

---

# Statistical concepts

## General Government Final Consumption Expenditure

The GGFCE deflator is used to convert raw financial data into constant (real) dollars (box 1).

### Box 1 GGFCE deflator formulas

#### GGFCE deflator re-base

The general formula used to re-base GGFCE deflators is:

$$N_t = 100 \times \frac{O_t}{B}$$

Where:

$N_t$  is the new index based in year  $t$ ;  $O_t$  is the current index for year  $t$

$B$  is the current index for the year that will be the new base.

#### GGFCE deflator application

The general formula for applying the deflator to convert nominal dollars to real dollars is:

$$R_t = \frac{D_t}{N_t} \times 100$$

Where, for year  $t$ :

$R_t$  is real dollars;  $D_t$  is nominal dollars;  $N_t$  is the new index

Raw or ‘nominal’ financial data are converted to ‘real’ dollars so that comparisons over time are not affected by inflation. (Not all financial data in the Report are deflated using the GGFCE deflator. The exceptions include some health chapters and the Vocational education and training chapter, which use service-specific deflators to calculate real dollars.)

The calculations to achieve constant (real) dollars are in two steps:

Step 1. Re-referencing of the GGFCE deflator.

The Report re-references the period where the GGFCE (published by the ABS) is at 100, as this Report requires a current year deflator (2015-16 = 100). The ABS publishes the GGFCE to the third most current year only (for example, if the current year is 2015-16, the available deflator is 2013-14 = 100). Table 1 shows how the GGFCE deflator is re-based.

**Table 1 Re-basing the GGFCE deflator<sup>a</sup>**

<i>Financial year</i>	<i>ABS index value (2013-14 = 100)</i>	<i>Calculation</i>	<i>Re-based GGFCE deflator (2015-16 = 100)</i>
2011-12	97.2	$97.2/103.5*100.0$	93.9
2012-13	98.7	$98.7/103.5*100.0$	95.4
2013-14	100.0	$100.0/103.5*100.0$	96.6
2014-15	101.0	$101.0/103.5*100.0$	97.6
2015-16	103.5	$103.5/103.5*100.0$	100.0

<sup>a</sup> Index values from ABS (2016) *Australian National Accounts: National Income, Expenditure and Product, June 2016*, Cat. no. 5206.0, table 36, Expenditure on Gross Domestic Product (GDP), Chain volume measures and current prices, Annual (Series ID. A2304687R).

Source: ABS (2016) *Australian National Accounts: National Income, Expenditure and Product, June 2016*, Cat. no. 5206.0, Canberra; table 2A.48.

Table 2A.48 contains GGFCE deflators for 2002-03 to 2015-16. Five GGFCE deflator series are published, from 2011-12 = 100 through to the latest year, where 2015-16 = 100.

Step 2. Transforming nominal dollars into constant dollars.

Nominal dollars are transformed into real dollars by dividing the nominal dollars with the GGFCE deflator for the applicable financial year and multiplying by 100. The deflator used may vary according to the most current year for which the particular financial data are available. For example, if the most current year for the data is 2014-15 then the data are deflated using the deflator series for 2014-15 = 100. If the most current year is 2015-16 then the data are deflated using the deflator series for 2015-16 = 100. Table 2 shows how the GGFCE deflator for 2015-16 = 100 is applied.

**Table 2 Applying the GGFCE deflator to derive constant (real) dollars<sup>a</sup>**

<i>Financial year</i>	<i>Nominal data</i>	<i>GGFCE deflator (2015-16= 100)</i>	<i>Calculation</i>	<i>Real data</i>
2011-12	6 200	93.9	(6 200/93.9)*100	6 603
2012-13	6 300	95.4	(6 300/95.4)*100	6 604
2013-14	6 350	96.6	(6 350/96.6)*100	6 573
2014-15	6 485	97.6	(6 485/97.6)*100	6 644
2015-16	7 020	100.0	(7 020/100.0)*100	7 020

<sup>a</sup> Index values from ABS (2016) *Australian National Accounts: National Income, Expenditure and Product, June 2016*, Cat. no. 5206.0, table 36, Expenditure on Gross Domestic Product (GDP), Chain volume measures and current prices, Annual (Series ID. A2304687R).

Source: ABS (2016) *Australian National Accounts: National Income, Expenditure and Product, June 2016*, Cat. no. 5206.0, Canberra; table 2A.48.

## Reliability of estimates

Data for some indicators in this Report are based on samples, either from surveys or from a selection of observations from, for example, administrative data sets. The potential for sampling error — that is, the error that occurs by chance because the data are obtained from a sample and not the entire population — means that the reported estimates might not accurately reflect the true value.

This Report indicates the reliability of estimates based on samples, generally by reporting either relative standard errors (RSEs) or confidence intervals (CIs). RSEs and CIs are calculated based on the standard error (SE). The larger the SE, RSE or CI, the less reliable is the estimate as an indicator for the whole population (ABS 2013).

## Standard error

The SE measures the sampling error of an estimate (box 2). (There can also be non-sampling error, or systematic biases, in data.) There are several types of SE. A commonly used type of SE in this Report is the SE of the mean (average), which measures how much the estimated mean value might differ from the true population mean value.

---

## Box 2      **Standard error**

The SE of a method of measurement or estimation is the estimated standard deviation of the error in that method. Specifically, it estimates the standard deviation of the difference between the measured or estimated values and the true values. Standard deviation is a measure of how spread out the data are, that is, a measure of variability.

The SE of the mean, an unbiased estimate of expected error in the sample estimate of a population mean, is the sample estimate of the population standard deviation (sample standard deviation) divided by the square root of the sample size (assuming statistical independence of the values in the sample):

$$SE_x = \frac{s}{\sqrt{n}}$$

Where:

$SE_x$  is the SE of the sample estimate of a population mean,  $s$  is the sample's standard deviation (the sample based estimate of the standard deviation of the population), and  $n$  is the size (number of items) of the sample.

Decreasing the uncertainty of a mean value estimate by a factor of two requires the sample size to increase fourfold. Decreasing SE by a factor of ten requires the sample size to increase hundredfold.

## **Relative standard error**

The RSE is used to indicate the reliability of an estimate (box 3). The RSE shows the size of the error relative to the estimate, and is derived by dividing the SE of the estimate by the estimate. As with the SE, the higher the RSE, the less confidence there is that the sample estimate is close to the true value of the population mean. A rule of thumb adopted in this Report is that estimates with an RSE between 25 and 50 per cent are to be used with caution and estimates with an RSE greater than 50 per cent are considered too unreliable for general use.

---

### Box 3      **Relative standard error**

The SE can be expressed as a proportion of the estimate — known as the RSE. The formula for the RSE of an estimate is:

$$RSE(x) = \frac{SE(x)}{x}$$

Where:

$x$  is the estimate and  $SE(x)$  is the SE of the estimate.

The resultant RSEs are generally multiplied by 100 and expressed as a percentage.

Proportions and percentages formed from the ratio of two estimates are also subject to sampling error. The size of the error depends on the accuracy of both the numerator and the denominator. One method for calculating the RSE of a proportion is expressed through the following formula:

$$RSE\left(\frac{x}{y}\right) = \sqrt{[RSE(x)]^2 + [RSE(y)]^2}$$

Where:

$x$  is the numerator, and  $y$  is the denominator, of the estimated proportion.

## Confidence intervals

Confidence intervals are used to indicate the reliability of an estimate. A CI is a specified interval, with the sample statistic at the centre, within which the corresponding population value can be said to lie with a given level of confidence (ABS 2013). Increasing the desired confidence level will widen the CIs (figure 1). CIs are useful because a range, rather than a single estimate, is more likely to encompass the real figure for the population value being estimated.

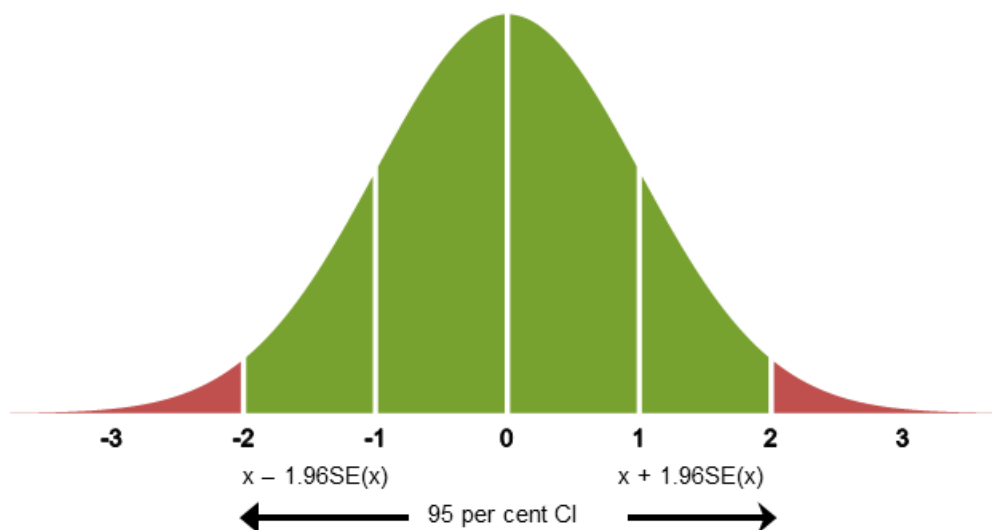
Confidence intervals are calculated from the population estimate and its associated SE. The most commonly used CI is calculated for 95 per cent levels of probability. For example, if the estimate from a survey was that 628 300 people report having their needs fully met by a government service, and the associated SE of the estimate was 10 600 people, then the 95 per cent CI would be calculated by:

- lower confidence limit =  $628\,300 - (2 \times 10\,600) = 628\,300 - 21\,200 = 607\,100$
- upper confidence limit =  $628\,300 + (2 \times 10\,600) = 628\,300 + 21\,200 = 649\,500$ .

This indicates that we can be 90 per cent sure the true number of people who perceive that their needs are met by a government service is between 607 100 and 649 500.

The smaller the SE of the estimate, the narrower the CIs and the closer the estimate can be expected to be to the true value.

Figure 1 Normal distribution with 95 per cent confidence intervals



■ It can be said with 95 per cent confidence that the population lies within the green area

Confidence intervals also test for statistical differences between sample results (box 4).

**Box 4 Using confidence intervals to test for statistical significance**

The CIs — the value ranges within which estimates are likely to fall — can be used to test whether the results reported for two estimated proportions are statistically different. If the CIs for the results do not overlap, then there can be confidence that the estimated proportions differ from each other. To test whether the 95 per cent CIs of two estimates overlap, a range is derived using the following formulas.

$$R_1 = \left( \frac{x_2}{y_2} - \frac{x_1}{y_1} \right) - 1.96 \sqrt{ \left( RSE \left( \frac{x_2}{y_2} \right) \times \left( \frac{x_2}{y_2} \right) \right)^2 + \left( RSE \left( \frac{x_1}{y_1} \right) \times \left( \frac{x_1}{y_1} \right) \right)^2 }$$

and

$$R_2 = \left( \frac{x_2}{y_2} - \frac{x_1}{y_1} \right) + 1.96 \sqrt{ \left( RSE \left( \frac{x_2}{y_2} \right) \times \left( \frac{x_2}{y_2} \right) \right)^2 + \left( RSE \left( \frac{x_1}{y_1} \right) \times \left( \frac{x_1}{y_1} \right) \right)^2 }$$

If none of the values in this range is zero, then the difference between the two estimated proportions is statistically significant.

For example, assume survey data estimated that 50 per cent of people for jurisdiction A perceived that their needs were met by government services, with a 95 per cent CI of  $\pm 5$  per cent, and 25 per cent of people for jurisdiction B, with a 95 per cent CI of  $\pm 10$  per

cent. These results imply that we can be 95 per cent sure the true result for jurisdiction A lies between 55 and 45 per cent, and the true result for jurisdiction B lies between 15 and 35 per cent. As these two ranges do not overlap, it can be said that the results for jurisdiction A and jurisdiction B are statistically significantly different.

## Variability bands

Rates derived from administrative data counts are not subject to sampling error but might be subject to natural random variation, especially for small counts. For mortality data, variability bands are used to account for this variation (box 5).

### Box 5 Variability bands

The variability bands to be calculated using the standard method for estimating 95 per cent confidence intervals are:

*Crude rate (CR)*

$$CI(CR)_{95\%} = CR \pm 1.96 \frac{CR}{\sqrt{\sum_{i=1}^I d}}$$

Where:

$d$  is the numerator of the estimated proportion

*Age-standardised rate (ASR)*

$$CI(ASR)_{95\%} = ASR \pm 1.96 \sqrt{\sum_{i=1}^I \frac{w_i^2 d_i}{n_i^2}}$$

Where:

$w_i$  is the proportion of the standard population in age group  $i$

$d_i$  is the number of deaths in age group  $i$

$n_i$  is the number of people in the population in age group  $i$ .

*Infant mortality rate (IMR)*

$$CI(IMR)_{95\%} = IMR \pm 1.96 \frac{IMR}{\sqrt{d_0}}$$

Where:

$d_0$  is the number of deaths in infants aged less than 1 year.

Variability bands accompanying mortality data should be used for the purpose of within jurisdiction analysis at a point in time and over time. They should not be used for

---

comparing mortality rates at a single point in time or over time between jurisdictions as they do not take into account differences in under-identification of Aboriginal and Torres Strait Islander people's deaths between jurisdictions.

Typically in this standard method, the observed rate is assumed to have natural variability in the numerator count (for example, deaths) but not in the population denominator count. Variations in Aboriginal and Torres Strait Islander people's death rates may arise from uncertainty in the recording of Indigenous status on the death registration forms (in particular, under-identification of Aboriginal and Torres Strait Islander people's deaths) and in the ABS population census, from which population estimates are derived. These variations are not considered in this method. Also, the rate is assumed to have been generated from a normal distribution (figure 1). Random variation in the numerator count is assumed to be centred around the true value — that is, there is no systematic bias.

## Population measures

Data are frequently expressed relative to population in this Report. For example, expenditure per person, or proportion of people who utilise a service or who benefit from a service. This enables comparison of data across populations of different sizes using relative numbers — standardised by population size — as distinct from absolute numbers.

ERP data are available quarterly — that is, at end March, June, September and December of each year. The mid-point ERP is typically used for the calculation of population rates in this Report — for example, the 30 June ERP for calendar year data (table 2A.1) and the 31 December ERP for financial year data (table 2A.2). As this Report presents annual data where available and appropriate, the mid-point ERP was adopted from a number of options predominately due to availability of ERP data.

This Report uses first preliminary ERP data wherever possible and replaces these with final rebased data when available.

## Estimated resident population rebasing and recasting

Where ERP data are reported they are based on the 2011 Census, backcast over the reported time series. Details of changes to ERP data are explained in the 2014 Report (SCRGSP 2014, pp. 2.26–27).

Some tables contain population data estimates and projections (tables 2A.13–14). Aboriginal and Torres Strait Islander population data up to 2011 are estimates and from 2012 are projections. Population data for all Australians for all years are estimates.



---

## Growth rates

This Report presents growth rates to facilitate meaningful comparisons of data movements over time (box 6). Two methods are generally used: *Average annual growth rate* (AAGR) is the uniform growth rate that would need to have applied each year for the value in the first year to grow to the value in the final year of the period of analysis; *Total growth rate* (TGR) is the growth rate between two periods/years, most commonly calculated by subtracting the value in the first period from the value in the last period, dividing the result by the value in the first period and multiplying by 100.

### Box 6 Growth rates

#### **Average annual growth rate**

The formula for calculating a compound AAGR is:

$$AAGR_{(t_0, t_n)} = \left[ \left( \frac{P_{(t_n)}}{P_{(t_0)}} \right)^{\left( \frac{1}{t_n - t_0} \right)} - 1 \right] \times 100$$

Where:

$P_{(t_0)}$  is the value in the initial period,  $P_{(t_n)}$  is the value in the last period and

$t_n - t_0$  is the number of periods (which will be one less than the total number of years).

#### **Total growth rate**

The formula for calculating the TGR is:

$$TGR = \frac{P_{(t_n)} - P_{(t_0)}}{P_{(t_0)}} \times 100$$

Where:  $P_{(t_0)}$  is the value in the initial period and  $P_{(t_n)}$  is the value in the last period.

The formula for calculating the TGR using a composite of growth rates between sub-periods within the overall period of analysis is:

$$TGR = \left( \prod_t (1 + r_t) - 1 \right) \times 100$$

That is, the TGR over the period is found by taking the product ( $\prod_t$ ) of each  $(1 + r_t)$  and deducting 1. This is multiplied by 100 so the growth rate is expressed as a percentage. If, for example, the sample ranges of growth rates are: 6 per cent in 2012-13 to 2013-14; 6 per cent in 2013-14 to 2014-15; 8 per cent in 2014-15 to 2015-16; then the total growth over the period 2012-13 to 2015-16 can be calculated as:

$$\begin{aligned} TGR &= [(1.06) \times (1.06) \times (1.08) - 1] \times 100 \\ &= (1.213488 - 1) \times 100 \\ &= 21.3 \text{ per cent.} \end{aligned}$$

---

## Age standardisation of data

### Rationale for age standardisation of data

The age profile of Australians varies across jurisdictions, periods of time, geographic areas and/or population sub-groups (for example, between Aboriginal and Torres Strait Islander and non-Indigenous populations). Variations in age profiles are important because they can affect the likelihood of using a particular service (such as a public hospital) or particular ‘events’ occurring (such as death, incidence of disease or incarceration). Age standardisation adjusts for the effect of variations in age profiles when comparing service usage, or rates, of particular events across different populations.

### Calculating age standardised rates

Age standardisation adjusts each of the comparison/study populations (for example, Aboriginal and Torres Strait Islander and non-Indigenous populations) against a standard population (box 7). The latest standard population used is the final 30 June ERP for the 2001 (AIHW 2015).<sup>1</sup> The result is a standardised estimate for each of the comparison/study populations.

The Review generally reports age standardised rates that have been calculated using either one of two methods, as appropriate. The direct method is generally used for comparisons between study groups, and is recommended by the AIHW (2011) for the purposes of comparing health and welfare outcome measures (for example, mortality rates, life expectancy, hospital separation rates and disease incidence rates) of the Aboriginal and Torres Strait Islander population and non-Indigenous population. The indirect method is recommended when the age-specific rates for the population being studied are not known (or are unreliable), but the total number of events is known (AIHW 2015).

The *direct method* has three steps:

- Step 1: Calculate the age-specific rate for each age group for the study/comparison group.
- Step 2: Calculate the expected number of ‘events’ in each age group by multiplying the age-specific rates by the corresponding standard population.
- Step 3: Sum the expected number of cases in each age group and divide by the total of the standard population.

The *indirect method* has four steps:

- Step 1: Calculate the age-specific rates for each age group in the standard population.

---

<sup>1</sup> See page 2.27 in the 2015 Report for the background on choice of year for the standard population and timeline for revision.

- Step 2: Apply the age-specific rates resulting from step 1 to the number in each age group of the study population and sum to derive the total ‘expected’ number of cases for the study population.
- Step 3: Divide the observed number of events in the study population by the ‘expected’ number of cases for the study population derived in step 2.
- Step 4: Multiply the result of step 3 by the crude rate in the standard population.

### Box 7 Direct and indirect age standardisation

The formula for deriving the age standardised rate using the *direct method* is:

$$SR = \frac{\sum(r_i P_i)}{\sum P_i}$$

The formula for deriving the age standardised rate using the *indirect method* is:

$$SR = \frac{C}{\sum(R_i p_i)} \times R$$

The formula for deriving the age standardised ratio using the *indirect method* is:

$$SR_a = \frac{C}{\sum(R_i p_i)}$$

Where:

$SR$  is the age-standardised rate for the population being studied

$SR_a$  is the standardised ratio for the population being studied

$r_i$  is the age-group specific rate for age group  $i$  in the population being studied

$P_i$  is the population of age group  $i$  in the standard population

$C$  is the observed number of events in the population being studied

$\sum(R_i p_i)$  is the expected number of events in the population being studied

$R_i$  is the age-group specific rate for age group  $i$  in the standard population

$p_i$  is the population for age group  $i$  in the population being studied

$R$  is the crude rate in the standard population.

Source: AIHW (2015).

Tables 2A.49–50 in the attachment contain examples of the application of direct and indirect age standardisation, respectively. Standardised rates are generally multiplied by 1000 or 100 000 to avoid small decimal fractions. They are then reported as age standardised rates per 1000 or 100 000 population (AIHW 2015).

---

## Calculating age standardised ratios

A variation of the *indirect method* is used to calculate age standardised ratios (box 7). These ratios express the overall experience of a study population in terms of a standard population, where the standard population is the population to which the study population is being compared.

## References

ABS (Australian Bureau of Statistics) 2013, Statistical Language — Statistical Language Glossary, [www.abs.gov.au/websitedbs/a3121120.nsf/00000000000000000000000000000000/1cf6fb476c3de1c7ca257b55002261f2!OpenDocument](http://www.abs.gov.au/websitedbs/a3121120.nsf/00000000000000000000000000000000/1cf6fb476c3de1c7ca257b55002261f2!OpenDocument) (accessed 25 September 2015).

AIHW (Australian Institute of Health and Welfare) 2011, *Principles on the use of direct age-standardisation in administrative data collections: for measuring the gap between Indigenous and non-Indigenous Australians*, Cat. no. CSI 12, Canberra.

— 2015, Age-standardised rate, METeOR, [meteor.aihw.gov.au/content/index.phtml/itemId/327276](http://meteor.aihw.gov.au/content/index.phtml/itemId/327276) (accessed 25 September 2015).

SCRGSP (Steering Committee for the Review of Government Service Provision) 2014, *Report on Government Services 2014*, Productivity Commission, Canberra.