided by the population at midyear, times 1,000. However, given the small population sizes of the Pacific Islands Countries and Territories, we will aggregate our data over 3-5 years and calculate CBRs for periods of time. If we aggregate over 5 years, we will divide the average number of births over this period by the population size at the midpoint of these 5 years.

*CBR of 5 year period = 1000 * (Average number of births over 5 years/Midpoint population)*

Example:
If our average number of births from 2003-2007 is 325 births, we will need to divide this by our midpoint population. Our midpoint is July 1, 2005. Let's assume the population was 15,645 on July 1, 2005. We then perform the calculation:

1000 * (325/15,645) to get a CBR of 20.8 so we can say there were 20.8 births per 1,000 population in 2003-2007.

The CBR is one of the easiest measures of fertility to understand, however, it has some limitations. It can be somewhat misleading due to the composition of its denominator. The midpoint population used in the denominator includes children, men, and women outside of childbearing ages. Therefore, the CBR is not only effected by the number of births, but also by the proportion of the population who cannot have children. A very young or very old population or one that has many more males than females would affect the CBR. Even when the frequency of having children among women of reproductive ages is the same in two countries, each country may have different crude birth rates. You can see this in the example below.

Assume the proportion of the population that is comprised of women of childbearing for Country A is 0.30 and for Country B it is 0.15. (That is, a larger proportion of the population of Country A is women of childbearing age compared to Country B.) For the sake of easy calculation, let's assume all of these women had one child each in the year we are performing our calculation. That means in County A there were 3,000 births and in Country B there were 15,000 births. One might think Country B would have a higher CBR because they had many more births, but when we divide by the midpoint populations, we see that Country

A actually has twice the CBR of Country B. The CBR is larger in populations where women of reproductive age comprise a larger proportion of its people.

Comparison of Crude Birth Rates by proportions of the population that are women of childbearing age for 2 fictitious countries

Country	Country A	Country B
Midpoint Population	10,000	100,000
Prop. that are women 15-49	0.30	0.15
Total women aged 15-49	3,000	15,000
Number of births	3,000	15,000
CBR	300	150

It's worth noting:

Two populations may have different crude birth rates even if fertility of women at each age is the same.

It is possible for one country to have a higher crude birth rate than another country even though fertility could be lower at each age.

You can calculate the crude birth rate from the tabulations you have already performed. Fill in the table below, indicating which years you aggregated over. You will need to make sure you have appropriate midpoint populations. Perform the interpolation exercises from Chapter 8 if you need to adjust your population data.

You can also calculate rolling averages for your CBRs and plot them on a line graph.

Table 5.6 : Crude Birth Rate with 95% Confidence Intervals by {#} year period, {Years}

Period	Crude Birth Rate
	{Rate (CI range)}

What does the data you just calculated tell you? Do you see any trends in your CBRs? If your CBR is increasing or decreasing, is this because the number of births is changing substantially (what does your data from the previous section tell you?), or is it because of changes in the population size – i.e. out-migration? Write a few short paragraphs summarizing your findings.

5.5 SEX RATIO AT BIRTH

The **sex ratio** is the ratio of males to females in a given population, expressed as the number of males for every 100 females. The **sex ratio at birth** is the ratio of male to female babies born. We also express the sex ratio at birth as the number of male babies born for every 100 female babies born. Sex ratios over 100

indicate that there are more males than females, and sex ratios under 100 indicate more females than males.

$$\frac{\text{Number of male births} \times 100}{\text{Number of female births}}$$

Example*:*

In a population where 25,643 male babies are born and 23,721 female babies are born, the sex ratio at birth is calculated as:

$$\frac{25{,}463}{23{,}721} \times 100 = 107.3$$

This means there are 107 (rounded) males per 100 females.

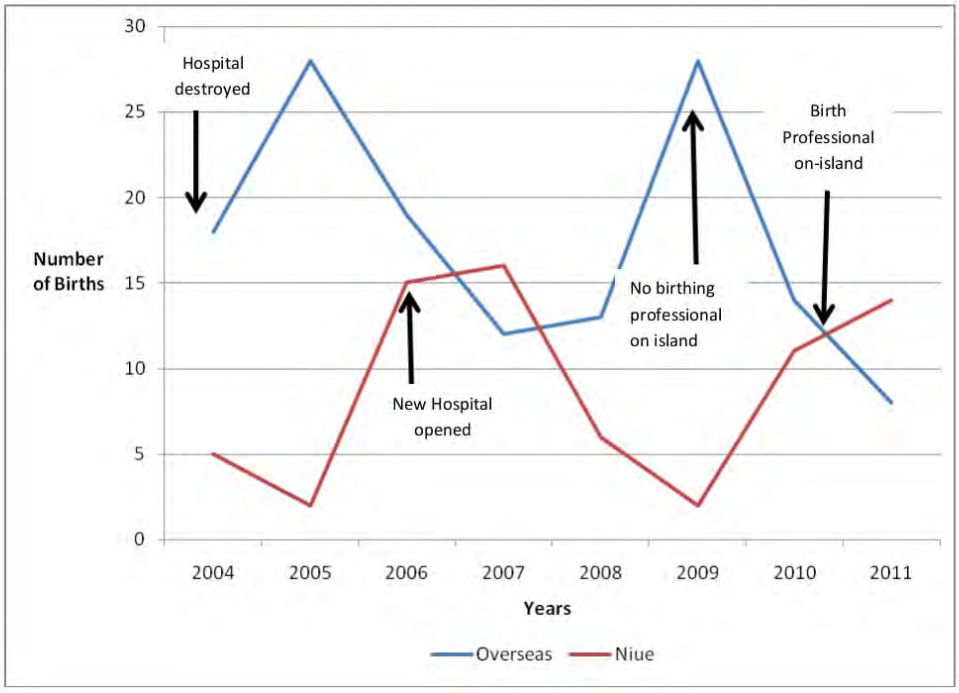
The normal sex ratio at birth for human babies is about 103-106. The slightly higher number of male births is believed to be due to factors influencing conception and intra-uterine mortality. The sex ratio can vary naturally between different countries, cultures, and geographic locations. However, in recent years, several factors have increased the sex ratio at birth, particularly in countries like China and India. The preference to have at least one son, and decreasing fertility (and thus fewer "chances" to have a son), have led to an increase in prenatal sex selections. Technology has abetted this phenomenon, allowing parents to determine the sex of their baby from an early age and selectively abort female foetuses. Aside from the ethical issues of sex selection, sociological problems can result from heavily imbalanced sex ratios at birth. When these babies grow up, there will be a shortage of eligible women (compared to men) for marriage, which can affect future fertility. We use the sex ratio at birth as a means to assess if prenatal sex selection is occurring in a given population.

When we have very small numbers of births, we should aggregate our data over at least 5-10 years to determine the sex ratio at birth. In the case of Niue, we can see that births were aggregated for the years 1987-2011. Perform these calculations and write a few sentences about what this means in your country. For example, you might say: The sex ratio at birth is {insert result}. This means that for every 100 live female births, there were {insert} live male births over the same time period. {Elaborate in context of country.}

5.6 PLACE OF BIRTH

Many births to Pacific Islanders occur overseas. Does your data source capture births that occur overseas? If so, how do these compare to the number of births occurring on the island? Are there factors that are influencing mothers' decisions on where to give birth from year to year, such as shown below for Niue?

Figure 1.6: Total Births by Place of Birth (Niue or Overseas): 2004 – 2011



Create another pivot table with years in your row heading and country of birth in your column heading. Use this data to make a graph similar to the one shown for Niue.

Additionally, some countries may be interested in where on-island births are occurring. If you have this information, create a second pivot table by location of birth by year. Locations might include hospitals, clinics, other health care centres, home births, or other places. You may want to show this data as a percent distributions, in which case you would change the field settings. Click on the 'show values as' tab and select '% of Row Total' from the dropdown list as you did for births by age of mother. You can use these results to make a pie chart to show where births are occurring in your country.

5.7 Other Risk Factors: Low birth weight, Prematurity & Antenatal Coverage

Some women are at higher risk of pregnancy and delivery complications which can also pose a risk to the infant's health. This is especially true for very young or for older mothers. To examine how many higher risk babies are being born, we might look at the percentage of babies with low birth weights, or the percentage of deliveries that occurred before the 37th week of pregnancy when gestation is not yet considered complete.

Examine your data set and determine if your country collects these data. If you do not have this information, is there another source that may? If this data is not being collected from any source in your country, is this something you might want to discuss with your CRVS or HIS committee?

Create a new pivot table to determine the percent distribution of births that are low-weight births, i.e. under 2500 grams (5.5 lbs). You will need to group your babies' weights into 2 categories: babies weighing under 2500 grams and those weighing 2500 grams or more. Don't forget to aggregate your data over 3 to

5 year intervals. Determine what percentage of babies were born with low birth weights and create a table similar to the one below.

Table 5.7: Percent Distribution of births by birth weight category, {Years}

Period	Birth weight		Total
	<2500g	2500g or higher	
			100
			100
			100

Create a new pivot table to determine the percent distribution of births that occur before 37 weeks gestation. You will need to group gestation age into 2 categories: babies born at less than 37 weeks of pregnancy and those born at 37 weeks or more. Perhaps you already have a field in your dataset that identifies if the child was born at preterm and you will not have to make categories by gestational age. Don't forget to aggregate your data over 3 to 5 year intervals. Determine what percentage of babies were born preterm and create a table similar to the one below.

Table 5.8: Percent Distribution of births by length of gestation in weeks, {Years}

Period	Gestation in weeks		Total
	<37	37 or above	
			100
			100
			100

Write a few paragraphs about your findings and what trends you may be seeing over time.

If you have data on the ante-natal care that was received from your health information, this is also a useful measure for health planning. Higher antenatal coverage should contribute to fewer infant and maternal deaths. This data is also frequently reported as part of a DHS survey.

Writing exercise 4: Birth tabulations

If you haven't already done so after each calculation, write some text to go with the tables you have created to describe births in your country. You may also want to describe whether births usually occur in a health facility or elsewhere and whether mothers are frequently referred overseas to give birth. Give average births per year, along with other interesting information such as whether there has been any significant change over time.

5.8 REVIEW

Births are useful for policy planning (how many immunizations needed each year, how many children entering school), and as denominators for later mortality calculations such as the infant mortality rate (IMR).

The number of births occurring in a country over a given period of time does not tell us anything about the fertility of the average woman in that country. A large number of births may simply reflect a large population of mothers. We need other measures to understand fertility.

It is useful to know how many births are occurring by age of mothers because when women are having their children at younger ages, the population grows faster because there is less time between generations. Additionally, very young or older mothers are more likely to have complicated pregnancies and risks associated with the infant's health and may require more health monitoring.

When examining vital events in small populations, the numbers of births and deaths are likely to "jump around" a lot over time. To smooth out some of this noise, we can calculate the average over a given period (3-5 years), or we can calculate rolling averages.

The crude birth rate (CBR) is the number of births per 1,000 population over a given period of time.

The CBR is one of the easiest measures of fertility to understand, however, it has some limitations. The midpoint population used in the denominator includes children, men, and women outside of childbearing ages thus the CBR is larger in populations where women of reproductive age comprise a larger proportion of its people.

The sex ratio at birth is the ratio of male to female babies born. We use the sex ratio at birth as a means to assess if prenatal sex selection is occurring in a given population.

Place of birth such as on-island or overseas can be useful for policy planners. Risk factors such as the percent of low birth weight babies and babies born preterm can also be useful to assess the health status of pregnant women and babies in a country.

6 MEASURES OF FERTILITY

We have already discussed births and crude birth rates in Chapter 9. This chapter will cover age-specific fertility rates (ASFRs) and the total fertility rate (TFR), both of which are less influenced by the age structure of a population, and can more accurately help us understand fertility.

In demography, the term **fertility** refers to the number of live born babies a woman (or a group of women) has actually had. This is in contrast to clinical medicine where the term fertility refers to the potential to bear children. Throughout this module we will use fertility in the demographic sense.

Fertility is important because it is the primary driver of population growth. Understanding fertility can help us plan for the future not just for maternity-related services, but for services needed to support the health and well-being of a growing population.

A variety of factors can influence a population's fertility including:

- The age structure of the population, i.e. what proportion of the population is comprised of women of childbearing age
- The age at which women marry
- Availability and use of family planning methods (contraceptives)
- The cultural preference for large or small families
- How much children are seen as an economic asset or as a cost
- The level of economic development
- The status of women including their education level, right to work, and cultural beliefs about their status and rights.

Data to estimate fertility can come from a variety of sources including vital registration, censuses, and household surveys. If data are accurate and reliable, fertility can be estimated directly from civil registration data. If civil registration data are not complete or are not reliable, indirect methods may be preferable to determine fertility.

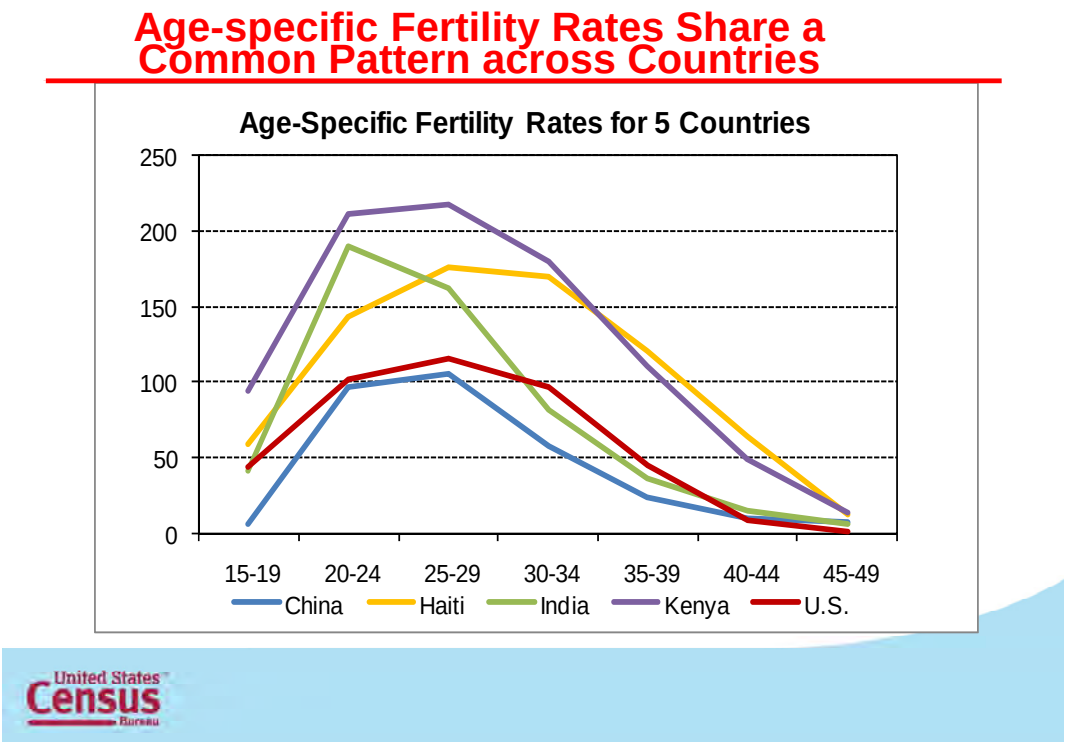
6.1 AGE-SPECIFIC FERTILITY RATES

When analysing fertility, it is helpful to know the fertility of females at particular age groups, also known as age-specific fertility rates. **Age-specific fertility rates (ASFRs)** are the number of births occurring to mothers of a certain age group per 1,000 women in that age group in a given period of time. ASFR's are usually calculated for women aged 15-49 in each 5-year age group. The advantage of ASFRs is that they are not affected by differences in the age distribution among women of childbearing ages.

Age-specific fertility rates generally follow a standard pattern; rates start from zero at very young ages where women are not yet able to bear children, then rise until they peak sometime in women's twenties. ASFRs then decline back to zero somewhere around 50 years of age. Variations in the pattern occur due to factors listed above, such as age at marriage, the prevalence of contraceptive use, etc.

ASFRs for 5 different countries are shown below from the U.S. Census Bureau's International Data Base.[3] You can see that fertility peaks in India much earlier than it does in the other countries. The figure also illustrates that Kenya has higher fertility in almost every age group compared to the other 4 countries.

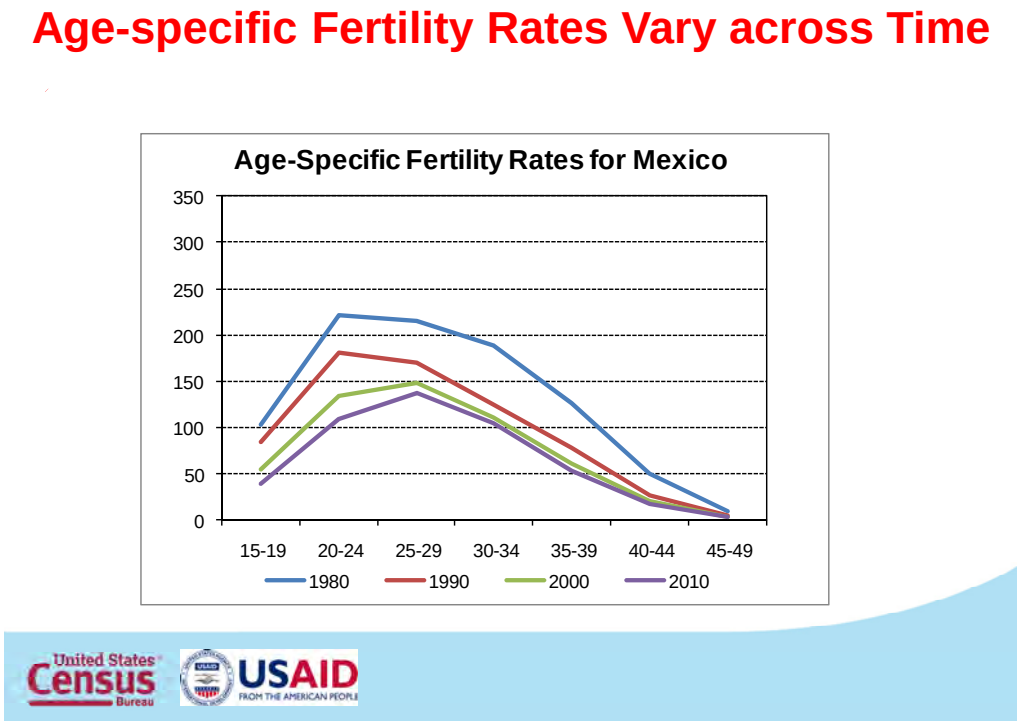
Figure 6.1 Example of trends in Age specific Fertility rates by country



ASFRs from 1980-2010 for Mexico are shown below. Data came from the U.S. Census Bureau's International Data Base. You can see that in 1980, 20-24 year olds had the highest fertility rates (this is where the curve peaks). However, by 2000, 25-29 year old women had the highest fertility rates. The peak of this graph had shifted to the right, meaning that women were waiting until they were older to give birth. Since each curve is smaller than the curve for the decade before it, fertility among all age groups was declining. By 2010, fertility had reached its lowest level yet.

3 http://www.census.gov/population/international/index.html

Figure 6.2 Example of trends in Age specific Fertility rates over time



Recall that:

ASFR = <u>No of births to women of age x</u>
Total number of women of same age (including women who had no children)

The following table from the U.S. Census Bureau's *Population Analysis with Microcomputers Volume I Presentation of Techniques*[4] shows how the calculation is applied to real world data. In Chile in 1983 there were 593,262 women aged 15-19. Those women had 36,784 births in that year. If we perform the calculation: (36,784 / 593,262) * 1,000 we get 62.0. This means that for every 1,000 women aged 15-19, 62 births occurred. Note that we will discuss the total fertility rate in the following section.

[4] Available at: https://www.census.gov/population/international/files/pas/PAMvI_Archive.pdf

Table IV-1. Age-Specific Fertility Rates and Total Fertility Rate for Chile: 1983

Age of women	Female population	Number of births	Fertility rate
(1)	(2)	(3)	(4) = (3)/(2) x 1,000
15-19	593,262	36,784	62.0
20-24	587,076	81,213	138.3
25-29	505,362	65,236	129.1
30-34	424,186	37,506	88.4
35-39	385,749	17,532	45.4
40-44	325,105	4,929	15.2
45-49	266,575	512	1.9
		Sum =	480.4
		Sum x 5 / 1,000 =	2.4

The total fertility rate in Chile in 1983 was 2.4 births per woman.

We will now perform these calculations on our data. However, because we are aggregating over 3-5 years, our denominators will be the midpoint population of women in each age group. For example, if your data covers the years 2007-2011, you will use population data for July 1, 2009 as this is your midpoint. This time you will only be including women in your denominators so if you have not already interpolated your census data for males and females separately, you will need to do so at this time.

Table 6.1: Age-specific fertility rates, {Years}

Age Groups	Periods		
	2007-2011	xxxx-xxxx	xxxx-xxxx
15 - 19			
20 -24			
25 - 29			
30 - 34			
35 - 39			
40 - 44			
45 - 49			

Graph your most recent ASFRs. When does fertility peak? Does it go up sharply and fall sharply or rise more gradually? If you have the data, graph your ASFRs over several periods and see how age-specific fertility has been changing over time. Does fertility peak at different age groups over time? Do some age groups now have higher fertility than they used to? Do some now have lower fertility?

Write a few sentences about the results of your findings.

6.2 ADOLESCENT/ TEENAGE FERTILITY RATE

The **teenage fertility rate,** or adolescent birth rate, refers to the number of births in a given period of time to females aged 15–19, divided by the number of all 15 – 19 year old females at the period midpoint. It is essentially the ASFR for 15-19 year olds. Adolescent birth rates are a key indicator in measuring the progress of target 3.1 of SDG 3: Reduce the global maternal mortality ratio to less than 70 per 100,000 live births... If you find that your adolescent births are higher than those in the Pacific Region (61.3)[5] discuss this in the context of your country and compare your rates to neighbouring countries.

6.3 TOTAL FERTILITY RATES

Although ASFRs adequately represent fertility by age group, they are difficult to use when comparing two different populations or when examining a given population over time. Furthermore, they are not an easily understandable measure of fertility. Therefore, when discussing fertility, we usually talk about the total fertility rate, or TFR.

The **total fertility rate (TFR)** is the average number of children a woman would give birth to during her lifetime if she were to pass through her childbearing years experiencing the present day age-specific fertility rates. The TFR is usually simply described as the average number of children per woman which makes it an intuitive measure of fertility.

The TFR is calculated by adding up all the age-specific fertility rates, multiplying this sum by five (the width of the age-group interval), and then dividing by 1,000.

TFR = (Sum of ASFR x 5) / 1,000

Recall the table discussed above that outlines births in Chile in 1983. We can see that summing all the ASFRs, multiplying this sum by 5, then dividing by 1,000 gives us the total fertility rate.

Table IV-1. Age-Specific Fertility Rates and Total Fertility Rate for Chile: 1983

Age of women	Female population	Number of births	Fertility rate
(1)	(2)	(3)	(4) = (3)/(2) x 1,000
15-19	593,262	36,784	62.0
20-24	587,076	81,213	138.3
25-29	505,362	65,236	129.1
30-34	424,186	37,506	88.4
35-39	385,749	17,532	45.4
40-44	325,105	4,929	15.2
45-49	266,575	512	1.9
		Sum =	480.4
		Sum x 5 / 1,000 =	2.4

The total fertility rate in Chile in 1983 was 2.4 births per woman.

[5] Source: http://www.unfpa.org/public/home/publications/pid/6526

Now, using the ASFR data you calculated in the previous section, find the TFRs for the periods your data covers and create a table similar to the one below.

Table 6.2: Total Fertility Rates (including 95% Confidence Intervals), {Years}

Period	Total Fertility Rate (TFR)
	{Rate (CI range)}

How is fertility changing over time in your country? Is this what you would expect? What factors do you think are influencing fertility? Write a few sentences about your findings. Comment on what the fertility rates mean in terms of population growth/ change and how this might affect the provision of services. Compare your findings to the most recent census and/or survey and discuss the plausibility of your finding. Do your results match the most recent census or DHS TFR?

Writing exercise 5: Fertility rates

Tidy up the paragraphs you have written on births, CBRs, ASFRs, TFRs etc. and link them all together in your report. Discuss them in the context of one another (i.e. perhaps you have lower births and a lower TFR or maybe births for older women have increased while those to younger women have decreased) and how they are related. Compare your findings to the findings from other sources such as censuses and the DHS and discuss why the results may be different.

FERTILITY IN THE PACIFIC

selected countries vital registration data

TOTAL FERTILITY RATE

The average number of children a woman will have in her lifetime

- 3.9 Nauru 2011-2013
- 3.4 American Samoa 2010-2012
- 3.1 Fiji* 2010-2012
- 2.8 Niue 2007-2011
- 2.7 Cook Islands 2006-2009

The average number of births for every 1,000 women aged 15-19 years old

TEENAGE FERTILITY RATE

- 20 Niue 2007-2011
- 29 Fiji 2010-2012
- 43 American Samoa 2010-2012
- 68 Cook Islands 2009-2013
- 97 Nauru 2011-2013

Information from the May 2014 CRVS DARW Course provided by Brisbane Accord Group (BAG) under the PVSAP.

* *Data from the Fiji Ministry of Health registration*

find out more at www.phinnetwork.org and www.spc.int/sdd

phin

SPC Secretariat of the Pacific Community

World Health Organization

UNSW

The University Of Queensland

QUT

UNFPA

unicef

DEATHS – BASIC TABULATIONS

The level of mortality is an essential measure of health outcomes. As health professionals and statisticians, we are interested in how old people are when they die, how this compares to accepted "norms" or other countries, and how this is changing over time.

The mortality of a population depends on various factors, including:

- Demographic composition of the population, i.e. the age and sex distribution;
- Quality and utilisation of health and medical services such as immunisation programmes, maternal and child health care, primary health care, etc.;
- Environmental conditions and availability of infrastructure such as housing, water supply, sanitation, waste disposal;
- Life style factors, such as abuse of alcohol and tobacco;
- Work-related dangers;
- Exposure to events outside individual control such as natural disasters, war;
- Socio-economic status, such as income and education.

There are a number of measures of all-cause mortality within a population:

- Absolute number of deaths
- Crude death rates
- Age-specific death rates
- Age-standardised death rates
- Life expectancy

In this chapter we will be looking at basic tabulations of deaths, and the crude death rate. These are the simplest measures of mortality, and the most commonly reported.

6.4 DISAGGREGATING DATA BY SEX

As we have discussed earlier, mortality patterns are different in men and women. Therefore while overall measures for both sexes are important to provide a national picture, it is also important that we disaggregate by sex for all key mortality indicators. We will do this throughout the following chapters.

6.5 NUMBER OF DEATHS

The most basic measure of mortality is the number of deaths. This is easily understood by decision makers, tends to resonate with the community, and in populations which are relatively stable (with little or slow population growth or age structure change) can be used to track changes in mortality over shorter periods.

As before, you can use a pivot table to create these basic tables or analysis.

Select the unit record data you want to include in the pivot table (if your headers are the top row you can do this simply by clicking between A and 1 in the top left corner).

Insert a new pivot table (in a new sheet) from the Pivot table button under the "insert" tab. On the right hand side of the new page, you will find a pivot table field list. Move the fields to the column and row headings as appropriate and use the "count of" function by moving a field that has a value for every record (such as name) into the values box to populate your table.

The first table you will need is the number of deaths by sex and year.

Table 6.3: Total Number of Deaths by Sex and Year, {Years}

Year	Female	Male	Total

You may notice that the number of deaths varies from year to year. Some of this variation is stochastic (random), but if there are years that look very different from those that come before or after them, consider what this means for your data.

If there are years where there is very low data, questions to ask may be:

- Am I missing any data (perhaps an office was closed for part of a year and no deaths were recorded, or some of the provincial offices forgot to file returns that year)?
- Was there a change in processes (i.e. there were more referrals that year and subsequently more deaths occurred overseas)?

At this point, if there are some years that look implausibly low to you, you may need to go back and check your data. Did all of it get included in the pivot table, is there an original data source you can use to work out the data.

In some cases, where data cannot be corrected and is clearly substantially under-reported, it may be necessary to exclude that year from further analysis in order not to skew the results.

If there is a very high amount of deaths compared to other years, questions to ask might include: was there a catch up campaign for registration or other change in procedures, and was there a natural disaster or disease outbreak. Again, years with extreme values may need to be excluded from further analysis. You should describe the pattern of deaths, and document possible reasons for any outliers.

An outlier is an extreme value which is not in keeping with the rest of the data when plotted.

Rolling averages

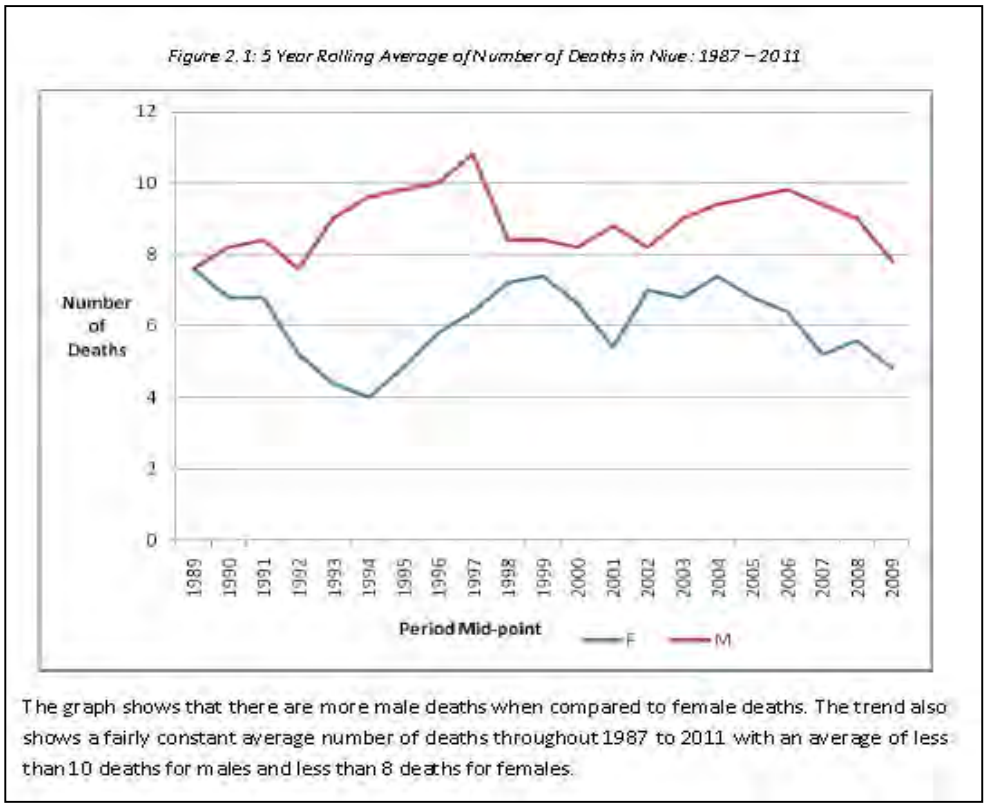
In most cases however, rather than excluding data, we can "smooth" out some of the stochastic variation and make it easier to interpret by creating a rolling average. This means that we still have an annual value, but the figures will not "jump around" so much. Copy your pivot table of deaths per year over to a new worksheet within your booklet (paste as values, and don't forget to label appropriately). Create a new column for average deaths (for both males and females) if you are using 3 year aggregations, starting on your second year of data the average value for this year (the mid-point) is the average of the original cell, the one before and the one after). You can use the average function in excel to do this.

=AVERAGE(cell1, previous cell, next cell)

A line graph may be useful to visualize your data.

With small numbers, this may still result in some stochastic fluctuation, as shown in the graph below for Niue. However, keeping in mind the scale, we can now see these fluctuations are quite small, and that there is no real trend occurring over the period shown in the graph.

Figure 6.3: Example of a rolling average of number of deaths from Niue



Other tables

Tabulations of total deaths and deaths by sex should also be generated in the same manner by major geographical area, ethnic group, or other population characteristics where relevant.

Deaths by age

Basic tabulations of the number of deaths (all-cause mortality) by sex by age group are important measures in their own rights, but also as the basis for all further calculations on age-specific and age-standardized mortality. You will need to repeat the above methods for creating pivot tables of deaths by age group and sex by moving age group into the row labels, and placing sex and years into the columns. You may then wish to combine the youngest age groups to give a table that has deaths from 0-4 as one group. Don't forget to save and label this table once created by copying and pasting it (as values) to another sheet.

6.6 CRUDE DEATH RATE (CDR)

The most frequently used measure of general mortality, the Crude Death Rate, is the number of deaths in a defined period (usually a calendar year) per 1,000 people. It is defined as "crude" because does not account for the age (and sex) composition of a population.

$$\text{CDR} = \frac{\text{Number of deaths in calendar year}}{\text{Mid} - \text{year population}} \text{ x } 1000$$

Thus, a crude death rate of 9.5/1000 in a population of 500,000 indicates there were 4,750 deaths per year in the total population (9.5/1000 X 500,000).

However, with all our measures, we will **aggregate** our data over several years to provide a more stable number. The formula for our CDR then becomes:

$$\text{CDR} = \frac{\frac{\text{Total number of deaths in period ofinterest}}{\text{number of years in the period of interest}}}{\text{mid} - \text{point population for the period of interest}} \text{ x } 1000$$

Or the average number of deaths per year divided by the mid-point population.

It is useful to calculate the crude death rate separately for males, females and both sexes combined. You should generally expect the CDR for males to be higher than for females. Deviations from this pattern could indicate that women and girls face severe disadvantages in terms of health and nutrition. Alternatively, there may be problems with data completeness and quality with systematic underreporting of female deaths.

The CDR would usually be reported with confidence intervals, which we will examine in the next chapter. It is important that both numerator and denominator refer to the same population in terms of geography and time. It is standard practice to take the size of the population at midyear as the denominator because population size may vary during the year (due to migration, births and deaths) and the midyear population serves as an estimate of the average population exposed to the risk of dying over the course of the year.

Interpreting the crude death rate

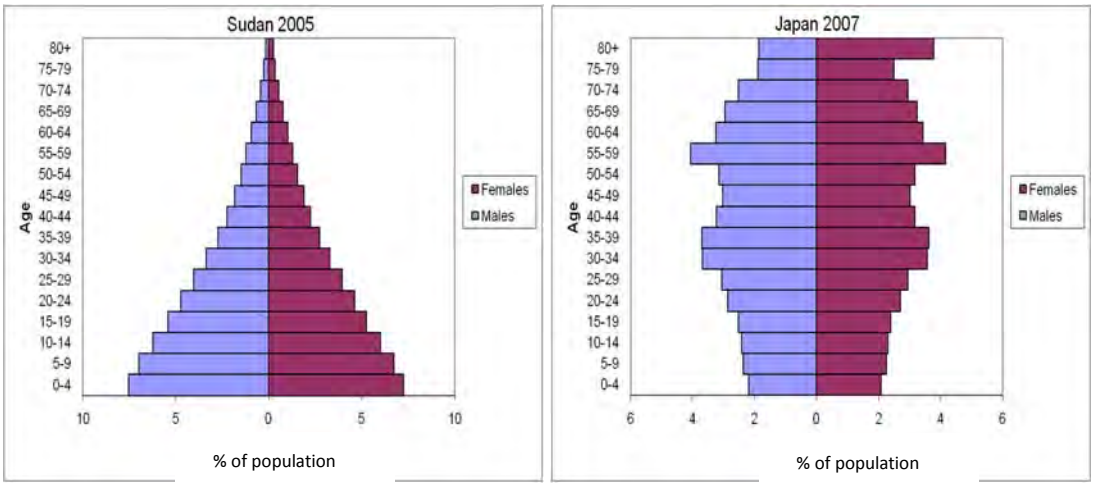
Extract from **UQ/ WHO CODMOD tool [AbouZahr et. al. 2010]** *http://www.uq.edu.au/hishub/wp13*

In practice, the risk of death in a given population group varies according to age and sex as well as patterns of socio-economic status, environmental and other factors. For example, populations with a large proportion of young children or a high proportion of elderly people will, other things being equal, have relatively higher crude death rates. This is because mortality risks are highest at very young and the oldest ages. In general, mortality rates are higher among males than females.

The use of population pyramids in helping to interpret Crude Death Rates is illustrated in Figure*. The CDR for Sudan in 2005 is estimated at 13 per 1000 population compared with 9 per 1000 population in Japan. This difference reflects the fact that Sudan has a high proportion of children aged below 4 years and this is precisely the age group where mortality rates are highest. By contrast, Japan has a much smaller*

percentage of population in this age group, although it has a large proportion of older people aged 60+, when death rates are also high. However, this is insufficient to counteract the effect of a large population of children in Sudan among whom death rates are comparatively high.

Figure 6.4: Population pyramids for Sudan 2005 and Japan 2007



Source: Calculated from UN Population Division estimates http://esa.un.org/unpp/index.asp.

Look back at the population pyramid that you drew earlier. Consider what your population structure looks like and what impact you would expect this to have on your crude death rate.

Lower limits for the crude death rate

Based on many decades of experience in calculating crude death rates, demographers have demonstrated that there is generally a lower limit for the CDR of around 5 per 1000. For example, over the past 20-30 years, Japan has consistently registered the lowest age-specific mortality rates in the world. Yet throughout this period, the CDR in Japan NEVER fell below 5 per 1000.

Any CDR below 5 per 1000 should be treated with extreme caution as such a figure is strongly suggestive of INCOMPLETE death registration.

Lab and writing exercise

Develop tables for:

- Deaths by sex by year
- Rolling average number of deaths by sex by year
- Deaths by age group by sex for each selected time period
- Deaths by sex by time period by other characteristics (as appropriate)

Make sure to label these correctly and paste them into your draft report.

Generate line graphs for any tables involving rolling averages. Again, label these correctly and paste into your draft report.

Write a paragraph(s) that describes both the variation in the results, and the overall trends.

Calculate the crude death rate for all deaths, and by sex, for each time period you have selected. Record these in an appropriate table.

7 AGE DISTRIBUTION OF DEATHS

In the previous chapter we looked at basic tabulations for deaths and mortality rates for the population as a whole. However, information on all-age mortality is not very useful for health planning or monitoring; what we really need to know is how many deaths occur in different age groups. We want to know how old people are when they die, and how this compares to the mortality in other countries and regions of the world. Additionally, we want to break down mortality by sex as men and women tend to die of different things at different ages.

In the chapter on population we learned that the crude death rate is heavily influenced by the age structure of the population. Countries with higher proportions of older people tend to have higher crude death rates. If we want meaningful comparisons of mortality between populations, we can use age-specific mortality rates and age-standardization to take into account age differences of the populations. This chapter will discuss these measures in more detail.

7.1 AGE-SPECIFIC MORTALITY

An **age-specific mortality rate** is the number of deaths per 1,000 people of a given age group in a given time period. These are often referred to ask the age-specific central death rates and are denoted by the symbol M_x.

$$\text{Age} - \text{specific death rate } 25-29 \text{ in } 2010 = \frac{\text{Deaths } 25-29 \text{ in } 2010}{\text{Mid} - \text{year population } 25-29 \text{ in } 2010} \text{ x } 1000$$

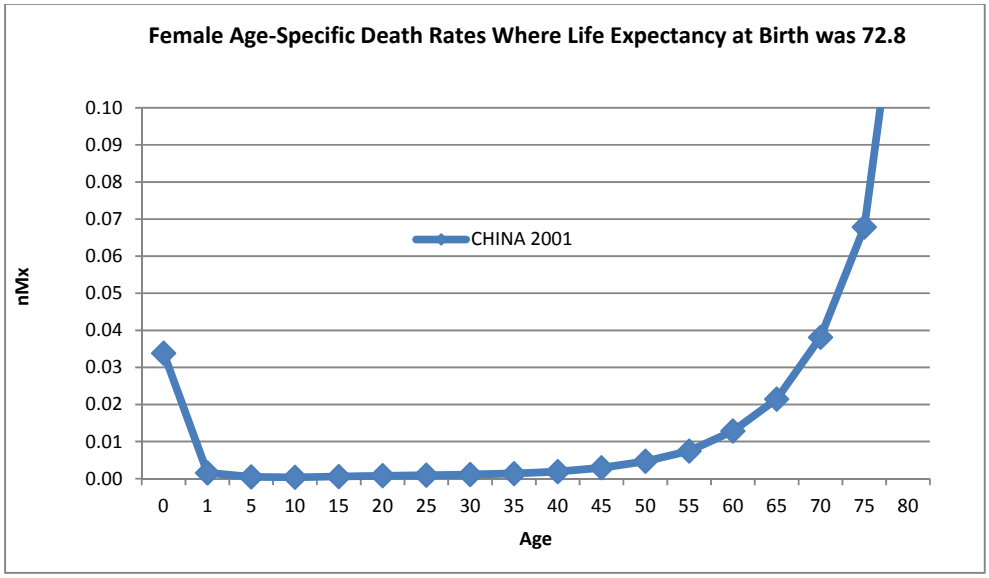
For example, suppose there are 2517 male deaths at ages 25-29 years in Australia in 2010, and the mid-year population of men aged 25-29 was 277,615, then the age-specific death rate for males aged 25-29 is:

$$= \frac{2517}{277{,}615} \text{ x } 1000$$

$$= 9.06 \text{ deaths per } 1000 \text{ males aged } 25-29$$

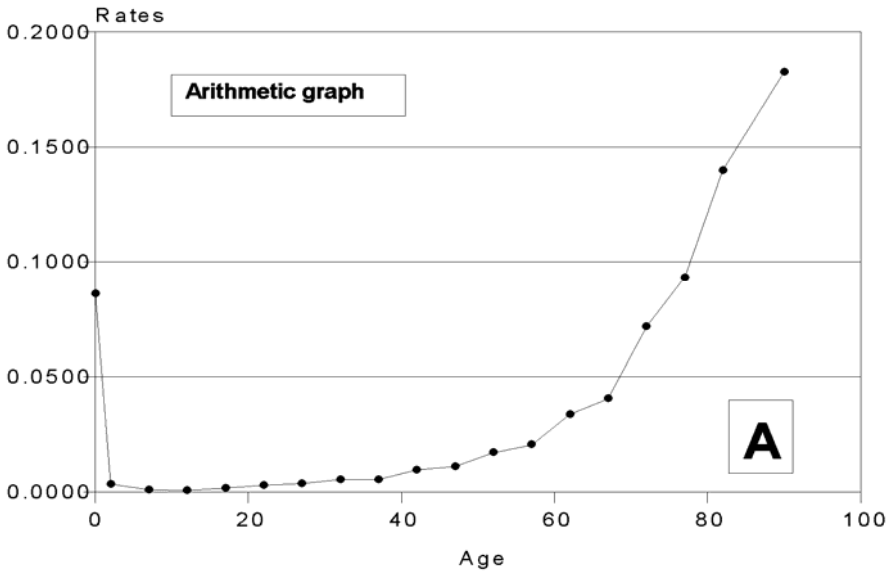
The typical pattern for age-specific mortality is J-shaped, as you can see in the figure below for China. Mortality is relatively high among infants and young children, after which it declines rapidly, reaching its lowest usually around the 10-14 year age group. It then gradually starts to edge up as young adult women are at risk for mortality due to childbirth and young adult men are at risk due to accidents and incidental causes such as suicide. Mortality continues to increase into the older adult age groups and generally starts to rise more rapidly among the oldest age groups in the population. As a general rule, mortality rates start to increase exponentially beyond age 35 or so. In the example below, mortality appears to increase quite rapidly once women reach age 65.

Figure 7.1 Example of Age-specific Death rate curve



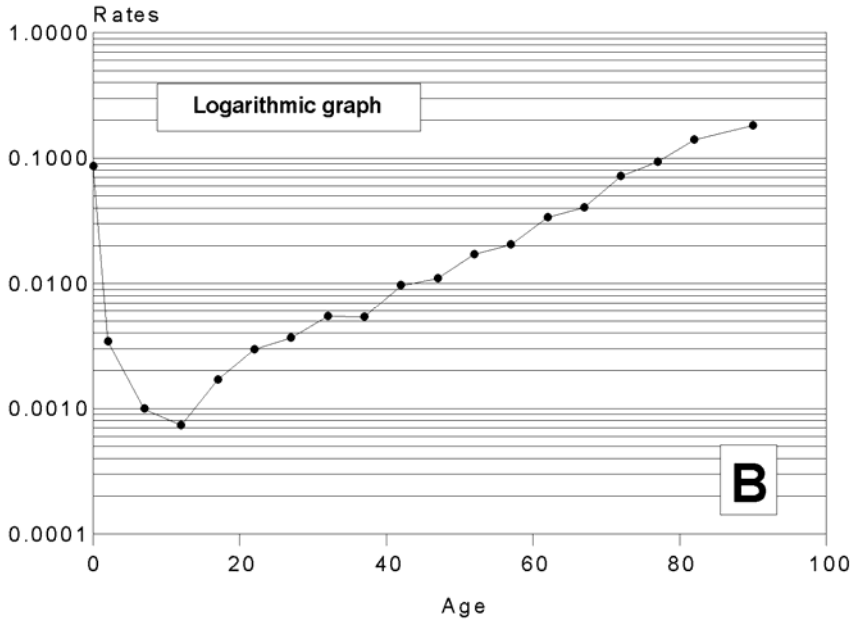
It may be difficult to pick out the changes in mortality between age groups on a graph using an arithmetical scale, such as the graph above. For this reason, we often graph age-specific death rates using a logarithmic scale. Below are two figures taken from the U.S. Census Bureau's *Population Analysis with Microcomputers Volume I Presentation of Techniques.*[6] Graph A (Figure 7.2) shows age-specific death rates graphed arithmetically while Graph B (Figure 7.3) has the same age-specific death rates graphed logarithmically. You can more clearly see which age group has the lowest and highest deaths rates and how things change over time in Graph B.

Figure 7.2 Example of Age-specific Death rate Arithmetic graph



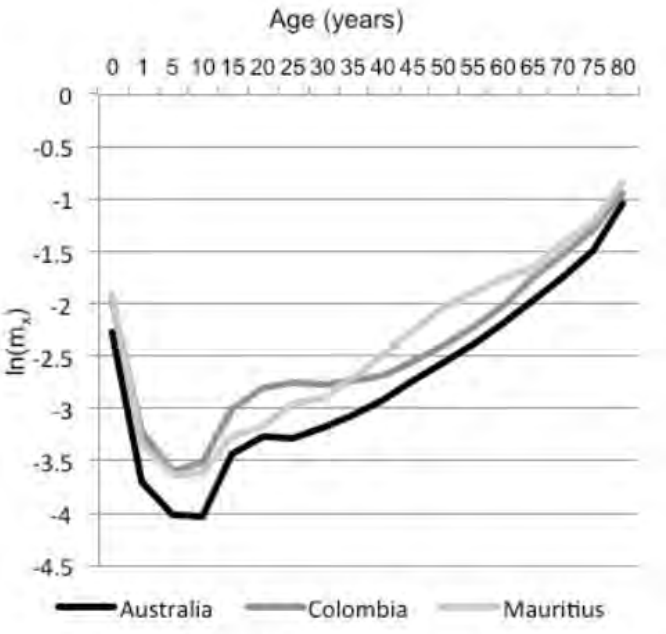
[6] Available at: https://www.census.gov/population/international/files/pas/PAMvI_Archive.pdf

Figure 7.3 Example of Age-specific Death rate Logarithmic graph



The figure and text below was taken from the University of Queensland Health Information Systems Knowledge Hub's Mortality statistics: a tool to improve understanding and quality.[7]

Figure 5 shows patterns of mortality across age for Australia, where death registration is complete, compared to Russia and South Africa, where death registration is less complete or essential information about the death is missing (e.g. unknown age or sex). In Australia, mortality rates are very low up to the age of about 15 years old, and although there is a small increase for males during the ages of 15–34 years due to accidents and other injuries, death rates only really begin to rise sharply after about age 55 years. This pattern is typical of most low-mortality populations. In Russia and South Africa, mortality in infants is relatively high (this is particularly marked in South Africa) but declines during childhood. In South Africa, there is a 'bump' in mortality during reproductive ages in both sexes, reflecting premature mortality due to AIDS-



Source: Institute for Health Metrics and Evaluation database

Figure 6 Log of male age-specific death rates for Australia, Mauritius and Colombia

[7] http://www.uq.edu.au/hishub/wp13

related illnesses. A similar bump may occur in females of reproductive ages in settings where maternal mortality is very high.

Comparing your data with this pattern can provide a simple check on the quality of the mortality data and indicate possible under-registration of deaths at certain ages. It is not the level of mortality that matters in this comparison but the relative age pattern of the ASMR among different age groups.

As noted above, beyond about 35 years of age, death rates rise exponentially with age. Therefore, the natural logarithm of the age-specific death rate (mx), written as ln(mx), should be a straight line as age (x) increases. Figure 6 shows examples of ln(mx) for three countries— Australia, Colombia and Mauritius—with very different patterns of mortality and variable quality of mortality data.

The primary purpose of preparing a graph of the log of the death rate at each age is to examine the data for irregular or implausible changes in ln(mx) from age to age. In countries with high maternal or injury mortality in young adults (especially males), death rates will rise steeply (i.e. ln(mx) will rise) around age 15 years, peak at age 25, and decline to a new low at about age 35 years old. Subsequently, the ASMR will rise linearly with age. Any other departure from this linear pattern in adult death rates suggests that deaths are being selectively (by age) underreported or that there is misreporting of the correct age of death. This is particularly common at older ages.

With this in mind, we can make the following observations from Figure 6 showing age-specific death rates for males:

- Australia —All deaths are registered and hence the ln(mx) increases smoothly in a straight line with increasing age (x), as would be expected. Note the slight bump around ages 15–25 years old, indicating an excess in injury-related deaths in this age group.
- Mauritius—Notice that in this case the ln(mx) does not increase linearly with age after about age 65, suggesting underreporting of deaths, particularly at the oldest ages.
- Colombia—Note the large bump in mortality at ages 15–34 years old due to accidents and other violent deaths. One would expect to see a similar large bump in the ln(mx) graph at these ages in countries with high AIDS-related mortality.

Thus, plotting the ln(mx) will help to identify if there are any age groups where deaths are being selectively underreported (e.g. older ages in Mauritius). In addition, by comparing the graph of ln(mx) for your population with a neighbouring country with good quality mortality data, it will be possible to assess whether, and to what extent, deaths are being systematically underreported at all ages. This will be the case if the graph for ln(mx) for your population is systematically lower than the graph for a neighbouring population.

Performing calculations using aggregated data

We will now calculate our age-specific mortality rates; don't forget to aggregate your data over 5 years. Start your calculations with the 5 year age groups listed below in the table. If your data cannot support the breakdown by 5 year age groups, as is the case in many Pacific Island Countries and Territories with small populations, you will need to aggregate up to 10 year age groups and make 75+ your upper limit for age (see the second table below). Once you perform your calculations, you will graph your data over time to

see how much change is occurring year to year and if aggregating age groups in needed. This will be outlined further in the following text. You can ask your instructors for guidance if needed.

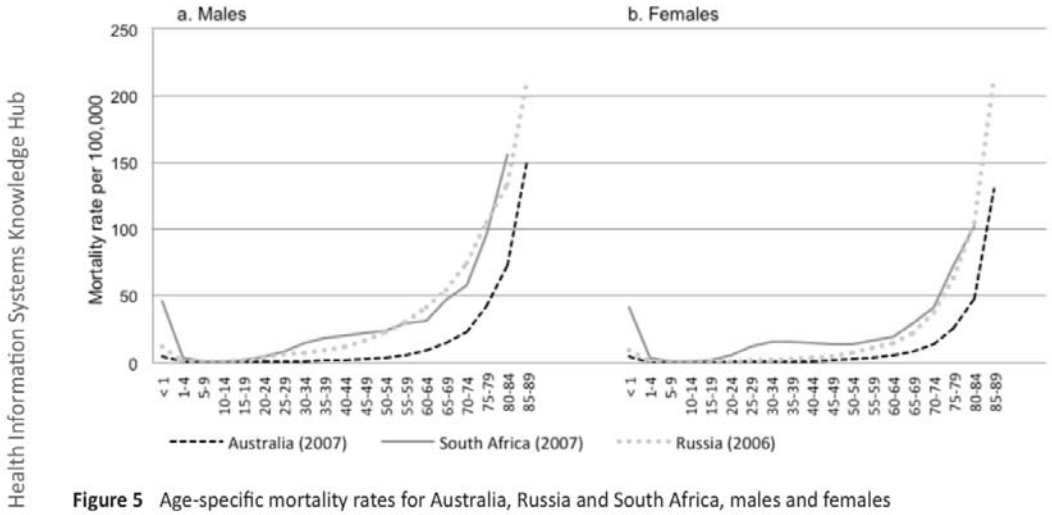


Figure 5 Age-specific mortality rates for Australia, Russia and South Africa, males and females

To perform the calculations, you will add up all your deaths over 5 years for a specific age group and sex. You will then divide that number by the midpoint population of that age group and sex and multiply this by 1,000. For example, let's assume during the period 1987-1991 there were 3 deaths in female children aged 0-4. We then divide 3 by 5 because we are averaging over a 5 year time span and we get 0.6. The midpoint of 1987-1991 is 1989; in 1989 let's assume there were 500 female children aged 0-4. Dividing 0.6 by 500 and multiplying by 1,000 gives us 1.2.

$$\frac{3 \text{ deaths}}{5 \text{ years}} = 0.6 = \text{average number of deaths in 0-4 year old girls from 1987-1991}$$

$$(0.6 / 500) * 1000 = 1.2 = \text{Age-specific mortality rate for 0-4 year old females from 1987-1991}$$

Perform the calculations for your data and complete the table below.

Table 7.1 Age Specific Mortality Rates (deaths per 1,000 people) by 5 Year Age Group, Sex, and period

Age group	Males			Females			Total		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
0-4									
5-9									
10-14									
15-19									
20-24									
25-29									
30-34									
35-39									
40-44									
45-49									
50-54									
55-59									
60-64									
65-69									
70-74									
75-79									
80-84									
85+									

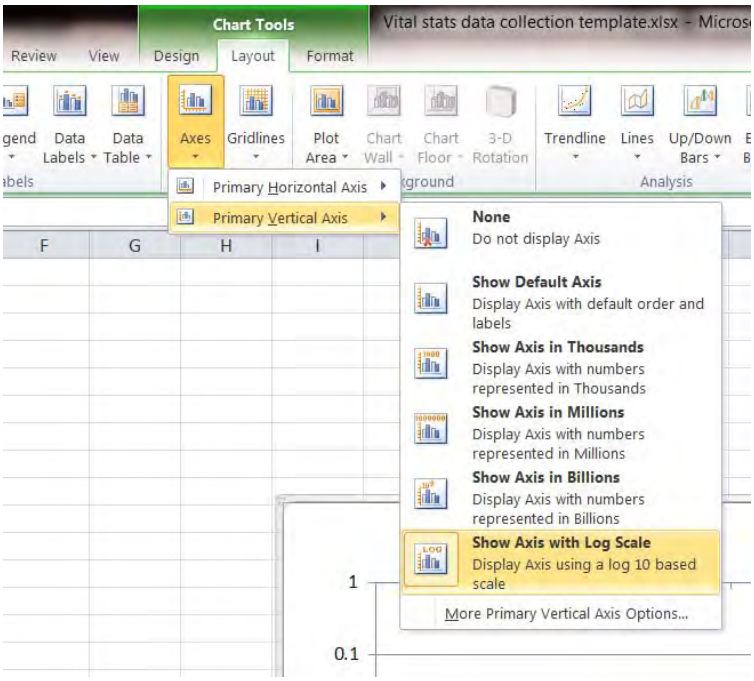
Next, graph your age-specific death rates on a line graph. Use different coloured lines for different periods and examine your data. Is there a lot of noise (fluctuations) between age groups and over time or can you make out a distinct J-shape pattern? If your data has age groups where zero deaths occurred or has large fluctuations you may need to aggregate your data into 10 year age groups and an upper age limit of 75+ and fill in the table below.

Table 7.2 Age Specific Mortality Rates (deaths per 1,000 people) by 10 Year Age Group, Sex, and period

Age group	Males			Females			Total		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
0-4									
5-14									
15-24									
25-34									
35-44									
45-54									
55-64									
65-74									
75+									

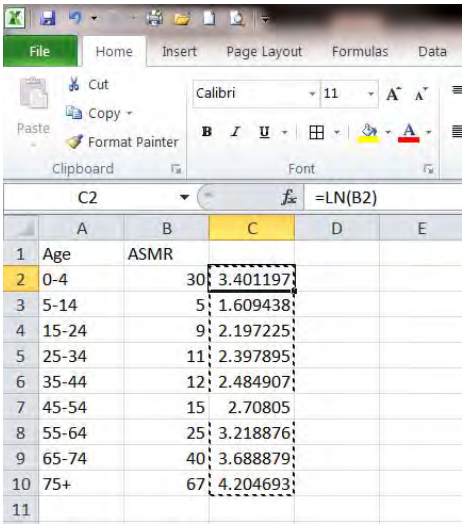
If you re-aggregated to 10 year age groups, graph your data again using a line graph and different coloured lines for different periods. Has your data become more understandable?

Next we will graph our data points on a logarithmic scale. To do this in Excel 2007, click on your graph then click on the layout tab => Axes => Primary Vertical Axis => Show Axis with Log Scale. In Excel 2013, (i) Click on the graph axis you want to change to a logarithmic scale (ii). On the "Format" menu at the top right of the window," Axis options" will be displayed.(iii) Select the "Logarithmic scale" tab under the Axis options tab



Does your data have the J-shaped pattern you would expect? Is there still a lot of fluctuation between age groups?

Now we will take the natural log of our age-specific mortality rates. To do this in excel, in a new column type: =ln(CELL) as shown in the picture below.



	A	B	C	D	E
1	Age	ASMR			
2	0-4	30	3.401197		
3	5-14	5	1.609438		
4	15-24	9	2.197225		
5	25-34	11	2.397895		
6	35-44	12	2.484907		
7	45-54	15	2.70805		
8	55-64	25	3.218876		
9	65-74	40	3.688879		
10	75+	67	4.204693		
11					

Graph the natural log of your age-specific mortality rates by age group.

Age specific mortality as a quality check

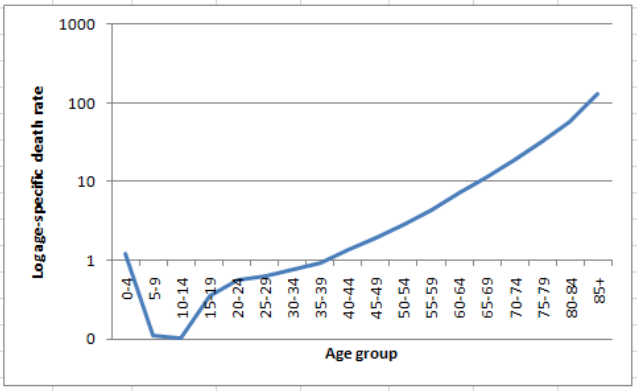
The primary purpose of graphing the natural log of ASMRs is to examine the data for irregular or implausible changes from age to age. In countries with high maternal or injury mortality in young adults (especially males), death rates will rise steeply (i.e. ln(mx) will rise) around age 15 years, peak at age 25, and decline to a new low at about age 35 years old. From about age 35 onwards, the ln(M_x) will rise linearly with age. Any other departure from this linear pattern in adult death rates suggests that deaths are being selectively (by age) underreported or that there is misreporting of the correct age of death. Age misreporting is particularly common at older ages. Note that in settings with high adult mortality from HIV/AIDS this analysis may be inappropriate, because high numbers of deaths at adulthood would mean that mortality rates may not increase exponentially.

The figure below shows examples of both good and bad quality data.

Figure 7.4 Examples of good and bad quality data

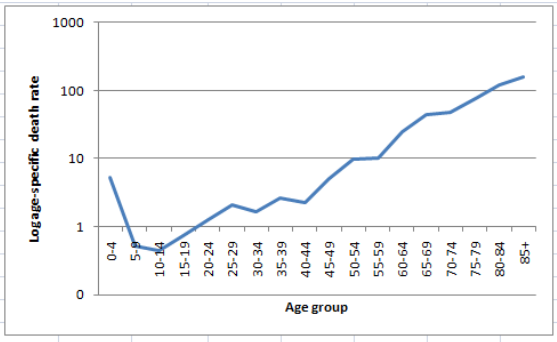
EXAMPLE OF VALID AGE STRUCTURE MORTALITY

Age	Population	Deaths	ASDR
0-4	1264507	1545	1.22
5-9	1321465	145	0.11
10-14	1392249	141	0.10
15-19	1388471	497	0.36
20-24	1431363	812	0.57
25-29	1361259	864	0.63
30-34	1508671	1141	0.76
35-39	1471707	1364	0.93
40-44	1536470	2059	1.34
45-49	1459226	2813	1.93
50-54	1334942	3796	2.84
55-59	1245336	5356	4.30
60-64	944942	6669	7.06
65-69	771804	8843	11.46
70-74	627027	11885	18.95
75-79	552987	18197	32.91
80-84	401156	23264	57.99
85+	315027	41297	131.09
Total	**20328609**	**130688**	



EXAMPLE OF INVALID AGE STRUCTURE MORTALITY

Age	Population	Deaths	ASDR
0-4	28440	152	5.36
5-9	26980	14	0.51
10-14	26120	12	0.45
15-19	22880	17	0.74
20-24	20750	27	1.28
25-29	23460	48	2.05
30-34	21690	36	1.67
35-39	19390	50	2.58
40-44	14850	33	2.22
45-49	10310	52	5.09
50-54	6950	68	9.78
55-59	4450	45	10.11
60-64	3280	80	24.39
65-69	2520	110	43.65
70-74	1920	91	47.37
75-79	1250	94	75.40
80-84	630	77	121.92
85+	300	48	158.71
Total	**236170**	**1054**	



You should graph all periods of time to check plausibility and completeness, but if some periods are missing data or have poor data quality they can be included in the appendices unless there is a specific story of interest to warrant their inclusion in the main body of the report (i.e. no physicians on the island so everyone went overseas for treatment etc.).

What does the shape of your graph tell you about age-specific mortality? Is there high mortality at the youngest age group? Does it fall around ages 5-14? Do you have a bump in ages 15-24 and if so, is this true for both males and females? What happens to your death rates from age 35 onwards? Do you see any underreporting? Are there age groups which are missing data? If you have age groups with no data, do you think this because we are more likely to report some deaths than others or just because there were no deaths in those ages due to low numbers of deaths?

Write a few paragraphs about how you aggregated your data, how the calculations were performed, what you saw when you graphed it, and some of the results you saw. Try to address the questions in the above paragraph.

7.2 AGE DISTRIBUTION OF DEATHS

The following is again from the University of Queensland Health Information Systems Knowledge Hub's *Mortality statistics: a tool to improve understanding and quality.*[8]

{W}e looked at the age and sex-specific mortality rates, and at how these vary at different levels of overall mortality. The objective {now} is to examine the age distribution of reported deaths. This age distribution should look quite different depending on the overall level of mortality in a population. The basic tabulations of data prepared {previously} can be used to prepare a chart showing the distribution of deaths by age group. You should use that same broad age group as shown in Figure 5 to tabulate your mortality data for this exercise. Your calculated distribution of deaths should then be compared with one of the expected distributions shown in Figure 7.5 *that most closely resembles the level of mortality in your population, as reflected in the infant mortality rate.*

To determine which of the four models is most relevant to your situation, use an independent estimate (derived from censuses or surveys, or estimated by the United Nations, WHO or other sources) of the infant mortality rate as follows:

- If your infant mortality rate is less than 20 per 1000, the age distribution of your reported deaths should be similar to that shown in panel A in Figure 7.5.
- If infant mortality is between 20 and 50 per 1000, the age distribution of your reported deaths should be similar to that shown in panel B in Figure 7.5
- If infant mortality is between 50 and 100 per 1000, the age distribution of your reported deaths should be similar to that shown in panel C in Figure 7.5
- If infant mortality is over 100 per 1000, the age distribution of your reported deaths should be similar to that shown in panel D in Figure 7.5

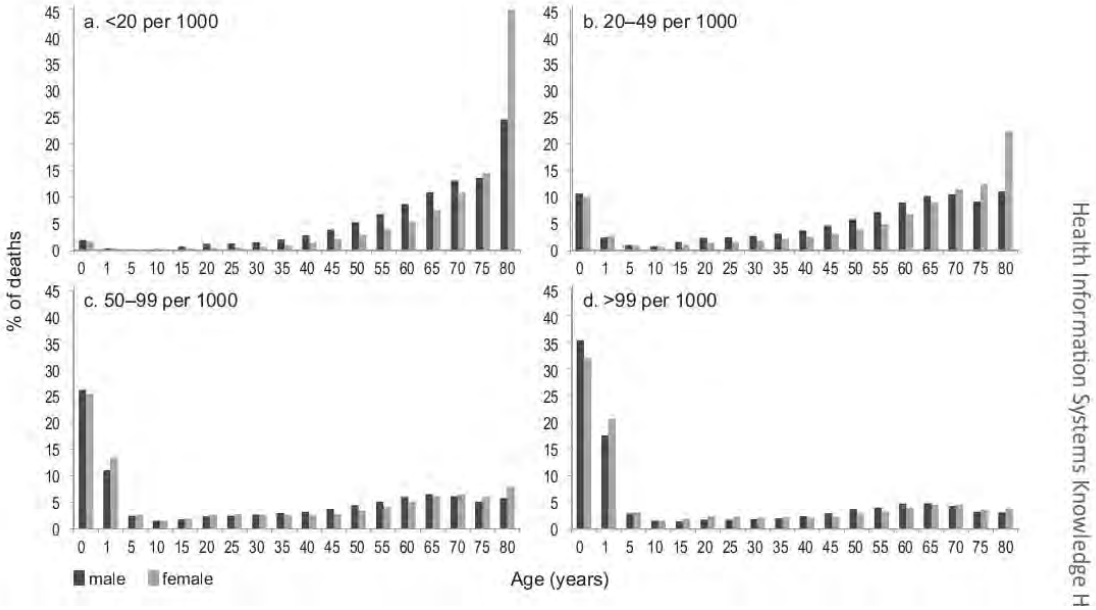
Significant departures from these model age distributions of deaths suggest that the reporting of deaths by age is selectively biased. One reason for such bias may be the way age at death is reported. For example,

[8] http://www.uq.edu.au/hishub/wp13

people tend to have a strong preference to report age at death as a number ending in 0 or 5 (e.g. 45, 50, 55). This is commonly known as digit preference or age heaping. In other instances, the age of the deceased person may be misreported; it is common for families to report that the deceased person was older than they actually were. This highlights the importance of checking the plausibility of age patterns of mortality, and to test for underreporting of deaths in certain age groups by plotting the graph of ln(mx) versus age (x), as described above.

An example of the application of this check on data quality is shown in Figure 9, which gives the reported age distributions of deaths calculated from civil registration data for Sri Lanka, and from the Sample Registration System (SRS) for India. Sri Lanka has an estimated infant mortality rate of 8 per 1000 (hence panel A should be used as the comparator) while the infant mortality rate for India is closer to 60 per 1000 (hence panel C is chosen). This comparison shows that the age distribution of deaths in Sri Lanka is very similar to what was expected (panel A), but in India, the SRS appears to have more deaths at ages 60–74 years and fewer deaths at ages 75+ than expected from a comparison with panel C. This may or may not reflect problems with misreporting of the age at death for older adults, and should be investigated further.

Figure 7.5 Example of age distribution of deaths by varying level of mortality



Check your data for incomplete age distributions

Create a figure in Excel showing the distribution of deaths by age (for each sex separately) and compare the pattern you see with that which would be expected given your level of infant mortality. You can use the image above to make comparisons.

Compare the age and sex distribution of your reported deaths with expected age–sex distributions based on your estimated level of infant mortality as shown in Figure 8. Departures from these expected patterns can be indicative of underreporting of deaths at certain ages for males or females. If, for example, you have very low infant and child mortality rates and also low mortality rates at the very oldest adult ages, you should suspect problems with the registration of adult deaths.

Graphing data in this way allows us to evaluate how complete our data is and whether there is enough information by age group to calculate summary measures of mortality (which we will cover in the next chapter). We need good age-specific mortality for infants and for the oldest age groups in order to do empirical calculations. If our data is insufficient we can "smooth" it using models but these techniques will not be covered in this course.

Comment on the plausibility of the pattern of deaths you see for males and females by age, and what this means about data quality and reporting completeness – including whether the data is good enough to use for life tables. What does your distribution look like? Is there any difference between males and females or over time?

7.3 MEASURES OF INFANT AND CHILD MORTALITY

Infant and child mortality is a key indicator for monitoring the health of population; as further underlined by goal three of the SDGs which all countries in the Pacific Region must report on.

The following is a good summary of infant and child mortality from the University of Queensland Health Information Systems Knowledge Hub's *Mortality statistics: a tool to improve understanding and quality.*[9]

Mortality among children under five years old, more than any other age group, reflects a range of economic, social and health conditions that all affect population health. Child mortality is therefore a key indicator for public health monitoring. Mortality in children under five can be divided into several components:

- neonatal mortality—mortality among infants aged less than 28 days old
- post-neonatal mortality—mortality in infants older than 28 days but less than 1 year old
- infant mortality—mortality among infants aged less than one year (neonatal and post-neonatal deaths)
- under-five mortality—mortality among children aged less than 5 years old.
- *Mortality in children aged* between 1 and >5 years is commonly referred to as child mortality.

Definition and calculation of under-five mortality indicators

Under-five mortality rate

The under-five mortality rate (U5MR) is defined as deaths in children aged 0–4 years in a given population over a specified time period divided by the total number of live births in that population over the same period.

$$\text{U5MR} = \frac{\text{Number of deaths in children aged less than five in a specified time period}}{\text{Number of live births in the same time period}} \times 1000$$

[9] http://www.uq.edu.au/hishub/wp13

However, because of the very different age pattern of mortality risks among children, it is usual statistical practice to transform the mortality rate in children under five into a probability of dying before age five, assuming that children would be subject to the ASMRs of that period. Thus, the U5MR is, strictly speaking, not a rate (i.e. the number of deaths divided by the number of population at risk during a certain period of time) but a probability of death, expressed as a rate per 1000 live births.

Infant mortality rate

The calculation of the infant mortality rate (IMR) is the same as for the U5MR with the exception that the numerator is the number of deaths in children aged less than one year old (i.e. died before their first birthday).

$$\text{IMR} = \frac{\text{Number of deaths in infants aged less than one year old in a specified time period}}{\text{Number of live births in the same time period}} \times 1000$$

Neonatal mortality rate

The calculation of the neonatal mortality rate (NNMR) is the same as for the IMR with the exception that the numerator only includes deaths in children less than one month (28 days) old.

$$\text{NNMR} = \frac{\text{Number of deaths in infants aged less than 28 days in a specified time period}}{\text{Number of live births in the same time period}} \times 1000$$

Neonatal deaths may be subdivided into early neonatal deaths, occurring during the first seven days of life, and late neonatal deaths, occurring after the seventh day but before 28 completed days of life.

Post-neonatal mortality rate

The calculation of the post-neonatal mortality rate (PNNMR) is the same as for the NNMR with the exception that the numerator only includes deaths in infants aged from 28 days to one year old.

$$\text{PNNMR} = \frac{\text{Number of deaths in infants aged between 28 days and one year old in a specified time period}}{\text{Number of live births in the same time period}} \times 1000$$

The reliability of under-five, infant and neonatal mortality estimates depends on the accuracy and completeness of reporting and recording births and deaths. It is essential to apply standard international terminologies and definitions to ensure comparability over time, and across areas or countries. These have been defined in the WHO ICD-10 (WHO 2007). Differences in IMRs, and especially NNMRs, can be greatly affected by the failure to apply the standard definition of live birth. In practice, underreporting and misclassification of under-five deaths are common, especially for deaths occurring very early in life, many of which are misclassified as stillbirths. In such cases, countries often do not record both the early neonatal death and the live birth. This is poor public health practice, as data on both events are critical to improve

maternal and child health services. An example of the calculation of the U5MR, IMR and NNMR based on birth registration and death data is given {right}.

Table 3 Child deaths by age and calculation of mortality indicators

	Male	Female	Total
Neonatal deaths registered	1 563	895	2 458
Infant deaths registered	2 075	1 677	3 752
Under-five deaths registered	3 980	3 456	7 436
Live births registered	191 263	182 275	373 538

Neonatal mortality rate (both sexes combined) = (2458/373 538)*1000 = 6.6 per 1000
Infant mortality rate (both sexes combined) = (3752/373 538)*1000 = 10.0 per 1000
Under-five mortality rate (both sexes combined) = (7436/373 538)*1000 = 19.9 per 1000

All of the rates discussed above take into account a slightly different age group. It's important when analysing and disseminating your data that you clearly state which age groups are included in your numerator. For example, sometimes the statistic being reported is "child mortality" but it is not clear if this is mortality for all children under age 5, or for children aged 1-4.

Figure 7.7 Defining mortality in early life

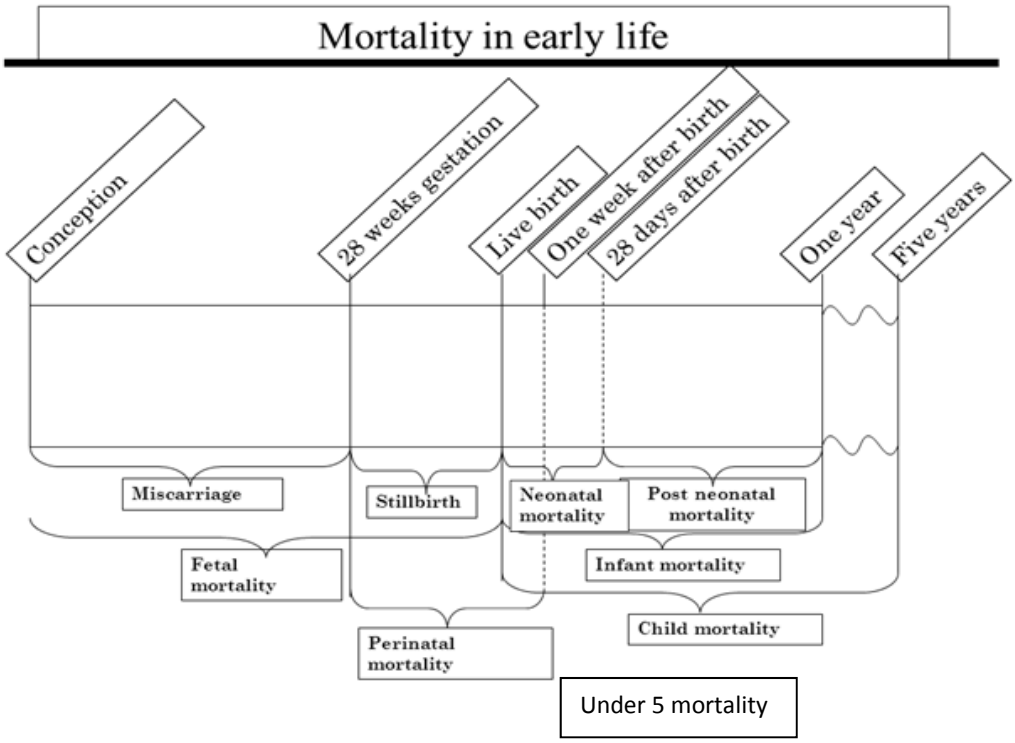


Figure adapted from: University of Queensland Health Information Systems Knowledge Hub, Tim Adair's "Use of existing health datasets workshop" manual.

Probability vs. Rates

You previously calculated the age-specific death rate for children aged 0-4 by taking the number of deaths to children in the age group divided by the midpoint population of children aged 0-4. However, we generally report infant and child mortality as a probability of dying, not as a rate. The under-five mortality rate (U5MR) is the probability (expressed as a rate per 1,000 live births) of a child born in a specified period dying before reaching the age of five if subject to current age-specific mortality rates. The denominators between these ASMR for those aged 0-4 and the U5M are quite different. The ASMR denominator takes into account all children in the population aged 0-4 while the U5M denominator is just live births that occurred in that year. This means that the ASMR denominator will be more than 4 times higher than U5M denominator because it includes all those born over the previous 5 years who are now aged 0-4, minus those who have died since birth.

For example, let's say there were 100 deaths in children under age 5 in 2011 and 1000 live births that same year. Let's say the midyear population of children aged 0-4 in 2011 was 4,700. The ASMR for children aged 0-4 is (100/4700) *1000 = 21.3. The U5M is (100/1,000)*1,000 = 100. The U5M is much higher than the ASMR because it has a much smaller denominator.

Alternative definitions and additional rates of child mortality

Infant mortality rate: Number of deaths at age less than 12 months divided per 1,000 live births.
Child mortality rate: Number of deaths at age 12 to 59 months per 1,000 children surviving to 12 months
Under-five mortality rate: Number of deaths at age less than 60 months per 1,000 live births.
Perinatal mortality rate: Number of perinatal deaths (still births plus deaths within one week of live birth per 1,000 live births.

For perinatal mortality, this is not a "fits-all" model, as some consider 28 weeks too late to begin counting a foetus as viable, and may use other intervals, such as 20 or 22 weeks. It depends largely on the hospital and intensive care available for premature babies. For developing countries, these divisions are fairly standard.

A Note about Neonatal Mortality
The neonatal mortality rate is the number of deaths in live-born infants during the first 28 days of life per 1,000 live births over a specified time period. Mortality during the neonatal period (the first 28 days of life) accounts for a large proportion of child deaths, and is considered to be a useful indicator of maternal and newborn neonatal health and care. Generally, As IMR falls (for example through major improvements in environment, immunisation and nutrition), and fewer deaths are attributed to infectious diseases and environmental influences, a greater proportion of infant deaths would be expected to occur in the neonatal period. The neonatal mortality rate, however, should not increase as this occurs.

Also, recall form Chapter 9 the definition of a live birth:

WHO Definition: The complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of the pregnancy, which, after such separation, breathes or shows any other evidence of life, such as beating of heart, pulsation of the umbilical cord, or definite movement of voluntary muscles, whether or not the umbilical cord has been cut or the placenta is attached.

Source: WHO, 2011, Health Statistics and Health Information Systems: Health Status Statistics – Mortality, http://www.who.int/healthinfo/statistics/indunder5mortality/en/

Calculating mortality rates for children in countries with small populations

Given the small population sizes of PICTs and the relatively rare occurrence of death in children under age 5, we may not be able to calculate rates for all the categories mentioned above. For our purposes, we will calculate the IMR, the under 5 mortality rate, and the neonatal mortality rate (if data is available to distinguish the age in days of children who died at age 0).

Let's assume the table below summarized our data for the 5 year period 2007-2011. Our total live births over that time were 516, we had 2 neonatal deaths, 6 infant deaths, and 8 under 5 deaths.

Table 7.3 Child Mortality for period 2007-2011

Year	Live births	Deaths in infants under 28 days	Deaths in infants under 1 year	Deaths in children under age 5
2007	100	1	3	3
2008	110	0	1	2
2009	98	0	1	1
2010	103	0	0	1
2011	105	1	1	1
Total	**516**	**2**	**6**	**8**

Neonatal mortality rate for a 5 year period: The sum over 5 years of deaths in live-born infants during the first 27 days of life per 1,000 live births over 5 years. Note: infants aged 28 days are not included.

$$\text{Neonatal mortality rate for a 5 year period:}$$

$$= \frac{\text{Sum over 5 years of all infant deaths aged less than 28 days}}{\text{Total live births over same 5 year period}} \text{ x } 1000$$

= (2 / 516) * 1000 = 3.9

We had 3.9 neonatal deaths per every 1,000 live births from 2007-2011.

Infant Mortality Rate (IMR) for a 5 year period: The sum over 5 years of deaths in infants under age 1 per 1000 live births over 5 years. Note that neonatal deaths are included in infant deaths and that they are not mutually exclusive.

Infant mortality rate for a 5 year period:

$$= \frac{\text{Sum over 5 years of all deaths of infants aged } 0-11 \text{ months}}{\text{Total live births over same 5 year period}} \text{ x } 1000$$

= (6 / 516) * 1000 = 11.6

We had 11.6 infant deaths per every 1,000 live births from 2007-2011.

Under 5 Mortality Rate for a 5 year period: The sum over 5 years of deaths in children under age 5 per 1,000 births over 5 years. Note that neonatal deaths and infant deaths are included in under 5 deaths and that these 3 categories are not mutually exclusive.

Under five mortality rate for a 5 year period:

$$= \frac{\text{Sum over 5 years of deaths of children aged } 0-59 \text{ months (under 5 years of age)}}{\text{Total live births over same 5 year periodve births}} \text{ x } 1000$$

= (8 / 516) * 1000 = 15.5

We had 15.5 deaths in children under age 5 per every 1,000 live births from 2007-2011.

Writing exercise: Infant and child mortality

After you have calculated all the measures above, write a brief summary of how you performed the calculations and what they mean. Has IMR or U5M changed over time? Why do you think that is?

7.4 ADULT MORTALITY

Adult mortality is an important factor in assessing the health of a population, especially the determination of how many adults are dying prematurely. We will discuss adult mortality in more detail in Chapter 15 and the following chapters.

7.5 AGE STANDARDIZED MORTALITY RATES

As discussed in previous sections, the age structure of the population can affect mortality indicators such as the crude death rate, making comparisons between populations unfeasible. In order to compare mortality between populations, or within the same population over time, we apply age-specific mortality rates from the population of interest to a standardized population. The direct standardization process eliminates the effect of the age structure by using a single age structure as a standard for all populations being compared. However, standardized crude death rates permit only the ranking, not the measurement, of levels of mortality between populations.

For the purposes of this course, we will use the WHO World Standard Population Distribution.[10] Applying this distribution is useful when comparing between countries, especially when our data does not cover more than one period in time. Below is the WHO World Standard Population Distribution table showing the projected average percent distribution of the world's population from 2000-2025. To standardize our population we would apply the age group percentage listed under the world average to our total population.

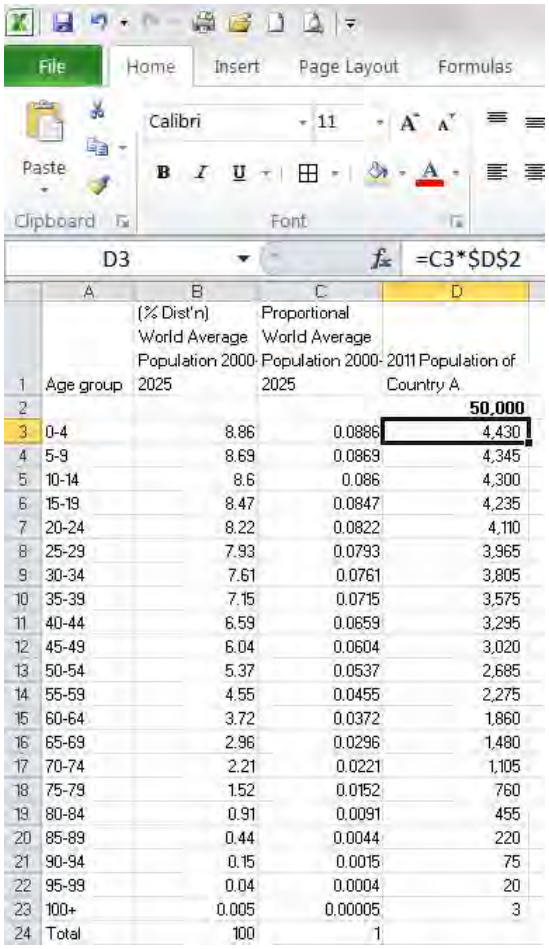
Table 7.4 WHO World Standard Population distribution

Table 4. WHO World Standard Population Distribution (%), based on world average population between 2000-2025

Age group	World Average 2000-2025
0-4	8.86
5-9	8.69
10-14	8.60
15-19	8.47
20-24	8.22
25-29	7.93
30-34	7.61
35-39	7.15
40-44	6.59
45-49	6.04
50-54	5.37
55-59	4.55
60-64	3.72
65-69	2.96
70-74	2.21
75-79	1.52
80-84	0.91
85-89	0.44
90-94	0.15
95-99	0.04
100+	0.005
Total	100

Let's imagine Country A has a population of 50,000. From the spreadsheet below, you can see how applying the WHO World Standard Population Distribution would distribute Country A's population by age group. From column B we see that 8.86% of the population would be children aged 0-4. This amounts to 4,430 children in Country A (0.0886*50,000=4,430).

[10] Available at *www.who.int/healthinfo/paper31.pdf*

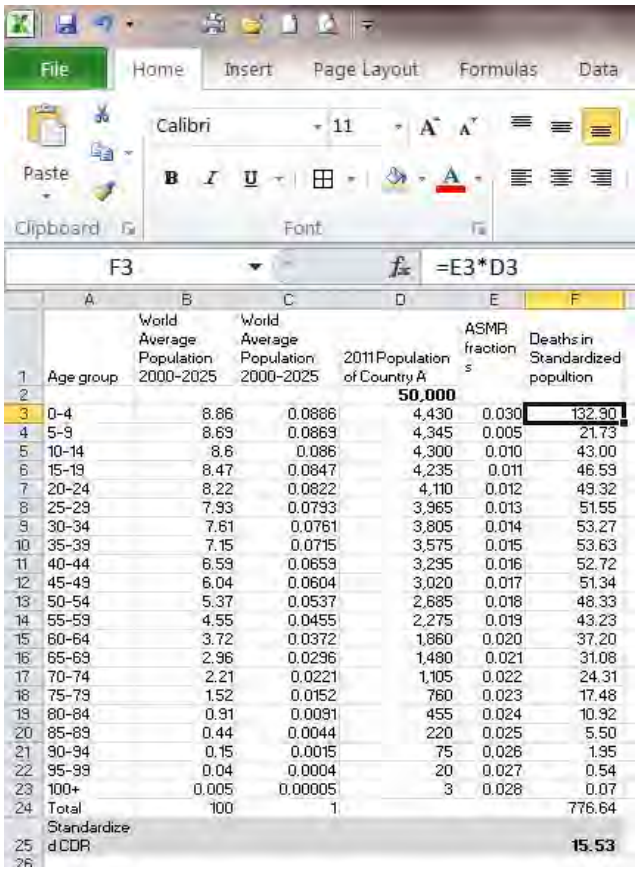


	A	B	C	D
1	Age group	(% Dist'n) World Average Population 2000-2025	Proportional World Average Population 2000-2025	2011 Population of Country A
2				**50,000**
3	0-4	8.86	0.0886	4,430
4	5-9	8.69	0.0869	4,345
5	10-14	8.6	0.086	4,300
6	15-19	8.47	0.0847	4,235
7	20-24	8.22	0.0822	4,110
8	25-29	7.93	0.0793	3,965
9	30-34	7.61	0.0761	3,805
10	35-39	7.15	0.0715	3,575
11	40-44	6.59	0.0659	3,295
12	45-49	6.04	0.0604	3,020
13	50-54	5.37	0.0537	2,685
14	55-59	4.55	0.0455	2,275
15	60-64	3.72	0.0372	1,860
16	65-69	2.96	0.0296	1,480
17	70-74	2.21	0.0221	1,105
18	75-79	1.52	0.0152	760
19	80-84	0.91	0.0091	455
20	85-89	0.44	0.0044	220
21	90-94	0.15	0.0015	75
22	95-99	0.04	0.0004	20
23	100+	0.005	0.00005	3
24	Total	100	1	

You will notice that the WHO World Standard Population Distribution goes all the way to 100+ or to 85+ for the oldest age group. If your ASMR upper age limit is 75+, you can use the sum of the age groups from 75-100+ for your final age group. In the case of the WHO World Standard Population Distribution this would be 3.065%, or the proportion 0.03065 for ages 75+. There is no need to distribute population to age groups that you do not have ASMRs for.

Once we have our population distributed according the WHO World Standard Population Distribution, we can multiply our age-specific death rates to the standardized population in each age group to get the number of deaths by age group. We sum up all these deaths and divide this sum by the population total to get a standardized crude death rate.

When applying our age-specific mortality rates to the standardized population, we multiply by the ASMR fraction we calculated before we multiplied the dividend by 1000. For example, if there were 30 deaths in a population of 1,000 children aged 0-4, then 30/1,000 =0.03. To get the ASMR we would multiply 0.03 by 1,000 and get a rate of 30 per 1,000, but in our case we will use the 0.03 as is. There were 4,430 children aged 0-4 in Country A's standardized population. We will multiply 4,430 by 0.03 to get 132.9 deaths. We will do this for each age group as shown to the right, then sum up all of these deaths.



	A	B	C	D	E	F
1	Age group	World Average Population 2000-2025	World Average Population 2000-2025	2011 Population of Country A	ASMR fractions	Deaths in Standardized popultion
2				**50,000**		
3	0-4	8.86	0.0886	4,430	0.030	132.90
4	5-9	8.69	0.0869	4,345	0.005	21.73
5	10-14	8.6	0.086	4,300	0.010	43.00
6	15-19	8.47	0.0847	4,235	0.011	46.59
7	20-24	8.22	0.0822	4,110	0.012	49.32
8	25-29	7.93	0.0793	3,965	0.013	51.55
9	30-34	7.61	0.0761	3,805	0.014	53.27
10	35-39	7.15	0.0715	3,575	0.015	53.63
11	40-44	6.59	0.0659	3,295	0.016	52.72
12	45-49	6.04	0.0604	3,020	0.017	51.34
13	50-54	5.37	0.0537	2,685	0.018	48.33
14	55-59	4.55	0.0455	2,275	0.019	43.23
15	60-64	3.72	0.0372	1,860	0.020	37.20
16	65-69	2.96	0.0296	1,480	0.021	31.08
17	70-74	2.21	0.0221	1,105	0.022	24.31
18	75-79	1.52	0.0152	760	0.023	17.48
19	80-84	0.91	0.0091	455	0.024	10.92
20	85-89	0.44	0.0044	220	0.025	5.50
21	90-94	0.15	0.0015	75	0.026	1.95
22	95-99	0.04	0.0004	20	0.027	0.54
23	100+	0.005	0.00005	3	0.028	0.07
24	Total	100	1			776.64
25	Standardized CDR					**15.53**

We see that there are 776.64 deaths in our standardized population. We then divide 776.64 by our population of 50,000 and multiply this by 1,000 to get a standardized crude death rate of 15.53.

It is worth noting that any two populations that have been standardized with the same age distribution will generate the same standardized crude death rates when identical ASMRs are applied to them. For example, if we had started with a population size of 77,000 and applied the WHO World Standard Population Distribution then applied these same ASMRs to the standardized population, we would have still come up with a standardized crude death rate of 15.53.

Lab & Writing exercise: Age-standardization

Because our data is aggregated over 3-5 years, we will need to standardize our midpoint populations. For example, if your ASMR covers 2007-2011, the midpoint would be the population on July 1, 2009. We will take the total population on July 1, 2009, and apply the WHO World Standard Population Distribution. We will then apply the ASMRs we calculated in the previous section to each age group to get the number of deaths for each age-standardized group. Sum up these deaths and divide by the total midpoint population to get an age-standardized crude death rate.

Perform these calculations and write a brief summary of what you did. Are the results plausible, what do they mean? How do these results compare to those of other countries? To the region as a whole?

Bonus exercise

Change your midpoint total population; you can pick any number of your choosing. Apply the WHO World Standard Population Distribution to this new population size and apply your same ASMRs to calculate the standardized crude death rate of this new population size. What do you find? How does this result compare to the standardized crude death rate of your true midpoint population?

Age standardization to compare trends over time within one country

If you have data spanning many years, you may want to compare standardized crude death rates over time within your country. To standardize for comparison in one country over time, you can use the age distribution of the most recent period midpoint population and apply this distribution to all the other period midpoint populations. You would then apply each period's ASMRs to the standardized midpoint population. (Note that this will not allow for comparisons between countries that have not standardized with the same age distribution.)

For example, if your most recent period is 2007-2011, your midpoint population would be the population on July 1, 2009. You would determine the percent distribution of the population by 5 year age group. Let's say in 2009, children aged 0-4 made up 10% of the population and those aged 5-9 made up 8% of the population. Your next period would be 2002-2006 with a midpoint population of July 1, 2004. Suppose the total population on July 1, 2004 was 10,000. Ten percent of 10,000 is 1,000 so you would then allocate 1,000 children aged 0-4. Eight percent of 10,000 is 800 so you would allocate 800 children to the age group 5-9 in 2004. Once you have completed this age distribution, you apply the ASMRs from the 2002-2006 period to your standardized population. If we assume the ASMR for 0-4 year olds in 2002-2006 was 0.03 we will get 0.03*1,000 = 30 deaths to children aged 0-4 in 2002-2006. Apply the ASMRs for 2002-2006 for each standardized age group, sum up the resulting deaths, and divide this total by the midpoint population, 10,000 in our case. This will give you a standardized crude death rate that you can compare to the crude death rate the 2007-2011 period.

Age standardization for other purposes

Standardization is most commonly applied to death rates. However, it can also be used for other indicators where risk varies by age. For example, if the risk of being admitted to hospital with renal failure varies by age, you should age-standardize renal failure admission rates when comparing two populations.

Indirect Age standardization

The techniques discussed in this chapter are direct methods of age standardization. When ASMRs are not available for each age group, it may be necessary to use indirect methods of age standardization. We will not cover indirect age standardization in this course, but we wanted to mention that it's a possibility when ASMRs are not complete.

The definition and steps of indirect age standardization from the U.S. Census Bureau's *Population Analysis with Microcomputers Volume I Presentation of Techniques:*[11]

"Indirect standardization of the crude death rate is accomplished by applying the age-specific death rates of a standard population to the age structure of the area whose crude death rate is to be standardized, and then multiplying the standard crude death rate by a ratio of the area's own number of deaths to the standard number of deaths."

The procedure consists of several steps:

(1) Apply the standard age-specific death rates to the population age structure of the area whose crude death rate is to be standardized, and take the sum of the resulting deaths by age. The result is the "standard" number of deaths.

(2) Divide the area's actual number of deaths by the standard number of deaths. The result is an "adjustment factor" representing the difference between the mortality of the population to be standardized and that of the standard population.

(3) Multiply the standard population's crude death rate by the adjustment factor. This will increase or decrease the standard crude death rate according to whether the mortality of the area is higher or lower than that of the standard.

[11] Available at: https://www.census.gov/population/international/files/pas/PAMvI_Archive.pdf

8 CONFIDENCE INTERVALS AND COMPARING DATA

Throughout the workbook and course, you will have noticed that when we talk about measures of fertility or mortality, we often refer to these as **estimates**. This is because there is nearly always an element of **uncertainty** in our calculations.

This concept is relatively easy to understand when we are talking about **samples** (such as for a survey or for a demographic surveillance site). In these cases, our sample is selected to be representative of a broader population. However, it is implausible that our sample will be EXACTLY like the population as a whole in every way. There is subsequently an element of uncertainty around the measures that we produce when we apply them to this broader population.

Every time we draw a sample from the population, no matter how much care we take to match the broader population characteristics, we will get some variation from the population as a whole. We are also likely to get some variation between each of the samples drawn. Our measures or results are therefore a result of the sample that we have drawn (as part of our study or survey).

Confidence intervals are a way to measure that uncertainty. If we were to use a 95% confidence interval, for example, this would mean that if we had been able to re-draw our sample and repeat our study 100 times (which of course we would never actually do in practice), in 95 of those studies we would end up with an answer within a given range – our 95% confidence interval. Let's use an example of a DHS survey which gave us an estimate of 23 deaths per 1,000 live births for infant mortality with a confidence interval of 20-25 deaths per 1,000 live births. If we were to repeat the DHS survey in the same population 100 times, in 95 of those times we would have estimated IMR to be somewhere between 20 and 25 deaths per 1,000. We have more certainty about this range of estimates than we do about the single point estimate. As such, confidence intervals are in some ways even more important than our point estimates.

As an aside –the statistical technique called "**bootstrapping**" does exactly this (virtually) – a computer program takes a larger set of data and repeatedly extracts samples from this larger set to run the required calculation, in order to establish the confidence interval.

Confidence intervals for population data

The use of confidence intervals is also very important for population-based estimates, not just for samples. This is especially true when dealing with small numbers of events where stochastic variation is likely to have an impact on the data. Although we are using the whole population of interest for these calculations, we can in fact treat the data as a **sample in time**.

Let's consider this concept in relation to a crude death rate:

Imagine there are 243 deaths in a population of 16,500 people in a single year. This would give a crude death rate of 14.7 deaths per 1,000 population. This data is drawn from midnight on the 31 Dec (0:00 am on the 1st January), to midnight on the 31 December the following year. But what if we repeated our calculations from 9pm on the 31st Dec to 9pm 31st December the following year, or we used data from 1am on the 2nd January to 1 am on the 2nd January the following year. As births and deaths could happen on any day at any time, both our numerator (the number of deaths) and the denominator (the population) may

be slightly different in each one of these time periods, meaning that even a matter of minutes and hours may produce different results in the same way that drawing a different sample from a population would vary in a survey. In a small population, even one event (one birth or death) could have a large impact on the calculations. We account for this uncertainty by treating our data as a sample in time.

Due to the uncertainty around measures of rare events, such as maternal mortality, infant mortality etc., these should always be reported with confidence intervals in order to not be misleading.

95% confidence intervals

The most common confidence interval used is the 95% confidence interval, although you will occasionally see 90% and 99% confidence intervals depending on the use of the data (for example, in a drug trial, we may wish to be more certain about the effect of a new drug and select a 99% confidence interval).

The bigger our sample size or population, the more certain we are of the results, and the smaller the confidence intervals will be.

The upper and lower bounds of a 95% confidence interval are the 95% confidence limits.

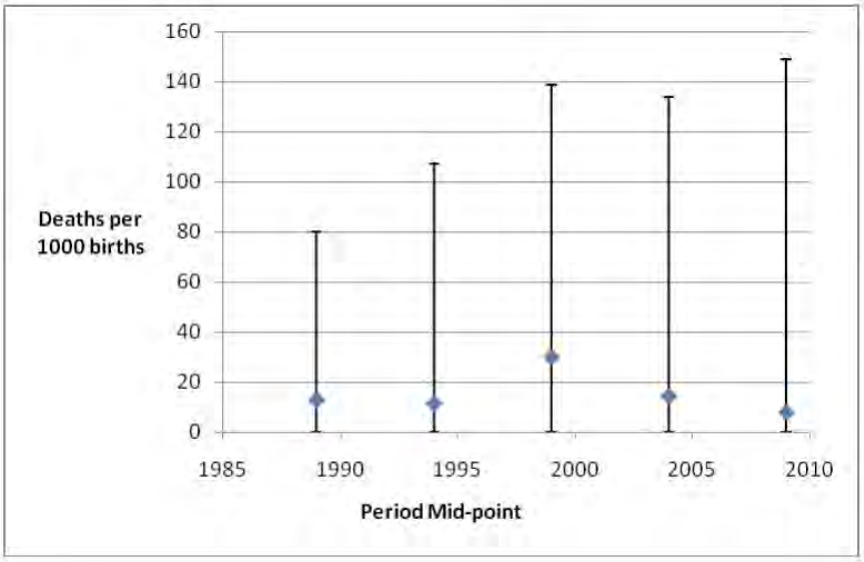
8.1 USING CONFIDENCE INTERVALS AND TRENDS FOR COMPARISONS

The 95% confidence interval is especially helpful when comparing two indicators, either between different populations or over time. This is because indicators of different values may in fact have overlapping 95% confidence intervals, which would mean that they are not statistically different. For example, if the infant mortality rate five years earlier was 60, with a 95% confidence interval of 50 to 70, and a later survey found an IMR of 45 with 95% confidence intervals of 35 to 55, then the confidence intervals overlap between the two surveys and we cannot say the IMRs are statistically different. Therefore, we cannot be certain that there has been a decline in infant mortality rates, even though our point estimates indicate otherwise. Hence, it is very important to report the confidence intervals.

A statistically significant difference is one where the confidence intervals of the measures being compared do not overlap. When reporting point estimates, you should also report their **level of certainty** - i.e. 95% certainty estimate falls within 'x-y' upper and lower bounds if 95% confidence intervals were used.

For example, have a look at the graph below of the IMR in Niue.

Figure 8.1 Infant mortality rate for Niue (1987-2011)



As can be seen in the graph, although IMR appears to have increased in the five year period around the year 2000, this was not a statistically significant result given the small population size and the subsequent uncertainty in the calculations. As such, we are unable to tell from confidence intervals alone whether there is a problem that needs to be investigated. The small populations seen in the PICTS will almost always generate vital statistics with confidence intervals that are not statistically significant, but this phenomenon is not useful to policy makers for monitoring health. It is therefore necessary to also take into account any long-term trends that may be occurring over time.

Looking at the long term trend in the IMR in Niue, we can see that there has been some slight downward trend since 2000, but that over the whole period of interest, there has been little evidence of an ongoing trend. It will be interesting to see whether the possible downward trend from 2000 is confirmed in the next time period (which will not be available until 2017 as the mid-point will occur in 2015).

You can see from the above example that it is not possible to examine a trend from 2 data points or even with three. We might find that we are over-interpreting small stochastic variations and our conclusions are not valid.

As a general principle, the more points we are able to work from, the more reliable our interpretation of trends is likely to be. It is good practice to use 4 or more points wherever possible to analyze trends over time.

8.2 COMPUTING CONFIDENCE INTERVALS

Computation of 95% confidence limits requires the standard error to be computed.

Where complex survey designs are used, the standard error is computed using advanced techniques.

Because CRVS data analysis is based on population data, there are two methods which can be used to generate confidence limits for most of the measures we are interested in. These are the **normal approximation of the binomial**, and **Poisson confidence intervals**.

Calculating confidence intervals using the normal approximation of the binomial

The normal approximation of the binomial can be used to calculate CI's for most rates (including crude death rates, age-specific mortality rates etc.). It cannot be used for small numbers - especially when there is less than 30 cases in the numerator. For anything less than 100 in the numerator, consider using the Poisson distribution.

The standard error for a proportion can be computed as:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

Where *p*=proportion and *n*=population.

For example, if we have 50 cases of diarrhoea from a population of 250 children.

p = 50/250 = 0.2 or 20% (rate of diarrhoea)

n = 250

$$SE = \sqrt{\frac{0.2(1-0.2)}{250}}$$

SE = 0.025

The 95% confidence interval is then computed as:

CI = p +/- $z_{1-\alpha/2}$ x SE

For a 95% CI this means

CI = p +/- 1.96 x SE

For the above example,

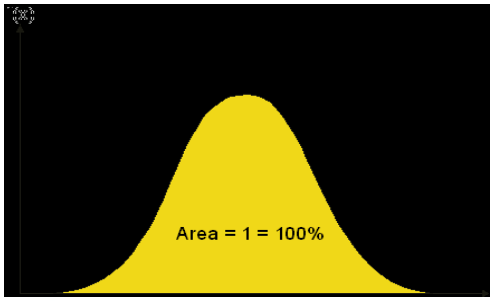
CI = 0.2 ± (1.96 x 0.025) = 0.15 to 0.25

Therefore, we are 95% confident the rate of diarrhoea, which we measured as 20%, lies between 15% and 25% in the population.

Further information

The value of Z comes from the normal distribution.

Remember, the total area under a normally distributed curve = 1.0



So each Half of the Area Under a Normally Distributed Curve = 0.5

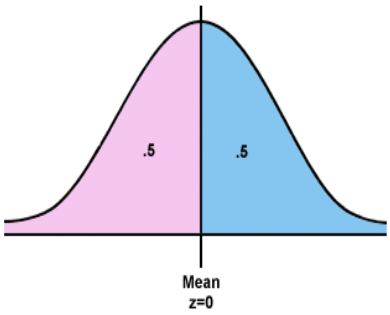


Figure source: http://daphne.palomar.edu/stat/nc1.gif

The α symbol comes from the confidence interval. If we want a 95% confidence interval for example, we want to include an area of 0.95 of the normal distribution. α is the area that we want to exclude or 0.05. As there are two equal sides of the normal distribution, we need to divide this in two to give a value $\alpha/2 = 0.025$

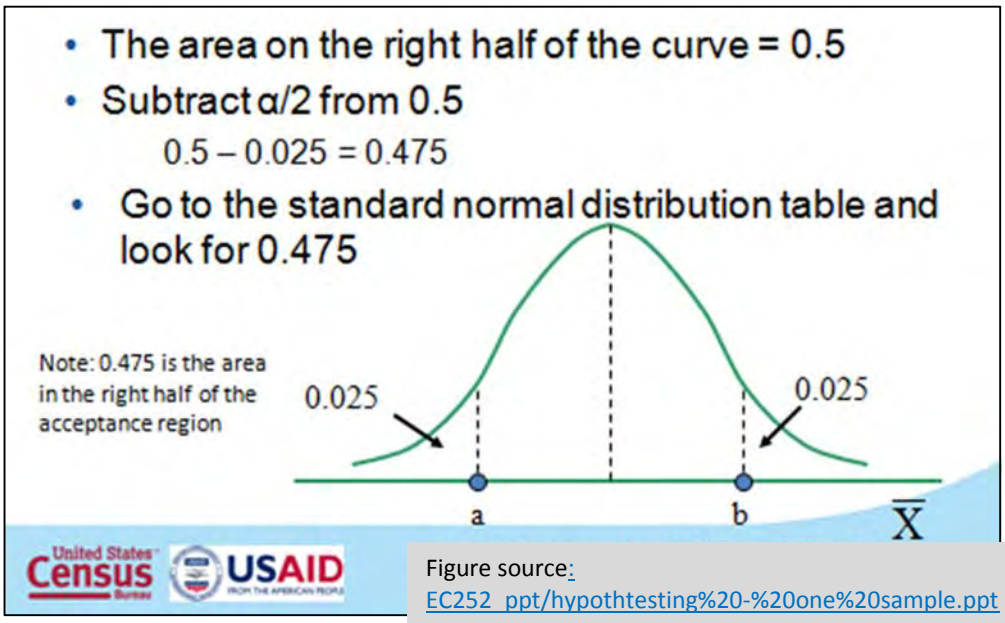


Figure source: EC252_ppt/hypothtesting%20-%20one%20sample.ppt

We can then look up the area on the positive half of the graph that we want – in this case 0.475 in the Standard Normal Distribution Table.

When we find the value, it is in row number 1.9 and heading 0.6. Adding these, we get 1.96.

$$z_{\alpha/2} = 1.96$$

Calculating confidence intervals based on the Poisson distribution.

The Poisson distribution is used when either:

- The numerator is very small (generally less than 100); and/or
- The numerator and denominator come from different sources

In the calculation of the infant mortality rate, the numerator - deaths in children aged less than 12 months - is divided by the number of live births in that year. The children who are at risk of dying aged less than 12 months, however, may have been born in the previous year. Additionally, children who were born in the current year (and which are therefore part of our denominator) remain "at risk" for dying for part of the following year before they reach 1 year of age. The numerator and denominator are therefore derived from different populations.

In IMR - The denominator – live births in the year – is used to approximate the size of the cohort of the children at risk in that year.

Calculating confidence intervals using this method is very easy.

We are in fact finding the confidence interval about the numerator – or the number of events.

We can do this simply by looking up the values on a Poisson distribution table as shown in the following page.

We then simply re-calculate our rates or measures again using the upper and lower confidence limits of the numerator.

Example:

For example, if our IMR was originally based on 20 deaths out of 2500 live births this would give us:

IMR =(20/2500) x 1000 = 8 deaths per 1,000 live births

If we look 20 up in the following Poisson table, we can see that the confidence limits for our number of deaths are 12.2165 and 30.8884. We then simply substitute these into the equation to get:

Lower limit 95% IMR = (12.2165 / 2500) x 1000 = 4.9
Rounded to 5 deaths per 1,000 live births

Upper limit 95% IMR = (30.8884 / 2500) x 1000 = 12.4
Rounded to 12 deaths per 1,000 live births

Our IMR is therefore 8 (95% CI: 5 – 12) deaths per 1,000 live births.

EXCEL

Poisson confidence limits for CRVS rates can be calculated in excel using the formulas below:

```
[LL] =IF(A1>0,(CHIINV(0.975, 2 * A1) / 2),0)

[UL] =IF(A1>0,(CHIINV(0.025, 2 * (A1+1) )/ 2)/, “NA”)
```

Note that these formulas only apply to data where the time frame is 1 year or data has been aggregated over time to give an average per year. If a different unit of time such as person-years (sometimes used in health studies) was to be used, the functions would need to be amended.

These have been programmed into the spreadsheets provided.

Table 8.1 Poisson lower (L) and upper (U) confidence intervals *(1 in 20)*

Source: www.nwpho.org.uk/sadb/ Poisson%20CI%20in%20spreadsheets.pdf

Count	L 95% CI	U 95% CI		Count	L 95% CI	U 95% CI
0	0.0000	3.6889		51	37.9729	67.0556
1	0.0253	5.5716		52	38.8361	68.1911
2	0.2422	7.2247		53	39.7006	69.3253
3	0.6187	8.7673		54	40.5665	70.4583
4	1.0899	10.2416		55	41.4335	71.5901
5	1.6235	11.6683		56	42.3018	72.7207
6	2.2019	13.0595		57	43.1712	73.8501
7	2.8144	14.4227		58	44.0418	74.9784
8	3.4538	15.7632		59	44.9135	76.1057
9	4.1154	17.0848		60	45.7863	77.2319
10	4.7954	18.3904		61	46.6602	78.3571
11	5.4912	19.6820		62	47.5350	79.4812
12	6.2006	20.9616		63	48.4109	80.6044
13	6.9220	22.2304		64	49.2878	81.7266
14	7.6539	23.4896		65	50.1656	82.8478
15	8.3954	24.7402		66	51.0444	83.9682
16	9.1454	25.9830		67	51.9241	85.0876
17	9.9031	27.2186		68	52.8047	86.2062
18	10.6679	28.4478		69	53.6861	87.3239
19	11.4392	29.6709		70	54.5684	88.4408
20	12.2165	30.8884		71	55.4516	89.5568
21	12.9993	32.1007		72	56.3356	90.6721
22	13.7873	33.3083		73	57.2203	91.7865
23	14.5800	34.5113		74	58.1059	92.9002
24	15.3773	35.7101		75	58.9923	94.0131
25	16.1787	36.9049		76	59.8794	95.1253
26	16.9841	38.0960		77	60.7672	96.2368
27	17.7932	39.2836		78	61.6558	97.3475
28	18.6058	40.4678		79	62.5450	98.4576
29	19.4218	41.6488		80	63.4350	99.5669
30	20.2409	42.8269		81	64.3257	100.6756
31	21.0630	44.0020		82	65.2170	101.7836
32	21.8880	45.1745		83	66.1090	102.8910
33	22.7157	46.3443		84	67.0017	103.9977
34	23.5460	47.5116		85	67.8950	105.1038
35	24.3788	48.6765		86	68.7889	106.2093
36	25.2140	49.8392		87	69.6834	107.3142
37	26.0514	50.9996		88	70.5786	108.4185
38	26.8911	52.1580		89	71.4743	109.5222
39	27.7328	53.3143		90	72.3706	110.6253
40	28.5766	54.4686		91	73.2675	111.7278
41	29.4223	55.6211		92	74.1650	112.8298
42	30.2699	56.7718		93	75.0630	113.9313
43	31.1193	57.9207		94	75.9616	115.0322
44	31.9705	59.0679		95	76.8607	116.1326
45	32.8233	60.2135		96	77.7603	117.2324
46	33.6778	61.3576		97	78.6605	118.3318
47	34.5338	62.5000		98	79.5611	119.4306
48	35.3914	63.6410		99	80.4623	120.5289
49	36.2505	64.7806		100	81.3640	121.6268
50	37.1110	65.9188				

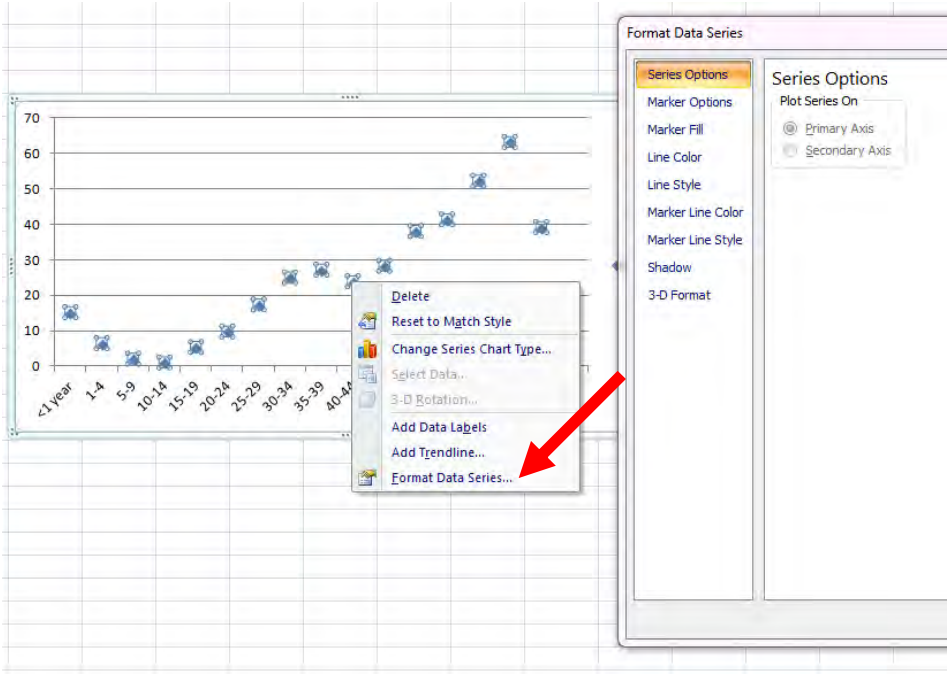
8.3 ADDING TREND LINES AND CONFIDENCE INTERVALS IN EXCEL (2007)

Start by creating a line graph of your data.

To add a trend line

First remove the point to point line that excel uses to create a plot diagram

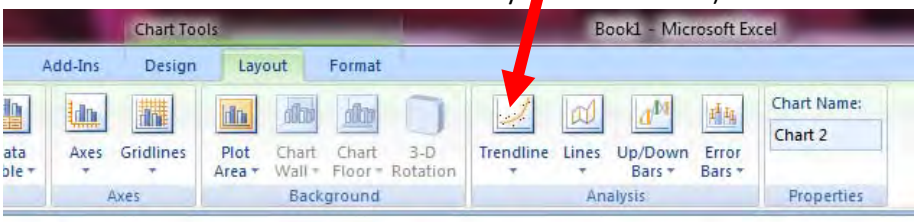
Right click on your data (on the line) to open the dialogue box and select format data series



Change the marker option to something that can be seen

Click on line style and select “no line”. In Excel 2013, click on the “Fill and Line” icon located on under the format data series on the top right of the screen), the select “No line”:

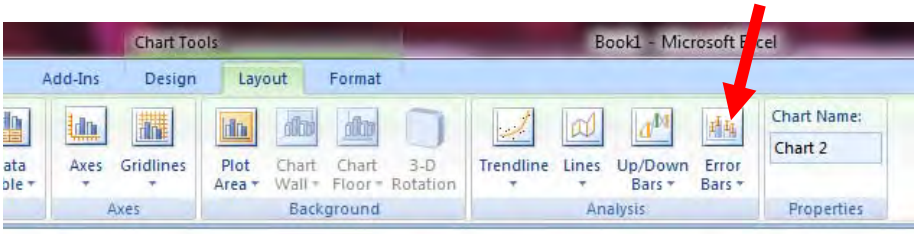
Again, click on your graph and select the “add trend line” option from layout tab under chart tools. In Excel 2013, click on the chart, three buttons appear just above the upper-right hand corner. Click on the options “trend line and select the kind of trend line you wish to insert)



You can then select the type of trend line you wish to insert. The most common options would be a linear trend or a moving average.

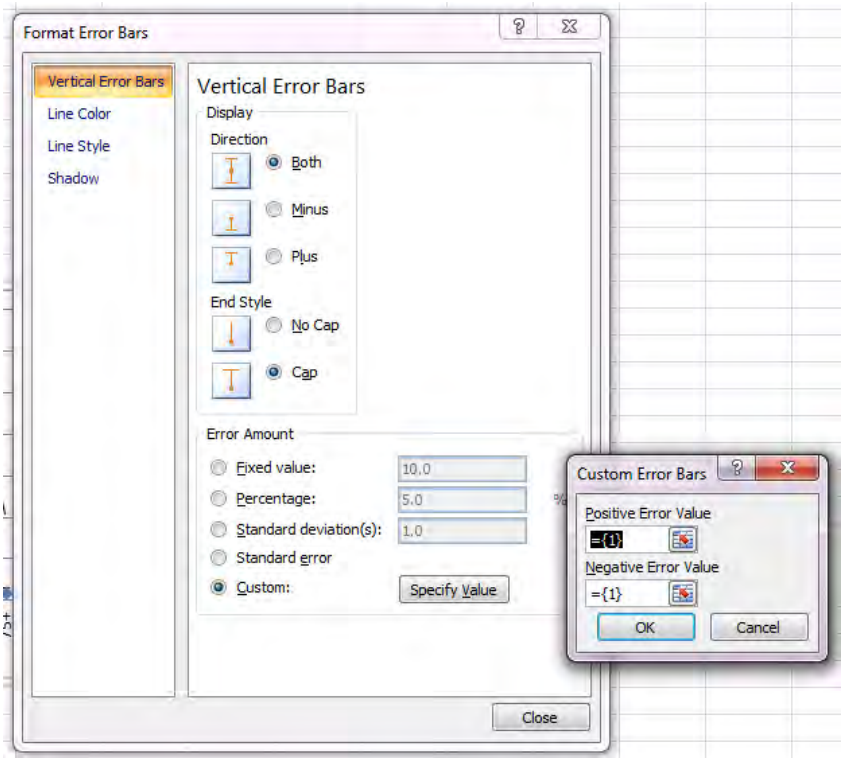
To add confidence intervals

You will also find the “error bars” button which can be used to add confidence intervals under the Layout tab in Excel 2007. In Excel 2013, when you click on the chart, three buttons appear just above the upper-right hand corner. Click on the “plus” button to add new chart elements – you will find error bars here. Check “Error Bars” and error bars will appear on your chart



Although excel has some standard error bars that it can add automatically, these are best avoided so that you can add in the confidence interval that you have calculated as described earlier in this chapter. Select “other options” in the box that comes up. This will take you to a dialogue box as seen below. .

For excel 2013 a task pane opens on the right side of the screen when you click on the error bars in the chart. This pane lets you customize the range and formatting of the error bars so that you are able to apply your own confidence intervals. Click on one of the error bars, and then on the “bars” icon in the task pane to see range options:

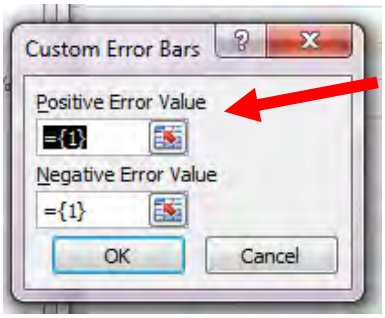


Select the custom option and click on “specify value”. This will open a box allowing you to set up the custom error bars.

The error value as used by excel is the difference between your point estimate and the confidence interval value you have determined. In order to graph these, you will first need to add columns for the upper and lower confidence intervals to your table of source data, and then to insert two additional columns for the positive and negative error values.

The positive error is given by the upper confidence interval minus the point estimate, while the negative error value is given by the point estimate minus the lower confidence limit.

Once you have set your source data table up to include these columns, you can select the upper and lower error values by entering the cell ranges in the boxes or click on the highlighted button (shown below) and select the appropriate cell range.



Lab – Confidence Intervals

Go back through the measures that you have already calculated for fertility and mortality and calculate confidence intervals for the following:

- Crude birth rate
- Total fertility rate
- Crude death rate
- Neonatal mortality rate
- Infant mortality rate
- Under 5 mortality rate
- ASMR

Start by selecting which method you would use for each given the data available (normal approximation of the binomial or Poisson), then use the supplied spreadsheets to calculate these.

Where appropriate, go back and add the confidence intervals to any graphs that you may have created for these measures.

Write a couple of sentences to record what methods you have used, and what the confidence intervals and review of trends shows for each of these. These can then be incorporated into your vital statistics draft report.

9 SUMMARY MEASURES OF MORTALITY

Adapted from course notes from Dr Timothy Adair and Dr Chalapati Rao, University of Qld & from: Carter K. 2013. Mortality and Causes of Death in the Pacific. PhD Thesis. UQ, Brisbane.

In the previous chapters we have looked at the limitations of measures such as crude death rates, and the need to account for the age structure of the population on the mortality rates. While we can do this by looking at age specific mortality, this is often too long or complex for policy makers who just want one single figure that can tell them how healthy (or not) the population is. We have looked at one way of generating this sort of summary statistic using age-standardised mortality rates. However age-standardised rates are only meaningful in comparison to other data which is standardized to the same population; they are not meaningful on their own. As such, they may not be as readily understood by policy and decision makers.

In this chapter we will look at several other summary measures of all-cause mortality which are commonly used, and how these are calculated.

Measures that we will look at include:

- Life expectancy at birth
- Life expectancy at 40
- Years of life Lost, and
- Adult mortality.

9.1 LIFE EXPECTANCY AND LIFE TABLES

Life expectancy at birth is one of the most commonly used measures to describe the health status of a population.

Definition:

Life expectancy is an estimate of the average number of years a person can expect to live, based on age-specific death rates in a given year.

In reality of course, people are exposed to different conditions over the course of their lives (and as mortality rates change over time), so that the estimate of how many years on average someone would live is hypothetical, and assumes that mortality rates throughout the course of their life stay exactly as they are now. The good news is of course, that as time passes, we hope that society becomes healthier, and in practice the average number of years lived should actually be higher than estimated by this measure.

However, it is generally understood well by policy makers, accounts for changing mortality rates across ages and is directly comparable to other populations.

Life expectancy is computed using a life table, which converts observed population based mortality rates (as you looked at earlier for age-specific mortality rates) into risk of dying at each age. It is conventionally calculated separately for males and females.

Life expectancy can be calculated at any age. The most common measure is life expectancy at birth (written as LE_0) which is the average number of years a person could be expected to live in total.

Life expectancy at 40 (LE_{40}), of the average number of years a 40 year old person could expect to live, is also included in the national minimum development indicators, and is useful for understanding the impacts of high adult mortality (such as from non-communicable diseases) on a population.

A complete life table uses individual years of age. However, there is usually insufficient data to do an analysis by single year age groups, and an abridged life table is used. Abridged life tables are based on assumption that death rates are similar at neighbouring ages. If data allows, it is useful to be able to leave the first year of life as a category on its own, followed by 1-4 years, and then 5 year age groupings if data allows. Countries with very small population numbers may need to use 0-4, then ten year age groupings in order to have sufficient data for stable populations.

Go back to the graphs of deaths by age group that you did in Chapter 13. Use your graphs to determine if there is sufficient information to calculate a life table, or if you will need to further aggregate your data (either by using broader age groups or longer time periods). Remember your data can have gaps but must have values for the highest and lowest age groups and should have sufficient data that the broad shape is recognizable.

9.2 SETTING UP A LIFE TABLE

Before we go further into our calculations – we need to take a look at the format and symbols used to describe each of the elements we are interested in.

Each element (such as mortality, probability, years lived) etc. is represented by a letter. N and x are used to note the number of years in the interval we are interested in, and the starting point (or age) of the interval. For example, the symbol below refers to mortality rate:

$$ {}_nM_x $$

Therefore, if we were to use the notation as follows: ${}_{10}M_5$, this would mean the mortality rate starting from the beginning of 5 years of age for 10 years – or mortality from 5 to 14 (at ages 5, 6, 7, 8, 9, 10, 11, 12, 13, and 14).

This same format is used for each element or measure in the life table.

An example life table is shown below:

Table 9.1 Example of life table drawn from Nauru

Age	X	Years in Interval	Linearity Adjustment	Reported Population	Deaths	Mortality Rate	Probability of Dying	Theoretical Cohort	Deaths in Interval x	Years Lived in Interval x	Cumulative Years Lived	Expectancy of Life at Age x
x		n	a	$_nN_x$	$_nD_x$	$_nM_x$	$_nq_x$	l_x	$_nd_x$	$_nL_x$	T_x	e_x
0-4	0	5	0.3	3269.331	39.485	0.012078	0.0579	100000	5794	479722	5281224	52.8
5-14	5	10	0.5	6702.471	6.075	0.000906	0.0090	94206	850	937812	4801502	51.0
15-24	15	10	0.5	5786.59	13.162	0.002275	0.0225	93356	2100	923064	3863691	41.4
25-34	25	10	0.5	3841.017	24.299	0.006326	0.0613	91257	5596	884586	2940627	32.2
35-44	35	10	0.5	2585.76	41.510	0.016053	0.1486	85661	12730	792958	2056040	24.0
45-54	45	10	0.5	1652.807	54.672	0.033078	0.2838	72931	20701	625806	1263083	17.3
55-64	55	10	0.5	529.6471	36.448	0.068816	0.5120	52230	26742	388595	637277	12.2
65-74	65	10	0.5	216.3351	17.212	0.07956	0.5692	25489	14508	182349	248682	9.8
75+	75	10	0.5	79.11543	11.137	0.140768	1.0000	10981	10981	66333	66333	0.0

Data from Carter, K., Soakai, T. S., Taylor, R., Gadabu, I., Rao, C., Thoma, K., & Lopez, A. D. (2011). Mortality trends and the epidemiological transition in Nauru. Asia-Pacific Journal of Public Health, 23(1), 10-23.

Looking at the table above – we will work through each of the columns to discuss what is shown. Note that while life tables may be set up slightly differently depending on the source you look at – they will have each of these elements in their calculations (even if they are not all shown)

Age - X

The first column is simply the age group that we are interested in, with the second column labelled X the age at the start of the age group.

In most abridged life tables it is good practice to set up using separate age groups for <1 year olds, 0-4 year olds, and then 5 year age intervals up to a final age group of 85+ or higher (depending on the data available). However, as shown in the table above, where populations are small (and therefore, there are only a few deaths in each age group even when the data is aggregated over several years) it is better to use a combined 0-4 year age group. It may also be necessary to use 10 year age aggregations (as shown here), or a lower final age group (such as 75+).

When setting the final open age group, it is important not to start this group at a lower age than what we estimate life expectancy might be as it will tend to lower our final estimates.

Years in Interval - n

This is the number of years in the age group used. For example, an age group of 0-4 would have 5 years in the interval. An age group of <1 year would have an interval of 1.

Linearity Adjustment - a

This is a co-efficient that is used to indicate when in the interval the deaths occur on average. Although deaths may occur at any time over this 5 year age interval, we assume that these average out. In most age groups, we assume that this means that on average, deaths in the interval occur half way through the interval, and "a" would then equal 0.5. For a 5 year age interval such as from ages 25-29, this will mean that while some deaths occur at each age (25, 26, 27, 28 and 29) and subsequently some people live longer through the time period than others, we can assume that on average, we can treat these deaths as though all of them happened at age 27 – half way through the time period (with a denoting a half or 0.5). This works because mortality rates at similar ages are very similar.

This assumption does not however work for ages under 5. As we have seen in our previous chapters, deaths in under 5 year olds are more likely to happen in infants under 1 year. Similarly, within the deaths aged less than 1 year, children are more likely to die in the neonatal period less than 28 completed days. Go back and have a look at the rates you calculated for U5M, IMR and NMR to see this effect. What this means is that we must use a lower adjustment factor to account for the deaths occurring earlier in the age group than we do with older ages.

It is convention to use 1a0=0.1 in low mortality countries and 0.3 in high mortality countries. For all 4a1, use 0.4.

This linear adjustment, may also be called a "separation factor" in some sources.

Probability of surviving $_nP_x$ (not shown on example table)

The inverse of the probability of dying is the probability of surviving. This is usually denoted by the letter "p" (but should not be confused with proportion which also uses this notation but means something entirely different– such as for our calculation of confidence intervals!)

Definition: probability of surviving between exact age x and x+n

$$_n p_x = 1 - _n q_x$$
$$\therefore {}_n p_x + {}_n q_x = 1$$

Theoretical Cohort l_x

Up to this point in the life table we are talking about real populations. However, if we think of our definition of life expectancy, we need to work out how long would people live on average if they were exposed to current conditions over their whole lives. Since this does not happen in reality, we now apply the current or real conditions we have just calculated for the probability of dying and probability of surviving each age group to a group of hypothetical (or imaginary) people. This is our theoretical cohort.

Definition: the number of people alive at exact age x.

This refers to an exact age at the start rather than an age interval (unlike other columns in an abridged life table)

The first value, l_0, is an arbitrary number called the radix (usually a round number such as 1, or 1000, or 100,000). We use 100,000 imaginary people in order to make the calculations easier!

Calculating l_x

First choose a radix then work down the table using:

$$l_x = l_{x-n} \square_n \ p_{x-n}$$

For the last (open-ended) interval, the number of persons dying is the same as the number alive at its start (because everyone has to die)

$$d_{x+} = l_x$$

l_x bears no relation to the actual number of individuals aged x to x+n in the real population and is only meaningful when related to the original radix

Everything we do from this point of the table, whether we are talking about deaths or years lived now refers to this group of hypothetical people.

Deaths in Interval - d

This is the number of deaths in the theoretical cohort within that age group. It is calculated by applying the probability of dying to the number of people alive at the beginning of the age group.

Years Lived in Interval $_nL_x$

Definition: the total number of person-years lived (contributed by those alive and those who died) between exact ages x and x+n

Every person who lives all the way through the age group contributes the number of years in the age group. In the example table, as we are using ten year age groups, each person who lives through the age group contributes 10 years lived (or in the first age group 5 years for living from 0-4 years).

Every person who dies part way through the age group contributes the time that they were alive within the age group. If we remember back to the linearity adjustment – for most ages this is half way through on average so they contribute 0.5 times the age interval to the years lived. For the younger age groups where deaths occur earlier on average, each person who dies contributes the years in the age interval multiplied by the linearity adjustment factors (a).

Calculating $_nL_x$

nLx is obtained by averaging lx and lx+n and multiplying by the interval width, n

$$_nL_x = n(l_{x+n} + {}_na_x \cdot {}_nd_x)$$

$$_nL_x = n(l_{x+n} + {}_na_x \cdot {}_nd_x)$$

As the last age group has no fixed length however, we cannot use this formula and must estimates how many years are contributed on average by each person before they die

Calculating L_{x+}

$$\therefore L_{85+} = \frac{d_{85+}}{M_{85+}} = \frac{l_{85}}{M_{85+}}$$

Cumulative Years Lived T_x

Definition: the total number of person-years lived after age x

The cumulative years lived refers to the average number of years lived by each person in total after the age at the beginning of age group. It is calculated by adding the $_nL_x$ function from the bottom (highest age interval) up.

Expectancy of Life at Age x - e_x

Definition: expected (average) number of years of life left for a person aged x

For the total number of people alive at age x (l_x), the total number of years left for them to live is given by T_x. This means we can work out how long each person will live on average by dividing T_x by lx. That is:

$$e_x = \frac{T_x}{l_x}$$

We can calculate this for any age in the life table, but the key measure that we are interested in is the average number of years a person will live in total – that is from their birth at x=0. Life expectancy at birth or e_0 is therefore given by the ratio:

$$e_0 = \frac{T_0}{l_0}$$

9.3 USING LIFE TABLES – KEY MEASURES

Now that we have calculated our life tables, there are several other key measures that we can calculate from these.

Life expectancy at 40

Life expectancy at 40 is a key indicator of the impact that premature adult mortality (from causes such as non-communicable diseases like cancer, heart disease or diabetes) is having on the population, especially if accurate cause of death data by age group is not available.

LE_{40} is included in the national minimum indicators for development that countries have committed to reporting.

This figure can be read directly from a life table if it has been set up using 5 year age intervals. If the life table uses ten year age groups, then LE 40 must be calculated by averaging life expectancy at 35 and 45. In the example table given this would be the average of 24 and 17.3 or 20.7 years. This means that at 40, on average people could expect to live a further 20.7 years (or to age 60.7)

Probability of surviving to age 40

The probability of surviving to age 40 is not an indicator commonly used in its own right, but is important in the calculation of the Human Development Index. If the table uses 5 year age groups this can be calculated by l_{40} or the number of survivors in our theoretical cohort at age 40 divided by the radix, or the original size of the cohort. Probability would often be reported as a percentage or per 1000 people, so would need to be multiplied by 100 or 1,000 respectively to do this.

Again, if the life table is in ten year increments, the average of l_{35} and l_{45} would be used to give l_{40} in the above calculation.

Adult mortality ${}_{45}q_{15}$

Adult mortality is another important measure of the impact of premature adult mortality. It is the probability of dying between the ages of 15 and 60 (15 plus 45 as we add the numbers from right to left in

the notation). It is also particularly useful as a comparison across countries and to other data sources as it is routinely calculated as part of census reporting.

To calculate adult mortality the total deaths in the theoretical cohort (not the real numbers of deaths in each age group) are added together and divided by the number of people who were alive at age 15. In other words

$$_{45}q_{15} = {}_{45}d_{15} / l_{15}$$

The deaths from ages 15 to 60 can be calculated by adding the cells for each age group in this interval. Again, if your life table uses 10 year age groups such as the one shown and x=60 falls at the midpoint of a ten year age interval, the deaths in this final age group should be halved before adding them to the other ages and completing the calculation.

In the table shown: our calculation would look like:

$$_{45}q_{15} = {}_{45}d_{15} / l_{15}$$

$$_{45}q_{15} = \frac{(2100 + 5596 + 12730 + 20701 + (0.5x\ 26742))}{93356}$$

$$= 0.589$$

$$= 60\%$$

In words, this would mean that for every 100 15 year olds in this country, 60 would die before reaching their 60th birthday.

Infant and under 5 mortality

As infant and child mortality are both measures of probability rather than population based rates (as discussed previously in this workbook), these can also be estimated from a lifetable if appropriate age groups have been used.

To calculate the IMR, the life table must have a separate age group for 0-1 year olds, in which case the probability of dying or $_1q_0$ is read directly off the table and multiplied by 1,000 in order to report the figure per 1,000 live births.

In a table set up this way, U5M would be calculated as

- $_5d_0$ = theoretical deaths in these age groups / divided by the radix
- = $(_1d_0 + {}_4d_1)$ /100,000

In a table with an initial age group of 0-4, U5M is the probability of dying in the first age group multiplied by 1000 to be reported as per 1,000 births.

In the example table supplied, $_5q_0$ is 0.0579 so the U5M would be estimated to be 57.9 deaths per 1,000 live births.

Years of Life Lost (YLL)

This is a measure of the time lost to early deaths in comparison to an "expected norm". It is calculated by adding up for each age group: the real deaths (d) in the age group multiplied by (the standard life expectancy minus the calculated life expectancy for each age group).

The choice the standard is usually

- a comparison population considered to have the "best possible" life expectancy (Japanese usually used for this standard). or
- a theoretical "potential limit to life"

9.4 CONFIDENCE INTERVALS FOR LIFE TABLES AND LIFE EXPECTANCY.

While there are several approaches for measuring confidence intervals for life expectancy, the most widely used approach (and approach that has been demonstrated as most appropriate for small populations which may have some missing information in the life table) was derived by Chiang and is commonly called the Chiang method. These are based on the variance around q (the probability of dying). We will not go through how these calculations are done in detail, but these calculations are built in to the provided spreadsheets and should be used to calculate a CI for your LE calculations.

Exercise

Using the templates provided, construct a life table for your data and calculate (with confidence intervals) LE0, LE40, U5M or IMR and adult mortality. Explain what your results mean in words, and discuss the plausibility of these results in comparison with previous reports or other sources such as the census.

9.5 MODEL LIFE TABLES

Extract adapted from: (adapted) from Manual for "Training in the Use of Existing Health Datasets" [Adair, T. 2012]. Additional material sourced Carter K. 2013. Mortality and Causes of Death in the Pacific. PhD Thesis. UQ, Brisbane.

Model life tables can be used in several ways: Firstly they can be used to "smooth" data where small numbers or incomplete reporting mean we do not have a full distribution of deaths by age (remember the log-graphs of age-specific mortality from chapter 13?).

Secondly, we can use either a single input (such as infant or child mortality), or two inputs (child mortality, plus adult mortality) to estimate the full life table and subsequently LE. This is frequently used in census and survey analysis using indirect measures of IMR or U5M (from questions on children ever born vs children surviving) and occasionally adult mortality 45q15 (from orphan-hood or widowhood questions).

Model life tables are life tables that are usually based on observed data in actual countries. The first models were developed by the United Nations in the 1950s.

Typically, model life tables have been derived from a group of countries presumed to have high quality data and which share distinctive age-patterns of mortality. What do we mean by distinctive age patterns? Remember that a particular level of life expectancy is consistent with any number of different age patterns of mortality. For instance, if two countries have the same life expectancy, one country might have both higher infant mortality and lower adult mortality than the other country.

A set of model life tables was developed by Coale and Demeny (1968) in the 1960's. These models classify the life tables into four different sets, labelled West, East, North, and South, according to the patterns of mortality in the predominant regions of Europe represented in the original data. In each of these sets, the life expectancy at age 10 was correlated with the probability of dying at different ages, and these correlations provided the basis for estimating a series of "nested" life tables at different overall levels of e(x) but with different age patterns of mortality.

At about the same time as the Coale-Demeny models appeared, Brass (Brass et al., 1968) developed the General Standard Model. In this model, life tables are related by a two-parameter logit system, based on a general pattern of mortality and a linear relationship between two sets of logits. By changing the two parameters of the linear relationship, the general pattern is changed in relation to the level and pattern of mortality. Brass also introduced his African Standard in the same book. The African Standard differs from the General Standard by its relatively high childhood mortality vis a vis infant mortality (Brass et al., 1968:113).

Coale-Demeny, in addition to developing regional model life tables, also derived stable populations corresponding to each of the four regions. Stable populations occur when mortality and fertility are constant and there is no migration. Stable populations were derived based on various combinations of mortality and fertility and have useful applications in demographic analysis.

Most of the empirical life tables used in constructing the models described above pertained to developed countries. In the 1980s, the United Nations published a new set of life table models, based on the mortality experience of developing countries with reliable information (Heligman, 1984 and United Nations, 1982). Empirical life tables for developing countries were grouped into four sets, plus a general set including all of them. These sets refer to the Latin American, Chilean, South Asian, Far Eastern, and General patterns.

While model life tables can be very useful, it is important that we recognize their limitations. When using model life tables, it is also good practice to always report what inputs you have used, and which model was used to derive your estimates.

A number of small area analysis studies in both the UK and NZ have demonstrated that "smoothing" of the age-specific mortality rates is not necessary to generate reliable measures of LE provided there is sufficient data in the early age groups and the last, open-ended category is not empty. The calculations are able to handle zero deaths in other age groups. As such, for analysing our CRVS data, we will focus on ensuring that there is appropriate aggregation of data across years, and reporting empirical measures.

There are a number of computer programs that can be used to generate model life tables.

One of the easiest to use, although it runs as an add on to STATA which not all offices would have, is the WHO modmatch program. Other commonly used examples include MORTPAK.

9.6 HEALTHY LIFE EXPECTANCY AND DISABILITY ADJUSTED LIFE YEARS

Healthy life expectancy (HALE) and Disability adjusted life years (DALYs) and are two composite measures used to provide a single statistic to account for health in the population based on both mortality, and morbidity.

The WHO definitions are as follows:

HALE:

The average number of years that a person can expect to live in "full health" by taking into account years lived in less than full health due to disease and/or injury

DALY:

One DALY can be thought of as one lost year of "healthy" life. The sum of these DALYs across the population, or the burden of disease, can be thought of as a measurement of the gap between current health status and an ideal health situation where the entire population lives to an advanced age, free of disease and disability. DALYs for a disease or health condition are calculated as the sum of the Years of Life Lost (YLL) due to premature mortality in the population and the Years Lost due to Disability (YLD) for people living with the health condition or its consequences:

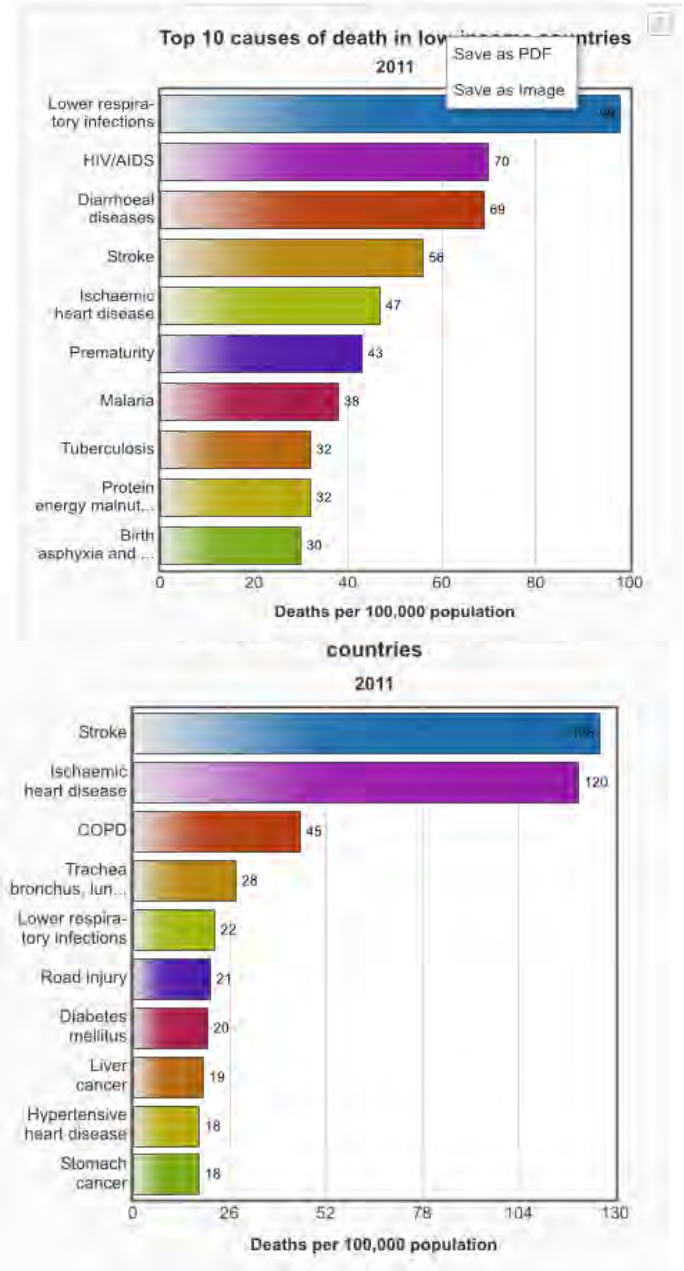
Both measures are central to the concept of being able to measure the Burden of Disease in a population. Although we will not be calculating these measures as part of this course, they are mentioned here for information. It is important to note that both rely on reliable mortality and morbidity data.

10 CAUSES OF DEATH

Adapted from: Carter K. 2013. Mortality and Causes of Death in the Pacific. PhD Thesis. UQ, Brisbane.

Understanding why people die is central to being able to identify interventions and policies to address and hopefully reduce premature mortality.

Figure 10.1: Global leading causes of death (WHO data, 2011)



If we look at data from the WHO on the leading causes of death in the world (shown to the left), we can see that the leading causes of death vary significantly from lower income countries to higher income countries. Infectious diseases top the list for lower income countries, while non-communicable diseases (stroke and ischemic heart disease) stand out as the primary causes of death in higher income countries.

As countries develop, we tend to see this shift from infectious diseases to non-communicable diseases. In some Pacific Island Countries, both infectious and non-communicable diseases are important, and the country is said to have a "double-burden" of disease.

While this sort of overall leading cause data is useful to give a broad level picture, it is not detailed enough to be very useful for policy planning and evaluation.

Causes of death vary significantly across ages, and in adults, vary significantly between sexes. It is therefore necessary to report cause of death separately for males and females and by broad age group in order to be meaningful.

Leading causes are also affected by how diseases are grouped. The international classification of diseases (ICD) provides a standard classification system for grouping diseases (coding and aggregation) in order to allow comparability.

10.1 Medical Certificate of Death

The medical certificate requires doctors to record "to the best of their ability" the causal sequence that lead to the death (starting from the immediate cause and working back through the chain of events) (WHO 2010). There is also space to record any other diseases that contributed to the death but were not strictly causal. In practice, certification has been found to be affected by many factors including the specialty of the physician, their experience and beliefs, as well as their education and training on the subject.

The international standard medical certificate has two parts. Part I is the causal sequence leading backwards from the immediate CoD in line one. Part II is for conditions that were present at the time of death, but did not directly lead to the death.

Figure 10.2: International standard medical certificate of death format (WHO)

Cause of death		Approximate interval between onset and death
I Disease or condition directly leading to death *)	a)..... due to (or as a consequence of)	.....
Antecedent causes Morbid conditions, if any, giving rise to the above cause, stating the underlying condition last	b)..... due to (or as a consequence of)	.....
	c)..... due to (or as a consequence of)	.....
	d).....	.....
II Other significant conditions contributing to the death, but not related to the disease or conditions causing it	.....	.....
	.....	.....
**This does not mean the mode of dying, e.g. heart failure, respiratory failure. It means the disease, injury, or complication that caused death.*		

Changes in the ICD rules over time have resulted in some shifts in patterns of underlying cause, making comparison between versions problematic.

10.2 The International Classification of Diseases

The international classification of diseases (ICD) is determined by an international panel and overseen by the World Health Organisation. It provides a standard classification for converting the text provided in a medical certificate into an code for the cause of death, and ensures that data is comparable over space and time. The current version of the ICD is version 10, although ICDv11 is currently under development.

Coding of causes of death

Standardised coding rules are used to establish the underlying cause of death from causes listed on the death certificate. This is done so to ensure comparable data over time and between countries. The International Classification of Diseases (ICD) is the standardised classification system used to classify causes of death.[12] Currently the 10th Revision being used (ICD-10). In ICD-10 deaths are classified according to an alpha-numeric code. In the ICD-10 deaths are classified according to chapters. Chapters include: A00-B99 Certain infectious and parasitic diseases, C00-D48 Neoplasms, etc.

ICD coders are generally Ministry of Health staff. They analyse a completed MCCD, assign the appropriate underlying cause of death (given the causes reported as leading to death) and report the ICD code to that underlying cause. Training is also important for ICD coding staff, to ensure the underlying cause of death is properly assigned.

Of the PICTs that currently code underlying cause of death, only the American territories (Guam, CNMI and American Samoa) do not use ICD 10. Deaths for these three jurisdictions are currently coded at the US national level using automated software based on ICD 9.

Underlying cause of death

The underlying cause of death should be used to determine which cause to code and tabulate your cause of death statistics. Using the underlying cause of death is based on the concept of where in the causal sequence you would INTERVENE to prevent that happening to someone else. It is not the same as Principal Cause or Diagnosis as diagnosed by hospital medical records.

Each death should be counted once only, and there are rules for extracting the underlying cause and code.

When Note: tabulating data for cause of death analysis, it has been common practice in the Pacific Islands to tabulate on the first line of the certificate. This is NOT a valid approach as it ignores the original sequence that led to the death occurring.

Remember to always cite the source your data came from in the report.

[12] WHO, International Classification of Diseases (ICD), http://www.who.int/classifications/icd/en/

ICD 10 General Mortality List 1

There are thousands of codes in the ICD, and for most uses, tabulating at the individual code level is too detailed to be general useful, particularly in small populations where the total number of deaths is low. As such, data is generally either tabulated to the chapter level (which is very broad), or more usefully for health planning, to the ICD General mortality list 1 (also known as the 104 list). This list is given in the appendices.

10.3 ACCURACY OF CAUSE OF DEATH DATA

The accuracy of cause of death data varies according to a number of factors. Of course the quality of information used to certify the cause of death will have a major impact on the accuracy of the cause of death. Some diseases more certified more accurately than others. For example, some diseases are more easily diagnosed than others that require more complex diagnostic procedures. The quality of data is also dependent on the certifier's training, knowledge, attitude and perceptions of the function of death certification. Some disease also associated with stigma may be under-reported, especially those with social implications such as HIV. External cause reporting may be subject to bias, especially where there are medico legal implications. Trends in physician certification can lead to spurious changes in death statistics over time, for example an increasing trend towards the reporting of diabetes as the underlying rather than associated cause with cardiovascular diseases such as stroke and ischaemic heart disease.[13] Other changes may induced by ICD revisions, especially in relation to guidelines for specific underlying causes.

Quality checks for cause of death data.

Before we can analyse cause of death data in a meaningful way, we need to make sure that the data quality is as good as can be.

As such there are a number of quality control checks that should be conducted on the data. Hopefully, we have reviewed and corrected our individual records in order to make sure we have removed stillbirths and duplicates and corrected missing data wherever possible – just as we did for all-cause mortality (and in most cases you will be using the same data set).

As we then start to tabulate our data (by either or both the ICD chapter and the General mortality List 1) we want to make sure that the total number of deaths is consistent with the total deaths that we used to calculate our all-cause mortality. If we are using a separate data set, we would need to explain why the totals are different.

[13] Adair T, Rao C., Changes in certification of diabetes with cardiovascular diseases increased reported diabetes mortality in Australia and the United States. 2010, Journal of Clinical Epidemiology. 63(2):199-204.

We also need to check for obvious errors in the data. Some of the key questions you should ask yourself include:

- Are there any adults in my data that died of perinatal or congenital causes?
- Are there any males in my data that died in childbirth?
- Are there any women that died of prostate cancer?
- Are there any men that died of breast or cervical cancer ?
- Did any children die of diabetes (it is possible for children to die of type I diabetes, but it is extremely unlikely that children will die of the more common type II diabetes)?
- Did anyone die of senility in the middle ages?

Note: it is possible for adults to die of congenital causes and for males to die of breast cancer, but this is extremely rare. Wherever possible, you should check the original source data.

11 MEASURES OF CAUSES OF DEATH

Measures of causes of death.

There are many ways to look at the distribution of causes of death in a population. The most important elements are however that our data is comparable over time and between countries. We should also be able to disaggregate data sufficiently to identify vulnerable populations.

As discussed in the previous section, the ICD provides a framework that allows comparisons to be made at a lot of different levels: from broad chapters to more specific causes. For the Pacific Islands the General Mortality List provides a good level of disaggregation to allow sufficient detail for planning while avoiding most of the problems created by disaggregating to categories that are too small.

IMPORTANT NOTE:

Since we are starting with small number of deaths anyway due to the size of the population, by the time we disaggregate data by age group and sex, we are talking about very small numbers - which if they get too small may not be meaningful. In order to deal with this, it is important to aggregate over at least several years (5 is better than 3). We will also use slightly broader age groups (<5, 5-14, 15-59, 65+) and for children where the causes of death do not differ greatly – will not disaggregate by sex. We will however disaggregate by sex for adults (including young adults) where causes of death are different for men and women.

Common measures used for COD include:

- Numbers
- Rates
- Proportional mortality

11.1 LEADING CAUSES OF DEATH

Perhaps the most common approach to reporting the causes of death is as a ranked list of the leading causes of death with the number of deaths by cause. This is frequently requested both by leaders and governments and for international reporting.

This approach is however not particularly meaningful when deaths are grouped across all ages and both sexes, as the causes of death will be particularly influenced by the age of our population. A population with a high proportion of elderly people will tend to have non-communicable diseases as the leading causes of death. This does not mean however that there is a major health problem from non-communicable diseases unless they are also contributing to significant premature mortality, which we cannot tell from a simple all-age list.

Nonetheless, as decision makers will request this information, it is likely that you will at some point want to present your data this way. If you do so, it is important to always report how many deaths Were attributed to other causes along with a total.

Tips for making your leading cause of death tables more useful are:

- Make sure the total number of deaths is reported
- Consider reporting separately for males and females; and by broad age group
- Lists are only comparable if they use the same level of categorisation. It is recommended that you tabulate and report your deaths using the General mortality list 1.
- It is useful to always report the proportion of deaths for which cause of death data was known alongside your tables.

Note: Cause of death is always more meaningful when reported by age group, and for adults by sex as the causes of deaths in these groups is very different. It is recommended that you use the following groups
< 5 years (combined sexes)
5-14 years (combined sexes)
15-59 years - by male and female
60+ years - by male and female

Adult groups may also be split into other smaller groupings, such as 15 year age groups from age 15 upwards.

Always make sure that you report the proportion of deaths for which no cause of death data was available, or for which the cause of death was ill-defined. Ill-defined deaths are those which either do not provide enough information to assign a cause or the cause assigned does not actually lead to death.

Examples include:

- Senility and old age
- Symptoms (cough, fever etc.)
- Injury without how the injury happened
- Mode of death where no cause is reported - i.e. cardiopulmonary arrest
- Depression

11.2 Proportional mortality

If you have reported your number of deaths and the total number of deaths for which cause is known, it is also possible to report a proportional mortality.

The proportional mortality is simply the number of deaths from that cause divided by the total number of deaths.

More often however, proportional mortality is reported as a proportion of deaths for which cause is known and valid. In this case the number of deaths from that cause are divided by the number of deaths which had a known and valid cause of death.

Again, proportional mortality should be reported by age, and by sex in adults.

11.3 CAUSE-SPECIFIC MORTALITY RATES

In previous chapters we looked at the calculation of crude death rates and age-specific death rates. A cause-specific mortality rate is calculated in the same way, with the number of deaths from that cause in a specific age group divided by the total population in the same age group. Again these would usually be calculated separately by males and females.

Although we are reporting by broad age group, if using wide adult age groupings, it is important to remember that these rates will still be affected by the age distribution of the population, and the older the population, the higher we would expect the rate of NCDs to be. For this reason cause specific mortality is often age-standardised for age groups such as 15-59 year olds in order to account for this influence.

Note; it is important that under count has been adequately corrected for in the number of deaths assigned to each cause for rates to be meaningful. If cause of death is not available for all deaths, then an estimated number of deaths from each cause should be derived by applying proportional mortality by age to the total number of deaths by age. In other words, deaths with unknown cause of death should be re-distributed as you did for deaths of unknown ages. Of course, it is very important that you document what you have done as this is another layer of estimation.

Age standardised cause-specific mortality

If you have proportional mortality and are able to estimate the number of deaths in each 5 year age group from the specific cause, you can calculate an aged-standardised cause specific mortality rate.

Steps in calculating an aged standardised cause specific mortality rate:

- First, decide on your standard population
- For each 5 year age group
    - record the proportion of deaths in the age group (use the broad age group that
    - corresponds to your 5 year age group) and the total deaths
    - multiply the proportion of deaths by cause by the total number of deaths to get estimated deaths by cause
    - divide deaths by cause by the population in that age group to get a mortality rate
    - multiply the cause specific mortality by the standard population to get an expected
    - number of deaths
- Add up all the expected deaths
- Divide this by the total of the standard population (not your own population)
- Multiply by 100,000 to report as a rate per 100,000 people (you could also report per
- 10,000)

11.4 PROBABILITY OF DYING FROM SPECIFIC CAUSES

In order to calculate the probability of dying from specific causes, it is easiest to use the life tables that we used to calculate life expectancy, but to ignore the column on life expectancy (which is not meaningful since we are not using all deaths but only some that we are interested in.

Probability of dying would normally be calculated for selected types of diseases (such as all of the major NCDs combined – and which we will look at in a later chapter), but can be done for any type of disease.

In order to use the life table spreadsheet to do this – we simply enter deaths due to the disease we want to calculate this measure for into the deaths column (instead of the total deaths in the population). Probability of dying can then be calculated in the same way as before- add up the number of theoretical deaths in the ages you are interested in (ndx) divided by the theoretical population (survivors – l) at the beginning of that age interval.

12 MATERNAL MORTALITY

Adapted from SPC's Data analysis and report writing course [Ryan, C. et. al. 2011]

Definition

The standard definition of a maternal death, as defined by the WHO is:

> The death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes.[5]

Maternal mortality can be identified from accurate cause of death data.

The sisterhood method is often used by surveys to measure maternal mortality, and is based on a set of questions to a woman about a deceased sister to estimate if it was a maternal death. A drawback of measuring maternal mortality is that it is, in statistical terms, a relatively rare event. So the use of sample surveys to measure maternal mortality may not result in a statistically reliable number of deaths.

> *In countries without comprehensive registration of deaths and good attribution of causes of death, … survey methods have to be used to estimate levels [of maternal mortality] and these are subject to wide margins of uncertainty and cannot be used to monitor trends. Therefore, there is consensus that for assessing trends over time, other indicators should be used.*
>
> *Report of the Inter-agency Expert Group on MDG Indicators, New York, April 2002*

Figure 12 shows the global differences in maternal mortality. Note that this is a very similar pattern to the one we saw for under-five mortality.

Figure 12.1: Maternal mortality ratio per 100,000 live births, 2000

[5] Source: WHO, 2011, Health Statistics and Health Information Systems: Maternal Mortality Ratio, http://www.who.int/healthinfo/statistics/indmaternalmortality/en/index.html

Maternal mortality ratio and maternal mortality rate

The two main indicators of maternal mortality are the maternal mortality ratio and the maternal mortality rate.

The maternal mortality ratio measures the rate of maternal death relative to the number of live births.

$$\text{Maternal mortality ratio} = \frac{\text{number of maternal deaths of women aged } 15-49 \text{ years}}{\text{number of live births}} \text{x} 100{,}000$$

The maternal mortality rate instead measures maternal deaths relative to the population of women of reproductive age (i.e., 15-49 years). This indicator does not account for the number of births in a population, but instead measures the risk of maternal death out of the women from ages 15 to 49 years.

The Maternal Mortality Ratio per 100,000 is one of the agreed indicators under the Sustainable Development Goals (SDG 3, target 3.1)

$$\text{Maternal mortality rate} = \frac{\text{number of maternal deaths of women aged } 15-49 \text{ years}}{\text{mid}-\text{year population of women aged } 15-49 \text{ years}} \text{ x } 100{,}000$$

Example

For example, if in a population in 2010 there were 1,200 maternal deaths of women 15-49 years, from 250,000 live births and a population of 500,000 women aged 15-49 years.

$$\text{Maternal mortality ratio} = \frac{1200}{250{,}000} \text{ x } 100{,}000$$

$$= 480 \text{ maternal deaths per } 100{,}000 \text{ live births}$$

$$\text{Maternal mortality rate} = \frac{1200}{500{,}000} \text{ x } 100{,}000$$

$$= 240 \text{ maternal deaths per } 100{,}000 \text{ population aged } 15-49 \text{ years}$$

Note that both the maternal mortality ratio and the maternal mortality rate may be shortened to use the notation MMR. To avoid confusion, it is important to always spell out which one you mean.

A Pacific example

For example, if in a given year there were 6 maternal deaths and a total of 3,867 live births, the maternal mortality ratio is calculated as:

$$\frac{6}{3{,}867} \times 100{,}000 = 155.15$$

Meaning

This means there were 155 (rounded) deaths per 100,000 live births. Note that this calculation gives rise to problems when applied to Pacific Island countries as the number of births is relatively small, and therefore the ratio expressed per 100,000 births may be misleading.

Reporting Maternal Mortality

For small countries in the Pacific where populations are fairly stable and where a maternal mortality ratio per 100,000 births may have very little meaning given the population sizes, it has been agreed that reporting the absolute number of maternal deaths (rather than a rate or ratio) is more appropriate for monitoring SDGs. In practice however, we have found that where countries do not report a ratio or rate, these will often be calculated and reported internationally anyway. In many of these instances, rates and ratios have been based on single years of data, and as such may prove either distracting or simply misleading if used in international comparisons or assessments of trends. Therefore, although recognizing that maternal mortality ratios may not be all that meaningful at a national or regional level, PICTs may wish to calculate these nonetheless so that they can retain control over what is reported about them. In this way PICTs can ensure that data is aggregated over a suitable time frame and compared against long term trends to reduce the potential for misleading inferences being drawn from their data.

Maternal mortality is a statistically rare event, even when mortality levels are high. Therefore, it is important that 95% confidence intervals are always reported with a maternal mortality rate or ratio, and data should be aggregated over a number of years for greater stability.

13 NON COMMUNICABLE DISEASES (NCDS)

Pacific Island Countries and Territories (PICTs) have among the highest burden of NCDs in the world. These diseases include diabetes, chronic respiratory illnesses, cancers, and cardiovascular diseases among others.

Mortality due to NCD's is a critical outcome indicator for NCD surveillance, and is vital information for monitoring the impact of any intervention measures.

13.1 INDICATOR OPTIONS FOR SURVEILLANCE OF NCD OUTCOMES

There are two groups of indicators for mortality due to NCD: primary or direct measures, and proxy indicators that do not measure deaths from NCD's directly, but which nonetheless show the effect of NCD's on overall mortality levels.

A] Primary Indicators

Probability of dying from NCD's

This measure is calculated from life tables constructed from the age-specific, cause specific mortality rates to determine a probability that a person will die from a specific type of disease over the period specified (i.e. between the ages of 30-70). As seen in the previous chapter, calculating the probability of dying from a specific disease (or group of diseases) requires a more complex analysis than the other measures proposed in this chapter. This indicator also requires a very high level of data coverage and quality.

Age-specific proportional mortality from NCD's

Age-specific proportional mortality from specific groups of NCDs should be reported separately for males and females. Proportional mortality shows the relative burden from NCDs compared to other causes and is easy to measure, but does not provide a measure of the overall impact of NCD related deaths.

Age-specific mortality rate from NCD's

Age-specific mortality rates from specific groups of NCDs should be reported separately for males and females. This provides a direct measure of the overall impact of NCD related deaths on the population, however cannot be used to provide a comparison either between countries or over time as it will be affected by the age structure of the population. This is a recommended indicator for PICTs. Countries may also find calculating rates from this group of specific NCD related causes for 10 year age groups useful in providing more detailed information for targeting specific interventions.

Age-standardised mortality from NCD's

For comparison over time and across countries, age standardized rates for the selected NCD's (as a total) could also be used as an indicator. This could be done for all ages, or for a selected adult age range. Of course, for comparison purposes across the region, all countries would need to standardize their data to the same population structure, and use the same age groups.

B] Proxy Indicators

Adult mortality (45q15)

Adult mortality is the probability of dying between ages 15 and 59 years inclusive. In demographic terms this is written as 45q15 (or the probability of dying from age 15 for the next 45 years – i.e. age

60). As the vast majority of deaths in this age group are due to non-communicable disease in most countries (excluding countries with high HIV prevalence, drought or natural disaster, or conflict), adult mortality is a useful indirect measure of the effect of non-communicable diseases on a population. It is more useful than life expectancy at birth (LE0) as it excludes the influence of deaths in younger age groups where deaths from infectious diseases and external causes are more common. Measures of adult mortality can be obtained from empirical data either directly through vital and civil registration systems, or indirectly through census analysis. As such it is already widely used and available across jurisdictions and is recommended for use in measuring NCDs in all PICTs.

Life expectancy at 40 (LE 40)

LE at age 40 is a useful indicator to demonstrate the impact of NCD's on life expectancy. Although less important than adult mortality, it provides additional depth to the understanding of the impact of NCD's and should be a standard indicator for all PICTs

Proportional mortality (by underlying cause of death) / against trends in Life Expectancy (E0)

Where data is poor and proportional mortality by cause is only available by broad category for all ages, trends in proportional mortality can be used in conjunction with trends in life expectancy at birth to make a judgment about what is happening with NCDs. As people live longer into old age, you would expect to see a greater proportion of deaths due to NCDs (people have to die of something). If however there is an increasing trend in proportional mortality due to NCDs in the absence of a corresponding increase in LE (over a number of years) - i.e. if LE at birth is either stable or declining, this would indicate that the increase in proportional mortality due to NCDs is possibly due to increasing NCD mortality in adults prior to old age. This indicator is fairly subjective and is best used only where more specific data is not available.

Indicators can all be calculated using the techniques as shown in the previous chapters

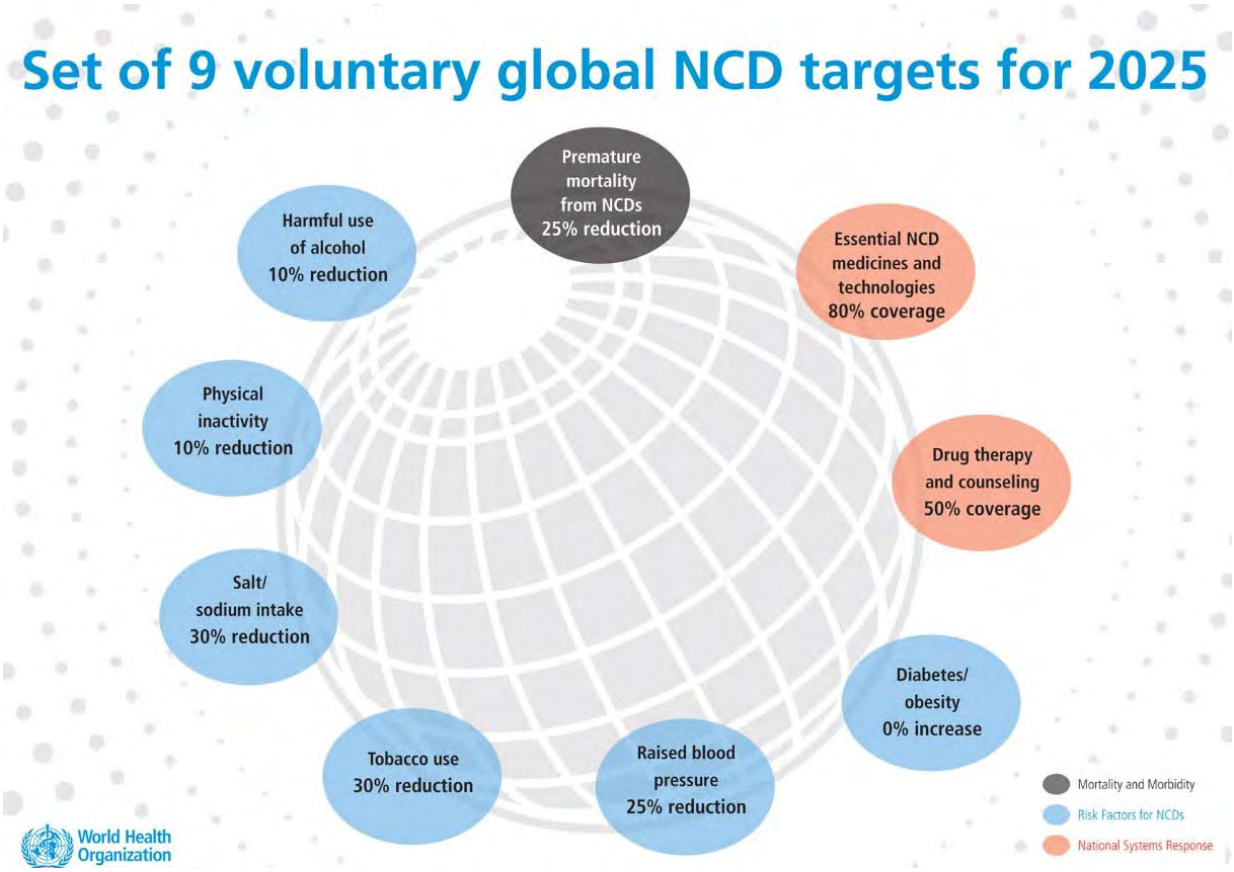
13.2 GLOBAL TARGETS AND INDICATORS

The World Health Organization has recommended nine (9) voluntary targets and twenty five (25) indicators for addressing NCDs by 2025. These have been adopted by countries at the World Health Assembly, and are also expected to be adopted as part of the post 2015 development agenda framework The first target relates to a reduction in premature mortality from NCDs, with an adopted target as follows:

Premature mortality from NCDs

Target: A 25% relative reduction in overall mortality from cardiovascular diseases, cancer, diabetes, or chronic respiratory diseases.

Figure 13.1 WHO NCD Targets for 2025



The indicator that has been selected to measure this target is:

Unconditional probability of dying between ages 30 and 70 years, from cardiovascular disease, cancer, diabetes, or chronic respiratory diseases.

13.3 Considerations for NCD indicators

Age Groups

The globally agreed indicator on probability of dying uses the age group 30-70 years of age.

As statisticians we would actually interpret to mean ages from 30 up to 70 years or 30-69 years inclusive. PICTs are recommended to produce this figure for comparison with international data.

However, as many PICTs have life expectancies in the mid 60's years, reporting the probability of dying before the age of 70 may have limited policy value over the short term when attempting to respond to and

prevent premature deaths from NCDs. As everyone dies eventually, including deaths from older age groups tends to result in over-representation of cardiovascular disease as deaths are attributed to non-specific causes such as Heart disease, and heart stopped. It should also be remembered that the primary public health concern is in preventing premature deaths from cardiovascular disease, and as such we are interested in the NCD related deaths in those people who do not reach old age rather than those who die of an NCD in old age. PICTs may therefore wish to also include additional indicators using a lower age group (such as NCD related deaths in 15-59 year olds), to monitor progress and inform policy at the local level.

One possibility for the Pacific would be 15-59 years inclusive, consistent with the standard measure of adult mortality. As discussed in the previous chapter, the impact of NCDs will vary at different ages and countries may also find it worthwhile to look at smaller age groupings such as 15-29, 30-44, and 45-59

Causes of death for NCDs

The global indicator that countries have agreed to discusses the probability of dying from: "cardiovascular diseases, cancer, diabetes, or chronic respiratory diseases" however this has not been further defined. In keeping with all other reporting of causes of death, Countries should include in this indicator deaths for which the underlying cause of death was one of these NCDs. Including deaths with causes such as "heart stopped" or "heart failure" as an immediate cause but an un-related underlying cause of death, for example, will artificially inflate the scale of the problem.

Most PICTs are working towards the 103 cause General Mortality List 1 as their standard tabulation for reporting. The categories: 1-026 – Neoplasms, 1-052 – Diabetes, 1-064 Diseases of the Circulatory System, 1-076 – Chronic Lower respiratory Disease, correspond to the diseases of interest as outlined in the global declaration.

There would also be advantages to including the following two categories: 1-080 – Diseases of the Liver, and 1-056 – Diseases of the genitourinary tract, in measurements for the Pacific Islands as most deaths in these categories will be due to chronic diseases (this is more certain for the former, but should also hold for the latter when deaths are categorised appropriately by underlying cause rather than intermediate or immediate causes which lead to kidney failure).

There is also a growing international push to also recognize the importance of other NCDs such as mental illness and musculoskeletal problems (such as chronic back and neck pain) in the overall burden of disease faced by countries. While these are certainly very important in terms of morbidity data, they are responsible for very few deaths, and therefore will not be considered further in this work on vital statistics analysis.

Recommended Measures of NCDs

As countries have committed to reporting against the voluntary targets as proposed by WHO, where data allows and countries have the required expertise, PICTs are encouraged to calculate and report the probability of dying from 30 years to 69 years inclusive using the following categories from the ICD general mortality list 1: 1-026 – Neoplasms, 1-052 – Diabetes, 1-064 Diseases of the Circulatory System, and 1-076 – Chronic Lower respiratory Disease. This should be done as a total, and for greater policy value, disaggregated by sex.

The draft Pacific reporting guidelines for CRVS have also proposed that countries produce proportional mortality and cause-specific mortality for the above disease categories, by sex, by broad age group (15-29, 30-44, and 45-59), and age-standardized mortality rates for 15-59 years by sex for these diseases.

Methods for these calculations were covered in the previous chapters.

PART 3: Communicating and Using Your Results

14 INTRODUCTION TO WRITING (1)

14.1 STATISTICS NEED TO TELL A STORY

Throughout the course, we will be looking at how we turn our data into measures or indicators that have meaning for decision makers and for the public. However, we need to go one step further than this and turn our data into information. While as data analysts, we may think the measures or indicators are clear and speak for themselves, it is our job to help people who may not be as statistically literate as ourselves to connect with and discover the meaning in our data. Many people, including senior decision makers, can either feel overwhelmed when confronted with pages and pages of numbers, or are simply too busy to wade through these to find the important information.

We need to use our data to tell the story of what we have found from our analysis. For example, if we were to calculate the adult mortality rate (the probability of dying between ages 15 and 59 inclusive) as 40% (which we know is very high) it is very hard for many people to tell if this is good or bad, or to interpret what this means. We need to provide some **CONTEXT** to help people connect with the data.
Consider how much more powerful is it to describe this result using the following: if the local high-school had 100 children aged 15 in year nine, 40 of those children would be likely to die before their 60th birthday. If people can relate to the story that the data tells, it is much more likely to inspire some action. Of course, we don't want to over-sensationalize our data or it will lose impact, but by putting our data in context we can help people to understand what the data is saying.

14.2 WHAT SORT OF INFORMATION DO WE NEED TO CONVEY

The story in your report should cover:

- **WHO** are we talking about.
- **WHAT** does the data say (about how many births there are, or how healthy we are etc.)?
- **WHERE** does this information apply.
- **WHEN** was this? Is it recent or has the situation changed.
- **HOW** do we know?

In field epidemiology (in the health sector), we use a very similar phrase and would describe **WHAT** we know about a population by **PERSON, PLACE,** and **TIME** (in other words WHO was affected, WHERE were they and WHEN did it happen).

14.3 STARTING TO WRITE

Extract (adapted) from SPC's Data analysis and report writing course [Ryan, C. et. al. 2011]

Many people feel that the hardest part of writing any document is starting to write. The reason for this is they think that the first thing they write should be the report that they give to their boss or the copy that will be published. It is unrealistic to believe that you can simply sit down and write a completed document, so where do we start?

… with a plan.

In practice, we use the same kind of planning for the different types of documents we prepare in the office. A letter, fax, email, discussion paper or report are all planned the same way, whatever the finished product might look like. Of course the degree of detail can be vastly different from a simple letter to a full report! First you should get familiar with this planning process, and then apply it in your written work.

Just remember that a report is one type of document, and while it is the most detailed and complex type, we still use the same principles that we apply to our other writing. Often when you are writing a report, the section headings and format you choose dictate its structure. In the material provided, you have been given a draft outline for a standard report on civil registration and vital statistics data to follow. You will be able to adapt this to suit your own country style and data as we go, but initially at least we will follow this structure as we start to put the report together.

Five steps to document writing

There are five basic steps that every good writer will go through to produce a quality document. Note that for more 'routine' documents you will not 'consciously' perform these tasks, as they will become habits (however each step will still get done)!

The steps are:

1. understanding the brief;
2. researching the information;
3. organising the information;
4. composing the rough draft; and
5. preparing the final draft.

Understanding the brief

Understanding the brief is an important task when starting on a document. It will determine where and what you research, the detail and the urgency of the work. Try and identify the:

- purpose – what is the topic of the document and why are you writing it?
- scope – how many and what issues do you want to cover?
- recommendations – are you required to make recommendations about any of the issues and what type of recommendation should you make?
- audience – who are you writing for and what degree of formality is required?

Brainstorming

Before you start it is often an advantage to sit down with a group of colleagues and have a "brainstorming" session. Brainstorms are a great way of getting different perspectives and ways to attack a problem. The basic idea of a brainstorming session is that you put forward the topic and some general information about it. From there, everyone gives their ideas on where to research, how to analyse, people who may be able to help, etc. In this type of session, you should encourage all ideas to be proposed, no matter how silly you may initially think they are.

Terms of reference

Sometimes you will be given Terms of Reference, (TOR), especially if you are preparing a report. The TOR will establish the brief by defining the subject, objectives, audience and scope of your report. If your TOR does not do this the first thing you have to do is clarify them so you understand your brief.

- *End of extract*

For the purposes of this workshop, the TORs are contained within the objectives of the workshop. That is for each country group of participants to **complete a vital statistics report (births, deaths and causes of death) that illustrates current levels and trends over time, and that can be used for planning and policy review purposes** (including the post 2015 discussions). The report should follow the agreed-upon regional guidelines for reporting vital statistics, and assist countries to meet their international reporting obligations.

What next?

We will look at the remaining steps in more detail later in the course, but as you look at your systems and analyse your data you are already carrying out step 2 – which is to research the information.

Once you are clear on what the report is to be about, it is very useful to document what you are doing and your results as you go rather than waiting until the end of the project. This is especially important if you are working on a very large project that is likely to take some time.

Write as you go…

By writing each section up as you go, you not only don't have to try to remember later on what you actually had to do (the methods), but you will also find the task at the end far less daunting because much of the writing is already done – you just need to tidy it up and organize it. Many people are overwhelmed by the task of writing a whole report, but by breaking it up and writing about each section as we go, we are able to break the task up into smaller, much more manageable tasks.

Throughout the workbook you will find writing exercises that will help you to write up each section as you complete it. Of course, we will need to review and edit these once we have all of the analysis completed, but much of your report will be done by then. Open a new folder in your flashdrive for your draft writing and clearly label each writing task with what it was about to help you organize the data easily later on.

14.4 PARAGRAPH STRUCTURE AND TOPIC SENTENCES

Adapted from Pudue Open Writing Lab (https://owl.english.purdue.edu/owl/resource/606/01/)

No matter how good your data is, unless your ideas are organized and a reader can follow what you are trying to say, you will find it difficult to get your message across to your target audience. In writing a report such as we will do in this workshop, we use a combination of headings, subheadings, and individual paragraphs to organise this information. The paragraph is therefore the basic building block for good writing.

A paragraph is a collection of related sentences dealing with a single topic.

The Basic Rule: Keep one idea to one paragraph

The basic rule of thumb with paragraphing is to keep one idea to one paragraph. If you begin to transition into a new idea, it belongs in a new paragraph. There are some simple ways to tell if you are on the same topic or a new one. You can have one idea and several bits of supporting evidence within a single paragraph. You can also have several points in a single paragraph as long as they relate to the overall topic of the paragraph. If the single points start to get long, then perhaps elaborating on each of them and placing them in their own paragraphs is the way to go.

Elements of a paragraph

To be as effective as possible, a paragraph should contain each of the following: Unity, Coherence, A Topic Sentence, and Adequate Development. As you will see, all of these traits overlap. Using and adapting them to your individual purposes will help you construct effective paragraphs.

Unity

The entire paragraph should concern itself with a single focus. If it begins with one focus or major point of discussion, it should not end with another or wander within different ideas.

> For example, let's take as an example the following sentence from the 2009 Kiribati DHS:
>
> **The first month of life is associated with the highest risk to survival. The neonatal mortality rate is around 26 deaths per 1,000 live births, implying that 26 out of every 1,000 infant deaths occur during the first month of life. As childhood mortality declines, post-neonatal mortality usually declines faster than neonatal mortality because neonatal mortality is frequently caused by biological factors that are not easily addressed by primary care interventions.**
>
> As you can see, the entire paragraph deals with aspects of neonatal mortality, even though several different points are raised within the sentence, including : how high the neonatal mortality rate is, the comparison with post neonatal mortality, and the factors associated with causing the mortality.

Coherence

Coherence is the trait that makes the paragraph easily understandable to a reader. You can help create coherence in your paragraphs by creating logical bridges and verbal bridges.

Logical bridges

- The same idea of a topic is carried over from sentence to sentence
- Successive sentences can be constructed in parallel form

Verbal *bridges*

- Key words can be repeated in several sentences
- Synonymous words can be repeated in several sentences
- Pronouns can refer to nouns in previous sentences
- Transition words can be used to link ideas from different sentences

Our example from the DHS above includes examples of both logical and verbal bridges.

As a second example, let's look at another paragraph from the Kiribati report as shown below:

> **Respondents were asked to consider a hypothetical situation independent of their current family size, and to report the number of children they would choose to have. Information on what women and men believe to be the ideal family size was elicited through two questions. Respondents who had no living children were asked, 'If you could choose exactly the number of children to have in your whole life, how many would that be?' Respondents who had children were asked, 'If you could go back to the time you did not have any children and could choose exactly the number of children to have in your whole life, how many would that be?'**
>
> As you can see, the last two sentences are constructed in parallel form, creating a logical link from one to the next. Key words such as family size and number of children have also been repeated in several sentences.

A topic sentence

A topic sentence indicates the idea of the topic that the paragraph is going to deal with.

A topic sentence contains:

1. A topic

e.g. The first month of life

2. A controlling idea (what you say about the topic)

e.g. … risk to survival

"The first month of life is associated with the highest risk to survival."

Although not all paragraphs have clear-cut topic sentences, and despite the fact that topic sentences can occur anywhere in the paragraph, t*he topic is best placed at or near the start of the sentence.*

In our second example, the topic sentence is the second sentence of the paragraph

Information on what women and men believe to be the ideal family size was elicited through two questions.

Our topic is: **what women and men believe to be the ideal family size***; and*

The controlling idea is: **was elicited through two questions**

The topic should not be a pronoun (he, she, his, this, it) as starting with a pronoun often makes the focus of the paragraph unclear.

Adequate development

The topic (which is introduced by the topic sentence) should be discussed fully and adequately. Again, this varies from paragraph to paragraph, depending on the author's purpose.

For technical reports, such as what we will be writing in this workshop, two or three sentences may be sufficient to adequately convey the key points that you want to say. For other forms of writing, you may need to include more information in order to tell your story. The important thing is to be as concise as possible, but not to leave your reader wondering where the rest of the information is.

How do I know when to start a new paragraph?

You should start a new paragraph when:

- When you begin a new idea or point. New ideas should always start in new paragraphs. If you have an extended idea that spans multiple paragraphs, each new point within that idea should have its own paragraph.
- To contrast information or ideas. Separate paragraphs can serve to contrast sides in a debate, different points in an argument, or any other difference.
- When your readers need a pause. Breaks between paragraphs function as a short "break" for your readers—adding these in will help your writing more readable. You would create a break if the paragraph becomes too long or the material is complex.
- When you are ending your introduction or starting your conclusion. Your introductory and concluding material should always be in a new paragraph. Many introductions and conclusions have multiple paragraphs depending on their content, length, and the writer's purpose.

Good paragraph structure will help you to put together your report, and help you to revise and edit your work. As paragraphs are essentially self-contained “mini-stories”, you can later move these around and order them in the way that flows best to make your point. We will look at editing and document flow later in the course.

14.5 Review

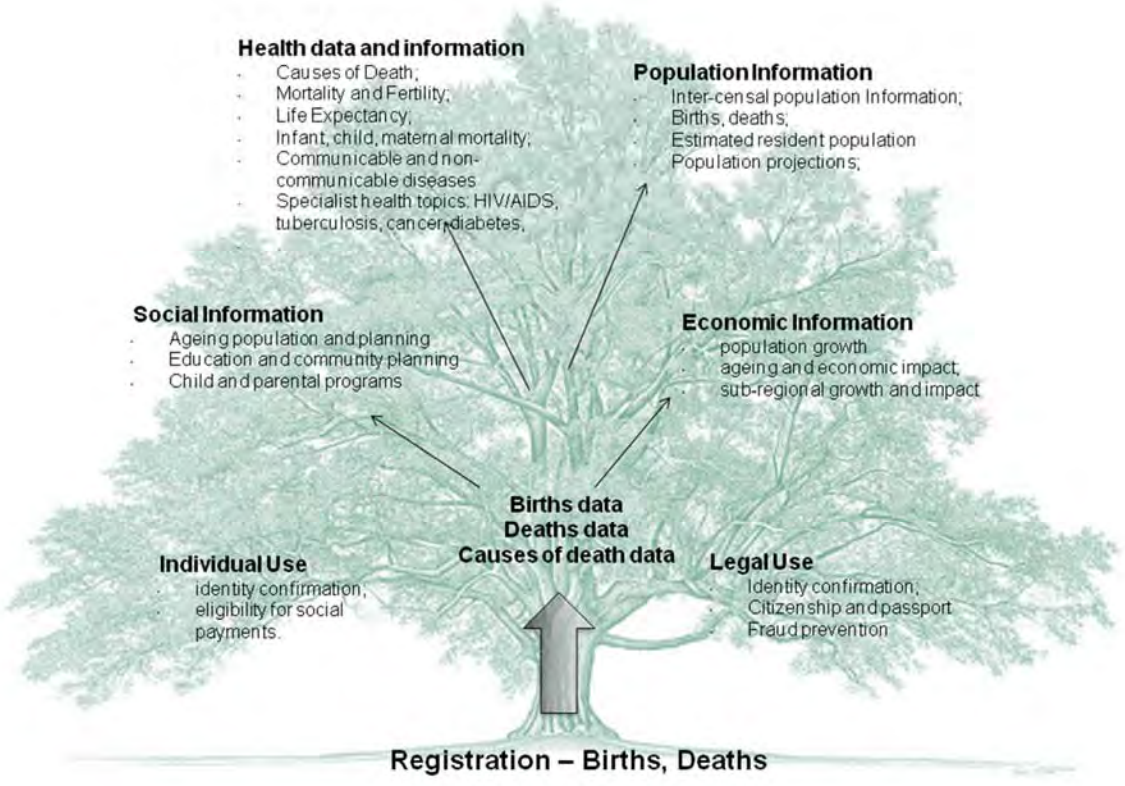
- Statistics need to tell a story. It is our job to help people who may not be as statistically literate as ourselves to connect with and discover the meaning in our data.
- We need to provide some **CONTEXT** to help people connect with the data.
- Your story should cover: **WHO, WHAT, WHERE, WHEN, and HOW** (or Person, Place, and Time)
- In field epidemiology (in the health sector), we use a very similar phrase and would describe **WHAT** we know about a population by **PERSON, PLACE,** and **TIME** (in other words WHO was affected, WHERE were they and WHEN did it happen).
- All writing should start with a plan. You should know roughly what you want to say and who you are writing for.
- Paragraphs are the basic building blocks of good writing. Each paragraph should have unity, coherence, a topic sentence and be adequately developed.
- A topic sentence should come early in the paragraph and includes a topic and a controlling idea.

15 Data Users and Their Data Needs

As we discussed at the beginning of this course, data on births, deaths and causes of death is essential in information for a range of development decisions. By taking a closer look at what the data is (or could be) used for, and by whom, we can get a better sense of their data needs. This not only helps us to ensure that we are producing the needed measures or indicators from our data, but also to better understand how we should disseminate this information.

The Australian Bureau of Statistics visualizes their birth, death and cause of death data as a tree; with the registration processes as the roots that provides the basis and support for their data, the data itself as the trunk or the centre of the process, and coming from this the range of different uses of this data as the branches and leaves.

Figure 15.1: The Australian Bureau of Statistics CRVS Data Uses Visualization



As you can see from the diagram, the broad range of data uses means that there is likely to be a lot of different organizations and individuals interested in CRVS data across a range of sectors. These may include government departments, researchers, and community organizations.

Writing exercise 22: Identifying data users

Within your country teams, use the following table to identify who your data users would be, and what data they need. Note the format that would be required as shown in the example. If possible, also note down when these stakeholders would require that information (i.e. is the data needed monthly, yearly, or just periodically when requested).

Table 15.1– Identifying your data users

User	What is the data required for?	What data is needed and in what format?	When do they need the data?	How should the data be communicated?
Example: University health researchers	Investigate the links between specific causes of death and potential exposures	Unit-record data (de-identified) on deaths including age, sex and cause (potentially by location and other variables)	Not routine – upon application only.	
Minister of Health	Budget decisions and setting priorities for the national health services	Core indicators only (with comparison data for context)	At least annually (when depends on the timing of the budget)	

15.1 TARGET AUDIENCES

Now that you have a clearer understanding of who your data users are, we can start to look at precisely who is our target audience for publication or dissemination of our findings. It might be helpful to think about which stakeholders must be informed, which stakeholder should be informed, and which stakeholders it would be nice to inform.

Writing exercise 22b: Targeting your message

Go back and look at the list of data users that you have just developed. Think about which category each of these users fits into in terms of whether you must, should, or would like to ensure that they have access to the data that they need.

Select the data users which you must communicate your data to and think about the best way to reach them. Will they read a report? Do they need a summary document? Do you need to tell them about your results in person? Record your decisions so that you can follow up when you return home.

16 STRUCTURING YOUR WRITING

Extract from SPC's Data analysis and report writing course [Ryan, C. et. al. 2011]

Earlier in the workbook, we discussed the importance of extracting the meaning from our data analysis in words in order to "tell a story" and assist people to connect with and interpret what the data means for them. We are essentially "describing the evidence" that policy makers need to make informed decisions.

As we discussed in Chapter 5, the five steps in writing a report are:

1 understanding the brief;

2 researching the information;

3 organising the information;

4 composing the rough draft; and

5 preparing the final draft.

Now that we are well underway with our analysis, we can step forward from "crunching numbers" or researching the information to take a closer look at the writing skills required to put together a concise, meaningful report.

16.1 ORGANISING YOUR INFORMATION

Organize your thoughts and start with an outline

An essential element in report writing is to present a logical and structured flow of information. Imagine if you were reading these notes and they kept jumping between writing skills and report writing without any clear structure or organisation. If you take the time it is usually quite easy to achieve a logical presentation of information.

Before you start, it may help to think about the outline which you are going to use and start to put together the headings, outlines and key points that you will want to make.

For example, if we look at the previous chapter, the outline that we started with looked something like the following:

- Introduction
- What do we intend to cover in this chapter
  - Why is knowing our data users and their needs useful
  - Visualizing our data users using the ABS data use tree
- Exercise – what do we want participants to do
- Defining target audiences
  - Why is this important/ How does it help?
  - Who would these include?
  - What should we do about it?
- Other reporting commitments

- What have countries signed up to already
- SDGs
- NMDIs

- What next?

In the final document, not all of these points will have its own heading, but as a starting point, this helps us work out what we want to cover in our writing and what our key messages will be.

You might find it useful to take the suggested headings for a national vital statistics report in the draft regional reporting guidelines, and record the key points that you would want to include under each of these (and indeed if there are any changes that you would like to make to the headings proposed). As you work through this list, think about the key messages that you need to get across and prioritise these. What were the outcomes of your data analysis on each section? Who do you want the information to get to?

Create an outline of the key points and messages that you want to convey.

While a list is the most common way that people use to pull this outline together – there are a number of different options available to help you organize your thoughts.

- written outlines, where you write a few sentences outlining your argument or points;
- mind maps where you list and order your ideas and main points;
- horizontal plans or fishbone diagrams where you follow an idea or topic through to its logical conclusion;
- triangle diagrams which are similar to fishbone diagrams, as they represent the development of ideas in a pictorial fashion; and
- tree diagrams, where you list out each of your main points in order of importance then add supporting information underneath each point.

Find the method that works for you

Different people prefer to use different tools. You should try each of these organising tools and find the one that you find easiest to work with. You will be surprised how they assist you organise your information.

Mind maps

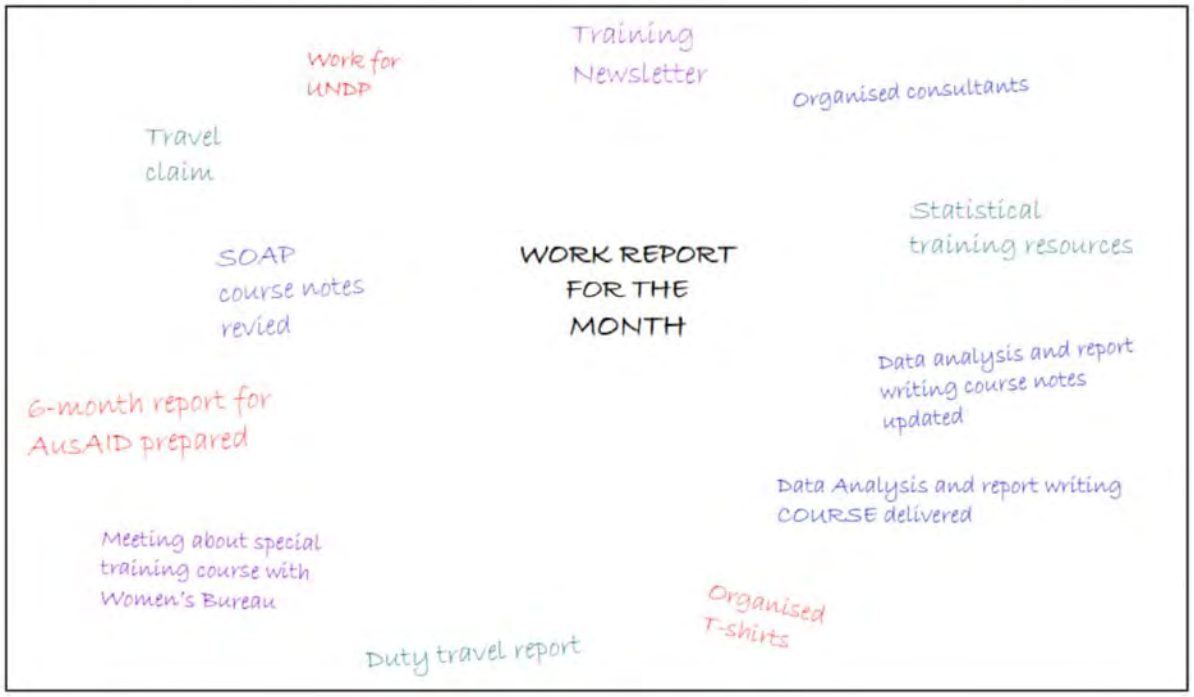
Mind maps are a useful way of listing and prioritising issues, especially if you are unsure of the issues involved and how these are related. In the example above the issues were quite clear and could just be listed out. It is not always that easy to identify the issues!

… list everything related to your topic.

When you make your mind map you write the topic of your document in the centre of the page. Then you write anything and everything related to that topic around it on the page, putting related topics near each other if possible. Experts say that it is a good idea to use different coloured pens for different topic areas in the mind map – as this stimulates creativity. You should write down all the topics you can think of – no matter how silly these may seem. Later you can decide if they will be included in your documents or not.

The following example is a mind map for a monthly work report. First of all the activities were all listed down, using the same colour for related activities.

Figure 16.1 Example of mind map



Most important issues first

Prioritising is a very important stage, as the reader expects to read the most important issues first. Imagine if you were reading a document and the important point was in the last line of the last page you would probably have stopped reading well before the end!

Group and prioritise ideas

The next step is to group your ideas into related topics or issues. At this stage some people choose to re-write their mind maps, but this is up to you. Then you number or letter the topics according to priority. The basic process used in the mind maps is very similar to that used in the written outline above, but initially it is not as structured.

Figure 16.2 Example of mind map with prioritised activities

Fishbone diagrams or horizontal plans

Solutions to a problem

A fishbone diagram is another technique used to establish issues relating to a problem or the subject of your research. Fishbone diagrams are especially good when you are planning a document which is finding solutions to a problem.

Can be used to identify the subject

Fishbone diagrams aim to identify a subject based on a problem or issue identified. The advantage of a fishbone diagram is that it forces you to clearly state all of your ideas and structure them in an explicit way. You can then start to research or write your report in a structured way. You can also easily add new 'branches' to a fishbone diagram when you find new information. Fishbone diagrams are also useful in a group discussion.

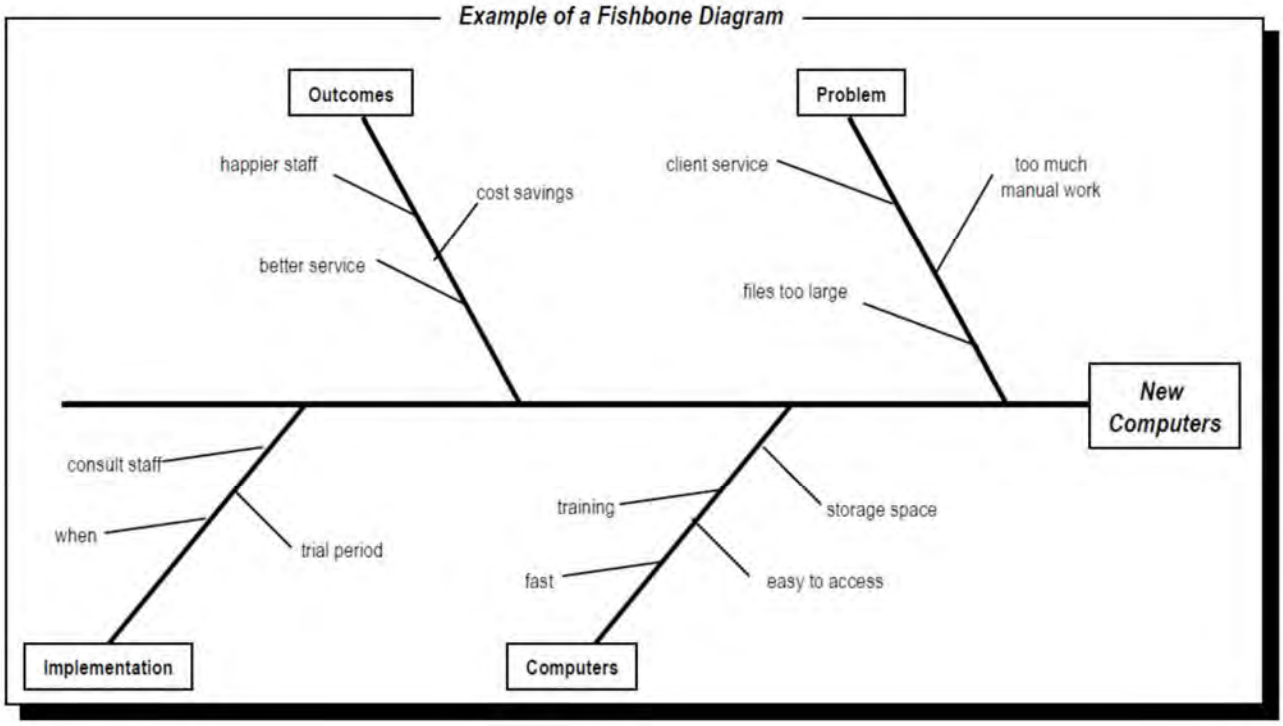
Factors in the problem are the 'bones'

In a small box on the right hand side of the page write the problem in as few words as possible. From the box draw a straight line across the middle of the page to the left hand side of the page. This line is the backbone. From the backbone draw stems out at an angle of about 45 degrees. Each stem identifies a

basic factor involved in the problem. Branches are then applied to each stem to represent the details affecting each basic factor all aspects of a problem

The purpose is to get everything down on paper so you can explore all aspects of the problem. Begin fishboning by drawing and labelling as many stems and branches as you can, working rapidly from one to the other as the thoughts tumble out. An example of a fishbone is:

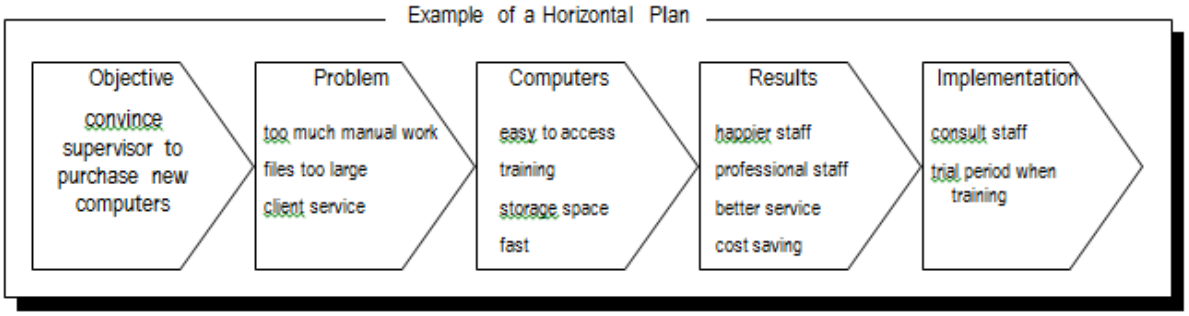
Figure 16.3 Example of Fishbone diagram



Horizontal plan

In a horizontal plan, jot down the major components of the project and write sub-points about these issues below the main components. Horizontal plans are more 'linear' than fishbone diagrams, in that one thing follows from another.

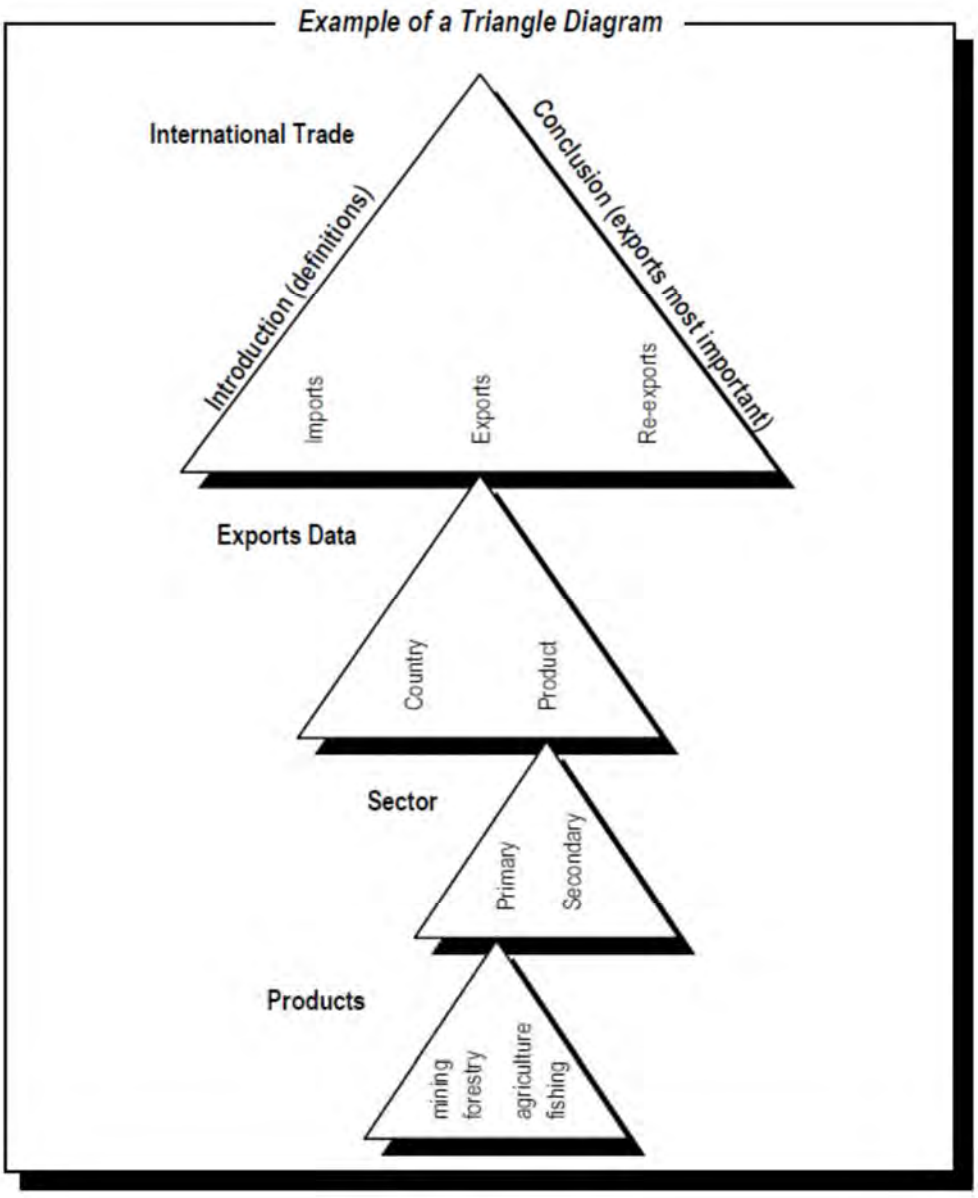
Figure 16.4 Example of Horizontal Plan



Triangle diagrams

Triangle diagrams are a good way to get to the 'bottom' of an issue, especially when your topic or subject deals with evaluating different alternatives. The discussions follow the arrows around the triangles. After each triangle you go into deeper detail about one aspect or point. An example is shown below for an analysis of international trade.

Figure 16.5 Example of Triangle Diagram



Introduction and conclusion

Along the sides of the triangle you write the introduction and the conclusion and the issues are written along the base. In the example the introduction and conclusion were written on the sides of the first triangle. Here the introduction will discuss the definitions of international trade and the conclusion will state that exports are most important in this discussion.

Issues on the base

The issues are written along the base of the triangle. In this example the issues in the first triangle are imports, exports and re-exports. The second triangle focuses on the components of exports – country and product; and so on to the last triangle.

Tree diagrams

Similar to mind maps

When you use a tree diagram as a planning aid you list your topic at the bottom of the page and each of the topics at the top. You then list the issues related to each topic as 'branches' around the topic. The process is similar to the way you construct a mind map.

Summary

Plan to structure your document

Before starting to write a draft report try one of these techniques. Find the one which 'works' best for you and use it whenever you have to plan a document. It might stop you from becoming lost or changing your argument halfway through a report.

Planning your document will often take a lot more time that the actual writing, and a good structure can help you avoid writing a lot of unnecessary words as you get around to making your point.

16.2 REPORTS – MORE INFORMATION

What is a report?

A report is an account given or opinion formally expressed after investigation, consideration, or collation of information.

Usage

Reports are used to:

record work done;

validate information;

circulate new ideas;

indicate further action;

assess a situation;

save duplication of effort;

cross-fertilise ideas; and

keep people informed.

Format

When we write reports we usually follow a format. This format is usually followed right from the planning stage.

For a short report, the suggested structure is:

Title Purpose
Background Issues
Conclusions
Recommendations

For a long report, the following structure is suggested:

Title
Table of Contents
List of Illustrations
Summary (or Executive Summary)
Introduction
Body Conclusions
Recommendations
Glossary
Bibliography and/or list of References
Appendix or Attachment

A detailed description of each of these and other possible components is given further below.

Classification system

When using one of these structures, you should use a classification system to allow points in your report to be referenced and help the reader digest the information. We should only ever use one type of classification system at a time in a report. The simplest classification system to use is to number the headings and subheadings as below:

Decimal system

1. PRIMARY INDUSTRY
    - 1.1 STAFF
        - 1.1.1 Recruitment
        - 1.1.2 Training
        - 1.1.3 Salary
        - 1.1.4 Promotion
    - 1.2 EQUIPMENT
        - 1.2.1 Costing
        - 1.2.2 Ordering
        - 1.2.3 Repair and Maintenance
2. SECONDARY EDUCATION
    - 2.1 STAFF
        - 2.1.2 Recruitment .....etc.

Detailed format of a formal report

(a) Cover

i. Title: A succinct, accurate and all-embracing description of the project. Avoid rhetorical questions or value judgment statements such as “Uranium - should it be mined?” or “Trailbikes - Environmental nuisance”. Titles should be no broader or narrower than the topic covered.

ii. Author’s name: include qualifications, organisation and professional affiliation if appropriate.

iii. Client’s name and organisation or intended audience, if appropriate.

iv. Date: generally month and year of completion.

iv. Stage: i.e. final, interim, draft, annual report etc.

v. Serial Number: sometimes there will be two or more parts to a report, so indicate appropriately.

vi. Classification Notice: i.e. confidential, restricted etc.

vii. Contract Number: include the file reference or contract number, if appropriate.

ix. Additional art work: optional – a photo or relevant image.

(b) Title Page

This is the first right hand page of a report and will cover all of the details on the front page. As well it might have the following information.

i. Individual Copy Number: in confidential reports this is important.

ii. Classification Details: i.e. to whom this report is restricted, details regarding copyright, etc.

Illustrations and art work, apart from your company's insignia or logo, are generally omitted from the title page.

(c) Letter of Transmittal (covering letter)

The Letter of Transmittal will reiterate broadly the criteria of the brief given by the person requesting the report. The letter may also give a very brief overview of the summary and an indication of the progress made - i.e. work completed, work remaining, etc.

The Letter of Transmittal is set out in standard letter form with all the components (recipient's address, letterhead, etc.) of a letter. It must be signed by the person responsible for the generation and transmission of the report, and should contain the project or contract work name and number, distribution details (in general terms if the report is of wide circulation), specific detail (names, etc.) if of limited distribution.

It does not bear the heading "Letter of Transmittal" because its function is obvious. It is bound in as an integral part of every report – not sent as a separate accompanying letter. (An accompanying letter should of course also be sent with any report.) It is the formal part of your presentation to your reader.

Letters of Transmittal are usually restricted to one page.

(d) Approvals

Often your report must be read and approved by one or more officials before it can be issued (usually your superior(s) in your organisation). Each person's name and title and any other customary identification must be given and their signature reproduced above the typed name. If there are a large number of approvals they are placed on a separate page. If there are only a few approvals they are placed on the title page or the inside front cover.

(e) Distribution

The primary recipient of a report must be informed of the location of all other reports. Sometimes it is necessary to supply a list of names (plus official title, company, location, number of copies sent and copy number sent). If this list is long it is placed on a separate page. If not, it appears on the title page.

(f) Preface

The preface introduces the report to the reader. It does not introduce the subject matter, that is the role of the introduction.

It will also state briefly and clearly, on a separate page, the subject and its relation to other reports in the program (if relevant), by whom the work project and report writing were authorised, and whether the report is final or one of a series. If the author's name is prohibited on the cover, as is the case with government reports, then the author's name may be given here.

(g) Foreword

Forewords are often inserted instead of a preface. They are written by a person other than the author and can contain praise of the author, which would, of course, be inappropriate in a preface. Forewords are more appropriate when the report has been sub-contracted by the consultant.

(h) Acknowledgments

You must carefully name each person or organisation that has contributed to the work program. Sometimes the nature of their contribution is acknowledged. Permission granted for the use of previously published material must be acknowledged here, except where this is adequately covered by the bibliography.

If the number of acknowledgments is large place them on a separate page. Otherwise place them in the preface or introduction, whichever seems most appropriate.

Keep acknowledgments brief and in good taste. Funny ones often fall flat.

(i) Table of contents

If your report is a large one, include a table of contents. List in sequence the material in your report, together with page numbers (surprisingly often overlooked because of typing problems). Indent headings where appropriate to make your table of contents as useful as possible.

(j) List of illustrations, diagrams, photographs, maps, etc.

If the list is long place it on a separate page. Otherwise put the list on the bottom of the Table of Contents.

(k) Summary (or executive summary)

This is a condensation of your whole report – i.e. introduction, discussion, conclusion, recommendations. It is intended for people who will probably not read your entire report.

(l) Abstract

Only include an abstract when the report is being published for a wider audience. Librarians will use the abstract in special abstract journals. Abstracts discuss the scope and treatment of a report. The abstract gives more of an overview of the report than the summary, and generally provides less detail than the summary.

(m) Introduction

This is the first part of the main text of the report. In some reports it will include the historical background on the subject and may cover a literature review of previous work (appropriately referenced as in research paper practice). It will outline the boundaries of the study, the limitations and problems.

A common fault of many students is to quote verbatim from other sources when almost always the information can be paraphrased and condensed into fewer words.

(n) Materials and methods (or scope and coverage)

It may not be necessary to use such a title to delineate this section of the report. You may use a more specific heading or you may integrate the section with the introduction. However, the broad method of

attacking the problem should be discussed, together with whatever minor detail of method seems appropriate. (Think of the people who will be reading your report – most readers don't want to be bored with all the details of how you do your work. They should be interested in the details such as the sample size of a survey so they can deduce the value of your recommendations.) Don't hide limitations in your method, and, where biases may have crept in, detail steps you made to ensure bias was minimised.

(o) Results

Leave all bulky data to appendices. Put significant results in this section. Consider how they can be best presented - tables, graphs, diagrams, pictures, etc. Pay attention to labels on diagrams and graphs.

(p) Discussion

Never leave your reader to draw his or her own interpretations of the results. You must point out what the results mean in the discussion - what is their significance. You may treat results and discussion together item by item, rather than all the results first followed by all the discussion.

(q) Conclusion

The boundary between discussion and conclusion is often hazy. In larger reports the conclusions will draw a lot of the discussion together and summarise it. Conclusions may include a discussion on the advantages and disadvantages of the alternatives suggested by the interpretations of the results. Do not include any ideas not supported by results and discussions.

(r) Recommendations

These are what you are suggesting should be done as a result of the discussion. You may arrange them in several ways - e.g. primary or secondary recommendations, or order of importance or urgency, or list them separately under different budget restraints. Do not include any recommendation not supported by the conclusions. Make sure your recommendations are practical. It is a common mistake of inexperienced investigators to dream up a wild imaginative idea that bears little chance of being accepted, or of working. If you have a recommendation that you feel will be a hard pill for you reader to swallow, pay particular attention to its justification in the discussion.

(s) Appendices or attachments

Include here all material which is superfluous to the main reader, but which may be helpful to someone else in the organisation. Don't put in your completed questionnaires (a blank sample will do) or other detailed research data. Sometimes photocopied articles of other similar studies may be appropriate to persuade your reader if you envisage difficulty with some recommendations.

Number the Appendices as I, II, III, etc. so you can reference them in the main body of the text. Don't use the appendices to "swell" the report and make it look bigger, and don't put material in that belongs elsewhere.

The modern trend is to economise in the text and put more in the appendices to ensure that more of the report text will be read.

(t) Addenda or corrigenda

Include in here any corrections or last minute alterations made after publication of the text. (u) Glossary

If you have terms that need definition or that the reader will not know, then include them in a glossary. If there are many of them, use a separate section at the back of the text. If only a few, put them somewhere near the beginning of the text. Don't use a glossary as an excuse for using unnecessary jargon in the text. Remember you have already picked up jargon that your reader probably won't understand.

(v) Bibliography

All sources of ideas and material not derived from your own results must belong to someone else, otherwise they are merely unsubstantiated speculation. You must therefore reference all sources in standard bibliographic form. Do not include books in your bibliography that you did not specifically use in the text. Such books would belong to a separate list called Further Reading.

Every punctuation mark has its function in a bibliography and every entry must be perfect so that a librarian can obtain the article for someone seeking that source. Don't make the mistake of leaving your bibliographical references until after you have completed the paper. It will only necessitate another trip to the library. Be careful to note sources on any photocopies for the same reason.

(w) Index

In very long reports it may be necessary to include an index to expand on the work of the Table of Contents.

16.3 COMPOSING THE ROUGH DRAFT

Just start writing

After you have prioritized points and have a rough outline of what your document will look like, you can then start to write a draft. Drafts are only rough versions of the document. Once again, do not be too concerned if the sentences don't follow all of the grammatical rules, just get the information down on paper.

Each point in your outline is likely to be either one paragraph on a single topic, or several paragraphs on related points. If you are having trouble getting started, one trick is to start by writing a topic sentence for the point that you are trying to make. We covered what makes a good topic sentence back in Chapter 5.

It is also not necessary to tackle each of the headings or points in your outline in order. Do try to be focused and stick to one major topic at a time as you go so that your thoughts will flow coherently, but if you find it hard to get started with the introduction or methods, move on and do one of the areas which you find easier to write about and the come back to these. Just try not to leave all the hard ones until last otherwise you will find it hard to find the motivation to get your report finished.

The good news is that because you have been writing notes as you go for each section of the analysis, you should now be able to move these paragraphs to where they fit in your outline and you will find that you already have a large proportion of your report written!

As you write – writing skills

Make sure your research is relevant

After researching, you should have a lot of information. Not all of your research has to be included into the final report. Go through all of your information and look at it in a critical way. Is it really relevant to the report? Does it add additional information? The idea is to reject unnecessary information!

Writing style

WRITING GUIDE - UQ BUSINESS SCHOOL: MAY 2010
BY DR NEAL WADDELL

Style: Eight Guidelines

Writing style considers authors' use of language for their audience. A good writing style takes into account how readers read. Style does not describe what writers must do, but what they should do to write clearly and precisely. The following eight guidelines are proven ways to enhance the readability of your writing:

1. Keep a good balance between verbs and nouns— because verbs (the action of a sentence) are so important, don't overpower them with too many nouns (the people or things involved in the action). When you turn verbs into nouns, you limit the action of their sentences.

a) Instead of, We have a belief that … , write, We believe that … .
b) Instead of, The accounts are a good source of information for investors, write, The accounts inform investors well.

In other words, strengthen your sentences by converting some nouns to verbs.

2. Use active not passive voice as much as possible — active voice means following a subject–verb– object flow in sentences and is much clearer because the subject (doer) begins sentences. Although some sentences may need to be in the passive voice, aim to write two out of every three sentences in active voice.

a) Instead of, You are advised by me, write, I advise you.
b) Instead of, Shares are bought by investors, write, Investors buy shares.

3. Use words as signposts to connect sentences to sentences, and paragraphs to paragraphs, and to provide easy reading. Signposts [also called lexical ties] clarify (e.g., for example), add (e.g., furthermore, as well), concede (e.g., nevertheless), sum up (e.g., in conclusion), contrast (e.g., on the other hand), verify (e.g., in fact), show cause (e.g., therefore). However, there are many others signposts or lexical ties that you can use.

4. Do not use there and it, in the form of there is/are/was and it is/was too many times — when used in this way, particularly at the beginning of sentences, there and it became dummy pronouns, in that they do not say who or what is performing the action of the verb is or was etc. If you say, There have been reports etc., you are not being clear.

- Instead of, There have been reports that our profits are rising, find the agent of the actions and write, Reports by Roth Research Inc. indicate that our profits are falling.

5. Write sentences of the best length for clear reading — try not to write sentences that are either too long or too short. Because readers are less likely to remember what you have written if the sentence is too

long, about 25 words is the best sentence length. Therefore, do not try to say too many things in the one sentence; don't be afraid to use full-stops.

6. Be definite in professional writing; don't hedge information — a statement such as: Chinese consumers could possibly behave differently this year says very little because of the hedging (underlined). Try writing so that you inform the reader of possible effects and outcomes. Instead of this above example, write: The behaviour of Chinese consumers will [outline some probably changes] if their Government [state the action].

7. Do not use and, but, or so too much when writing sentences — they join statements of equal value. Sometimes it is better to use joining words like because, although, despite etc., which allow you to show which of the joined statements is more important:

Instead of: We have just begun sales but we should still profit well this year.

write: Although we have just begun sales, we should still profit well this year.

In this example, the underlined clause is shown to be more important because the clause beginning with Although is made a condition of the underlined clause. The writing is thus more interesting and easier to follow.

8. Do not use:

a) pretentious words, such as, impecunious circumstances (use short of money); inclement weather conditions (use bad weather); etc.

b) too many words, such as: due to the fact that (use because); At this point in time (use now); Until such time as (use when); I find it within my capacity to undertake the research (use I can do the research); etc.

c) words that unnecessarily modify others, such as: end product; future plans; another alternative, recur again; prove conclusively; 10 am in the morning; etc.

d) words that say the same thing more than once in a sentence, such as, biography of her life; pregnant mother to be; full and complete; etc.

16.4 Turning Data into Information

What helps turn data into information?

Comparisons

Comparisons are an important analytical tool as they provide context to help the reader understand the data. Telling a decision maker there were 100 deaths last year means very little, however telling them there were 100 deaths, compared to 20 deaths the year before, and 30 deaths during the same period for all other countries in the region, indicates either 1) there is an issue with your data, or 2) something serious has happened to mortality rates in your country.

Comparisons can be made between countries and regions, but also internally between different locations, gender or ethnic groups. The position of your country can be summarised in terms of a rank, percentile, standard deviations from the mean, or position compared with country mean or median. Benchmarking, comparing your position with good performers (or 'gold standards') is another common method used in comparing.

Comparisons are important as they provide information on:

- Identifying contextual changes
  - Demographic, economic, social and political factors
- Progress assessment
  - Compared to targets
  - Compared to peers
- Equity analysis
  - Trends in equity gaps by key stratifiers
- Efficiency analysis
  - Results by inputs
- Performance = summarising and interpreting the results

Consult existing records

There are many ways to obtain information about a topic. The most popular way is to consult with existing records or documentation. Often the existing records or documentation is in the form of published papers, books or previous reports. Other alternatives may be documentation of previous tasks and minutes from meetings. If it is possible, try to get points of view from opposing sides. This will lead to a balanced report.

Check your sources

If you are using information from another source, such as a book or a report, remember to check:

- the standing of the author;
- the date when the information was compiled or revised;
- the purpose for which the information was produced and how closely it relates to your purpose (is it biased in any way?); and
- the extent to which the information is supported by facts and figures that can be checked from other sources.

Using Graphs and Diagrams

Graphs and tables should illustrate the key messages you wish to convey to the reader. Too much information in a table or graph will confuse the reader! Each table and graph should make a point, much like a paragraph.

It is common practice to refer to tables or graphs in text by their table or figure number. For example:

- As shown in figure 1, ……..
- It can be seen from table 2 that…..
- Figure 2 refers to …..

These references imply that the reader has yet to come across the specific tables or figures, and so the text should come before the presentation of the tables and graphs. It is also acceptable to reference the table or figure number in parenthesis so that it does not obscure the main point of the message (e.g. see figure 1).

It is generally not useful to restate the numbers presented in a table or graph in text as the reader will have already have seen the table. However in cases where the size of the values is small (e.g. less than 50) then the frequency count should only be reported.

It is good practice to explain the size of the values in terms of percentages as this is a concept the audience easily understands. Remember to describe both the largest and lowest values, and also the most common values. Also, mention any unusual values that don't fit the expected pattern.

Example:

Figure 11 looks at the educational pattern by attainment. The figure shows that females tended to be less educated than their male counterparts. Females are more likely to have below high school attainment (68 percent) than males (38 percent).

Johnson BS A Summary Analysis of Selected Indices of the Preliminary Results of the Kosrae 2000 census.

It may be useful to describe a normal distribution in terms of an average and standard deviation. These statistics are often presented in the text, rather than in a table, as their meaning does require some explanation. For non-normal data you may want to describe the shape of the distribution and its skewness.

Interpreting The Results

The interpretation of statistical data requires experience and a common sense approach. To a large extent the interpretation of results will be influenced by the needs of the audience, the objectives of the research and the quality of the data presented. The interpretation of your findings are presented in the discussion section of the report. You must interpret your results for the readers so that they can understand the meaning of your findings. The two most important questions to answer are:

? What do my results mean?
? What are their implications?

The results of the analysis should be interpreted for the reader in terms of the objectives set out in the introduction. The key questions to be answered are:

- What are the overall findings from the analysis?
- How reliable or valid are the findings?
- What explanation can be provided for the findings?

A summary of the results should be provided in the discussion section. These may be presented in dot point fashion. The summary should refer to the objectives of the study and show how the results are relevant to meeting these objectives. It is important to ensure that the explanations you present are supported by the statistical data.

16.5 Sources and referencing

Adapted from Referencing, The University of Newcastle, Australia (http://www.newcastle.edu.au/service/library/foundation-portal/referencing-whydo.html)

Referencing is a way of showing what sources you have used, giving credit to the original authors. We use referencing when writing for a number of reasons, including:

- to avoid plagiarism
- to give credit to the original source of an idea, piece of information or resource
- to support your analysis, conclusions or recommendations with the work of another author or source
- to demonstrate you have researched your topic
- to help readers of your work find the original source of information or ideas you have used.

TIP: It is much easier to reference as you go rather than trying to remember all of your references at the end of the document.

In Harvard referencing, the in text citation would look something like:

Sophisticated searching techniques are important in finding information (Berkman 1994)
OR
Berkman (1994, p. 25) claimed that …
OR
Berkman (1994, pp. 30-35) agrees that …

The reference would then be given in the references list as:

Berkman, RI 1994, Find It fast: how to uncover expert information on any subject, HarperPerennial, New York.

Using Vancouver referencing, in text citation would be marked by a number or in superscript. If, in your text, you cite a piece of work more than once, the same citation number should be used. Example:

Recent research {1} indicates that the number of ……

OR

Recent research[1] indicates that the…….

The reference list is then given with the references in the order they appeared in the text as follows:

1. Butler SW. Secrets from the Black Bag. London: The Royal College of General Practitioners; 2005.

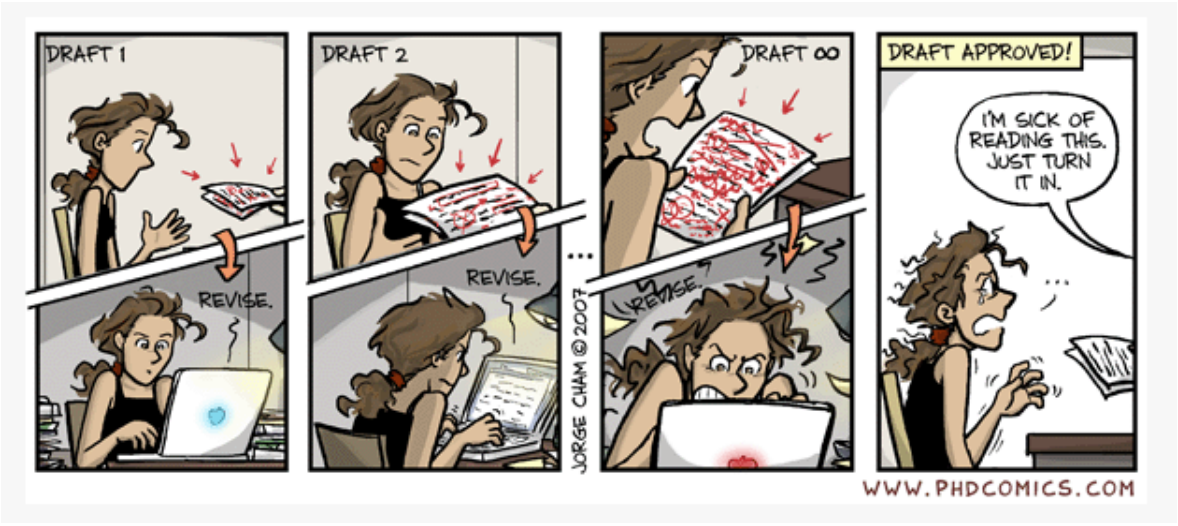
It is best practice to reference in Harvard as you write so that references can be moved later on without having to re-number the whole document as you do in Vancouver, unless all your references are managed in referencing software such as Endnote. There are a number of free referencing software programs available off the internet. These include: Mendelay (http://www.mendeley.com/features/), Zotero (http://www.zotero.org/) and Readcube (http://www.readcube.com/?locale=en)

Why use referencing software?

- Save time creating and managing your references
- You can type the references yourself or import them from a database or catalogue
- You can then create a bibliography for your thesis, assignment or journal article in your preferred citation style

16.6 Revision

Revision is often the most frustrating part of writing a document! At this stage you have collected all your information, said what you want to say, and all you want to do is hand your report over so you don't have to look at it anymore….



However, revising your document is a key step (or steps) in making sure that you are publishing a document that will get your message across and that people will want to read.

At this first step of revision you should both revise and edit your document:

1. Revising is concerned with the overall organisation and logic of the document. Therefore, when you revise, you should:

a) Put ideas in order
b) Keep readers' understanding in mind
c) Choose the right format/design for the message
d) Look for gaps in detail or the underlying research
e) Cover the major goal of the document, that is, to solve, describe, analyse, justify, and/or explain.

The main purpose of your first revision should be to ensure that the document does what is claims that it is doing, and that the order or structure of the sections and paragraphs makes sense.

2. Editing is a question of style and clarity at sentence and paragraph level so that the connections are clear in the text. Therefore:

a) Locate style problems quickly
b) Edit for a forceful style by ensuring your writing is active with strong verbs
c) Be sure you say what you mean and mean what you say
d) Strive for clarity by revealing the logic of your thinking to your reader
e) Ensure you have used signposts to reveal your logic.

Things to look out for in particular:

- Waffle words
  - Generally speaking, writers can basically rely in the main on certain fundamental techniques to structure their text
  - ~~Generallyspeaking~~, writers can ~~basically~~ rely ~~inthemain~~ on ~~certain~~ fundamental techniques to structure their text
  - Writers can rely on fundamental techniques to structure their text
- Repeated meanings
  - The end result was shorter in length than we had hoped, but we plan to increase the text in the future. Our writing methods and techniques enabled us to achieve our aims and objectives
  - The ~~end~~ result was shorter ~~inlength~~ than we had hoped, but we plan to increase the text ~~inthefuture~~. Our writing methods ~~andtechniques~~ enabled us to achieve our aims ~~andobjectives~~
  - The result was shorter than we had hoped, but we plan to increase the text. Our writing methods enabled us to achieve our aims
- Complex words
  - Use familiar words (keep it plain not posh!)
- Jargon: be careful when...
  - Assigning precise/specific meaning to words that others may interpret differently (e.g. the use of the word 'significant' has a very precise statistical meaning that your reader might not be familiar with)

– Using expressions that aren't used in everyday speech – you might need to define them

Source: Biotext 2011. Successful science writing and editing

These steps will often need to be repeated several times to "polish" the document and make sure it is as good as it can be.

If you can, walk away from the document for one or two days (long enough to travel home from this course) before you do your final revision. We need to come back to the document with fresh eyes! Otherwise, we find we "skim" over the document. When we do this - we tend to read what we think we have said or what we intended to say rather than what is actually on the page!

16.7 Preparing the final document

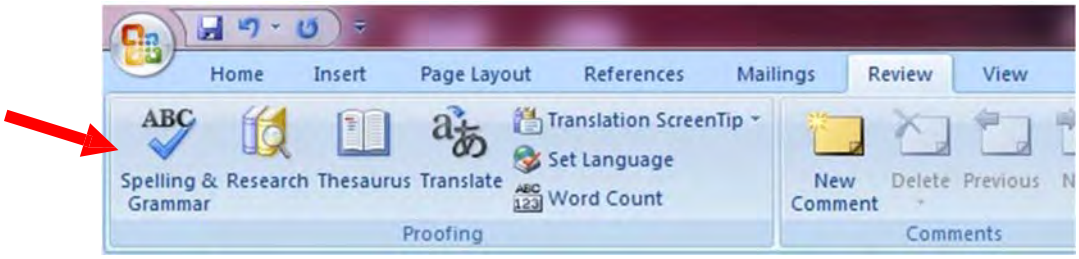
When you have finished all your revisions and edits, you should do a final check to proofread before the document is considered complete.

3. Proofreading is very close attention to detail; concerned with the typographical and visual layout of the document and the technical aspects of its language. Therefore:

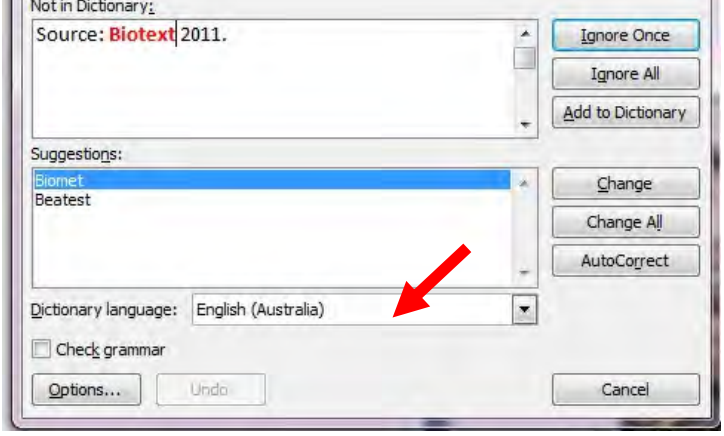
a) Know the conventions of English and be consistent

b) Know what typographical and visual aspects are needed and be consistent c. Avoid cramping your writing by not allowing sufficient white space.

Make sure that you have done a final spelling check – there is nothing worse than a published document with hundreds of spelling errors. You can do this using the spell check tool in word (under the review tab).



Remember though – that word may use different versions of English so you should pick the appropriate one for your country! You can do this by clicking on the spelling and grammar button and selecting a different dictionary.



Beware of spell- and grammar-checkers; they draw attention to obvious errors and 'typos' but none of them is perfect.

You should also read through to look for over/under punctuation, grammatical mistakes and the ease of following the argument.

Final checks

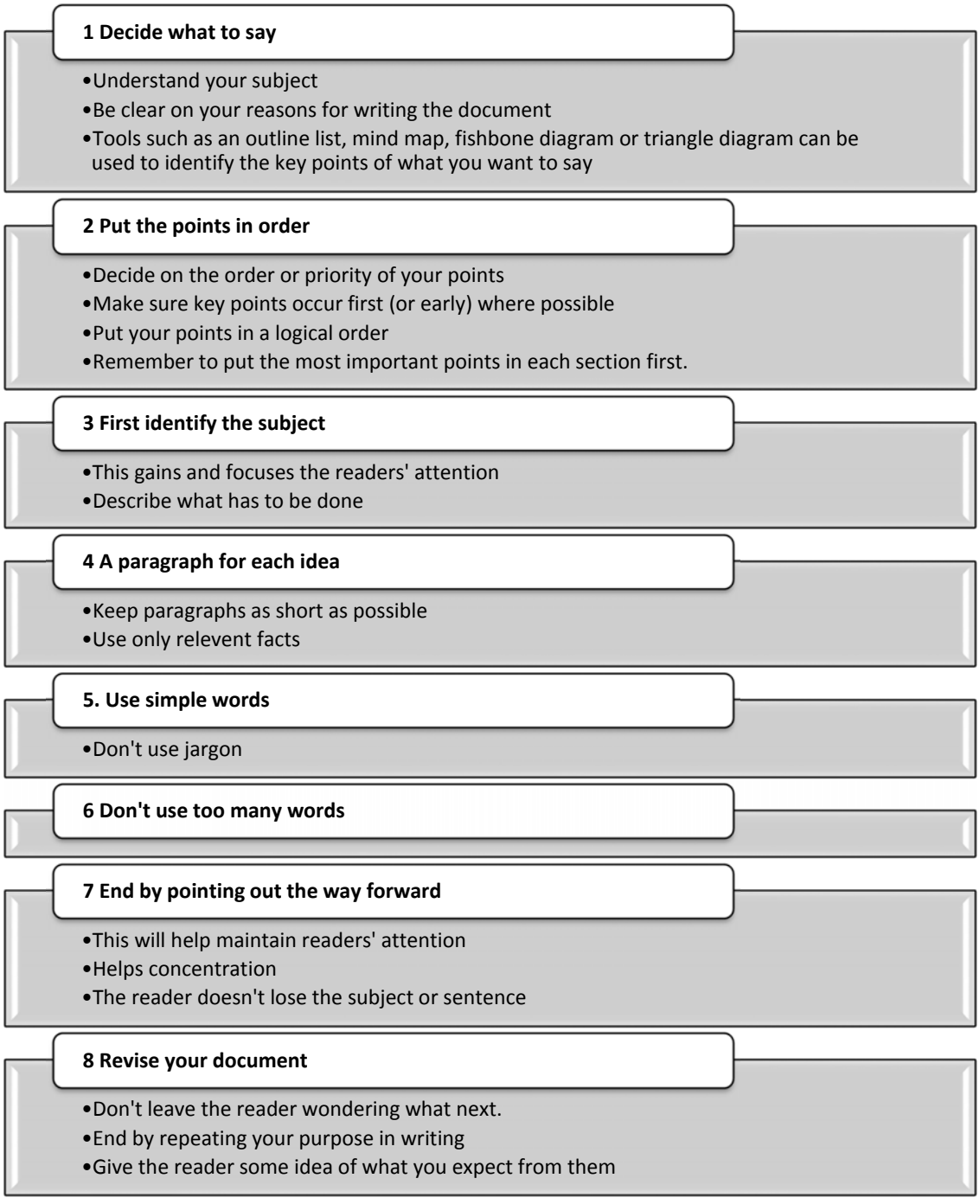
? Look for things like passive sentences and abstract nouns.
? Can you shorten any sentences?
? Have you used words that your reader won't understand?
? Have you written clearly?
? Have you met your deadline?
? Is all the material referenced appropriately?

Once you are happy with your final document, it is also always a good idea to ask someone else to proof read it as well – as they may see things that you have missed, and will be able to point out to you where the document is not clear.

Characteristics of effective documents

- Consistent
- Concise
- Structured
- Clear
- Informative
- Interesting
- Attractive
- Readable
- Relevant
- Well presented
- Grammatically correct
- Easily available

Figure 16.6 Key things to remember when writing a document



17 CREATING TABLES FOR REPORTS

Extract from SPC's Data analysis and report writing course [Ryan, C. et. al. 2011]

As you work through this course, you will generate a number of tables for analysis. Many of these will need to be included in the main report, while others may be attached as appendices for further information. In this section, we will take a brief aside to look at the rules for generating and publishing tables so that they can be understood.

Statistical tables

In practice, when we are preparing statistical reports or publications we apply a number of guidelines to the layout and format of frequency distributions. We call these formatted frequency distributions statistical tables. Tables are used to present univariate (for example, a table of age distribution) or bivariate (for example, a table with age categories represented in the rows and sex categories represented in the columns) data. They are also used to show more than two variables – such as tables with age, sex and region.

Definition

A table is an arrangement of data in a number of rows and columns. The simplest form of a table is a column or row of numbers representing the number of units falling in the categories of a single variable and is called a *one-way classification table*.

Guidelines

The guidelines for published or released statistical tables are:

- Have a reference to the table (such as a table number);
- Have a clear title;
- Have rows and columns clearly labelled;
- Specify the units of the data in the table (for example, kg);
- Include the source of the data;
- Use vertical and horizontal lines to separate the labels from the data themselves;
- Usually do not have the columns separated by vertical lines or rows by horizontal lines – this splits the table up too much;
- Space the table entries so that the table is easy to read;
- Use summary statistics (e.g. sub-totals, means) to provide additional summary information;
- Include footnotes to explain any strange features in the data;
- Use appropriate rounding (usually to one or two decimal places); and

Make sure that you have not breached confidentiality by disclosing personal or commercially sensitive information.

Formatting tables

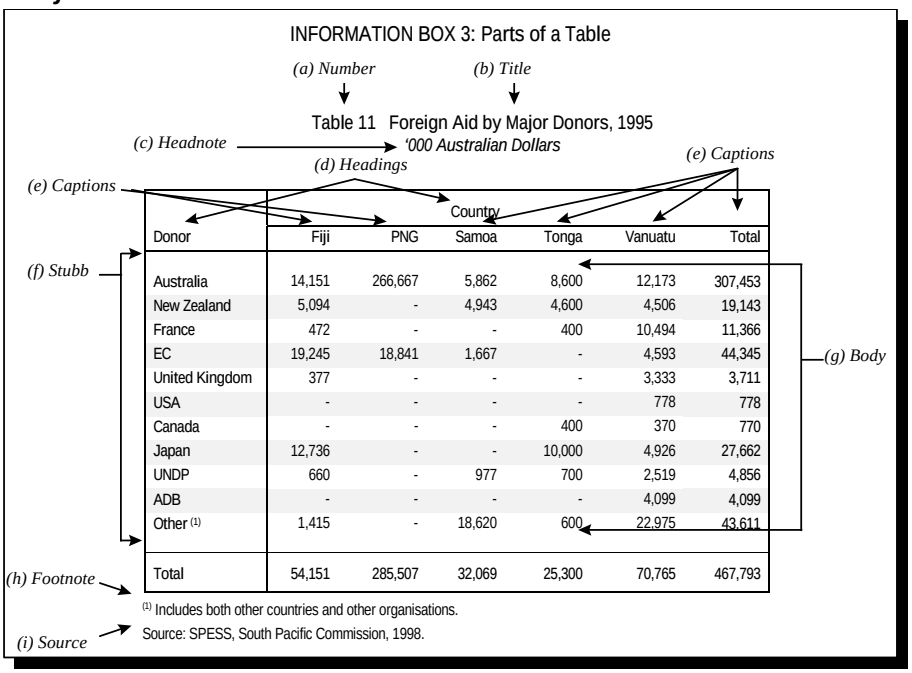
The following principles will help you format your table:

Put numbers most likely to be compared with each other in columns.

Where practical, put columns with larger values at the left of the table, and columns with smaller values at the right of the table.

A table set out following these guidelines will be much easier to read and understand.
When in doubt keep tables SIMPLE.

Figure 17.1 ***Parts of a table***



(a) table number

This identifies the table and precedes the title. If any report or publication contains more than one table they should all be numbered.

(b) title

The **title** is placed above the main body of the table. It should be brief and concise but fully self-explanatory. In some cases several lines of title are necessary. If a title is too long, then an **abbreviated title** may be used above the full title.

(c) headnote

In some tables it may be necessary to include a **headnote**. This is usually printed in smaller types than the title and provides supplementary information about the table or a substantial section of it. **Headnotes** are often used to specify the units of the data in the table, or the survey the data was collected in.

(d) headings

The variables in the rows and columns of the table should be defined by a **heading**.

(e) captions

The **caption** is the designation at the top of each column and it explains what each column represents.

(f) stub

This is the left hand column and its caption. It indicates the description of each row in the table.

(g) body

The **body** of the table includes the numerical information that is placed in appropriate cells governed by row and column headings. A **cell** is the intersection of one row and one column.

(h) footnotes

These provide explanations concerning individual numbers or column or rows of numbers. They are placed at the bottom of the table and are usually in smaller type. They are denoted by either letters of the alphabet or numbers and should run left to right down the page. A **new set of footnotes** should be provided with each table. Only in cases where lengthy repetition will be avoided should the words "See footnote .. to table .." be used instead of repeating a footnote.

(i) source notes

If statistics are collected from a secondary source then this should be acknowledged below the title or more usually below the footnotes.

Rounding

Reasons for rounding
Rounding is often the first step in simplifying and summarising statistical data. Good rounding is essential if a table is going to be easy to understand.

There is often a fear that "accuracy is being lost" when rounding is done. There are two arguments against this which you should consider:

Accuracy versus understanding

Just because a computer can produce a number with lots of decimal places in it does not mean the number is really that accurate. *A number is no more accurate than the instrument which measured it.*

Even if the number is super–accurate and merits lots of digits, and the alternative is a rounded number which can be understood at a glance, you would normally opt for the rounded option.

Rules

The following are the general guidelines for performing rounding:

Numbers **less than 5 are rounded down** Numbers **greater than 5 are rounded up** If the number is 5 then in CRVS analysis **we would round this up**

Example
9.**3** is rounded to 9.0
9.**6** is rounded to 10.0
9.**5** is rounded to 10.0

In Table 6.1 you will see that if you add up the rounded numbers of females you get 51,500 but if you round the raw number total (51,583) you get 51,600. The general guideline here is that the rounded totals should be consistent with the unrounded totals – if you were presenting only the rounded numbers the total should be 51,600.

Table 17.1 Population by State, Federated States of Micronesia, 1994

State	Un-rounded		Rounded to the nearest '00	
	Males	Females	Males	Females
Yap	5,565	5,613	5,600	5,600
Pohnpei	17,253	16,439	17,300	16,400
Kosrae	3,806	3,511	3,800	3,500
Chuuk	27,299	26,020	27,300	26,000
Total	**53,923**	**51,583**	**54,000**	**51,500**

Source: 1994 FSM Census of Population and Housing, Detailed Social and Economic Characteristics Report, 1996.

For population data or other data involving large numbers, we may wish to round our data for clarity.

However, for CRVS data, the basic unit of interest is a single person or individual, so we would tend to round up or down to a whole number in order for our information to be meaningful, but would tend not to round any further as we are dealing with rare events and small numbers.

Calculating percentages

Often it is helpful to present your data as percentages. To change an amount to a percentage divide it by the total and multiply by 100.

Percentage

$$\text{Percent} = \frac{\text{amount}}{\text{total}} \times 100$$

Guidelines

Generally we do not have more than two decimal places with percentages. The total of the percentages should add up to 100. You should also take care to say, either in the column title or as a footnote, what number was used for the total – especially if the overall total was not used. For example:

Table 17.2 : Population by State, Federated States of Micronesia, 1994

State	Number		Percent of total population	
	Males	Females	Males	Females
Yap	5,565	5,613	5.3	5.3
Pohnpei	17,253	16,439	16.4	15.6
Kosrae	3,806	3,511	3.6	3.3
Chuuk	27,299	26,020	25.9	24.7
Total	**53,923**	**51,583**	**51.1**	**48.9**

Source: 1994 FSM Census of Population and Housing, Detailed Social and Economic Characteristics Report, 1996. End of extract

18 GRAPHICAL PRESENTATION OF DATA

Extract (adapted) from SPC's Data analysis and report writing course [Ryan, C. et. al. 2011]
With additional material from: Making Data Meaningful [United Nations Economic Commission for Europe, 2009]

A picture is indeed worth a thousand words, or a thousand data points. Graphs (or charts) can be extremely effective in expressing key results, or illustrating a presentation. An effective graph has a clear, visual message, with an analytical heading. If a graph tries to do too much, it becomes a puzzle that requires too much work to decipher. In the worst case, it becomes just plain misleading. In this topic the main types of graphical presentation are presented along with the types of data the graphs are most suitable for depicting statistical data.

Good statistical graphics:

- • Show the big picture by presenting many data points;
- • Are "paragraphs" of data that convey one finding or a single concept;
- • Highlight the data by avoiding extra information and distractions, sometimes called "non-data ink" and "chart-junk";
- • Present logical visual patterns.

When creating graphics, let the data determine the type of graph. For example, use a line graph for data over time, or a bar graph for categorical data. To ensure you are not loading too many things into a graph, write a topic sentence for the graph.

18.1 GENERAL RULES WHEN PREPARING GRAPHS

We must always be aware that different graphs show different aspects of the same data. As we prepare graphs there are a number of general guidelines to follow to ensure that the graph can be understood and interpreted.

- ☆ Graphs should have a clear, self-explanatory title.
- ☆ The units of measurement should be stated.
- ☆ Use only one unit of measurement per graphic.
- ☆ Graphs should be simple and not too cluttered.
- ☆ All axes should be carefully labelled.
- ☆ Include the source of the data.
- ☆ The scale on each axis should not distort or hide any information. That is, the graph should show the data without changing the message of the data.
- ☆ Graphs should clearly show any trend or differences between the data.
- ☆ Start the Y axis scale at zero- Use 'scale breaks' for false origins – where the scale on the axis does not start at zero to avoid misinterpretation. A scale break is indicated by a 'wiggle' on the axis.

(Note – there are a few exceptions to this such as when graphing life expectancy at birth where by convention the scale should start at 40).

☆ The type of presentation needs to be chosen carefully. This includes the size and shape of the plotting symbol (e.g. asterisks, dots or squares) and the method, if any, of connecting points (straight line, curve, dotted line, etc.). Make your chart clear by:

- Using two-dimensional designs for two-dimensional data;
- Using solids rather than patterns for line styles and fills;
- Avoiding data point markers on line graphs.

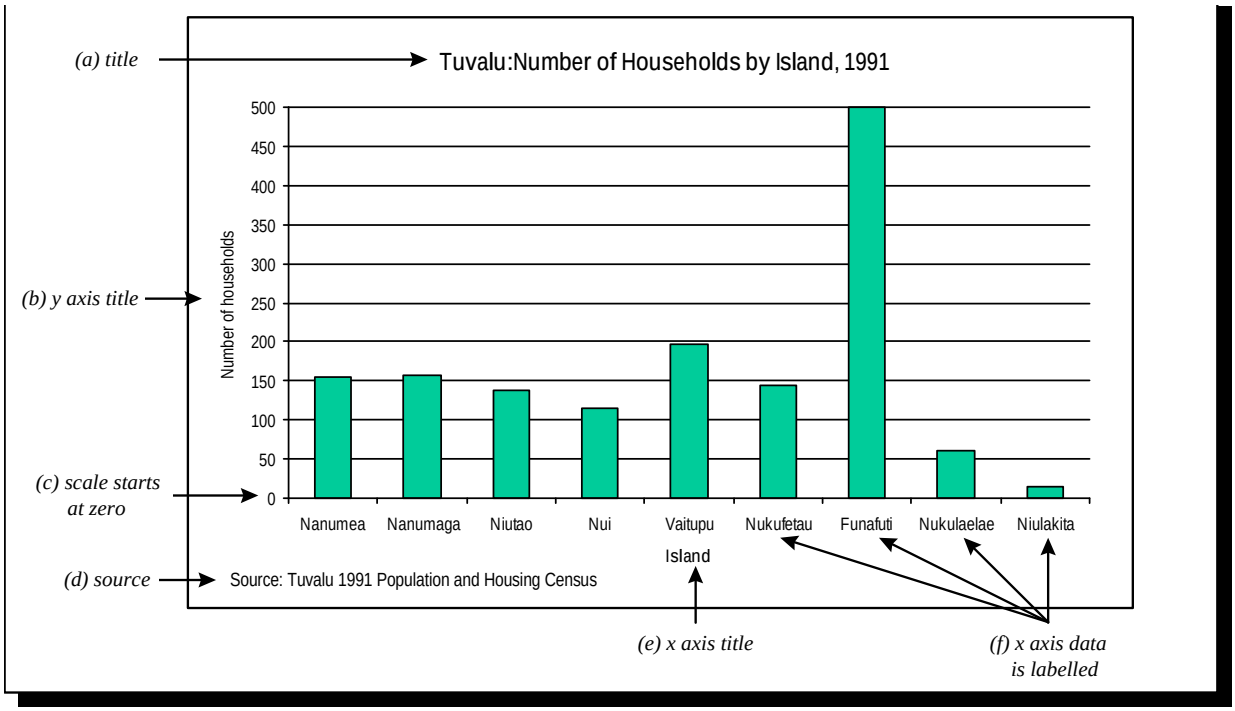
☆ Graphs should be accurate in a visual sense – if one value on the chart is 15 and another 30, then the second value should appear to be twice the size of the first.

☆ Make all text on the graph easy to understand by,

- Not using abbreviations;
- Avoiding acronyms;
- Writing labels from left to right;
- Using proper grammar;
- Avoiding legends except on maps.

▪ A trial and error approach can be very helpful in improving graphs. It is unlikely that you will find the right graph the first time.

Figure 18.1 Parts of a graph



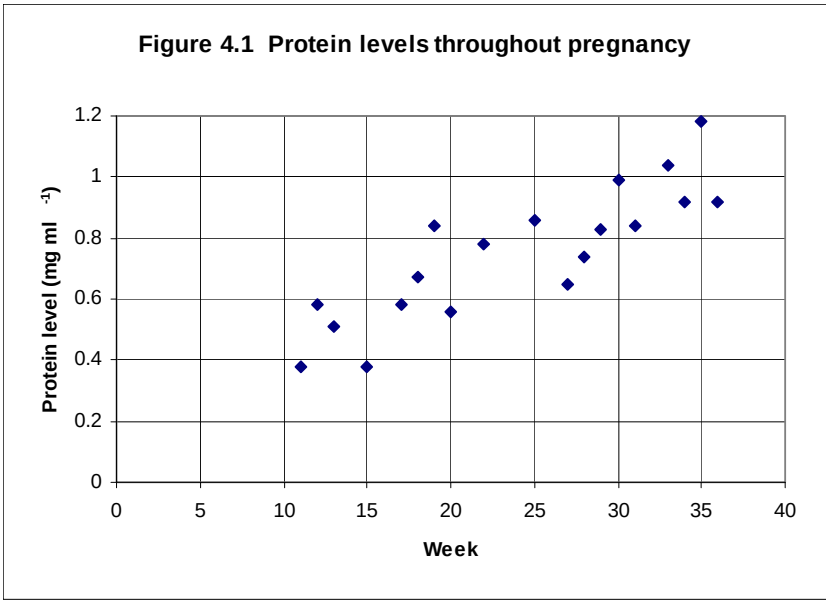
18.2 TYPES OF GRAPHS

Scatter diagrams

simple to see relationships clearly

One of the simplest forms of graphical presentation is the **scatter diagram**. Scatter diagrams are usually used to show the relationship between two variables (bivariate data), to detect any outliers or to detect clusters of observations.

Figure 18.2 Example of scatter plot diagram



Source: Illustrative data only.

outliers – the unexpected

Figure 4.1 depicts quantitative bivariate data, where each observation consists of two measurements – time in weeks and a protein level. It shows the change during pregnancy in the protein levels. We can see the **outliers** – that is, the observations which are higher or lower than the general pattern of the other observations.

Bar charts

qualitative and quantitative data

Bar charts are used to describe the distribution of **qualitative and quantitative data** where the data is grouped into **equal sized class intervals**. *The frequency of the variable is represented by the height of the bar*.

comparisons

Bar charts are good to use when you want to make a comparison between two variables – such as differences between males and females. With nominal data, the order of the categories on the axis is arbitrary, although the variables are often ordered from largest to smallest in frequency.

vertical or horizontal

Bar charts can be vertical or horizontal, although they are usually vertical with the variable categories on the *x* axis and their frequencies on the *y* axis.

TIP

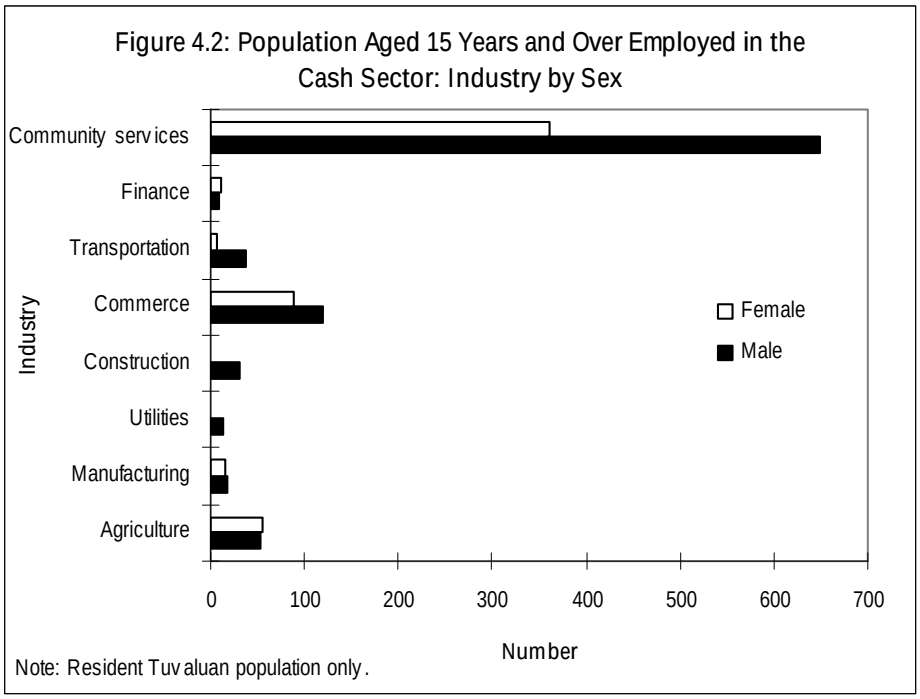
Vertical bar charts are better to show the frequency distribution across the variable categories.

Horizontal bar charts are used when the labels for the variable categories are too long to fit neatly on the *x* axis. It is also easier to read the scale on horizontal bar charts.

Example

The chart used in Information Box 4 is a bar chart (vertical). Figure 4.2 is a horizontal bar chart. Imagine how small the labels would have to be to fit neatly on the horizontal axis!

Figure 18.3 Example of bar graph



Source: Tuvalu 1991 Population and Housing Census.

Pie charts

slice of the pie

Pie charts are one of the most commonly used graphs for **qualitative** data. Each "slice" of the pie represents the relative class frequency of that category. They are simple to construct, but are most effective when you have **five or fewer classes** in your data.

small number of classes

Generally we use pie charts when there is a small number of classes and we are interested in the relative frequency of *each category* (rather than comparisons with other variables).

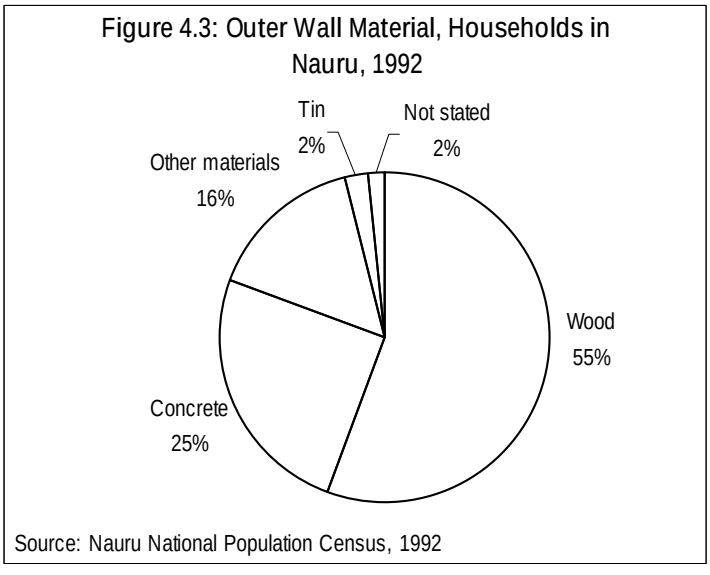
TIP

An important guideline when constructing pie charts is that the segments should be **ordered by size.** This makes it easier to read and to understand the information.

Example

Figure 4.3 below contains data from the Nauru 1992 Census. In this chart the percentage of each variable is displayed. This is a useful option when creating pie charts since there is no scale to read from as in bar charts.

Figure 18.4 Example of pie chart



Line graphs

Independent and dependent variables

Line graphs are used for bivariate data. They have two axes – the horizontal (*x*) axis on which the scale for values of the **independent variable** is recorded, and the vertical (*y*) axis on which the scale for the **dependent variable** is recorded. The independent variable is the variable on which predictions are based, while the dependent variable changes in relation to the dependent variable. Time is an example of an independent variable – it depends on nothing. For example, when we look at population counts from successive censuses, we could use a line graph. Census year would be the independent variable, and the population count would be the dependent variable.

Common usage

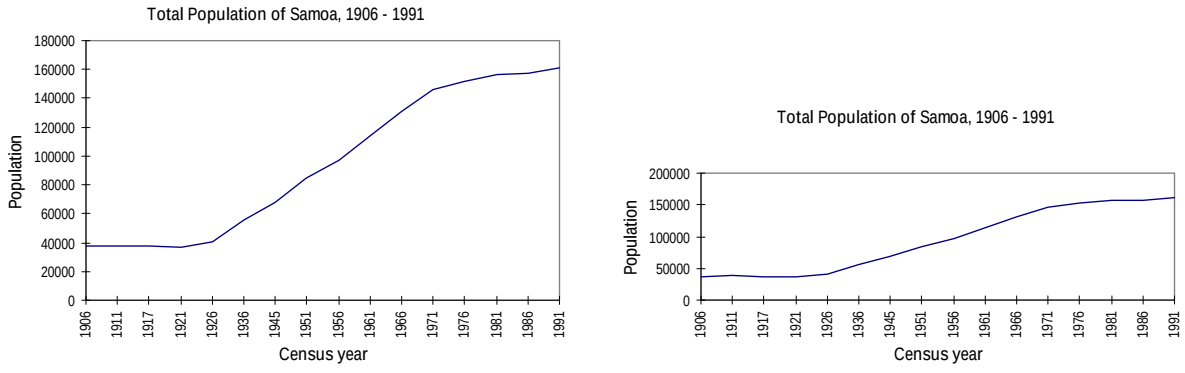
Line graphs are commonly used to display data *over a time period* and when you want to show *trends* in data. Care must be taken when preparing line graphs that the scale on the vertical (*y*) axis does not give the wrong impression about the data.

Example

Figure 4.4 contains two line graphs with the same data, but different scales are used on the *y* axis. The second graph gives a much different impression than the first. Looking at the first graph you would assume that the population increased quite radically from the 1930s – 1960s, but the second graph does not give the same impression. It is particularly important when comparing different populations to keep the same scale.

Figure 18.5 Example of line graph

Figure 4.4: Total Population of Samoa, 1906 – 1991



Source: Census of Population and Housing, 1991, Western Samoa

Histograms

Area of each bar is proportional to frequency

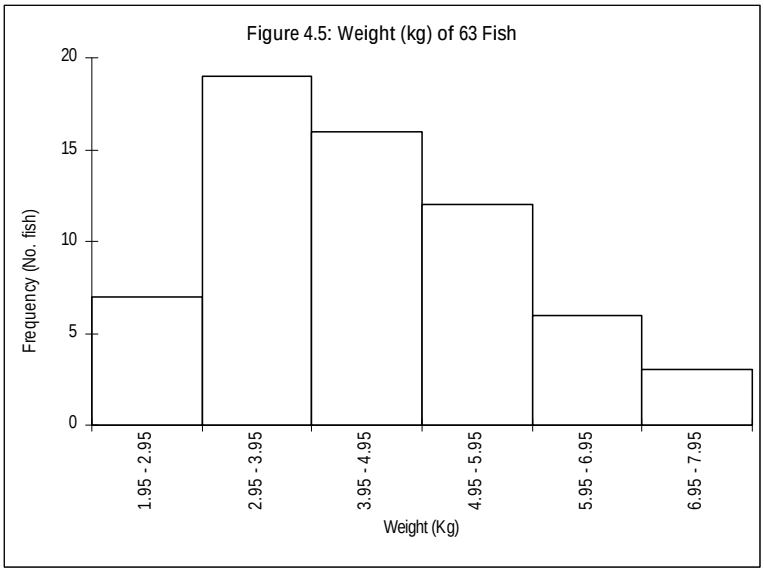
The most common method of representing a frequency distribution for **continuous data** (or data that is assumed to be continuous) is by using a **histogram**. A histogram is a vertical bar chart with the *area of the*

bars (not necessarily the heights of the bars) proportional to the class frequencies (whereas for bar charts both the height and area of the bars is proportional to the class frequency).

Different from bar charts

Histograms differ from bar charts because the vertical bars are always side by side, without gaps. This is to reflect the continuous nature of the data. Generally a histogram has equal width bars, although when class intervals are not equal the width of the bars has to be adjusted.

Figure 18.6 Example of bar chart



Source: Table 3.1

Area is important

You must remember when plotting a continuous distribution, it is the **area of each rectangle** (and not the height) that is **proportional to the frequency**. It is only in the case of equal class intervals (as in Figure 4.5) that the frequency is proportional to the height of the rectangle.

Unequal classes

Let us look at another example of a histogram. The histogram in Figure 4.6 (from Table 4.1) shows how we should deal with a distribution with unequal class intervals and with open-end class intervals. Remember it is the **area of the bars** which is important in a histogram. This means when we have unequal classes we have to make sure the area of each bar accurately represents its frequency. The class 100 – 199 is twice the size of the 50 – 99 class, so in the histogram the width of the 100 – 199 rectangle has to be twice the width of the 50 – 99 rectangle.

height = frequency density

To ensure the area of the rectangle reflects the class frequency, the height of the rectangle must then be the frequency density.

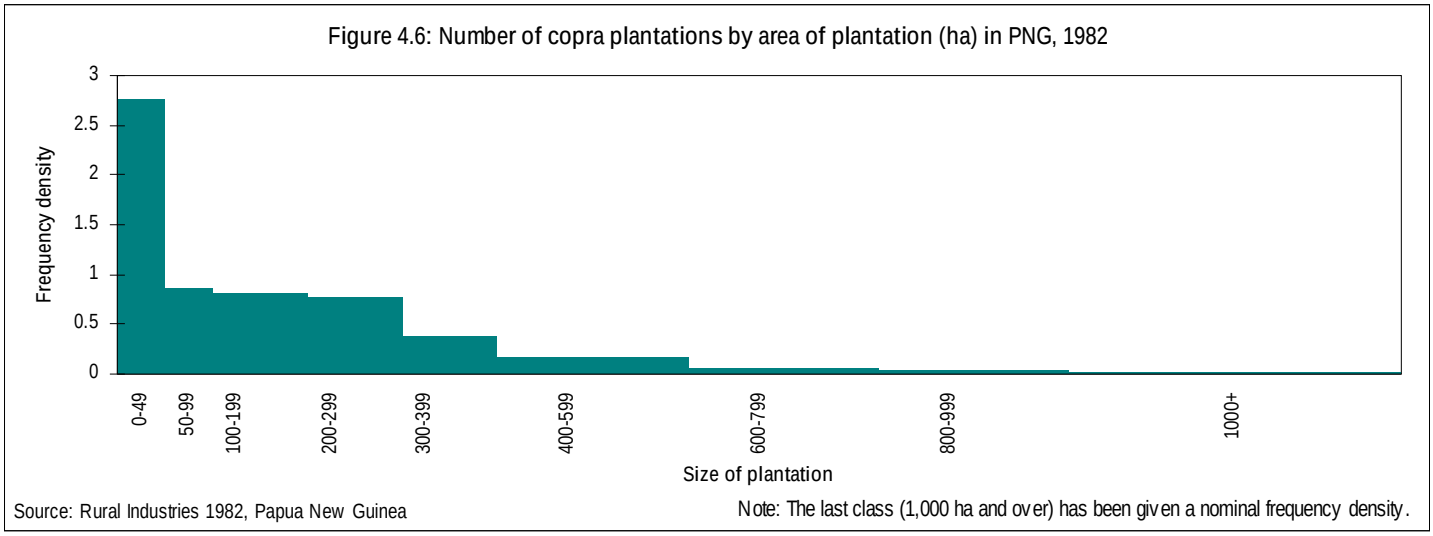
Note that the method for dealing with open-end class intervals has been footnoted, to ensure the reader is informed of the method used.

Table 18.1 Number of Copra plantations by area of plantation in PNG. 1982

Class Group (ha)	Class Interval	Class Frequency	Frequency Density (per ha)
0 - 49	50	138	2.76
50 - 99	50	43	0.86
100 - 199	100	81	0.81
200 - 299	100	77	0.77
300 - 399	100	38	0.38
400 - 599	200	34	0.17
600 - 799	200	13	0.07
800 - 999	200	6	0.03
1,000 and over		8	–
Total		**438**	

Source: Rural Industries 1982, Papua New Guinea.

Figure 18.6 Example of bar graph with unequal classes



Open ended classes

In Figure 4.6 there is also a problem in deciding how to deal with open-ended classes, in this case the '1,000 and over' class. We do not know the width or the interval of this class and we have no way of making even a reasonable guess. In some situations is it possible to assume an upper class limit – for example if we were dealing with data on age at marriage we could assume an upper limit of 60, since it is unlikely that people over 60 would marry.

In Figure 4.6 we have no way of knowing the upper limit of the 1,000 and over class so we cannot calculate the height of the rectangle to represent this part of the total frequency. It would be wrong to leave it out altogether, so we have to decide what to do with this class.

Two ways of dealing with open ended classes

The problem can be dealt with in one of the two ways. Firstly, we can assume an upper limit for the distribution and draw the rectangle accordingly (such as with the age at marriage illustration). The second alternative is to draw the rectangle with a nominal height but leave it open and footnote the method applied, as in Figure 4.6 above. This would then indicate an open-ended interval. This second method, however, only works when the frequency in the open-ended class interval is small (as is the case in Figure 4.6), but this should be the case (large frequency open ended classes should always be avoided).

Discrete data

When we draw a histogram for **discrete data**, we have two possible approaches. Let us look at Table 4.2 given below:

Table 18.2 Employment size group by number of paid employees, 1996

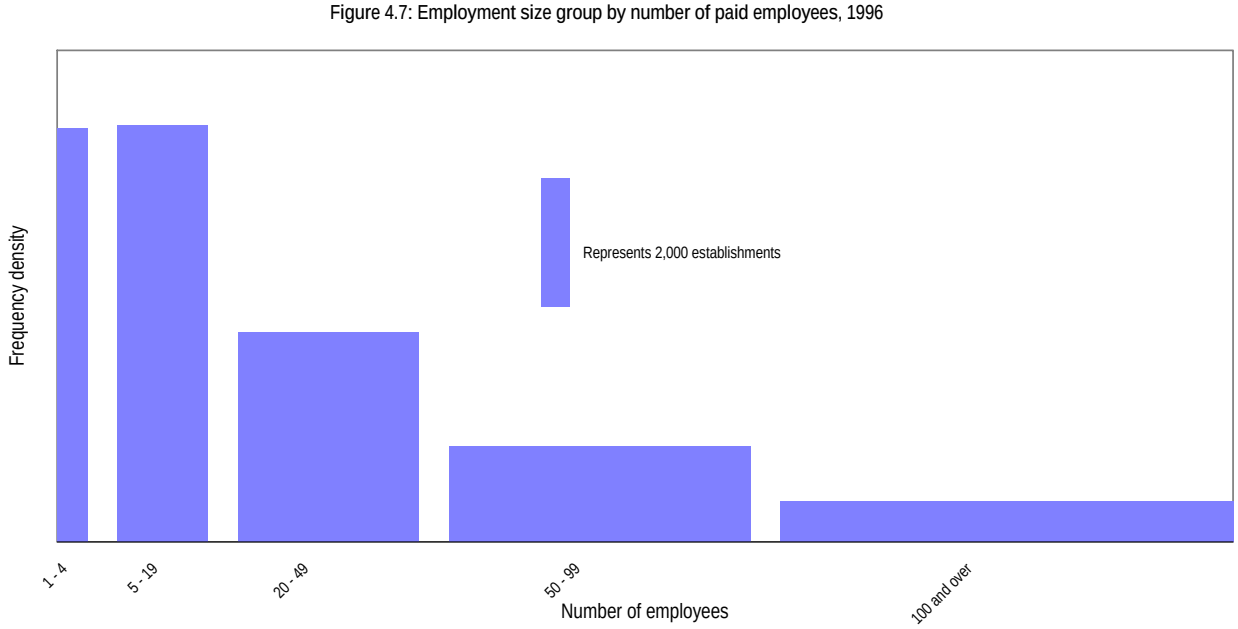
Employment size group	**Wage and salary earners**	**Earners per enterprise** = freq * class mid-point	
1–4	5,053	=5,053 ÷ 4 =	1,263.25
5–19	15,260	=15,260 ÷ 15 =	1,017.33
20–49	15,312	=15,312 ÷ 30 =	510.40
50–99	11,714	=11,714 ÷ 50 =	234.28
100 and over [1]	62,742	= 62,742 ÷ 400 =	156.85
Total	**110,081**		

Assumed upper limit of 500
Source: 1996 Annual Employment Survey, Fiji Bureau of Statistics, Table 2, p. 15.

Two alternatives

The first approach is to draw a histogram with the rectangles separated by spaces. This signifies to the reader that the data is discrete. This approach is shown in Figure 18.7.

Figure 18.7 Example of histogram drawn from discrete data



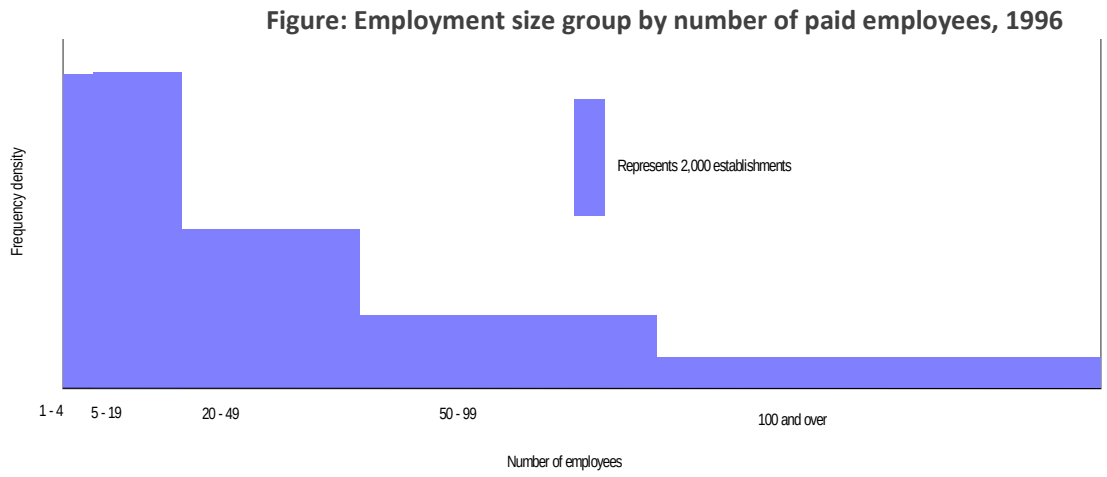
Source: 1996 Annual Employment Survey, Fiji Bureau of Statistics

Note: The 100 and over class has been given a nominal frequency density

Second alternative

The second approach is to proceed as if the data were, in fact, continuous. If we have a distribution similar to that given in Table 4.2, then the unit of measurement (in this case single persons) is small compared with the range of values we have observed. In such a case, we can draw an approximate histogram by proceeding as if the data were continuous. The histogram is drawn in Figure 4.8 and looks somewhat like Figure 18.7 since the horizontal scale is divided into unequal intervals. You must realise that *Figure 18.7 is only an approximation* because the data are discrete and so we cannot, strictly speaking, represent them by a continuous rectangles. For the purpose of representing the general shape of the distribution, however, the diagram is quite adequate.

Figure 18.8 Example of histogram drawn from an assumption of continuous data



Source: 1996 Annual Employment Survey, Fiji Bureau of Statistics

Note: The 100 and over class has been given a nominal frequency density

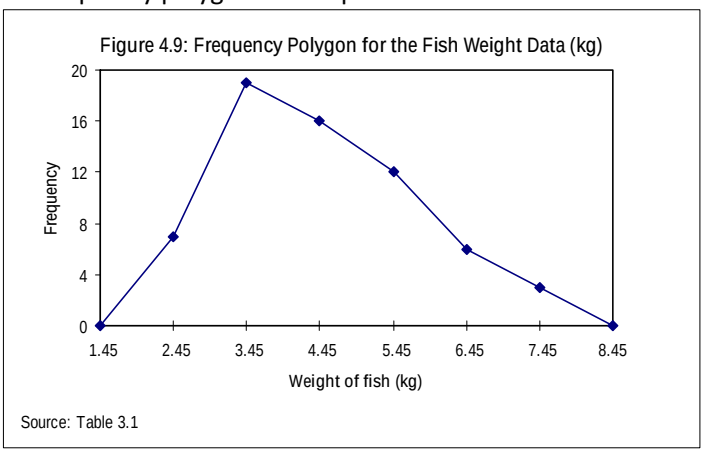
Frequency polygons

Frequencies plotted at class midpoint

An alternative type of graph, suitable for **continuous data**, or for *discrete data which can be considered to be approximately continuous*, is a **frequency polygon**. In this graph the frequencies of each class are plotted at the class midpoint, and successive points joined up by straight lines.

An example of a frequency polygon with equal class intervals is given in the figure below.

Figure 18.9 An example of a frequency polygon with equal class intervals



Unequal class width

When the class intervals are unequal lengths then, as in the case of a histogram, the frequency densities or relative frequency densities are to be used.

Must start and end at the x axis

The beginning and end of the polygon should be extended to the horizontal axis, to the mid-points of the classes below and above those covered by the distribution.

For a discussion of the graphical presentation of cumulative frequency distributions, refer to the following section ‘More on Frequency Distributions’.

Cumulative frequency distributions

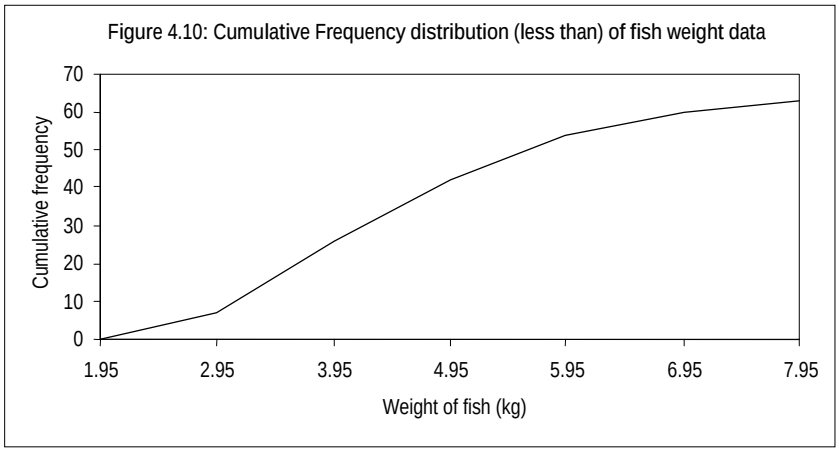
Ogives

Just as for some purposes it is better to construct a cumulative frequency distribution, so we find it useful to represent this graphically. A cumulative frequency distribution of the fish data is given in Table 4.3. A graphical representation of the cumulative frequency distribution (either less than or greater than) is called the **ogive** and the less than ogive of the Table 4.3 data is given in Figure 4.10.

Table 18.3. Cumulative frequency distribution of fish weight data (kg)

Class Group	Class frequency	Cumulative Frequency (less than)	Cumulative Frequency (greater than)
2.0 - 2.9	7	7	63
3.0 - 3.9	19	26	56
4.0 - 4.9	16	42	37
5. 0 - 5.9	12	54	21
6.0 - 6.9	6	60	9
7.0 - 7.9	3	63	3
Total	**63**		

Figure 18.10 **Cumulative frequency distribution of fish weight data**



Source: Fish Data from Table 4.3

greater than ogive

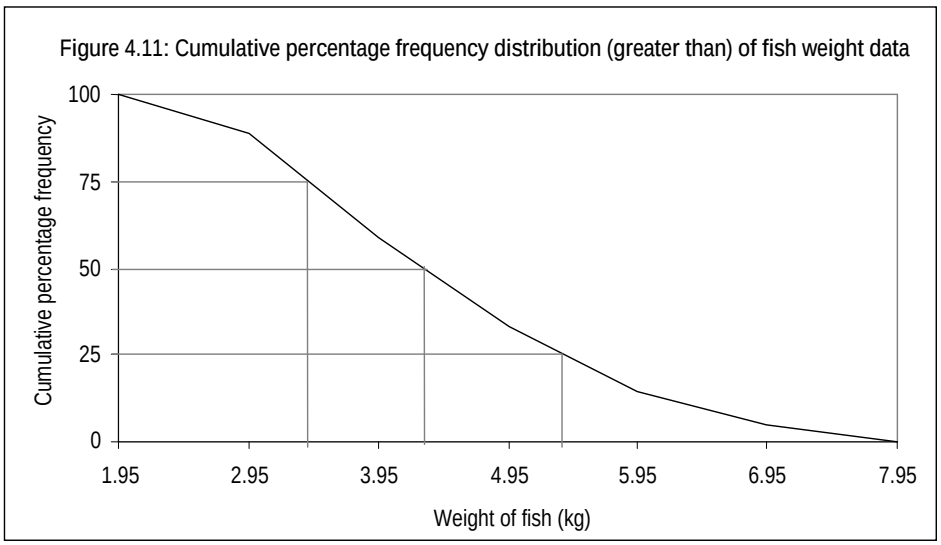
The line in Figure 4.10 represents the cumulative frequency distribution (less than). We can also draw a similar diagram for the cumulative frequency distribution (greater than). This is shown in Figure 4.11. The data used is that from Table 4.4 over.

Table 18.4 : Cumulative frequency distribution (greater than) of fish weight data (kg)

Class Group	Class frequency	Cumulative Percentage frequency (greater than)
2.0–2.9	7	100.0
3.0–3.9	19	88.9
4.0–4.9	16	58.7
5.0–5.9	12	33.3
6.0–6.9	6	14.3
7.0–7.9	3	4.8
Total	**63**	

Source: Table 3.1

Figure 18.11 **Cumulative frequency distribution (greater than) of fish weight data**



Source: Table 4.4

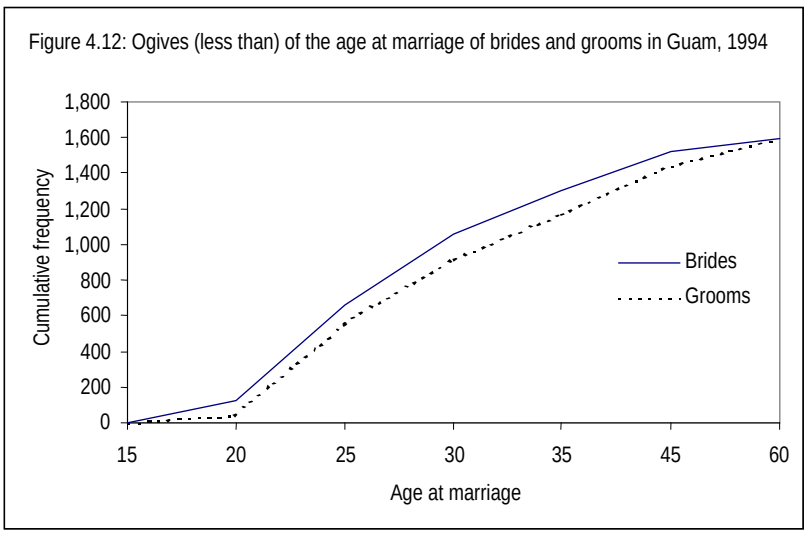
Cumulative frequencies show distribution of the data

In Figure 4.11, instead of plotting the actual frequencies on the vertical axis, we have plotted the cumulative percentage frequencies. We can clearly see from this diagram that 100 per cent of the observations were greater than 1.95 kilograms, 75 per cent are estimated to be greater than 3.4 kilograms, 50 per cent are estimated to be greater than 4.4 kilograms and 25 per cent are estimated to be greater than 5.4 kilograms. This is the graphical representation of quartiles, will be discussed in detail in the section "More on measures of location".

Two distributions

Finally, let us look at the ogive of the age at marriage of brides and grooms data from Table 3.14.

Figure 18.12 Example of ogive curve



Comparisons between variables and observations

Because the two distributions have the same base, it is helpful to show their cumulative frequencies on the same diagram. In Figure 4.12, we can clearly see the 'lag' effect, that is, the difference of age at marriage of brides and grooms. Grooms tend to be older than brides as can be seen from the graph above.

18.3 EXCEL (2007)

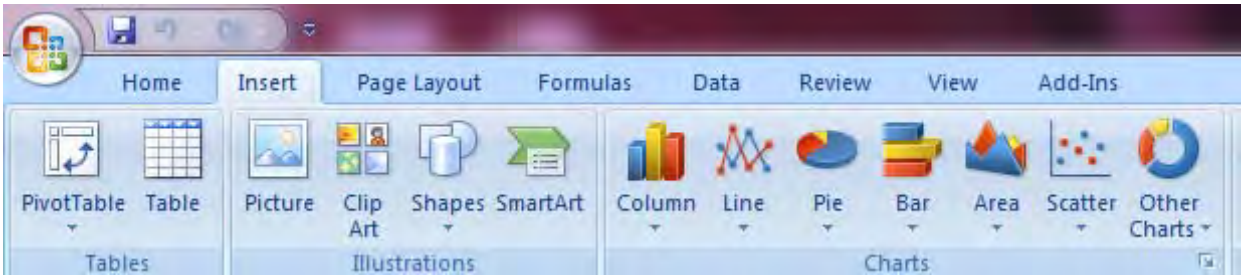
Creating charts

It is easy to create and format charts in Excel using the Chart Wizard ( ). Charts can be *embedded* on the worksheet next to your data, or they can appear as a *chart sheet*. A chart sheet is a sheet in a workbook that contains only a chart. The position of the chart depends on your preference.

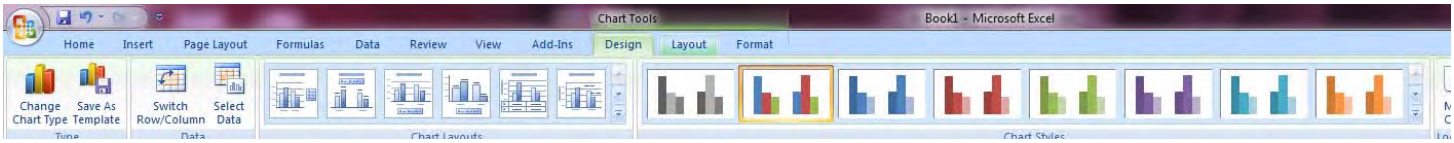
Create a chart

The main thing to be aware of when creating a chart in Excel is to select the correct *x* and *y* variables. For example, when basing a chart on a table containing Male, Female and Total variables, you wouldn't usually chart the 'Total' column. Generally you do not chart the total column if you are charting the parts which make up the total. If you include the total it will affect the scale of the chart, and change the emphasis of data items.

1. Select the cells you want to chart.
2. Under the INSERT tab, select the type of chart you want to create

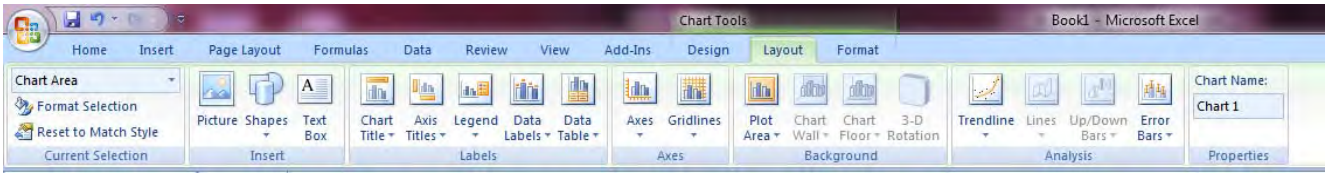


1. To edit the chart, click on you graph and the Chart tools options will appear in your header. Tools to edit the chat can be found in the "designs" tab when using Excel 1013
2. Use the Design tab to change the colours, and set an overall feel for the graph. You can also go back and try a different chart type on the far left of this screen if you don't feel you have selected correctly.



1. Under the layout tab, you can go in and edit the title, axis, labels and gridlines to set the graph up the way you want it to be. Make sure your chart has a title, labels for each axis and the units

are shown. A chart should be able to be read as a standalone item. In excel 2013, a dialogue box that will appear on the right side of the screen will be useful in making these adjustments. Additional options can be found by simply right clicking on the chart.



TIP	As a general rule, if you want to change the look of something on your chart select the object and double click on it. This will open a dialog box that should contain options for formatting.

Graphing exercise 1:

Go back and look at the tabulations you have developed with your birth data. Identify appropriate graphs for each of these and draw these using excel.

19 INFOGRAPHICS

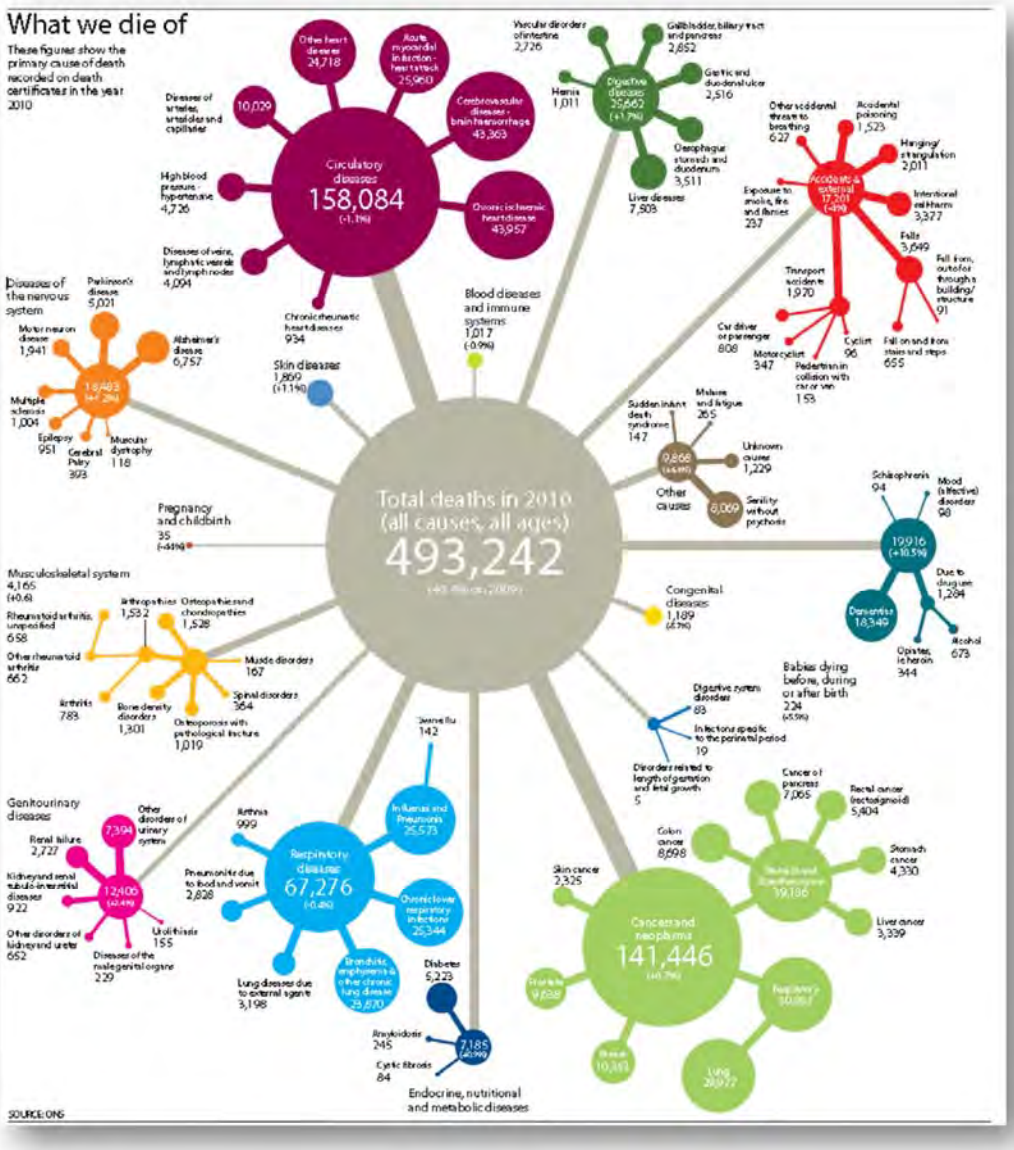
Once you have reliable, robust data to work with, it is possible to do a lot more to make sure that your information is getting the attention it deserves from politicians, planners, and from the public.

Infographics are an extremely useful tool for conveying complex information in a way that makes it readily understandable to a broader audience.

Below are some good examples of mortality infographics to give you some ideas of how your data can be used.

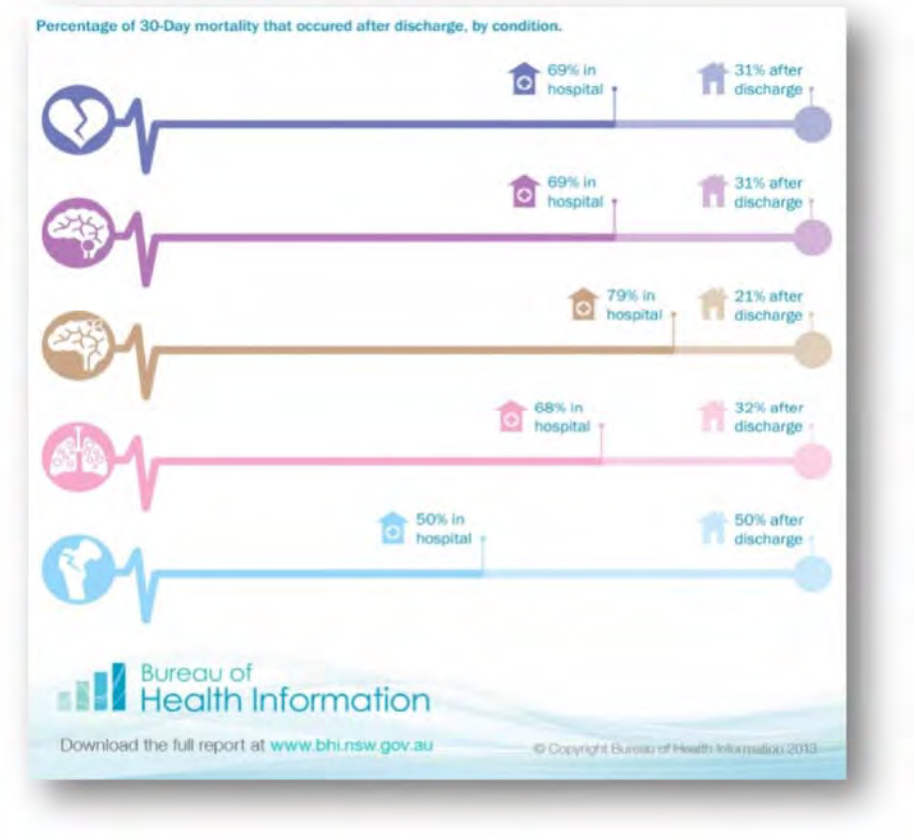
Guardian Newspaper, 2011 – UK mortality data 2010

Figure 19.1 Example of infographic: causes of death



NSW Health, Hospital Mortality

Figure 19.2 Example of infographic: death by place of death and cause



Lifetime risk of Cervical Cancer, IHME

Figure 19.3 Example of infographic: % lifetime risk of death from cervical cancer

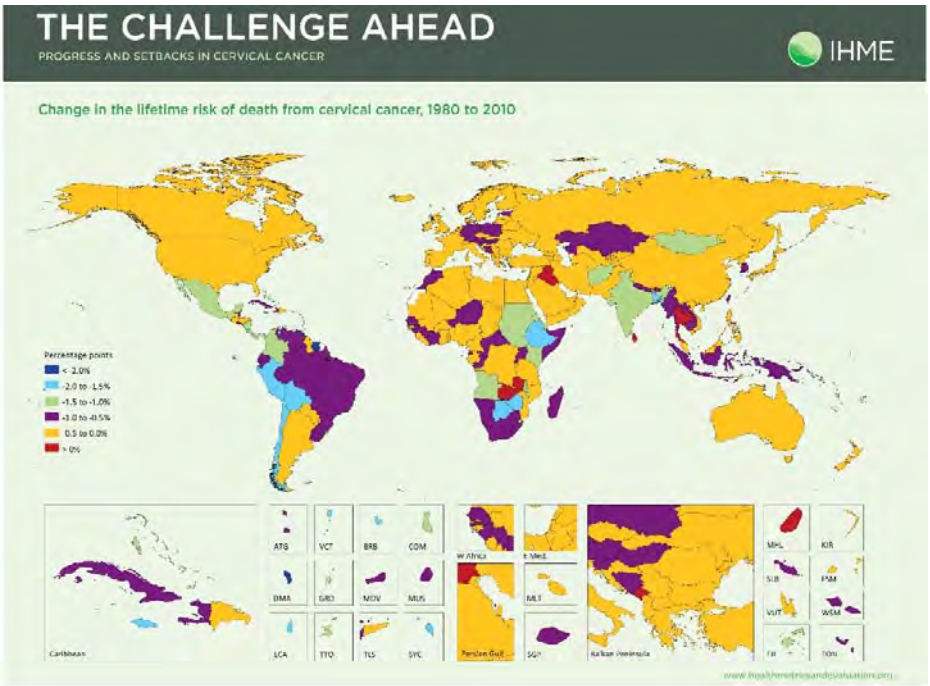
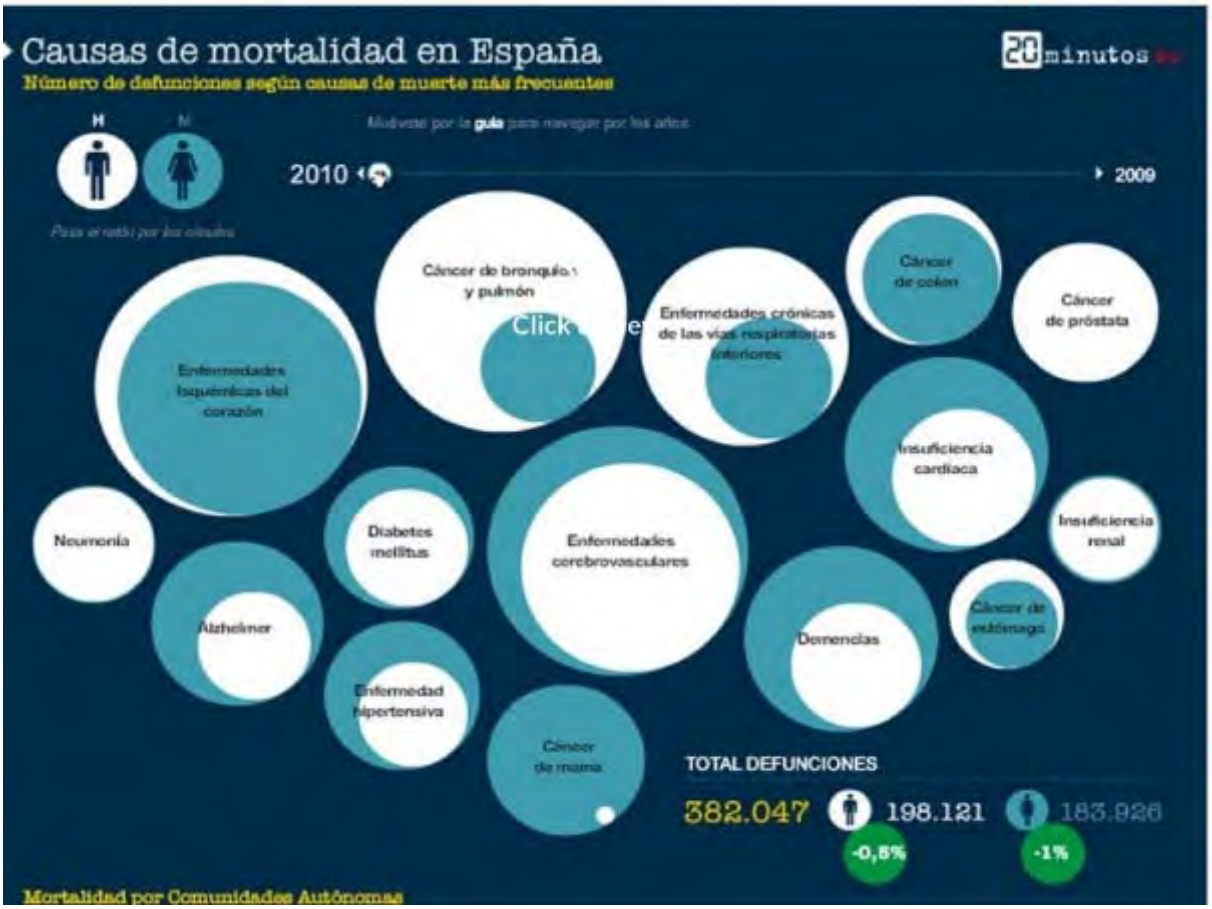


Figure 19.4 Causes of mortality in Spain – comparison of men and women



Exercise: Design an infographic

- Select one of the key messages from your vital statistics and causes of death draft report
- Consider who needs the information and what you are trying to tell them
- Think about how you could display this data in a way that makes it interesting for your intended audience.

Remember to make sure that the information is clear, accurate, and concise.
If you have a great idea – share it with your trainers or have a look at some of the data visualization tools on the web such as http://www.makeuseof.com/tag/awesome-free-tools-infographics/ ; https://www.canva.com/templates/infographics/

20 SPECIAL TYPES OF WRITING

20.1 POLICY BRIEF/ MINISTERIAL BRIEF

Adapted from Young E and L Quinn, The Policy Brief, available at http://www.policy.hu/ipf/fel-pubs/samples/PolicyBrief-described.pdf and Bennett G and Jessani N, The Knowledge Translation Toolkit, SAGE Publications: New Delhi, 2

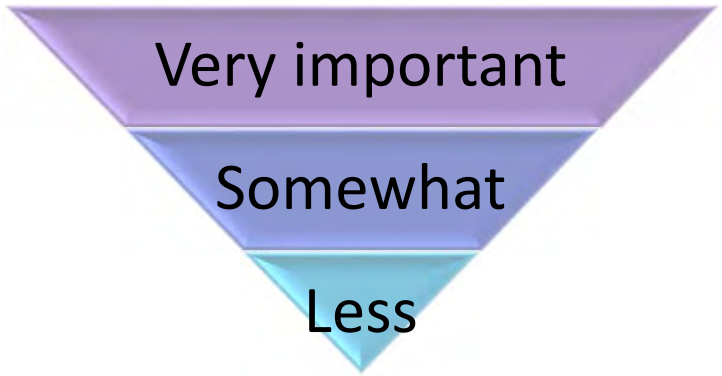
Policy (or Ministerial) briefs are a method of communicating with officials, politicians, development partners, decision-makers and managers. This type of audience will generally have limited time to read any document and also limited technical understanding of any complex terms. So the brief must be just that: a brief (concise) document that shows the importance, relevance and urgency of the issue – all in less than two pages. The policy brief should pass the 'cornflakes' or 'elevator' test: could a politician read the document and understand the main points in the time it takes to eat a bowl of cornflakes? If you were in an elevator and had 30 seconds with a decision-maker, could they read your policy brief before they get out at their floor?

Overall, a policy brief should:

- Identify a problem
- Show evidence that informs policy on the matter
- Propose solutions
- Present a recommendation
- Move the reader from problems to possibilities to policies

As with writing a media release or newspaper article, policy briefs should be written using an 'inverted pyramid' approach (see Figure 1). With this approach, all of the most important information is provided at the very beginning, often all in the first paragraph. All other paragraphs support what is written in the first, with the least important information presented last. This way, if your reader only reads one paragraph (the first), they will still understand the main point(s) of the policy brief.

Figure 20.1: Inverted pyramid writing



An effective policy brief should:

- Address the needs of the audience
- Be written in a professional tone
- Use strong and reliable evidence
- Limit the focus to a particular problem or issue
- Be concise, using short and simple sentences and paragraphs
- Avoid using jargon or acronyms
- Make sure the recommendations are practical and realistic. A policy brief should also have the following components:
- Title. Short and interesting
- Summary. An overview of the problem, its importance, and reasons why action is necessary
- Statement of the problem. Focus on 'what', 'how' and 'when'
- Background and/or context to problem. The essential facts to convince an audience that the problem requires urgent attention and action
- Pre-existing policies. An overview of what has been done in the past or what is in place now (if anything) to address the problem
- Policy options. Some alternative ways to address the problem
- Critique of policy options. The pros and cons of each alternative (from the perspective of the target audience
- Policy recommendations. A preferred alternative, providing a convincing argument and reasons for action.

20.2 WRITING FOR PUBLICATION (JOURNALS AND SCIENTIFIC PAPERS)

Publishing a scientific paper in a journal is an important way of sharing your information with a specific audience, in this case, academics and researchers. Journals are an important communication method as they are available internationally in a number of formats, including hardcopies and electronically. The benefits of publishing in a journal include:

- Highlighting important issues to academics and researchers for further investigation
- Providing a 'Pacific voice' on an international level
- Promoting the use of local data, as opposed to models and estimates
- Creating links with universities in your country or region.

Each journal has its own very specific criteria about word length, writing style, references, etc. It is important you follow these criteria carefully, as shown in Table 1. Examples of journal articles on mortality in the Pacific are provided in the appendices.

Figure 20.2 Sections in a scientific paper

What did I do, in a nutshell?	**Abstract**
What is the problem?	**Introduction**
How did I solve the problem?	**Materials and Methods**
What did I find out?	**Results**
What does it mean?	**Discussion**
Who helped me out?	**Acknowledgements** (optional)
Whose work did I refer to?	**References**
Extra Information	**Appendices** (optional)

Source: Home Department of Biology. 2003. "The Structure, Format, Content, and Style of a Journal-Style Scientific Paper," in *How to Write a Paper in Scientific Style and Format*. Lemistan, Maine: Bates College. Available online at http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWsections.html (accessed October 14, 2010).

A note on Authorship

1. Everyone who is listed as an author should have made a substantial, direct, intellectual contribution to the work. For example (in the case of a research report) they should have contributed to the conception, design, analysis and/or interpretation of data. Honorary or guest authorship is not acceptable. Acquisition of funding and provision of technical services, patients, or materials, while they may be essential to the work, are not in themselves sufficient contributions to justify authorship.
2. Everyone who has made substantial intellectual contributions to the work should be an author. Everyone who has made other substantial contributions should be acknowledged.
3. When research is done by teams whose members are highly specialized, individuals' contributions and responsibility may be limited to specific aspects of the work.
4. All authors should participate in writing the manuscript by reviewing drafts and approving the final version.
5. One author should take primary responsibility for the work as a whole even if he or she does not have an in-depth understanding of every part of the work.
6. This primary author should assure that all authors meet basic standards for authorship and
7. Should prepare a concise, written description of their contributions to the work, which has been approved by all authors. This record should remain with the sponsoring department.

21 INCORPORATING CRVS DATA INTO POLICY AND PLANNING DECISIONS

While analysing and publishing our data is important, our ultimate aim is to ensure that the data is used as evidence to inform policy and planning decisions.

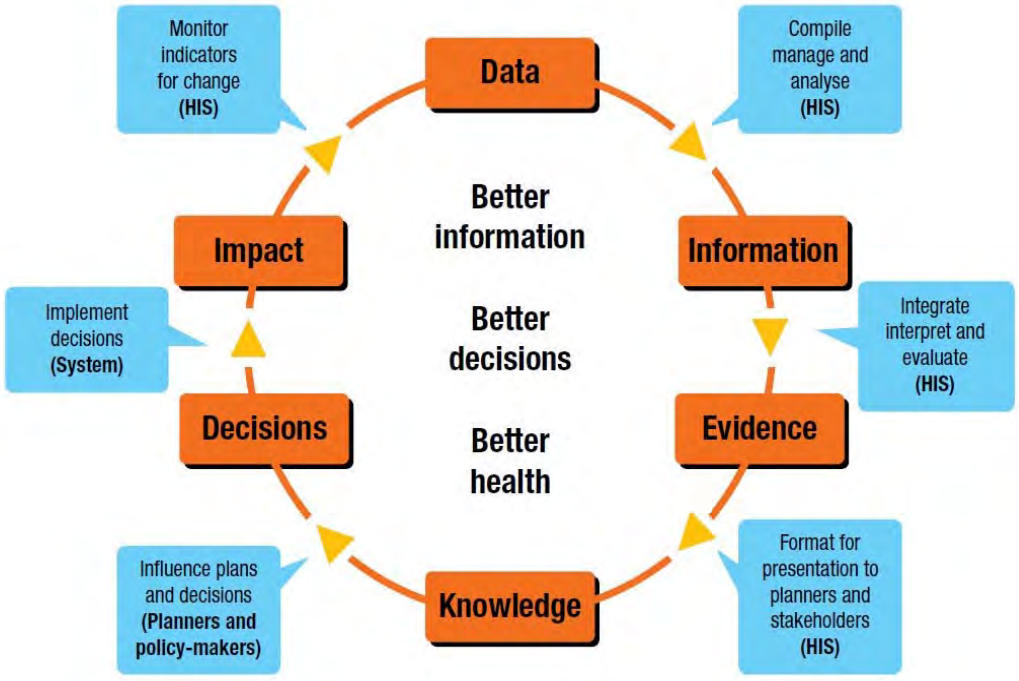
To be useful for policy and planning, data must be

- Trusted,
- Available, and
- Timely

Policy makers must have confidence that the data is both accurate and meaningful.

Figure 24.1 shows the Health Metrics Network data cycle which demonstrates how data becomes evidence and is used in decisions to create impact. While the diagram was written for health data, this applies across all sectors.

Figure 21.1 HMN cycle of data generation, knowledge brokering and use of evidence



A key aspect of this cycle is the integration, synthesis, analysis and interpretation of information from multiple sources. Formatting or packaging of information for decision makers is a core function of any information system in demonstrating the value of the data it contains. Such information products should guide decision makers by providing an analysis of the potential consequences of decisions or scenarios (or consequences of inaction).

21.1 LINKING DATA TO PLANNING PROCESSES

Integrating CRVS data into formal planning and decision making processes is an important step in ensuring information and evidence are used. One way of doing this is by having CRVS-related measures as a goal, strategy or target within relevant sections of a plan. Another way is to clearly identify which indicators and measures can be addressed by data from a CRVS system (or HIS, or other routine data source, depending on what data you are analysing).

Linking National Statistics Development Strategies (NSDSs) to CRVS

National Statistics Development Strategies

At the end of 2003 PARIS21 made a recommendation to all the national and international actors for each developing country to devise and adopt a National Statistics Development Strategy (NSDS) with a view to supporting the millennium development goals (MDGs) and policies to combat poverty. In February 2004 this recommendation was endorsed, and indeed made stronger, at the second international round table on "Results-based Management" in Marrakesh. With the significant increase in the number of goals and targets to be monitored, the SDGs make demand for more robust NSDSs and underline the need for strengthening existing statistical systems; and more importantly vital statistics systems.

A National Statistics Development Strategy (NSDS) may be defined as a coherent set of interrelated decisions taken by the national authorities concerning what will be done over the next 4-5years in order to make available better statistics and better analyses of these statistics, thereby responding to the priority needs of national and international decision-makers and to those of civil society.

While almost universal principles underlie both strategic management and public statistics, an NSDS is part of a specific national context and time frame. When it comes to the actual design of an NSDS, there are three separate phases:

- drawing up an overall diagnosis based on a detailed review of the situation and establishing general guidelines;
- selecting strategic objectives for results and products, together with capacity-building strategies which will make it possible to achieve the results and deliver the products;
- planning the actions which will be carried out until the end of the envisaged period, including monitoring and assessment.

The Pacific Ministers of Health meeting and the regional conference of the Pacific Heads of Planning and Statistics in 2013, both endorsed the importance of linking CRVS improvements and the use of data from routine reporting systems such as CRVS to the development of NSDSs.

Coordinating HIS and CRVS improvements with the NSDS

Most Pacific countries now have multi-department committees established for CRVS and for HIS. These may be separate committees or may in fact be a single committee (working group or steering group etc.) that addresses both sectors. Committees have been encouraged to develop formal terms of reference and

seek official endorsement for the assessment and improvement work, and many countries have indeed done so.

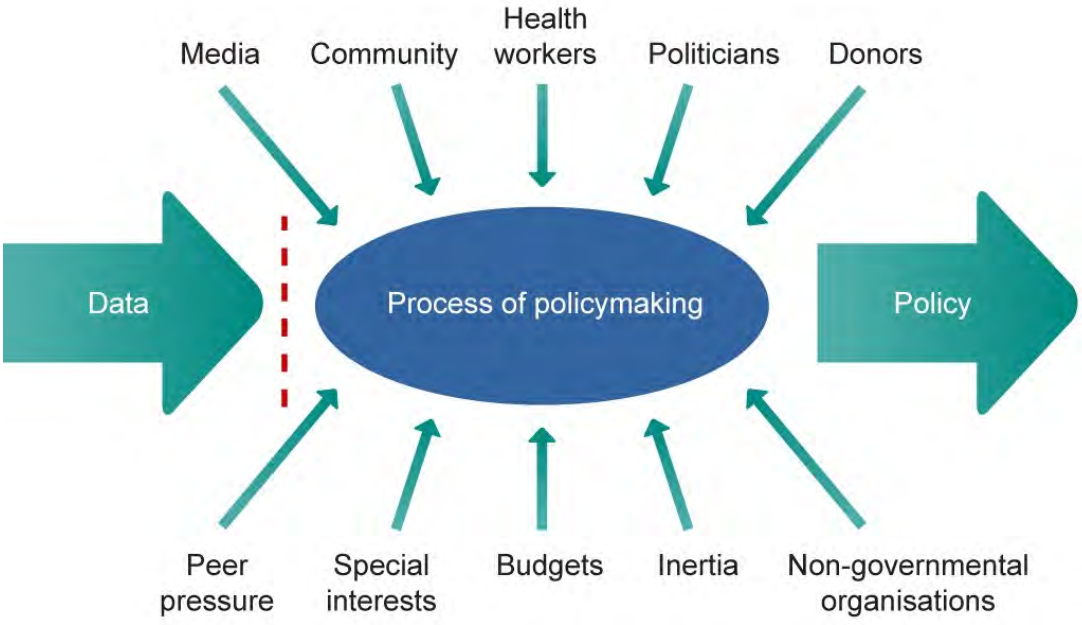
Whether or not the committee has been formalised, these groups have been identified as key stakeholders in these sectors and should be represented on the NSDS team, while avoiding setting up a process which duplicates their functions (of assessment, prioritisation, and planning for improvement, and then overseeing the implementation of that plan). In small countries, it may be appropriate that a joint national CRVS/HIS committee could also act as the planning committee for the social sector of the NSDS development (with the addition of education if not already represented). The role and function of the committee should be reflected in their terms of reference.

The NSDS process consists of an assessment of a broad range of sectors for which statistics are either collected or are required. For CRVS, this should be drawn from the results of the comprehensive CRVS assessment wherever this has been conducted. Further, the production and dissemination of vital statistics should be integrated as a priority into the plan.

21.2 WHY DATA IS NOT USED

Policymaking is often fragmented and decisions are sometimes difficult to make because of several players and competing interests. As a result, decision making on important issues can be delayed or compromised (see Figure below).

Figure 21.2 The policy making process (adapted from Lippeveld, Sauerborn and Bodart 2000)



Many decision makers often comment that data is not used for policy or planning because:

- The data is inaccessible, unclear and of poor quality
- Statistics are not relevant to their immediate needs
- Statistics are often out-of-date by the time they are published

- Information is 'lost in translation' because it is not directed in formats they can understand and use.

Additionally, people with the ability and authority to use information often:

- don't know about the information,
- don't understand the information,
- don't care about the information, or
- don't agree with the information.

Understanding which of these four reasons is preventing your decision makers from using your information will impact the strategies you need to take to change the situation.

21.3 Bridging the gap between data and decision making

In addition to whether specific indicators based on vital statistics are written into the plan itself, there are three key elements to consider in ensuring that data is useful and used in planning and policy decisions. These are:

- The format of the data,
- The perception about the reliability of the data source, and
- The timeliness of the data.

Bridging the gap between the availability of data and the use of information and evidence to guide decision making requires a number of key steps. As discussed previously, the analysis and interpretation of data is an important first step in understanding policy implications of the data and identifying required actions. One of the most important analytical tools you can use is to compare your data to other sources such as from the census or DHS. Assessing the overall performance of your CRVS system, or a specific indicator (such as child mortality) is another way of summarizing and interpreting your data. This can be done by:

- identifying contextual changes such as demographic, economic, social and political factors
- assessing progress compared with targets, or compared with countries in the region
- analysing the data in terms of trends in equity gaps by key stratifiers
- analysing the data in terms of efficiency (results by inputs).

The next important step is the communication and dissemination of findings to stakeholders with an interest in the information. When communicating with decision makers, it is important to get the right information to the right person, at the right time, and in the right way. Decision makers don't have much time, so a big report with lots of tables will not help them in making decisions. Rather, information should be communicated as short executive summaries, key messages and recommendations, actionable items, and visuals. Health summary bulletins are one example of such an information product. They usually contain information on key health indicators on a specific program area, and if targeted at decision makers, should also include policy recommendations.

Policy briefs highlight actionable recommendations for decision making in a short (two to six page) format. The typical format identifies a problem, proposes a solution and presents a compelling and feasible

recommendation. This format is ideal for conveying specific evidence-based policy recommendations. Chapter 21 provides additional information on how to develop a policy brief.

One of the most important aspects of effective communication is to develop a dissemination schedule and plan your communications around related events. Most, if not all, departments will have a set planning and review period each year linked to the timing of budget allocations and decisions. This process is also likely to be linked to a strategic plan or framework, either at the departmental or national level. These may include national health strategies, national strategies for development and national statistics strategies etc. Linking CRVS data to these planning processes and documents will help ensure that the data is used to inform decisions. The most direct way to ensure this linkage is to incorporate specific monitoring and evaluation indicators into the plan or its review.

Along with these three key steps, it is also important to build trust with decision makers, so they know that the data you provide is reliable. Interpretation of the data must clearly show the strengths and limitations of the evidence (which is why we started the course with data quality and understanding our systems). And finally, you need to make sure the quality of the analysis and quality of the writing has been checked for errors or inconsistencies.

22 MAKING DATA AVAILABLE

Once the report writing is complete, there should be a focused effort on making the data tables, the data trends, and the data commentaries available by every means necessary, and with advances in technology over the past few years there are numerous free tools which allow users to quickly and easily to just this.

In essence the 'report' will be read by groups such as thematic experts, government ministers, and students. But what about everyone else? The general public and the average internet browsers are less likely to read reports in great detail, so we must consider other methods of dissemination to ensure the correct messages are getting across.

Online reporting at SPC

In terms of web platforms, there are two main systems for making data available at SPC: The Pacific Regional Information System (PRISM), and the Joomla Content Management System. PRISM deals with official data from Pacific Island Country and Territory (PICT) National Statistics Offices (NSOs). This project has been in operation since 2001, and has provided PICTs with training and capacity building to ensure NSOs have the skills to release timely statistics via their own website. In 2010 the Joomla Content Management System was introduced, which is a free open-source platform that allows users to manage their content in a user-friendly way. SPC has facilitated sub-regional workshops as well as in-country training on a regular basis - and all websites can be viewed from the following link - www.spc.int/prism.

Document management for online sources

One of the first aspects of this system to understand is Document Management. All government departments have a certain level of public-facing documents that need to be available to everyone. These can include reports, newsletters, factsheets, excel spreadsheets, presentations, posters, strategic plans, and even job vacancies. This provides the base of a website which you can then build from. Ideally this is a direct reflection of a folder structure, as this makes the system very user- friendly to maintain. For example the Solomon Islands Stats Office disseminates their documents in this fashion: www.spc.int/prism/solomons/index.php/sinso-documents - where users can easily expand and collapse the folders and access the documents they need.

The next crucial aspect is to produce pages with online tables and charts, with 'easy-to-access' downloads. These pages allow you as administrators to publish key aspects of your reports as single web pages. Here is a good example from the Cook Islands Stats Office:

www.mfem.gov.ck/statistics/social-statistics/vital-stats-pop-est

Reporting indicators in summary tables

From here it is extremely important to report the main indicators from the report as a summary table, to allow regional agencies to easily complete and publish NMDI and SDG indicators for regional and global comparisons to be made. The NMDI website at www.spc.int/nmdi displays all the key indicators for the Pacific region. The themes of Population & Development and Public Health (including sub-themes such as Vital Statistics, Child Health and Maternal Health) make up the bulk of the data set, and will in turn feeds into the global SDG indicators. It is crucial that the correct indicators (with correct sources and caveats) are reported, or there is an inherent risk of unofficial (and un-sourced) estimates being used.

Other reporting platforms

Finally here are some other platforms to consider for making your data available especially in the Pacific region:

- Online mapping applications to show sub-regional trends. For example Solomon Islands: www.spc.int/prism/prism_mapping_solomons
- Social media - the use of Twitter to provide regular updates and bite-sized chunks of information. Please follow us at www.twitter.com/prismstats
- Infographics - an excellent medium for displaying and disseminating data: e.g. (discussed more in the next section)

23 INTERNATIONAL AND REGIONAL REPORTING REQUIREMENTS

Countries have agreed to provide data to satisfy a variety of international and regional reporting requirements. These requirements are usually reported as indicators which are useful for policy makers and development practitioners to monitor progress against health and development goals. Indicators are a quantitative way to simplify information so that policy makers can use data to prioritize resources. This allows them to determine which areas are improving, getting worse, or staying the same over time.

Standardized indicators allow regional comparisons to be made so that disparities across regions and even countries become apparent. Highlighting geographic areas and populations in need allows development partners to use their resources more effectively. Indicators can be used to showcase a country's commitment and efforts towards improving their citizens' health and well-being, demonstrating that that country has a reliable track record in using donor funds for development. Indicators can also be used to highlight the areas where targets are not being met and appeal to donors to invest in these areas.

23.1 SUSTAINABLE DEVELOPMENT GOALS (SDGS) REPORTING REQUIREMENTS

With the conclusion of the 2015 Millennium development goals[14], all United Nation member states including Pacific Island countries and territories have agreed to a more robust development framework dubbed the Sustainable development goals. The SDGs carry forward all of the health related MDGs, and has included additional targets to measure non-communicable diseases, the proportion of children who have their births legally registered, and specific targets on causes of death attributed to road traffic accidents. Table 23.1 provides a list of key health related targets and the specific CRVS data needed for their monitoring.

Table 23.1: UN Sustainable Development Goals and data requirements

SDG 3: GOOD HEALTH AND WELL-BEING	
Target	**Data needed**
3.1 By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births.	▪ Deaths due to maternal causes (deaths by cause – ICD summary tabulation list) ▪ Number of births
3.2 By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under 5 mortality to at least as low as 25 per 1,000 live births	▪ Deaths by age
3.3 By 2030, end the epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water-borne diseases and other communicable diseases.	▪ Deaths by cause (ICD summary tabulation list)
3.4 By 2030, reduce by one third premature mortality from non-communicable diseases through prevention and treatment and promote mental health and well-being.	▪ Deaths by cause (ICD summary tabulation list) – disaggregated by age group and sex ▪ Life expectancy (calculated by deaths by age group and population)
Target	Data needed

[14] UN Millennium Project. 2005. Investing in Development: A Practical Plan to Achieve the Millennium Development Goals. United Nations Development Programme, New York.

3.6 By 2020, halve the number of global deaths and injuries from road traffic accidents.	▪ Deaths by cause (ICD summary tabulation list) – by age group
3.d Strengthen the capacity of all countries, in particular developing countries, for early warning, risk reduction and management of national and global health risks.	▪ Deaths by cause (ICD summary tabulation list) – by age group, sex, and geographic sub-region

Source: SPC, 2016: Civil Registration and Vital Statistics (CRVS) and the Sustainable Development Goals (SDGs), http://www.pacific-crvs.org/docs.

23.2 WHO REPORTING REQUIREMENTS

A summary of WHO reporting requirements from the University of Queensland HIS Hub 2011 Health Information Systems Short Course:

The World Health Statistics series is WHO's annual compilation of health-related data for its 193 Member States, and includes a summary of the progress made towards achieving the health-related MDGs and associated targets. Taken together, these indicators provide a comprehensive summary of the current status of national health and health systems in the following nine areas:

- life expectancy and mortality;
- cause-specific mortality and morbidity;
- selected infectious diseases;
- health service coverage;
- risk factors;
- health workforce, infrastructure and essential medicines;
- health expenditure;
- health inequities; and
- demographic and socioeconomic statistics.

WHO presents World Health Statistics 2011 as an integral part of its ongoing efforts to provide enhanced access to high-quality data on core measures of population health and national health systems. Unless otherwise stated, all estimates have been cleared by WHO following consultation with Member States and are published here as official WHO figures. However, these best estimates have been derived using standard categories and methods to enhance their cross-national comparability. As a result they should not be regarded as the nationally endorsed statistics of Member States which may have been derived using alternative methodologies.

23.3 National Minimum Development Indicators (NMDIs)

The NMDIs are a powerful tool for public health improvement in the Pacific Region through improved accessibility of country-based data for decision making and priority setting. In all, there are 206 indicators across 6 broad themes (e.g. health, agriculture, fisheries, communication and infrastructure, population and development, human development), and 27 sub-themes (e.g. education, labour force, poverty, gender, youth) or thematic applications (e.g. child health, macro- economic aspects of fisheries).[15] The NMDI website has over 2,000 data records and allows user to export tables, charts, and maps. Below is a table of the NMDI indicators related to CRVS.

The NMDI data is used for a range of purposes including informing development partners and donor agencies how countries are faring in terms of development, and where key priorities for assistance may fall. They are also useful for advocating regional needs and priorities at a global level to ensure that Pacific concerns are reflected in international decisions and priority setting. To aid in this endeavour, the NMDI database has been linked to the UN Statistics reporting to ensure that as far as possible, empirical country data is available and used to inform these high level discussions.

Additionally, and perhaps most importantly, the NMDIs provide a platform that encourages countries to make sure that core data is made available for decision makers, and can be readily compared (or "benchmarked") against other countries in the region.

Exercise:

Create a summary table for the front of your report with a list of indicators that you are required to report and that are important for your country. You can use the table in the Niue Vital Statistics Report as an example. Once you have tabulated all of the indicators necessary for reporting, fill in your Table with the appropriate data. You can enter in the indicators you have already calculated such as the Under 5 mortality rate, the IMR, the adolescent birth rate etc.

---

[15] http://www.spc.int/nmdi/

NMDI

Indicators related to CRVS

Major Group	Sub Group	Indicator	Name
Population & Development	Population	PD-POP-1.1	Population size
Population & Development	Population	PD-POP-1.2	Population growth
Population & Development	Population	PD-POP-1.3	Rate of Natural increase
Population & Development	Population	PD-POP-1.4	Urbanization
Public Health	Vital Statistics	PH-VS-1.1	Under five mortality rate
Public Health	Vital Statistics	PH-VS-1.2	Infant mortality rate
Public Health	Vital Statistics	PH-VS-1.3	Neonatal mortality rate
Public Health	Vital Statistics	PH-VS-1.4	Life expectancy at birth
Public Health	Vital Statistics	PH-VS-1.5	Life expectancy at age 40 (e/40)
Public Health	Vital Statistics	PH-VS-1.6	Adult mortality rate, (45q15)
Public Health	Vector Borne Diseases	PH-VBD-1.3	Death rates associated with malaria per 100,000 population
Public Health	Communicable Diseases (TB)	PH-CD-1.3	TB Mortality rate (Deaths from TB given year per 100,000 population)
Public Health	Child Health	PH-CH-1.4	Low birth weight babies
Public Health	Maternal Health	PH-MH-1.1	Maternal Mortality Ratio per 100,000
Public Health	Maternal Health	PH-MH-1.2	Maternal Mortality (Number of Deaths)
Public Health	Maternal Health	PH-MH-1.3	Proportion of births attended by skilled health personnel
Public Health	Maternal Health	PH-MH-1.5	Adolescent birth rate (Teenage fertility rate)
Public Health	Maternal Health	PH-MH-1.6	Antenatal care coverage (% of women who had at least one antenatal session with a skilled provider)
Public Health	Maternal Health	PH-MH-1.7	Total Fertility Rate

PART 4: End material

24 CONCLUSIONS AND NEXT STEPS

Vital statistics are central to health and development monitoring and planning

As we have discussed throughout this workshop, data on births, deaths and causes of death are central to a vast range of information needed by governments and development partners to: identify health priorities; monitor and compare progress; plan for social services such as schools, clinics, infrastructure etc.; support a secure and safe society through legal protection; and to provide population denominator data for a vast range of development indicators. This data is therefore important to a broad range of stakeholders including politicians, government departments, communities, and development partners.

Accessible, timely and reliable data for decision making

Given the importance of the data, it is important that planners and decision makers can access this data when they need it, that it is meaningful, and that they have confidence in what the data is telling them. PICTs across the region have made significant progress over the last few years in improving their CRVS systems, but we need to make sure this then leads to an improvement in data accessibility and quality. During the course, we have reviewed the importance of understanding where our data comes from, and the impacts that the collection system may have on fertility and mortality measures. There is a need to develop a stronger practice both in analysing and reporting our data, but also in clearly outlining the strengths and limitations of what we are able to produce so that we can build confidence in the data available. Comparing and contrasting data to other sources is also essential in providing context for our figures.

Real data for real decisions

When CRVS data is not readily available, PICTs often rely heavily on census and survey estimates, estimates from international sources, or make decisions with very little evidence to inform these. While estimates of fertility and mortality serve an important function in highlighting the potential scale of a problem or calling attention to issues which would otherwise go unnoticed, they are not a substitute for real data. Estimated data cannot adequately monitor changes over time or account for differences in local conditions that may not be built into models. There is a real risk that policy decisions based on incorrect estimates could lead to poorer health and development outcomes, and that policy makers may not understand the uncertainty that is associated with estimated statistics. While censuses and surveys can produce some of the information required, they do not capture causes of death, and frequently rely on a range of models to impute fertility and mortality based on the age of the respondent. For adult mortality in particular, we know that many of these models under-estimate the real mortality impacts of NCDs. The only way to reduce the uncertainty of health and development statistics is to improve the collection, analysis, and dissemination of empirical data, sourced from a robust civil registration and vital statistics system (CRVS).

A need for meaningful, measurable goals and targets reviewed as trends rather than point estimates.

The small population sizes in PICTs mean that for health and mortality indicators in particular, small stochastic (random) variations in the number of births and deaths from year to year can have a significant impact on annual rates; and subsequently on the interpretation of how countries are progressing. Responding to these small insignificant fluctuations can detract attention from overall progress (or lack thereof) made over a sustained period. Similarly, the small numbers can result in targets which are implausible and are subsequently meaningless or which require a higher standard and outcome from small island states than is achievable in even the most developed countries.

In localising the post 2015 development agenda for PICTs, there must be a shift to ensuring targets are meaningful and theoretically achievable, along with a shift in emphasis from reporting annual figures which may be misleading, to looking at data aggregated over several years and trends over time in order to generate a realistic picture of country progress. Similarly, while reporting using internationally agreed age groups and indicators is essential for international comparison purposes,

Countries with complete or near-complete reporting of deaths including cause of death should also be able to establish a baseline for cause based indicators. Countries which do not have this information available endeavour to establish an empirical baseline for indicators based on cause of death through improved civil and vital registration as soon as possible. For countries that do not have complete reporting of deaths by cause, it may be possible to use proportional mortality by cause as derived from major population centres or hospital data as a substitute until systems improve sufficiently to collect this data, understanding of course that this data may not be representative at a national level.

The use of modelled estimates for a baseline as outlined in the working paper is not recommended for PICTs for several reasons. Firstly, the disease groupings used to generate these estimates appears to be from the Global Burden of Disease categories used by WHO, and are not directly comparable to the ICD (or the 103 cause group categories). As such countries would not be measuring exactly the same set of causes as indicated in the model baseline. Secondly, the intent of the data modelling was to establish the scale of the overall global burden from NCDs rather than to set baselines. There was very little pacific data used in the models (see attached extract from the WHO Report on NCD's) and while the models used to generate the estimates appear to hold up well at a global scale, they have not been validated at a country level for this region.

24.1 NEXT STEPS

Having finished working through the material in this workbook and the workshop, each country group should have at least a draft outline of a national vital statistics report based on their routinely collected civil registration and vital statistics data. The next step for each country group is to complete this draft report and to share it with their national CRVS committee or task force.

It is important that the work that you have started over these two weeks is completed, and made available for policy and planning use. We would like to see each country commit to publishing these reports within 2 months of the end of this workshop. As we have discussed in the workshop, this could be through either a web publication or printed version of the final report (or both), and does not have to be a major undertaking.

We would also encourage each country group to discuss what other approaches they may wish to consider in making the information synthesized from these reports available to key data users as discussed in 17

and 24. Committees may also wish to discuss how frequently this type of report may be useful in the local context, and ensure it is included in departmental publication schedules as appropriate.

Finally, we hope this work has highlighted both the importance of CRVS, and the possibilities of how this data could be used. It is also likely that in assessing and analysing the data from these systems, you may have identified a number of system issues which could be improved. We strongly encourage you to work with your national CRVS committees to complete (or update if necessary) a national assessment of your CRVS system and develop a national plan for improvement. This plan should be formalized and endorsed at an appropriate level of government as recommended by the Pacific Ministers of Health meeting and the Regional conference of Heads of Planning and Statistics in 2013. We would also strongly encourage you to publish and make these plans available to development partners in order to assist us to provide appropriate support.

REFERENCES

AbouZahr, C., Mikkelsen, L., Rampitage, R., Lopez, A. Mortality statistics: a tool to improve understanding and quality. Health Information Systems Knowledge Hub, School of Population Health, The University of Queensland, 2012. Available at: http://www.uq.edu.au/hishub/wp13

AbouZahr C, Cleland J, Coullare F, Macfarlane SB, Notzon FC, Setel P, et al. The way forward. Lancet. 2007;370(9601):1791-9. Epub 2007/11/22.

Adair, T. 2012. Manual for "Training in the Use of Existing Health Datasets". UQ HIS Hub, Brisbane

Ahmad O, Boschi-Pinto C, Lopez AD, Murray CJL, Lozano R, Inoue M. Age standardization of rates: a new WHO standard. Geneva, World Health Organization, 2001. Available at: http://www.who.int/healthinfo/paper31.pdf

Arriaga, E.E., Johnson, P.D., Jamison, E. Population Analysis with Microcomputers: Presentation of Techniques. Vol. 1. Washington, D.C.: Bureau of the Census, 1994. Available at: https://www.census.gov/population/international/files/pas/PAMvI_Archive.pdf

Australian Bureau of Statistics. Data Quality Framework. https://www.nss.gov.au/dataquality/aboutqualityframework.jsp

Baiden F, Bawah A, Biai S, Binka F, Boerma T, Byass P, et al. Setting international standards for verbal autopsy. Bulletin of the World Health Organization. 2007;85(8):570-1. Epub 2007/09/05.

Bairagi R, Becker S, Kantner A, Allen KB, Datta A, Purvis K. An evaluation of the 1993-94 Bangladesh Demographic and Health Survey within the Matlab area. Asia Pac Popul Res Abstr. 1997(11):1-2.

Biotext 2011. Successful science writing and editing

Byass P. Who needs cause of death data? PLoS medicine. 2007;4(11):1715.

Carter K. 2013. Mortality and Causes of Death in the Pacific. PhD Thesis. UQ, Brisbane.

Carter KL, Rao C, Lopez AD, Taylor R. Mortality and cause-of-death reporting and analysis systems in seven pacific island countries. BMC public health. 2012;12(1):436

Carter K, Cornelius M, Taylor R, Ali SS, Rao C, Lopez AD, et al. Mortality trends in Fiji. Aust N Z J Public Health. 2011; 35(5):412-20. Epub 2011/10/07.

Carter, K., Soakai, T. S., Taylor, R., Gadabu, I., Rao, C., Thoma, K., & Lopez, A. D. (2011). Mortality trends and the epidemiological transition in Nauru. Asia-Pacific Journal of Public Health, 23(1), 10-23.

Haberkorn G. Vital registration and health programme monitoring in Pacific Island countries–some myths and realities. Public Health Surveillance in the Pacific Noumea: Secretariat of the Pacific Community. 1998:91-104.

Hancock A, 2013, The Role of International Standards for National Statistical Offices. Classifications and Standards, Statistics New Zealand

Lilienfeld AM. Practical limitations of epidemiologic methods. Environmental health perspectives. 1983;52:3-8. Epub 1983/10/01.

Lopez AD, AbouZahr C, Shibuya K, Gollogly L. Keeping count: births, deaths, and causes of death. Lancet. 2007;370(9601):1744-6. Epub 2007/11/22.

New Zealand Statistics – Data Quality Framework

Niue Statistics Unit, Statistics for Development Programme, Secretariat of the Pacific Community (SPC). Niue Vital Statistics Report: 1987 – 2011. Niue Statistics Unit, Department of Treasury, Niue, 2012. Available at: http://www.spc.int/prism/niue/index.php/niue-documents/doc_download/120-vital-a-health-statistics

Mahapatra P, Shibuya K, Lopez AD, Coullare F, Notzon FC, Rao C, et al. Civil registration systems and vital statistics: successes and missed opportunities. Lancet. 2007.

Mathers CD, Fat DM, Inoue M, Rao C, Lopez AD. Counting the dead and what they died from: an assessment of the global status of cause of death data. Bulletin of the World Health Organization. 2005;83(3):171-7.

Obermeyer Z, Rajaratnam JK, Park CH, Gakidou E, Hogan MC, Lopez AD, et al. Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974-2006. PLoS medicine.7(4):e1000260.

Pacific Community (SPC), 2016. Civil Registration and Vital Statistics (CRVS) and the Sustainable Development Goals (SDGs) Available at: http://www.pacific-crvs.org/docs

Pudue University. 2014. Open Writing Lab (https://owl.english.purdue.edu/owl/resource/606/01/)

Riley I. Demography and the epidemiology of disease in Papua New Guinea. Papua and New Guinea medical journal. 2009;52(3-4):83-95. Epub 2009/09/01.

Rowland DT. Demographic Methods and Concepts. New York: Oxford University Press; 2003.

Ryan, C. et. al. 2011. Workbook: Data analysis and report writing course. Secretariat of the Pacific Community.

Sauerborn, R. 2000. 'Chapter 3: Using information to make decisions'. In, T. Lippeveld, R. Sauerborn and C. Bodart (eds.), Design and implementation of health information systems. World Health Organization: Geneva.

Setel PW, Macfarlane SB, Szreter S, Mikkelsen L, Jha P, Stout S, et al. A scandal of invisibility: making everyone count by counting everyone. Lancet. 2007.

Stanton C, Abderrahim N, Hill K. An assessment of DHS maternal mortality indicators. Studies in family planning. 2000;31(2):111-23. Epub 2000/07/25.

Taylor R, Bampton D, Lopez AD. Contemporary patterns of Pacific Island mortality. International journal of epidemiology. 2005;34(1):207-14.

Taylor R. Small area population disease burden. Aust N Z J Public Health. 2001;25(4):289-93. Epub 2001/09/01.

Tomas P, Summers L, Clemens M. Migrants Count. Five steps toward better migration data. Washington: Centre for Global Development; 2009.

UN Statistics Division. Principles and recommendations for a vital statistics system: United Nations, Department of Economic and Social Affairs, Statistics Division; 2001

United Nations. Manual X: Indirect Techniques for Demographic Estimation New York: United Nations; 1983.

United Nations. Handbook of vital statistics systems and methods. Volume II: Review of national practices. New York: 1985.

United Nations Economic Commission for Europe, 2009. Making Data Meaningful. New York

United Nations Population Fund (UNFPA). How universal is access to reproductive health? A review of the evidence. New York: UNFPA, 2010. Available at: http://www.unfpa.org/public/home/publications/pid/6526

Vanuatu Ministry of Health, UNICEF. Vanuatu Multiple Indicator Cluster Survey 2007. Port Vila: Government of Vanuatu, 2008.

Waddell N. 2010. Writing Guide - UQ Business School

WHO. 2011. International Classification of Diseases Volume 10 (version 2). World Health Organization. Geneva

PART 5: Appendices

ICD General Mortality List 1

List code	Disease	ICD Codes
1-001	Certain infectious and parasitic diseases	A00–B99
1-002	Cholera	A00
1-003	Diarrhoea and gastroenteritis of presumed infectious origin	A09
1-004	Other intestinal infectious diseases	A01–A08
1-005	Respiratory tuberculosis	A15–A16
1-006	Other tuberculosis	A17–A19
1-007	Plague	A20
1-008	Tetanus	A33–A35
1-009	Diphtheria	A36
1-010	Whooping cough	A37
1-011	Meningococcal infection	A39
1-012	Septicaemia	A40–A41
1-013	Infections with a predominantly sexual mode of transmission	A50–A64
1-014	Acute poliomyelitis	A80
1-015	Rabies	A82
1-016	Yellow fever	A95
1-017	Other arthropod-borne viral fevers and viral haemorrhagic fevers	A90–A94, A96–A99
1-018	Measles	B05
1-019	Viral hepatitis	B15–B19
1-020	Human immunodeficiency virus [HIV] disease	B20–B24
1-021	Malaria	B50–B54
1-022	Leishmaniasis	B55
1-023	Trypanosomiasis	B56–B57
1-024	Schistosomiasis	B65
1-025	Remainder of certain infectious and parasitic diseases	A21–A32, A38, A42–A49, A65–A79, A81, A83–A89, B00–B04, B06–B09, B25–B49, B58–B64, B66–B94, B99
1-026	Neoplasms	C00–D48
1-027	Malignant neoplasm of lip, oral cavity and pharynx	C00–C14
1-028	Malignant neoplasm of oesophagus	C15
1-029	Malignant neoplasm of stomach	C16
1-030	Malignant neoplasm of colon, rectum and anus	C18–C21
1-031	Malignant neoplasm of liver and intrahepatic bile ducts	C22
1-032	Malignant neoplasm of pancreas	C25

1-033	Malignant neoplasm of larynx	C32
1-034	Malignant neoplasm of trachea, bronchus and lung	C33–C34
1-035	Malignant melanoma of skin	C43
1-036	Malignant neoplasm of breast	C50
1-037	Malignant neoplasm of cervix uteri	C53
1-038	Malignant neoplasm of other and unspecified parts of uterus	C54–C55
1-039	Malignant neoplasm of ovary	C56
1-040	Malignant neoplasm of prostate	C61
1-041	Malignant neoplasm of bladder	C67
1-042	Malignant neoplasm of meninges, brain and other parts of central nervous system	C70–C72
1-043	Non-Hodgkin's lymphoma	C82–C85
1-044	Multiple myeloma and malignant plasma cell neoplasms	C90
1-045	Leukaemia	C91–C95
1-046	Remainder of malignant neoplasms	C17, C23–C24, C26–C31, C37–C41, C44–C49, C51–C52, C57–C60, C62–C66, C68–C69, C73–C81, C88, C96–C97
1-047	Remainder of neoplasms	D00–D48
1-048	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	D50–D89
1-049	Anaemia	D50–D64
1-050	Remainder of diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	D65–D89
1-051	Endocrine, nutritional and metabolic diseases	E00–E88
1-052	Diabetes mellitus	E10–E14
1-053	Malnutrition	E40–E46
1-054	Remainder of endocrine, nutritional and metabolic diseases	E00–E07, E15–E34, E50–E88
1-055	Mental and behavioural disorders	F01–F99
1-056	Mental & behavioural disorders due to psychoactive substance use	F10–F19
1-057	Remainder of mental and behavioural disorders	F01–F09, F20–F99
1-058	Diseases of the nervous system	G00–G98
1-059	Meningitis	G00, G03
1-060	Alzheimer's disease	G30
1-061	Remainder of diseases of the nervous system	G04–G25, G31–G98
1-062	Diseases of the eye and adnexa	H00–H59
1-063	Diseases of the ear and mastoid process	H60–H93
1-064	Diseases of the circulatory system	I00–I99
1-065	Acute rheumatic fever and chronic rheumatic heart diseases	I00–I09
1-066	Hypertensive diseases	I10–I13
1-067	Ischaemic heart diseases	I20–I25
1-068	Other heart diseases	I26–I51
1-069	Cerebrovascular diseases	I60–I69

1-070	Atherosclerosis	I70
1-071	Remainder of diseases of the circulatory system	I71–I99
1-072	Diseases of the respiratory system	J00–J98
1-073	Influenza	J10–J11
1-074	Pneumonia	J12–J18
1-075	Other acute lower respiratory infections	J20–J22
1-076	Chronic lower respiratory diseases	J40–J47
1-077	Remainder of diseases of the respiratory system	J00–J06, J30–J39, J60–J98
1-078	Diseases of the digestive system	K00–K92
1-079	Gastric and duodenal ulcer	K25–K27
1-080	Diseases of the liver	K70–K76
1-081	Remainder of diseases of the digestive system	K00–K22, K28–K66, K80–K92
1-082	Diseases of the skin and subcutaneous tissue	L00–L98
1-083	Diseases of the musculoskeletal system and connective tissue	M00–M99
1-084	Diseases of the genitourinary system	N00–N99
1-085	Glomerular and renal tubulointerstitial diseases	N00–N15
1-086	Remainder of diseases of the genitourinary system	N17–N98
1-087	Pregnancy, childbirth and the puerperium	O00–O99
1-088	Pregnancy with abortive outcome	O00–O07
1-089	Other direct obstetric deaths	O10–O92
1-090	Indirect obstetric deaths	O98–O99
1-091	Remainder of pregnancy, childbirth and the puerperium	O95–O97
1-092	Certain conditions originating in the perinatal period	P00–P96
1-093	Congenital malformations, deformations and chromosomal abnormalities	Q00–Q99
1-094	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	R00–R99
1-095	External causes of morbidity and mortality	V01–Y89
1-096	Transport accidents	V01–V99
1-097	Falls	W00–W19
1-098	Accidental drowning and submersion	W65–W74
1-099	Exposure to smoke, fire and flames	X00–X09
1-100	Accidental poisoning by and exposure to noxious substances	X40–X49
1-101	Intentional self-harm	X60–X84
1-102	Assault	X85–Y09
1-103	All other external causes	W20–W64, W75–W99, X10–X39, X50–X59, Y10–Y89
1-901	SARS	U04

WHO WORLD STANDARD POPULATION DISTRIBUTION

From: AGE STANDARDIZATION OF RATES: A NEW WHO STANDARD, GPE Discussion Paper Series: No.31, EIP/GPE/EBD, World Health Organization 2001

Table 4. WHO World Standard Population Distribution (%), based on world average population between 2000-2025

Age group	World Average 2000-2025
0-4	8.86
5-9	8.69
10-14	8.60
15-19	8.47
20-24	8.22
25-29	7.93
30-34	7.61
35-39	7.15
40-44	6.59
45-49	6.04
50-54	5.37
55-59	4.55
60-64	3.72
65-69	2.96
70-74	2.21
75-79	1.52
80-84	0.91
85-89	0.44
90-94	0.15
95-99	0.04
100+	0.005
Total	100

25 MAKING DATA MEANINGFUL

United Nations Economic Commission for Europe. 2009. “Making Data Meaningful”. United Nations. New York and Geneva

Part 1: A guide to writing stories about numbers

http://ec.europa.eu/eurostat/documents/64157/4374310/32-UNECE-making-data-meaningful-Part1-EN.pdf/82cf87d2-4bc9-431c-be96-2257cf77c810

Part 2: A guide to presenting statistics

http://www.unece.org/fileadmin/DAM/stats/documents/writing/MDM_Part2_English.pdf

26 EXAMPLES OF WRITING ON CRVS DATA FOR JOURNALS

Carter, Karen, Margaret Cornelius, Richard Taylor, Shareen S. Ali, Chalapati Rao, Alan D. Lopez, Vasemaca Lewai, Ramneek Goundar, and Claire Mowry. "Mortality trends in Fiji." *Australian and New Zealand journal of public health* 35, no. 5 (2011): 412-420.

http://onlinelibrary.wiley.com/doi/10.1111/j.1753-6405.2011.00740.x/abstract;jsessionid=43153A37F0BB54348E4F903D2B36E0C8.f03t02?deniedAccessCustomisedMessage=&userIsAuthenticated=false

Carter, Karen, Taniela Sunia Soakai, Richard Taylor, Ipia Gadabu, Chalapati Rao, Kiki Thoma, and Alan D. Lopez. "Mortality trends and the epidemiological transition in Nauru." *Asia-Pacific Journal of Public Health* 23, no. 1 (2011): 10-23.

http://aph.sagepub.com/content/23/1/10.short