



Australian Government

Australian Institute of
Health and Welfare

Cancer statistics for small geographic areas

User guide and methods manual

The logo for the Australian Institute of Health and Welfare (AIHW), consisting of the letters 'AIHW' in a bold, sans-serif font. Each letter is filled with a different color: 'A' is teal, 'I' is green, 'H' is blue, and 'W' is purple. The letters are slightly shadowed and appear to be floating above a white background.

AIHW



Australian Government

**Australian Institute of
Health and Welfare**

Cancer statistics for small geographic areas

User guide and methods manual

Australian Institute of Health and Welfare
Canberra

Cat. no. CAN 108

The Australian Institute of Health and Welfare is a major national agency whose purpose is to create authoritative and accessible information and statistics that inform decisions and improve the health and welfare of all Australians.

© Australian Institute of Health and Welfare 2019



This product, excluding the AIHW logo, Commonwealth Coat of Arms and any material owned by a third party or protected by a trademark, has been released under a Creative Commons BY 3.0 (CC-BY 3.0) licence. Excluded material owned by third parties may include, for example, design and layout, images obtained under licence from third parties and signatures. We have made all reasonable efforts to identify and label material owned by third parties.

You may distribute, remix and build upon this work. However, you must attribute the AIHW as the copyright holder of the work in compliance with our attribution policy available at www.aihw.gov.au/copyright/. The full terms and conditions of this licence are available at <http://creativecommons.org/licenses/by/3.0/au/>.

A complete list of the Institute's publications is available from the Institute's website www.aihw.gov.au.

Suggested citation

Australian Institute of Health and Welfare 2019. Cancer statistics for small geographic areas: user guide and methods manual. Cat. no. CAN 108. Canberra: AIHW.

Australian Institute of Health and Welfare

Board Chair
Mrs Louise Markus

Chief Executive Officer
Mr Barry Sandison

Any enquiries relating to copyright or comments on this publication should be directed to:

Australian Institute of Health and Welfare
GPO Box 570
Canberra ACT 2601
Tel: (02) 6244 1000
Email: info@aihw.gov.au

Published by the Australian Institute of Health and Welfare

This publication is printed in accordance with ISO 14001 (Environmental Management Systems) and ISO 9001 (Quality Management Systems). The paper is sourced from sustainably managed certified forests.



**Please note that there is the potential for minor revisions of data in this report.
Please check the online version at www.aihw.gov.au for any amendments.**

Contents

User guide to AIHW cancer statistics for small geographic areas	1
Introduction.....	1
The structure of the tables—how to find your way around	1
Reporting periods	3
General principles for using and interpreting small area cancer data	4
Potential issues with small numbers and the need for caution	4
Why are rates higher in some areas than they are in others?	6
Using and interpreting the data.....	7
A note about the treatment of small numbers and wide confidence intervals	7
Populations, median age and counts	7
Rates.....	8
Data and methods used to calculate small area cancer statistics.....	15
The data	15
ABS population data	15
ABS correspondences	15
The Australian Cancer Database	15
The National Mortality Database.....	16
Methods	16
Population	16
Annualised cancer incidence	16
Crude rate	17
Age-standardised rates.....	17
Over/under expected number of new cases.....	18
Median age at diagnosis.....	18
Trends	18
Projected incidence in 2019.....	19
Cancer mortality statistics	20
Prevalence	20
Suppression of statistics pertaining to small numbers.....	21
Confidence in estimates	21
Why measures of confidence have been used.....	21
How measures of confidence have been used.....	22
How measures of confidence have been calculated	23

Exploration of the use of indirect age-standardisation in reporting small area cancer incidence and mortality, and trends	28
The use of indirect age-standardisation in reporting small area cancer incidence and mortality.....	28
The first method.....	28
The second method.....	30
Conclusion.....	32
The use of indirect age-standardisation in reporting small area cancer incidence and mortality trends.....	32
Background	32
Method	32
Results	33
Conclusion.....	35
Appendix A: Data quality statements	36
Acknowledgments.....	37
Abbreviations	37
References.....	38
List of tables	39
List of figures	39
Related publications	40

User guide to AIHW cancer statistics for small geographic areas

Introduction

This user guide and methods manual provides important information about the data presented in the [Cancer statistics for small geographic areas product](#), what should be considered in interpreting the data and the methods used.

Small area cancer statistics are subject to volatility as a consequence of small counts in many of the geographic areas presented. Caution should be exercised in interpreting the data presented in the tables. This user guide aims to assist readers in getting the most out of the small area cancer incidence, mortality and prevalence data presented in the tables.

Methods used to calculate rates, trends and projections, and confidence intervals around these are detailed from page 15.

Reasons for the use of indirect age-standardisation and an assessment of the potential for introduced error through its use are provided from page 28.

The structure of the tables—how to find your way around

Data are presented for 2010–14, which is the most recent 5-year period for which national cancer data are available by small geographic areas. A five-year block is used to provide enough data for meaningful reporting in most areas, while being as up-to-date as possible.

Data are also presented for 2005–14, the most recent 10-year period, to allow reporting for more of the smaller geographic areas (where there might have been insufficient data to report for 2010–14).

Typically, a user will access the more recent 2010–14 data in the first instance, and only access the 2005–14 data if the former were not considered reliable enough.

Some areas will be so small as to prevent reporting of many or most, statistics, even when reporting for a 10-year period.

Cancer data are presented in four Excel workbooks, relating to:

1. 31 PHNs in the 5-year period 2010–14
2. 333 SA3s for the 5-year period 2010–14
3. 31 PHNs in the 10-year period 2005–14
4. 333 SA3s for the 10-year period 2005–14

Each workbook contains 20 spreadsheets, presenting data for all cancers combined and each of the following 18 most common cancers:

- Brain cancer
- Breast cancer
- Bladder cancer
- Cervical cancer
- Colorectal cancer

- Head & neck (with lip) cancer
- Kidney cancer
- Leukaemia
- Liver cancer
- Lung cancer
- Lymphoma
- Melanoma of the skin
- Ovarian cancer
- Pancreatic cancer
- Prostate cancer
- Stomach cancer
- Thyroid cancer
- Uterine cancer.

The workbooks also contain a spreadsheet describing the population in each area for the years 2005 to 2014, and the projected population for 2019.

The two Excel workbooks detailing small area cancer data for SA3s also include a lookup-table listing the SA3s in each of the PHNs, including details of which SA3s are shared between PHNs.

The information presented in each spreadsheet relates to population, and to cancer incidence, mortality, and prevalence.

Data are presented for males, females, and persons.

Specifically the following statistics are presented:

- Populations:
 - the population of males, females, and persons
 - the median age of the population
- Incidence of each cancer:
 - the annualised number of new cases of cancer diagnosed
 - the crude rate at which new cases of cancer are diagnosed (that is, crude incidence rate)
 - the direct age-standardised incidence rate
 - the Standardised Incidence Ratio (SIR)
 - the difference between the number of observed and expected new cases
 - the median age at diagnosis
 - the trend in the crude incidence rate (only for PHNs)
 - the trend in the SIR (only for PHNs)
 - the projected number of new cases expected to be diagnosed in 2019 (only for PHNs and ‘all cancers combined’)
 - the projected crude incidence rate in 2019 (only for PHNs and ‘all cancers combined’)
- Mortality for each cancer:
 - the annualised number of deaths due to cancer

- the crude death rate due to cancer
- the direct age-standardised death rate
- the Standardised Mortality Ratio (SMR)
- the difference between the number of observed and expected deaths;
- the median age at death
- the trend in the crude death rate (only for PHNs)
- the trend in the SMR (only for PHNs)
- the projected number of deaths due to cancer in 2019 (only for PHNs and 'all cancers combined')
- the projected crude mortality rate in 2019 (only for PHNs and 'all cancers combined')
- Prevalence of each cancer:
 - the 1-year prevalence
 - the 5-year prevalence
 - the 10-year prevalence
 - the 33-year prevalence.

Due to the volatility that can affect small area statistics, rates, trends, projections, and other calculated statistics are presented as ranges within which there is 95% confidence that the underlying statistic lies. An estimate for each statistic is also provided, but this should always be interpreted with reference to the range.

Cells have been coloured to indicate statistical significance; that is, where there is:

- 95% confidence that rates are different from national rates
- 95% confidence that rates are increasing or decreasing
- 95% confidence that the number of observed cases is different from the number expected.

Reporting periods

At the time of analysis, the most recent cancer incidence and cancer mortality data that could be reported nationally, by PHN and SA3 relate, respectively, to the calendar years 2014 (for incidence) and 2016 (for mortality). To enhance the ability to interpret the incidence and mortality data together, reporting for both data sets relates to the same time-periods. Therefore, the most recent mortality data available (2015 and 2016) has not been reported.

Statistics have been presented for the 5-year period 2010–2014 in the first instance. Reporting for a multi-year period increases the number of counts upon which rates can be calculated, resulting in less volatility in calculated rates, which is especially useful when reporting for smaller geographic areas and when reporting for less common cancers.

To enhance the opportunity to report reasonable estimates for less common cancers and for the smaller areas, cancer incidence and mortality have also been calculated and presented for the ten-year period, 2005–2014. The longer time period approximately doubles the number of counts upon which rates can be based, but some of the data included is older and therefore less timely. However, when able to be compared, there appears to be strong similarity between rates for the shorter and longer time-periods.

General principles for using and interpreting small area cancer data

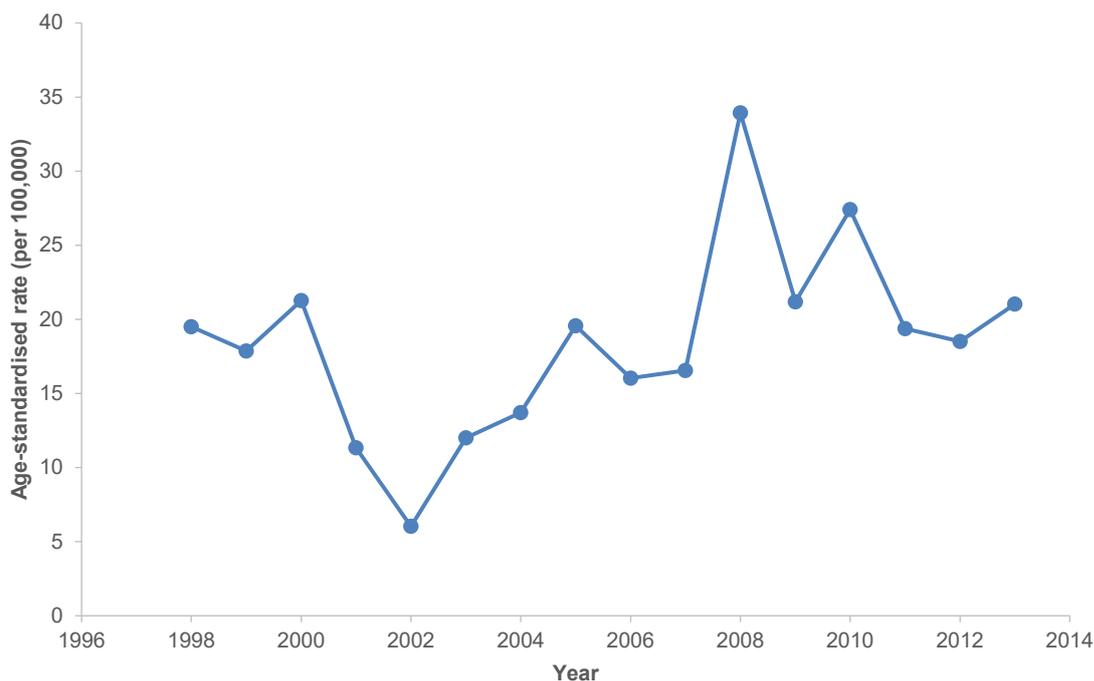
Potential issues with small numbers and the need for caution

Reporting rates for large populations, where there are large numbers of cases upon which to calculate reliable rates, is reasonably straightforward. But reporting rates for small areas relies on the use of small (and frequently very small) numbers of cases to describe cancer in those areas.

Rates based on small numbers can be subject to volatility over time; and the smaller the number of cases, the more opportunity there is for chance events to affect the calculated rates. For example (and as discussed on page 21), a rate based on three cases, could by chance be a rate based on 2 cases or 1 case depending on how chance affects the exact time at which they are diagnosed—which in this example would have a very significant effect on calculated rates.

Another way to demonstrate the volatility of small area cancer data is to consider the year-to-year volatility of incidence. For example, assume that in a population of 100,000 people, there were 21 new cases of a particular cancer diagnosed in 2010, 15 in 2011, 24 in 2012, 27 in 2013, and 18 in 2014, with a total of 105 for the 5 years, or an average of 21 for the 5 year period 2010–14. The reported (best estimate of the) rate is 21 new cases per 100,000 per year, on average, but we know that the rate in each of the individual years was not 21 (it varied from 21 to 15, to 24, to 27, to 18 per 100,000 from year to year). However, it is unlikely that the underlying rate of cancer in the population would change so dramatically from year to year. It is much more likely that chance affects the timing of individual cancer diagnoses, and that the underlying rate of cancer is relatively stable (or increases/decreases slowly) over time. Figure 1 illustrates this volatility with real data for one of the less common cancers (liver cancer), for a population about the same size as an average PHN (Aboriginal and Torres Strait Islander males).

Figure 1: Year to year variability in liver cancer incidence rates, Aboriginal and Torres Strait Islander males, 1998 to 2013



Source: AIHW 2018.

The take home message from Figure 1 is that cancer incidence rates in small populations are volatile, and estimates of the rate are subject to a certain amount of variability and uncertainty.

Because of the uncertainty around these rates, the rate reported for a small population is more appropriately considered an estimate within a range of possible values. For this reason, rates and trends in the tables have been reported with 95% confidence intervals to provide the reader with two values between which we can be 95% confident that the underlying rate in the population falls.

For example, there might be 95% confidence that the underlying rate in an area lies somewhere between 10 and 20, with the estimate of the rate being 15. If the estimate for the national rate was 22, and there was 95% confidence that the underlying rate lies somewhere between 21 and 23, then we would report the national rate as 22 and, in this case, there is sufficient evidence to conclude (with 95% confidence) that the area rate is lower than the national rate (because the confidence intervals for the two estimates do not overlap).

If the 95% confidence intervals for the rate in another small area ranged from 12 to 22, then there is not sufficient evidence to conclude, with 95% confidence, that the area rate is statistically significantly different from the national rate (because the confidence intervals (12 to 22 for this small area, and 21 to 23 for the national rate) overlap, and it is possible that the underlying rate in each of the areas could be the same; for example, 21 or 22).

However, if there is 95% confidence that rates in each of 100 small areas are statistically significantly higher than the national rate, then it is possible for about five of them (5%) to be mistakenly identified as having rates that are higher than the national rates when in fact their rates are not actually different. Consequently, the reader is encouraged to think of the rates and trends in the tables as 'likely' as opposed to 'absolute'. The reader can assume that

around 5% of the rates and trends presented as 'statistically significantly different', might not be.

There is no practical way of avoiding this dilemma. If confidence was increased to 99%, then the criteria would be so stringent, that many truly different rates could be missed, and if confidence was to be increased to 100, then no areas would be considered to have rates that were different from national rates (or different from those in other small areas). As it is, there will be areas whose rates are truly different from the national rate but which, due to chance, fail to be identified as statistically significantly different at the 95% confidence level. For example, a trend with confidence intervals of -0.1 and 25.0 would be considered to not be statistically significantly different from zero because the confidence interval includes zero. However, if a slightly lower level of confidence was acceptable, then the trend might be found to be statistically significantly different from zero (for example, at the 94.5% level of confidence).

For these reasons, all calculated statistics should be considered indicative rather than precise so should be interpreted with reference to confidence intervals.

The reader is therefore advised to be cautious in drawing conclusions. All rates and their relationship with rates in other areas should be considered with reference to as much local and contextual information as possible.

Why are rates higher in some areas than they are in others?

When we look at the small area data tables, some rates are higher and some lower than the national rates, and a lot have rates that are similar to, or not statistically significantly different from, the national rates.

Why are rates higher in some areas than they are in others?

For most cancers, the causes are not fully understood. However, some factors that place individuals at greater risk for cancer are well recognised (AIHW 2019). These factors can be broadly categorised as family history and genetic susceptibility, demographics (such as older age), behavioural risk factors, biomedical risk factors and environmental risk factors.

The rates of many diseases, including cancer, vary strongly with age (rates usually increasing with increasing age). Age-standardisation is a method of removing the influence of age when comparing populations with different age structures. Some of the rates presented have therefore been age standardised to remove the influence of age on comparisons between populations with different age structures. Further information is provided on pages 8–10 about age-standardisation.

Behavioural risk factors include those that are modifiable through changes in individual behaviour, such as diet, tobacco smoking and drinking alcohol.

Biomedical risk factors are bodily states that have an impact on a person's risk of disease, and include diabetes and excess body fat.

The risk of developing some cancers is associated with exposure to certain substances, pollutants or energies. For example, the risk of developing skin cancer increases with increasing exposure to ultraviolet radiation.

The prevalence of risk factors varies considerably between PHNs and would be expected to influence rates. Modelled estimates (AIHW 2017) show that smoking rates in individual PHNs range from 9% to 22% of the adult population, risky alcohol consumption ranges from 11% to 22%, adequate fruit intake from 46% to 55%, physical inactivity from 59% to 75%. Diabetes and obesity also vary considerably, with diabetes prevalence ranging from 3.0% in

one PHN to 5.5% in another, and obesity in males; for example, ranging from 19% of the population in one PHN to 40% in another.

Known and common risk factors are responsible for a considerable proportion of cancers in Australia. Tobacco smoking is estimated to be responsible for 22% of the cancer burden in Australia, high body mass 7.2%, dietary effects (for example, low fruit or fibre consumption, high red and processed meat consumption) 7%, physical inactivity 6.4%, sun exposure 4.5%, alcohol use 3.3%, and occupational exposure (for example, to asbestos, diesel engine exhaust, second-hand smoke, silica, and so forth) 2.4% (AIHW 2019).

Exposure to ultraviolet radiation is a major risk factor for melanoma with opportunity for considerable inter-regional variation. Exposure to a range of carcinogens (IARC 2019) such as diesel exhaust, asbestos, soot, and so forth may also vary regionally.

The relatively large geographic size of PHNs and SA3s reduces the likelihood that any cancer clusters due to point source exposures to carcinogens would be noticeable in the tables.

Any consideration of why cancer rates are higher in some areas than in others should consider the prevalence and effect of known risk factors in the first instance.

Using and interpreting the data

A note about the treatment of small numbers and wide confidence intervals

Numbers and rates based on fewer than 20 cases or deaths have been suppressed. Also, numbers of deaths and associated mortality rates have been suppressed if there were fewer than 20 new cases of cancer in that population, in the reporting period.

Where numbers and rates have been suppressed, the trends for these areas have also been suppressed because they will be based on extremely small numbers and likely to be especially unreliable.

Due to the small number of cases and deaths in some areas (especially for less common cancers), some estimates can have relatively wide confidence intervals, indicating a higher level of uncertainty about that rate, trend, or projection.

To reduce the likelihood of misinterpretation, all calculated statistics (rates, trends, over/under expected counts, and projections) have been presented accompanied by their upper and lower 95% confidence limits, and those statistics that are statistically significantly different from their comparison statistic have been highlighted to differentiate them from less reliable statistics.

Populations, median age and counts

Populations and median age have been included as contextual information. More cases of cancer and more deaths are to be expected from larger populations, while fewer cases of cancer and fewer deaths are to be expected from younger populations.

Average annual numbers of new cases of cancer (and deaths due to cancer) are simple counts of the average number notified to the state and territory Cancer Registries (and state and territory Registers of Births, Deaths and Marriages) each year over the reporting period. Whereas the incidence is the number of cancers diagnosed (similar to, but not necessarily the same as the number of people diagnosed with cancer), deaths relate to the number of people who die due to cancer.

Guidance on the interpretation of rates, trends, over/under expected counts, projections, and cancer prevalence is presented below.

Rates

Three rates are presented:

- the crude rate
- the direct age-standardised rate
- the indirect age-standardised rates (standardised incidence ratios (SIRs) and standardised mortality ratios (SMRs)).

Each of these three rates will describe a slightly different aspect of the experience of cancer in each population.

The reliability of rates is influenced by the number of cases upon which they are based.

Confidence in the rate will be low if the rate is based on a small number of diagnoses or deaths, and high if it is based on a large number.

Rates are expressed as three numbers, which should be read together: the lower 95% confidence limit, the estimate and the upper 95% confidence limit. The reader can be 95% confident that the underlying rate in the population lies somewhere between the lower and upper 95% confidence limits, and is more likely to be close to the estimate than it is to be close to either of the confidence limits.

Where confidence intervals for an area don't overlap with those for the nation, cells have been coloured to indicate that rates are statistically significantly different from national rates at the 95% level of confidence. Rates that are statistically significantly higher are coloured orange; rates that are statistically significantly lower are coloured blue.

Where rates are statistically significantly different from national rates, the reader can be 95% confident that they are different from the national rates, and that there is a 95% probability that the rate in any such population is different from the national rate.

If a rate is not statistically significantly different from the national rate, inferences should not be made because rates may not be meaningfully different from the national rate.

Any crude or directly age-standardised rates for PHNs or SA3s with confidence intervals that do not overlap can be considered different from one another at the 95% level of confidence (that is, there is a 95% probability that they are different from one another).

Indirect age-standardised rates should only be compared with the standard (that is, the national rate), and comparisons between PHNs or SA3s using indirect age-standardised rates should be avoided because of the effect of the error (when comparing between areas) inherent in the method. In this work, this error has typically been found to be small, but it is evident (see page 28).

Crude rates

The crude rate is calculated as the number of cases in the time period divided by the population at that time (see page 17).

A high crude rate indicates that either the underlying age specific cancer rates are high, or that the average age of the population is high or there could be a combination of these factors.

While crude rates are valuable for expressing the total amount of disease per unit of population, they do not take account of differences in the age and sex structure of the populations when making comparisons.

The risk of cancer is strongly associated with age, with older people much more likely than younger people to be diagnosed with cancer.

To allow for differences in the age and sex structure of populations, rates can be age-standardised by either the direct or the indirect method.

Direct age-standardised rates

Direct age-standardisation is the most stable way to adjust for differences in the age structure of different populations (see page 17) to enable comparisons of underlying rates with the effect of differing age structures removed. As it relies on the calculation of stable age-specific rates in the study population, it is most effectively used to age-standardise rates in larger populations, but is less effective when populations are small (as is the case for many of the smaller populations described in the tables).

Where there are sufficient numbers of cases for the method to be reliable, direct age-standardisation is the preferred method for comparing between populations as it preserves the consistency between the comparison populations.

Unlike crude rates, direct age-standardised rates are artificial constructs. They describe the number of new cases of cancer that would have been diagnosed if the rates of cancer in the study population had applied to the 2001 Australian population (the standard population). For example, a direct age-standardised incidence rate of 250 indicates that age-specific cancer rates in a study population, if applied to the 2001 Australian population, would have resulted in 250 new cases of cancer being diagnosed for every 100,000 people in that standard population. This direct age-standardised rate can be compared with direct age-standardised rates for other populations which are based on the same standard.

Standardised incidence rates and standardised mortality rates

Standardised incidence rates (SIRs) and standardised mortality rates (SMRs) are a comparison of the number of observed cases or deaths divided by the number of expected cases or deaths (see page 17).

A ratio of 1.00 indicates that there are exactly the same number of cases (or deaths) as expected, while a ratio of 1.08 indicates that the number of cases or deaths is 8% higher than expected, and a ratio of 0.80 indicates that the number of cases or deaths is 80% of that expected (that is 0.8 times the number expected).

While they are ideal for comparing with the standard population, and are particularly useful when working with small populations, SIRs and SMRs for individual areas should not be compared with those for other individual areas. The purpose of indirect age-standardisation is to adjust for differences in the age structure of two different populations and it is only valid when comparing the study population (for example, that of a PHN) with the standard population (for example, the Australian population in 2010–14).

Where possible, comparisons between areas such as PHNs and SA3s should be based on the direct age-standardised rate, not the SIR or the SMR.

Box 1: Worked example of interpretation of direct age-standardised rates and standardised incidence rates

Melanoma incidence rates for females living in the Tasmanian PHN.

From the 2010–14 tables:

1. The direct age-standardised rate (DASR) is 43.6 (95% CI 40.3 to 47.1)—which compares with 40.1 (95% CI 39.6 to 40.6) for Australian females.
2. The standardised incidence rate (SIR) is 1.083 (95% CI 1.003 to 1.167)—which compares with 1.00 for Australian females.

DASR

We can be 95% confident that the underlying rate at which new cases of melanoma is diagnosed among Tasmanian females is between 40.3 and 47.1 new cases per 100,000 population per year, with the 'best' estimate being 43.6 new cases per 100,000 population per year. We can be 95% confident that the national rate lies somewhere between 39.6 and 40.6 new cases per 100,000 population per year. Even though the best estimate of the rate for Tasmanian females is 43.6 and for Australian females is 40.1, the difference between the Tasmanian rate and the national rate would not be considered statistically significant at the 95% confidence level, because the confidence intervals overlap.

Note that the DASR can be compared with any other DASRs presented in these tables because the 2001 Australian population has been used as the standard in all cases. For example, the DASR for melanoma diagnosis in Tasmanian females (95% confidence interval 40–47) could be compared with that for females living in Western NSW (95% confidence interval 38–47; not statistically significantly different), or for females living in Northern Queensland (95% confidence interval 51–58; statistically significantly higher than for Tasmania).

SIR

We can be 95% confident that the rate at which new cases of melanoma are diagnosed among Tasmanian females lies somewhere between 1.003 and 1.167 times the rate at which new cases of melanoma are diagnosed among Australian females. This rate is therefore considered statistically significantly higher than the Australian rate of 1.00. We can therefore be 95% confident that the incidence of melanoma among Tasmanian females is between 0.3% and about 17% higher than the rate among Australian females generally with the best estimate being that it is 8.3% higher (SIR is 1.083) than among Australian females in general.

Commentary

Note that the two rates above (the DASR and the SIR) appear to be in disagreement. The DASR narrowly misses out on being considered statistically significantly higher, while the SIR is narrowly considered to be statistically significantly higher.

For a comparison between the PHN and the national rates, particularly when the numbers of cases are relatively small, the indirect method (SIR) will be slightly better.

In this example, our suggestion as to the best interpretation is that the Tasmanian rate is probably (with 95% confidence) slightly higher (by between 0.3% and 16.7%) than the Australian rate.

Over/under expected counts

Over/under expected counts, is the difference between the observed and expected number of new cases (or deaths). A positive number indicates more cases than expected, while a

negative number indicates fewer cases than expected. A value of 20 indicates that on average, there were 20 more cases than expected each year, while a value of –15 indicates that on average, there were 15 fewer cases than expected each year.

Over/under expected counts is a by-product of the calculation of SIRs and SMRs, and can be used to gauge the size of the additional burden due to higher cancer incidence or mortality (see page 18).

Over/under expected counts are presented with 95% confidence intervals. For example, in the period 2010–14, there is 95% confidence that there were between 236 and 279 more new cases of melanoma than expected in the north coast PHN, with the best estimate being 257.

Box 2: Worked example of interpretation of over/under expected cases

Over/under expected new cases of melanoma among persons living in the North Coast PHN.

According to the 2010–14 tables there were 257 (95% CI 236 to 279) more cases of melanoma diagnosed than expected, on average each year.

Based on the observed and expected number of cases in 2010–14, there were, with 95% confidence, between 236 and 279 more cases of melanoma diagnosed each year in the period 2010–14 (with the best estimate being 257 new cases), than if age-specific rates were the same as for Australians generally. This indicates that, had the Australian rates of melanoma applied in this PHN, there would have been about 257 fewer people in the North Coast PHN diagnosed with melanoma, annually.

Median age at diagnosis and at death

These two statistics provide a sense of the average age at which cancer is diagnosed and at which death due to cancer occurred. A low median age at diagnosis could be due to either the early onset of cancer or earlier detection. The median age at diagnosis and at death for 'all cancers combined' ranges considerably between different SA3s from 52 to 81 years and from 54 to 84 years, respectively.

Trends

Methodology is described from page 18.

Trends have been calculated to describe whether rates have changed over time and to allow the calculation of the projected 2019 incidence and deaths for all cancers combined.

Four trends for the period 2006–14 have been reported. These relate to changes in:

- Crude incidence rates
- SIRs
- Crude death rates
- SMRs.

Population age structures can change over time, and as cancer incidence is influenced by age, this can affect the crude rate of cancer. Trends in the crude rate describe the change in the overall rate of cancer diagnosis and death, and are used in the calculation of the projected new cases and deaths for 2019.

For example, a trend in the crude incidence rate of 10 new cases per 100,000 population per year, indicates that during the period 2006–14, the crude rate at which new cases of cancer were diagnosed increased by 10 cases per 100,000 population per year, on average, each year. If there were 500 new cases of cancer diagnosed per 100,000 population in 2006, then, eight years later in 2014 the trend would predict there to have been 580 ($500 + [10 \times 8]$) new cases of cancer diagnosed in that year.

The trend in the age-standardised rate has been assessed using indirectly rather than directly age-standardised rates. The direct method was not used because of the small number of cases available in many areas for the calculation of rates in each of the years 2006–2014. Instead, the indirect method was used to assess trend. As these types of trend comparisons can be affected by changes in the population of interest over time, details of method verification and an assessment of the potential error are included from page 32).

The trend in the SIR or SMR is potentially subject to error related to changes in the age structure of the individual populations over time. In the most extreme case tested, the size of the error was as much as 20% of the value of the trend (which had the effect of depressing the trend from around 0.0275 points per year to 0.0225 points per year). Errors in most cases tested were substantially less than this and relatively small (see page 34).

The trend in the SIR or SMR describes changes in the rate of cancer while accounting for changes in the age structure of the population (that is, largely removes the effect of an ageing population on the calculated trend).

For example, a trend in the SIR of -0.01 points per year indicates that, during the period 2006–14, the observed number of new cases of cancer (or deaths) decreased by 1 % per annum relative to the number expected each year (crudely akin to the rate decreasing by 1% per annum). If the SIR had been 1.00 in 2006 (the same as the rate in the standard (national) population in the period 2006–14), then a trend of -0.01 would result, eight years later, in the rate becoming 0.92 (that is, 92% of the rate in the standard population in the period 2006–14, indicating that rates decreased by 8% to become 0.92 times the original rate).

Trends in the crude and age-standardised rates describe slightly different concepts, and can sometimes appear to be in conflict. For example, the crude rate for a population could increase by 2 new cases per 100,000 population per year, each year, at the same time as the SIR decreased by 0.01 point per year. This apparent contradiction can occur if the population is ageing (which is typically the case for most parts of Australia)—there can be a decrease in age-specific cancer incidence rates (which tends to decrease overall cancer rates), but the average age of the population is increasing (which tends to increase overall cancer incidence rates); that is, the incidence of cancer for each age group is declining, but there are more people in the older age groups where the incidence of cancer is highest, causing the overall crude rate to increase.

The tables provide the reader with an estimate of the trend, expressed as a range. This range is defined by the upper and lower 95% confidence limits, and includes a 'best estimate'.

Estimates of the trend are based on a 'line of best fit' through the nine estimates of the rate of cancer in each of the years from 2006 to 2014 inclusive. The more cancers that were diagnosed each year, the less that chance will influence the trend, and the greater the

confidence in the trend. Conversely, a trend based on a small number of diagnoses each year may be more influenced by chance events, and is less reliable.

Confidence in the trends has been indicated by colouring cells for which the confidence intervals for the trend are statistically significantly different from zero (at the 95% confidence level). In some cases, particularly in smaller areas and for less common cancers, the confidence intervals are wide indicating considerable uncertainty in the trend estimate.

Trends have not been reported for SA3s because of the wide confidence intervals around many of those trends.

Box 3: Worked example of interpretation of trends

Melanoma incidence trends for persons living in the North Coast PHN.

From the 2010–14 tables:

- trend in the crude rate is 4.1 (95% CIs 2.0 to 6.2)—which compares with 0.7 (95% CIs 0.4 to 0.9) for Australian persons
- trend in the SIR is 0.041 (95% CIs 0.009 to 0.072)—which compares with 0.006 (95% CIs 0.001 to 0.011) for Australian persons.

Trend in the crude rate:

The rate at which cases of melanoma were diagnosed increased by between 2.0 and 6.2 cases per 100,000 population per year, each year, over the period 2006–14 (with 95% confidence and the best estimate being an increase of 4.1 cases per 100,000 population per year, each year, over this period). This increase was statistically significant (at the 95% confidence level) and greater than the increase in the national rate of between 0.4 and 0.9 new cases per 100,000 population per year, each year. If this trend in the North Coast PHN continued beyond 2014, a diagnosis rate of 121 per 100,000 per year in 2010–14 would increase to around 150 per 100,000, per year by 2019. As the changes described are in relation to crude rates, it is unclear whether they are due to an ageing population or due to an increase in the rate of melanoma in the population independent of age.

Trend in the SIR:

Between 2006 and 2014, the rate at which melanoma was diagnosed in the North Coast PHN, increased by between 0.009 and 0.072 (with 95% confidence and the best estimate being 0.041 points) annually. At this rate, the SIR of 1.74 in 2010–14 would be projected to increase to just over 2.00 in 2019; that is, double the Australian rate in 2010–14 (but not necessarily double the Australian rate in 2019 because the Australian SIR also increased between 2006 and 2014, by between 0.001 and 0.011 points (with 95% confidence and the best estimate being 0.006 points) each year).

Projected number of new cases and deaths in 2019

Methodology is described on page 19.

Projections are estimates and are subject to a number of limitations and assumptions.

The projected estimates presented in this report may differ from those published by states and territories, due to differences in the underlying methodologies. In particular, state and territory cancer registries have access to additional information that could lead their projections to be more accurate for geographic areas within their jurisdiction.

The number of new cases (and deaths) projected to occur in each area in 2019 have been calculated by projecting the 2006–14 trend in the crude rate, to 2019, and then applying that rate to the projected population in that year.

Projections for 2019 have been presented to provide estimates of 'current' cancer incidence and mortality.

Because of the uncertainty around projections for smaller areas and individual cancer types (as indicated by the wide confidence intervals), projections are provided only for PHNs and 'all cancers combined', where they are considered more reliable.

Projections should be interpreted carefully with reference to the methodology and to the confidence intervals provided.

Projected crude incidence and death rate in 2019

The projected crude incidence and mortality rates are based on the 2006–2014 trend for the crude rate, extrapolated to 2019 (as for the projected numbers of new cases and deaths in 2019 above).

It assumes that the trend evident between 2006 and 2014 continues unchanged until 2019.

As for the previous section on projected numbers of cases and deaths, projected crude rates should be used cautiously, and with reference to the 95% confidence interval around the estimate.

Prevalence

Prevalence is expressed as the number of people still alive, who have been diagnosed with cancer in a specified previous time period (the past year, 5 years, 10 years or 33 years).

Data and methods used to calculate small area cancer statistics

The data

The data sources used in this work comprise:

- ABS population data
- ABS correspondences
- the Australian Cancer Database
- the National Mortality Database.

ABS population data

Population data were used to derive rates of cancer incidence and mortality. The population data were sourced from the ABS using the most up-to-date estimates available at the time of analysis.

To derive estimates of resident populations, the ABS uses the 5-yearly Census of Population and Housing data and adjusts it as follows:

- All respondents in the Census are placed in their state or territory, Statistical Local Area and postcode of usual residence; overseas visitors are excluded.
- An adjustment is made for persons missed in the Census.
- Australians temporarily overseas on Census night are added to the usual residence Census count.

Estimated resident populations are then updated each year from the Census data, using indicators of population change, such as births, deaths and net migration. More information is available from www.abs.gov.au.

Population projections for 2019 are taken from the small area population projections to 2027 in the GEN aged care data (Australian Government 2012). These projections have been developed by the ABS for the Australian government and hosted by the AIHW.

ABS correspondences

The geographic identifier in both the Australian Cancer Database (ACD) and National Mortality Database (NMD) is the SA2. ABS concordances were used to allocate 2016 SA3s and PHNs to both the ACD and NMD, on the basis of the reported SA2 of the residential address at the time of diagnosis (in the case of the ACD), and on the basis of the reported SA2 of the residential address at the time of death (in the case of the NMD).

The Australian Cancer Database

All forms of cancer, except basal and squamous cell carcinomas of the skin, are notifiable diseases in each Australian state and territory. This means there is legislation in each jurisdiction that requires hospitals, pathology laboratories and various other institutions to report all cases of cancer to their central cancer registry. An agreed subset of the data collected by these cancer registries is supplied annually to the AIHW, where it is compiled

into the ACD. The ACD currently contains data on all cases of cancer diagnosed from 1982 to 2015 for all states and territories with the exception of 2015 New South Wales data.

At the time of analysis, 2014 was the most recent year for which coding of the small geographic areas, known as SA2s, was complete for all states and territories. SA2s are the key geographic entities used to report for the larger SA3s and PHNs. Consequently, 2014 is the most recent year for which it is currently possible to report cancer incidence for PHNs and SA3s.

Cancer reporting and registration is a dynamic process, and records in the state and territory cancer registries may be modified if new information is received. As a result, the number of cancer cases reported by the AIHW for any particular year may change slightly over time and may not always align with state and territory reporting for that same year.

For more information on the ACD please see the [ACD 2015 Data Quality Statement](#).

The National Mortality Database

The AIHW National Mortality Database (NMD) contains information provided by the Registries of Births, Deaths and Marriages and the National Coronial Information System—and coded by the ABS—for deaths from 1964. Registration of deaths is the responsibility of each state and territory Registry of Births, Deaths and Marriages. These data are then collated and coded by the ABS and are maintained at the AIHW in the NMD.

For the purposes of small area data reporting, rates are calculated for the year in which a death occurred rather than the year in which it was registered.

Methods

Not all records include a geographic location, and some geographic locations are not recognised by correspondences used to allocate PHN or SA3. Consequently, a small fraction of cancer notifications and deaths are not allocated to a PHN or a SA3. Overall, this applies to an average of 0.16% of all cancer notifications and 0.22% of cancer deaths in the period 2010–2014, and 0.20% of all cancer notifications and 0.20% of cancer deaths in the period 2005–2014.

Methods are described here in the order in which statistics are presented in the tables.

Population

In order to provide the reader with contextual information, each spreadsheet documents the size and the median age of the population for the reported period. The populations and median ages presented were derived from ABS Estimated Resident Populations (ERPs).

Annualised cancer incidence

Annualised cancer incidence for the period 2010–2014 is presented as the ‘Average annual number of new cases (2010–2014)’.

The average number of new cases of cancer notified to cancer registries annually, in the period 2010–14 provides the reader with an understanding of the volume of new cases each year.

Annualised incidence for 2005–2014 are also provided (to be used when numbers of new cases in the 2010–14 period are too small to provide adequate estimates).

Crude rate

The crude incidence rate is calculated as the number of new cases diagnosed in the period, divided by the number of person years of exposure in the period, multiplied by 100,000. The resultant rate is the number of new cases diagnosed per 100,000 population per year.

The number of person years of exposure is calculated as the sum of the estimated resident population for each of the years in the period.

Age-standardised rates

More meaningful comparisons between populations can be made using age-standardised rates (ASRs), with such rates adjusted for age in order to facilitate comparisons between populations that have different age structures; for example, between the populations of different PHNs. This standardisation process effectively removes the influence of age structure on the summary rate.

Two methods are commonly used; direct and indirect age-standardisation, which are described below.

Direct age-standardisation

Direct age-standardised rates have been calculated using the following method:

1. Age-specific rates in each study population are calculated for each 5-year age group, by dividing the number of cases or deaths in each of the age groups by the population in each of the age groups
2. These age-specific rates are applied to the standard population (Australian population of persons as of 30 June 2001) to calculate the number of cases or deaths that would have occurred in each age group of the standard population
3. The number of cases in each age group are summed and then divided by the total of the standard population to give the age-standardised rate for the population of interest
4. Finally, this is expressed per 100,000 population.

Direct age-standardisation is the preferred method for comparing between populations, as it preserves the consistency between the comparison populations.

However, as direct age-standardisation relies on calculation of age specific rates in the local population, it is sensitive to small cell sizes which are common with less common cancers and when dealing with small areas such as SA3s.

Indirect age-standardisation

Indirect age-standardisation involves a comparison between the number of cases that occurred and the number that would have been expected if standard age specific rates had occurred in the study population.

Standardised Incidence Ratios (SIRs) and Standardised Mortality Ratios (SMRs) have been calculated using the following method:

1. National age-specific rates (national age specific rates for the period 2010–14 for the tables relating to 2010–14, and national age specific rates for the period 2005–14 for the tables relating to 2005–14) for each sex are calculated for each 5-year age group, by dividing the number of cases or deaths in each age group, by the population in each age group.

2. These national age-specific rates are applied to each 5-year age group for each sex in the study population to calculate the expected number in each age group, which are then summed to form 'E', the total number of cases that would have occurred in the study population—if the rates in that study population had been the same as in the standard (national) population.
3. This expected number 'E' is compared with 'O', the observed number of cases in the study population, by dividing O by E, to yield the ratio of observed to expected cases.

An SIR of 1.00 indicates that as many cases occurred as would have been expected. An SIR of more than 1.00 indicates that more cases occurred than expected. An SIR of less than 1.00 indicates that fewer cases occurred than expected.

SIRs and SMRs for persons are calculated as the sum of the observed cases for males and females, divided by the sum of the expected cases for males and females.

The advantage of indirect age-standardisation is that it is more reliable than direct age-standardisation when working with small populations. All that is required from the study population is knowledge of the population, its age structure, and the total number of cases of cancer diagnosed in the period.

Its disadvantage, however, is that it is not recommended as a means of making comparisons between multiple areas because differences in age structure of the individual populations can affect comparisons. In practice however, we have found the level of influence in this work to be relatively small (see page 28).

The SIRs and SMRs presented in the work should only be used to compare rates in specific areas with those in Australia. SIRs and SMRs should not be used to make direct comparisons between individual PHNs or SA3s.

Over/under expected number of new cases

The over/under expected number of new cases of cancer is calculated by subtracting the expected number of new cases of cancer (calculated during indirect age-standardisation) from the observed number of new cases of cancer. A positive number indicates more cases than expected, a negative number indicates fewer cases than expected.

Median age at diagnosis

The median is the middle point in the distribution of ages at which cancer diagnosis was made or death occurred, in which half the ages are above the median and half are below.

Trends

Cancer incidence and death rates may change over time. Trends have been reported for PHNs to provide an understanding of the way in which cancer incidence may have changed over time.

Confidence intervals around the trends for SA3s were considered too wide, in the most part, to allow reliable reporting so have not been presented.

Trends were also used to produce estimates of the projected number of new cancer cases and of deaths in 2019. (At the time of writing, the latest cancer incidence data available nationally, capable of being described at the SA3 level, relates to 2014.)

Two trends were calculated:

- the trend in the crude rate
- the trend in the SIR and SMR.

Trend in the crude rate

The trend in the crude rate describes the change in the overall rate of diagnosis of new cases of cancer. This trend reflects the ageing of the population as well as any changes in the underlying age specific rates of cancer in the population.

Trends have been described for the period 2006–2014, a period during which rate changes at the national level appear to be relatively consistent.

Crude incidence rates, for the period 2006–14, were modelled as a linear function of diagnosis year using ordinary least squares regression.

The slope of the regression line has been reported as the trend.

Due to the relatively small size of the populations of some areas and the relative rarity of some cancers, there are a number of trends for which the confidence intervals are relatively large and the trend is therefore uncertain.

To provide the reader with a means of gauging the level of confidence in trends and projections, confidence intervals are presented.

Trend in the indirect age-standardised rate

The trend in the age-standardised rate describes the change in the rate of cancer in the population, after accounting for changes in the age structure of the population.

Direct age-standardisation would be the most methodologically appropriate means of age-standardising rates for use in determination of trend, with no opportunity for slight changes in the age structure of the population to affect comparability over time. However, for some cancers and areas, the number of new cases notified each year is small, compromising the effectiveness of direct age-standardisation in this case.

Indirect age-standardisation has been used in preference to direct age-standardisation because it is more reliable where there are small numbers of cancer notifications each year. However, comparisons between the nine indirect age-standardised rates could be slightly affected by any gradual changes in the age structure of the population over the nine years.

The methods have been tested and the magnitude of the error found to be small. For PHNs with statistically significant slopes, any error introduced was typically a very small percentage (0.5%–4%) of the value of the slope, and in by far the worst example, equivalent to 20% of the value of the slope (reducing the trend from 0.0275 points per year to 0.0225 points per year (see page 32).

SIRs and SMRs for each year in the period 2006–2014 have been calculated using the national age-specific rates for males and females in the period 2006–2014 as the standard. These rates were then modelled as a linear function of diagnosis year using ordinary least squares regression.

Projected incidence in 2019

The projected incidence in 2019, is the number of new cases of cancer expected to be diagnosed in 2019, based on 2006–14 trends and expected demographic changes.

The number of new cases projected to occur in each area in 2019 have been calculated by projecting the 2006–14 trend in the crude rate to 2019, and then applying that rate to the projected population in that year.

The projected population count was derived by aggregating SA2 counts in the GEN population projection data to the PHN level.

This strategy was applied for males, females, and persons. That is, the person estimate is not the sum of the male and female estimates.

To assist in interpretation, 95% confidence intervals were fitted around the 2019 projected incidence (see page 27).

Due to the uncertainty around projections for smaller areas and less common cancers (as indicated by wide confidence intervals), projections are provided only for PHNs and only for 'all cancers combined', where they are considered more reliable.

Projected crude incidence rate in 2019

The projected crude incidence rate is the crude rate at which new cases of cancer are predicted to be diagnosed in 2019.

Projected crude incidence rates were calculated by projecting the 2006–14 trend in the crude rate, to 2019.

Crude rates were projected separately for males, females, and for persons.

To assist in interpretation, 95% confidence intervals were fitted around the projected 2019 crude rates (see page 27).

Due to the uncertainty around projected rates for smaller areas and less common cancers (as indicated by wide confidence intervals), projected rates are provided only for PHNs and only for 'all cancers combined', where they are considered more reliable.

Cancer mortality statistics

Cancer mortality statistics have been presented in the tables. These include:

- Annualised cancer deaths 2010–14
- Crude death rate
- Direct age-standardised death rate
- Standardised Mortality Ratios
- Excess annual number of deaths
- Median age at death
- Trends
- Projected deaths in 2019
- Projected crude death rate in 2019.

These have been calculated using the same methodology as for incidence described above.

Cancer mortality statistics have been sourced from the National Mortality Database (NMD).

Prevalence

Limited-duration prevalence is expressed as N-year prevalence throughout the small area data tables. N-year prevalence on a given index date—where N is any number 1, 2, 3 and so on—is defined as the number of people alive who had been diagnosed with cancer in the past N years.

Reported prevalences comprise:

- 1-year prevalence is the number of living people who were diagnosed in the past year to 31 December 2014

- 5-year prevalence is the number of living people who were diagnosed in the past 5 years to 31 December 2014. This includes the people defined by 1-year prevalence
- 10-year prevalence is the number of living people who were diagnosed in the past 10 years to 31 December 2014. This includes the people defined by 1-year and 5-year prevalence
- 33-year prevalence is the number of living people who were diagnosed in the past 33 years to 31 December 2014. This includes the people defined by 1-year, 5-year and 10-year prevalence.

Prevalence relates to the number of people diagnosed with cancer, not the number of cancer cases. An individual who was diagnosed with two separate cancers will contribute separately to the prevalence of each cancer. However, this individual will contribute only once to the prevalence of all cancers combined. For this reason, the sum of prevalence for individual cancers will not equal the prevalence of all cancers combined.

Suppression of statistics pertaining to small numbers

Reporting cells with small numbers risks erroneously reporting extremely high or low rates that are due to chance, rather than genuine differences from the national rates.

In order to increase protection against these issues, any cell containing a count based on less than 20 individuals or events, or a cell containing a rate or other statistic based on a count of less than 20 individuals or events, will be suppressed (that is, not reported).

For example, while 20 new cases of cancer reported in the 5-year period 2010–14 would be reported (as an average of 4 cases per year, along with a rate), 19 new cases of cancer reported in the 5-year period 2010–14 would not be reported (and nor would the rate).

In such situations, the data for that area may be reported (if adequate numbers of cases are available) in the supporting small area cancer tables that describe cancer for the 2005–14 period.

Confidence in estimates

Why measures of confidence have been used

The purpose of confidence intervals is to provide the reader with a sense of the confidence they should have in the rate, trend or other statistic presented.

If the reader is interested only in the cancer rate within a specific time-period, then the published rate will suffice without the need to reference confidence intervals.

However, the rate for a specific time-period may be misleading because of the opportunity for chance variation in the population between different time-periods. To illustrate, consider a small area in which three new cases of cancer are diagnosed within a five-year time period:

- one case is diagnosed on the first day of the time period
- one case is diagnosed somewhere in the middle of the time period
- one case is diagnosed on the last day of the time period.

This example is extreme, but serves to illustrate the point that, if by chance, diagnosis of the first case had occurred a day earlier, and if diagnosis of the third case had occurred a day later, then the number of cases diagnosed in the five-year period would have been reduced from 3 to 1, and the reported rate of cancer in that area reduced dramatically to a third.

Another way to demonstrate the volatility of small area cancer data is to consider the year-to-year volatility of incidence. For example, assume that in a population of 100,000 people, there were 21 new cases of a particular cancer diagnosed in 2010, 15 in 2011, 24 in 2012, 27 in 2013, and 18 in 2014, with a total of 105 for the 5 years, or an average of 21 for the 5 year period 2010–14. The reported (best estimate of the) rate is 21 new cases per 100,000 per year, but we know that the rate in each of the individual years was not 21 (it changed from 21 to 15, to 24, to 27, to 18 per 100,000 from year to year). A rate based on data from 2009–2015 would be different again, as would one based on data from 2011–2014. However, we know that the underlying rate of cancer in the population is unlikely to change so dramatically from year to year; it is much more likely that chance affects the timing of individual cancer diagnoses, and that the underlying rate of cancer is relatively stable (or increases/decreases slowly) over time. Figure 1 illustrates real previously published year-to-year volatility of liver cancer incidence for Aboriginal and Torres Strait Islander males, a population of similar size to many PHNs.

Such volatility is much more common in small populations, and rates in larger populations are substantially less subject to large fluctuations from one period to the next.

The use of confidence intervals provides the reader with a means of gauging the reliability of, and interpreting, rates, trends and other statistics.

How measures of confidence have been used

In this work, two rates are considered to be different if the confidence intervals around the estimate do not overlap—which is a relatively conservative approach (that is, this approach yields fewer statistically significantly different comparisons than other possible methods).

Crude and direct age-standardised rates

If the 95% confidence intervals for the rate in a small area overlaps the confidence interval for the national rate, then the small area rate can be considered not statistically significantly different from the national rate.

If the two 95% confidence intervals do not overlap, then the small area rate can be considered to be statistically significantly different from the national rate.

SIRs and SMRs

If the 95% confidence intervals for the SIR or SMR includes 1.00, then the small area SIR or SMR can be considered to not be statistically significantly different from the national rate.

If the 95% confidence interval for the SIR or SMR does not include 1.00, then the small area rate can be considered to be statistically significantly different from the national rate.

Over/under expected new cases and deaths

If the 95% confidence intervals for the estimate includes zero, then the estimate is considered to be not statistically significantly different from zero.

If the 95% confidence intervals for the estimate in the small area does not include zero, then the estimate is considered to be statistically significantly different from zero.

Trend in the crude rate, SIRs and SMRs

If the 95% confidence intervals for the trend includes zero, then the trend is considered not statistically significantly different from zero.

If the 95% confidence intervals for the trend in the small area does not include zero, then the trend is considered to be statistically significantly different from zero.

Projected numbers and rates of new cases of cancer and deaths in 2019

All projections are presented with 95% confidence intervals.

Confidence intervals around projections are relatively wide (indicating substantial uncertainty for estimated projections) when calculated for smaller areas and less common cancers. Consequently, projections have been provided only for PHNs (not SA3s) and only for 'all-cancers combined', where they are considered to be more reliable.

How measures of confidence have been calculated

This section provides derivations of confidence limits for crude rates, direct and indirect age standardised rates, projections and the over/under expected number of cases. The assumption underlying all of these formulae is that incidence and mortality counts, disaggregated by sex 5-year age group and year, are realisations of independently distributed Poisson random variables.

Confidence intervals are also derived for trends, in the crude and indirect age standardised rates, obtained from linear least squares regression modelling. These formulae are based on the assumption that the modelling error terms are normally distributed. More specifically, we assume these residuals are independent and identically distributed normal random variables.

Confidence intervals for crude rates

An annual incidence, or mortality, count X is assumed to follow a Poisson distribution with (unknown) mean parameter θ . By equation (1) of Dobson et al. (1991), lower and upper 95% confidence limits for this parameter, X_L and X_U , are given by:

$$X_L = 0.5 * C(0.025|2X)$$

and

$$X_U = 0.5 * C(0.975|2X + 2)$$

where $C(\alpha|\mu)$ is quantile α of the cumulative Chi-squared distribution with μ degrees of freedom.

Suppose incidence count X is observed in a population of size N . Then the crude rate R , per 100,000 population, is given by:

$$R = \frac{X}{N} \times 10^5.$$

This measure has mean parameter:

$$\frac{\theta}{N} \times 10^5.$$

The 95% confidence limits for this parameter, R_L and R_U , are given by:

$$R_L = \frac{50,000 * C(0.025|2X)}{N}$$

$$R_U = \frac{50,000 * C(0.975|2X + 2)}{N}.$$

Confidence intervals for direct age-standardised rates

In this section, the formulae for the 95% confidence limits of a direct age-standardised rate (calculated per 100,000 population) are described. Let:

X_i = observed count for age group i

N_i = Estimated Resident Population count for age group i

c_i = relative population frequency of age group i , in the standard Australian population

$$w_i = \frac{c_i}{N_i} \times 10^5.$$

Then the direct age-standardised rate per 100,000 population, D , corresponding to the total observed count, $X = \sum_{i=1}^{18} X_i$, is given by:

$$D = \sum_{i=1}^{18} c_i \frac{X_i \times 10^5}{N_i} = \sum_{i=1}^{18} w_i X_i.$$

The count X_i is assumed to be the realisation of a Poisson random variable whose mean parameter we will denote by θ_i . Consequently:

$$E(D) = \sum_{i=1}^{18} w_i \theta_i \text{ and } Var(D) = \sum_{i=1}^{18} w_i^2 \theta_i.$$

To obtain a 95% confidence interval for $E(D)$ we consider the following linear transform of the total observed count:

$$T = a + bX$$

where constants a and b are defined so that:

$$E(T) = E(D) \text{ and } Var(T) = Var(D).$$

That is,

$$a = \sum_{i=1}^{18} w_i \theta_i - \left[\frac{\sum_{i=1}^{18} w_i^2 \theta_i}{\sum_{i=1}^{18} \theta_i} \right]^{0.5} \sum_{i=1}^{18} \theta_i = \sum_{i=1}^{18} w_i \theta_i - \left[\sum_{i=1}^{18} \theta_i \sum_{i=1}^{18} w_i^2 \theta_i \right]^{0.5}$$

$$b = \left[\frac{\sum_{i=1}^{18} w_i^2 \theta_i}{\sum_{i=1}^{18} \theta_i} \right]^{0.5}.$$

Since $X = \sum_{i=1}^{18} X_i$ is the sum of independent Poisson random variables, it is also a Poisson random variable with parameter $\sum_{i=1}^{18} \theta_i$. So, by equation (1) of Dobson et al. (1991), lower and upper 95% confidence limits for this parameter, X_L and X_U , are given by:

$$X_L = 0.5 * C(0.025|2X)$$

and

$$X_U = 0.5 * C(0.975|2X + 2)$$

where $C(\alpha|\mu)$ is quantile α of the cumulative Chi-squared distribution with μ degrees of freedom.

Therefore, a lower 95% confidence limit for $E(T) = E(D)$, T_L , is given by:

$$T_L = \sum_{i=1}^{18} w_i \theta_i - \left[\frac{\sum_{i=1}^{18} w_i^2 \theta_i}{\sum_{i=1}^{18} \theta_i} \right]^{0.5} \sum_{i=1}^{18} \theta_i + \left[\frac{\sum_{i=1}^{18} w_i^2 \theta_i}{\sum_{i=1}^{18} \theta_i} \right]^{0.5} X_L$$

$$= \sum_{i=1}^{18} w_i \theta_i + \left[\frac{\sum_{i=1}^{18} w_i^2 \theta_i}{\sum_{i=1}^{18} \theta_i} \right]^{0.5} (X_L - \sum_{i=1}^{18} \theta_i).$$

In estimation, we replace the parameter θ_i with the corresponding observed value X_i . That is, the lower 95% confidence limit is estimated using:

$$\widehat{T}_L = D + \left[\frac{\sum_{i=1}^{18} w_i^2 X_i}{X} \right]^{0.5} (X_L - X).$$

Similarly, the upper 95% confidence limit, T_U , is estimated using:

$$\widehat{T}_U = D + \left[\frac{\sum_{i=1}^{18} w_i^2 X_i}{X} \right]^{0.5} (X_U - X).$$

Confidence intervals for SIRs and SMRs

Consider cancer incidence, or mortality, cross-classified by geographic area, site, five-year time period and sex category (males, females or persons).

Let:

X = observed number of cases

E = expected number of cases based on national age – sex specific rates

R = SIR (or SMR) = $\frac{X}{E}$.

The 95% confidence interval for R is obtained by assuming that X is the realisation of a Poisson random variable. Consequently, the confidence limits for X are given by equation (1) of Dobson et al. (1991). Specifically, the 95% confidence limits are:

$$\frac{1}{2} C(0.025|2X) \text{ and } \frac{1}{2} C(0.975|2X + 2)$$

where $C(\alpha|\mu)$ is quantile α of the cumulative Chi-squared distribution with μ degrees of freedom.

Therefore, the 95% confidence limits for R are:

$$\frac{1}{2E} C(0.025|2X) \text{ and } \frac{1}{2E} C(0.975|2X + 2).$$

Confidence intervals for over/under expected cases

Consider cancer incidence, or mortality, cross-classified by geographic area, site, time-period and sex category (males, females or persons).

Let:

X = observed number of cases

E = expected number of cases based on national age – sex specific rates

n = number of years in the time period.

Then:

$$\text{Excess cases per year} = \frac{X}{n} - \frac{E}{n}.$$

Confidence limits for this measure are obtained by following the argument of the previous section.

The 95% confidence limits for the annualised number of cases, X/n , are:

$$\frac{1}{2n} C(0.025|2X) \text{ and } \frac{1}{2n} C(0.975|2X + 2)$$

Hence, the corresponding limits for *over/under expected cases per year* are:

$$\frac{1}{n} \left(\frac{c(0.025|2X)}{2} - E \right) \text{ and } \frac{1}{n} \left(\frac{c(0.975|2X + 2)}{2} - E \right)$$

Confidence in the trend for the crude incidence rate

An ordinary least squares regression model for the crude incidence rate R_t , as a function of diagnosis year t , was used to estimate the trend in this rate. This model was developed using the nine rates in the period 2006–14.

Specifically, the mathematical form of the model is:

$$R_t = \beta_0 + \beta_1 t + \epsilon_t$$

where the residuals ϵ_t are assumed to be independent and identically distributed normal random variables with mean zero and variance σ^2 . Model parameter β_1 is interpreted as the trend in the crude incidence rate.

To ensure arithmetic precision in the estimation of β_0 and β_1 , the large numeric values of diagnosis year were zeroed at 2006. That is, the observation window 2006–14 was coded as $t = 0, 1, \dots, 8$. The parameter estimates were obtained using the methodology discussed in Section 11.4 of Mendenhall (1983); a summary of which is given below.

Let:

$$\bar{R} = \text{mean observed crude rate}$$

$$\bar{t} = \text{mean of } t = 0, 1, \dots, 8, \text{ that is, } \bar{t} = 4$$

$$SS_R = \sum_{t=0}^8 (R_t - \bar{R})^2$$

$$SS_t = \sum_{t=0}^8 (t - \bar{t})^2 = 60$$

$$SS_{tR} = \sum_{t=0}^8 (t - \bar{t})(R_t - \bar{R}).$$

Then, from page 406 of the above-mentioned reference:

$$\widehat{\beta}_1 = \frac{SS_{tR}}{SS_t} = \frac{SS_{tR}}{60}$$

$$\widehat{\beta}_0 = \bar{R} - \widehat{\beta}_1 \bar{t} = \bar{R} - \frac{SS_{tR}}{15}.$$

As mentioned above, the residuals ϵ_t are assumed to be independent and identically distributed normal random variables. Also, the observation window 2006–14 contains nine data points and there are two model coefficients, β_0 and β_1 . For these reasons, the 95% confidence interval for $\widehat{\beta}_1$ is based on the Student's t distribution with seven degrees of freedom. Specifically, from page 417 of Mendenhall (1983), this interval is given by:

$$\widehat{\beta}_1 \pm \frac{t_{0.025} \times \hat{\sigma}}{\sqrt{SS_t}} = \widehat{\beta}_1 \pm \frac{t_{0.025} \times \hat{\sigma}}{\sqrt{60}}$$

where

$t_{0.025}$ = upper 2.5% quantile of the Student's t distribution with 7 degrees of freedom

$$\hat{\sigma} = \sqrt{\frac{1}{7} \sum_{t=0}^8 (R_t - \widehat{\beta}_0 - \widehat{\beta}_1 t)^2}.$$

Confidence in the trend for the SIR or the SMR

The trend in the SIR or SMR, and its 95% confidence interval, were estimated using the technique described in the previous section. That is, the same formulae were applied with R_t set to the SIR or SMR in year t . Also, the same observation window, 2006–14 (zeroed at 2006), was used.

Confidence around projected 2019 incidence, deaths and crude rates

Projected incidence and mortality crude rates for 2019 were obtained by extrapolating the trends in these rates, which were estimated over the nine year period from 2006 to 2014. The corresponding predicted counts were calculated by multiplying the projected rates by the appropriate ABS population projection.

The 95% confidence limits on a predicted count were obtained by multiplying the limits on the rate by the population projection. The confidence limits on the predicted rate are derived below.

Details on trend estimation, taken from Mendenhall (1983), are given above in the section titled 'Confidence in the trend for the crude incidence rate'. In particular, it was mentioned that the following model was fitted to the crude rate.

$$R_t = \beta_0 + \beta_1 t + \epsilon_t$$

Model estimation was undertaken using the observation window 2006–14, which was zeroed at $t = 2006$. Hence, predictions for 2019 were obtained by putting $t = 13$. More specifically, the formula for the predicted rate is:

$$\widehat{R}_{13} = \widehat{\beta}_0 + 13\widehat{\beta}_1.$$

The variance of this prediction is given by:

$$\begin{aligned} Var(\widehat{R}_{13}) &= Var(\widehat{\beta}_0) + (13)^2 Var(\widehat{\beta}_1) + 2 \times 13 \times Cov(\widehat{\beta}_0, \widehat{\beta}_1) \\ &= Var(\widehat{\beta}_0) + 169 \times Var(\widehat{\beta}_1) + 26 \times Cov(\widehat{\beta}_0, \widehat{\beta}_1) \end{aligned}$$

where

$$Var(\widehat{\beta}_0) = \frac{\sum_{t=0}^8 t^2}{9 \times SS_t} \hat{\sigma}^2 = \frac{204}{9 \times SS_t} \hat{\sigma}^2 = \frac{68}{3 \times SS_t} \hat{\sigma}^2$$

$$Var(\widehat{\beta}_1) = \frac{\hat{\sigma}^2}{SS_t}$$

$$Cov(\widehat{\beta}_0, \widehat{\beta}_1) = -\frac{\bar{t}}{SS_t} \hat{\sigma}^2 = -\frac{4}{SS_t} \hat{\sigma}^2.$$

That is,

$$Var(\widehat{R}_{13}) = \frac{263}{3} \times \frac{\hat{\sigma}^2}{SS_t} = \frac{263}{180} \times \hat{\sigma}^2.$$

The 95% confidence interval for \widehat{R}_{13} is based on $Var(\widehat{R}_{13})$ and the upper 2.5% quantile of the Student's t distribution with seven degrees of freedom, $t_{0.025}$. Specifically, this interval is equal to:

$$R_{13} \pm t_{0.025} \times \left(Var(\widehat{R}_{13})\right)^{0.5}.$$

Exploration of the use of indirect age-standardisation in reporting small area cancer incidence and mortality, and trends

The use of indirect age-standardisation in reporting small area cancer incidence and mortality

Direct age-standardisation provides the most reliable and convenient means of quantifying the underlying cancer incidence and death rates, independent of any differences in the age structure of populations, and making comparisons between areas. However, it does not perform effectively in small populations.

Indirect age-standardisation works well in small (or large) populations, but while it is an effective means of comparing rates in a study area with that in a comparison (standard) area, it is potentially problematic when used to compare between several study areas.

If the aim is to report cancer incidence and mortality for small areas then which method should be used? If indirect age-standardisation were used, how misleading could comparisons between areas be?

Two methods for exploring this question were used:

1. comparison of SIRs and direct age-standardised incidence rate-ratios
2. comparison between SIRs calculated for multiple small areas using one common set of age-specific rates.

The first method

Ideally, direct and indirect age-standardised rates should tell the same story. An SIR of 1.00 indicates that the rate of cancer in the study population is identical to that in the standard population, after accounting for differences in the age structure of the populations. Ideally, the direct age-standardised rate in the same study population would also be identical to that in the standard population.

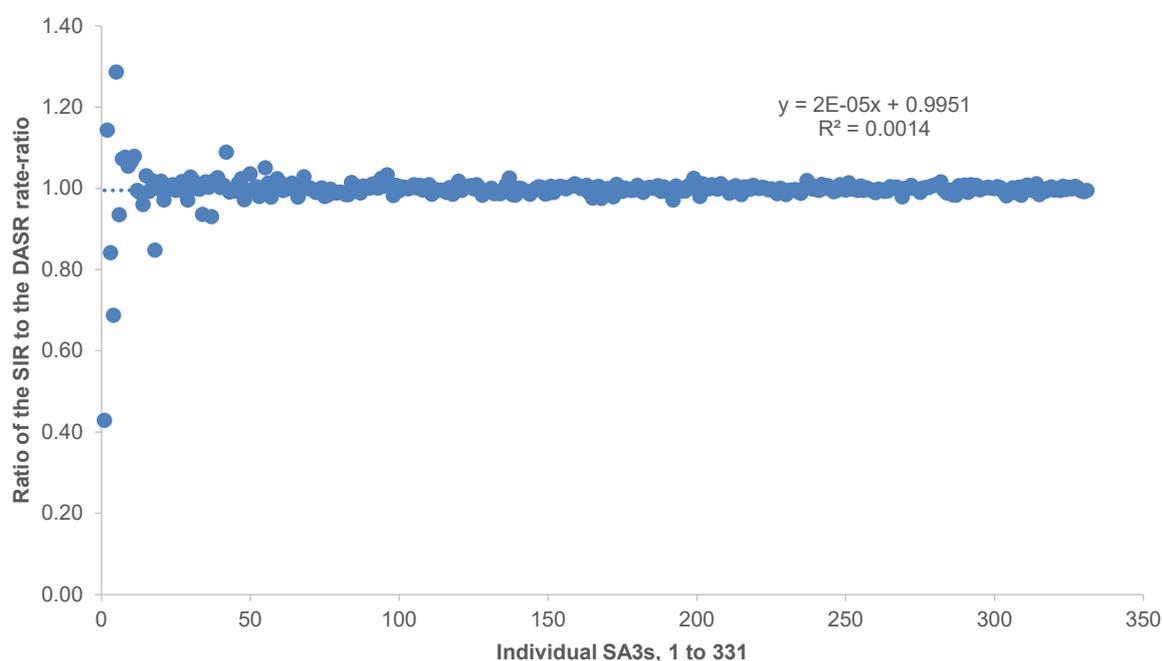
However, two issues may affect direct comparison between the two:

1. the poor performance of direct age-standardisation in describing rates based on small numbers
2. the error inherent in the indirect method that complicates comparison between study populations.

It is important to gauge the size, direction and relevance of the potential error in using the indirect method.

Figure 2 compares the SIR with the direct age-standardised rate (DASR) rate-ratio for each SA3, sorted so that SA3s with the smallest number of diagnosed cases of cancer (typically those with the smallest populations) are on the left hand side of the figure, with those SA3s with the greatest number of cancer diagnoses (typically those with the largest populations) on the right hand side of the figure. (The DASR rate-ratio is calculated as the ratio of the DASR for an SA3 divided by the DASR for the standard population. A rate-ratio of 1.00 indicates that the rates in the two populations are identical.)

Figure 2: Comparison of the SIR with the direct age-standardised incidence rate-ratio, for males resident in each of 331 SA3s, diagnosed with any cancer in the period 2010–14, sorted by incidence



Note: Rates that are based on fewer cases are on the left hand side of the figure, while rates that are based on more cases are on the right hand side of the figure. Two SA3s in which no males were diagnosed with cancer in the period have been omitted.

Source: AIHW analysis of the AIHW Australian Cancer Database.

Ideally, all values of the ratio would be 1.00, but this is not the case because of the influences discussed previously.

On the left hand side, there appear to be substantial differences between the SIR and the DASR rate-ratios. Rates for those thirteen SA3s for which the difference between the SIR and the DASR rate-ratio is more than 7%, are typically based on less than ten cases annually or are otherwise based on fewer than 80 cases annually. Two of these SA3s had both SIRs and DASRs that were statistically significantly different from the national rates (based on 71 and 68 newly diagnosed cases of cancer), in one case (based on 34 cases) the DASR was statistically significantly different from the national rate, but the SIR was not, and in all other cases neither SIRs nor DASRs were statistically significantly different from the national rate.

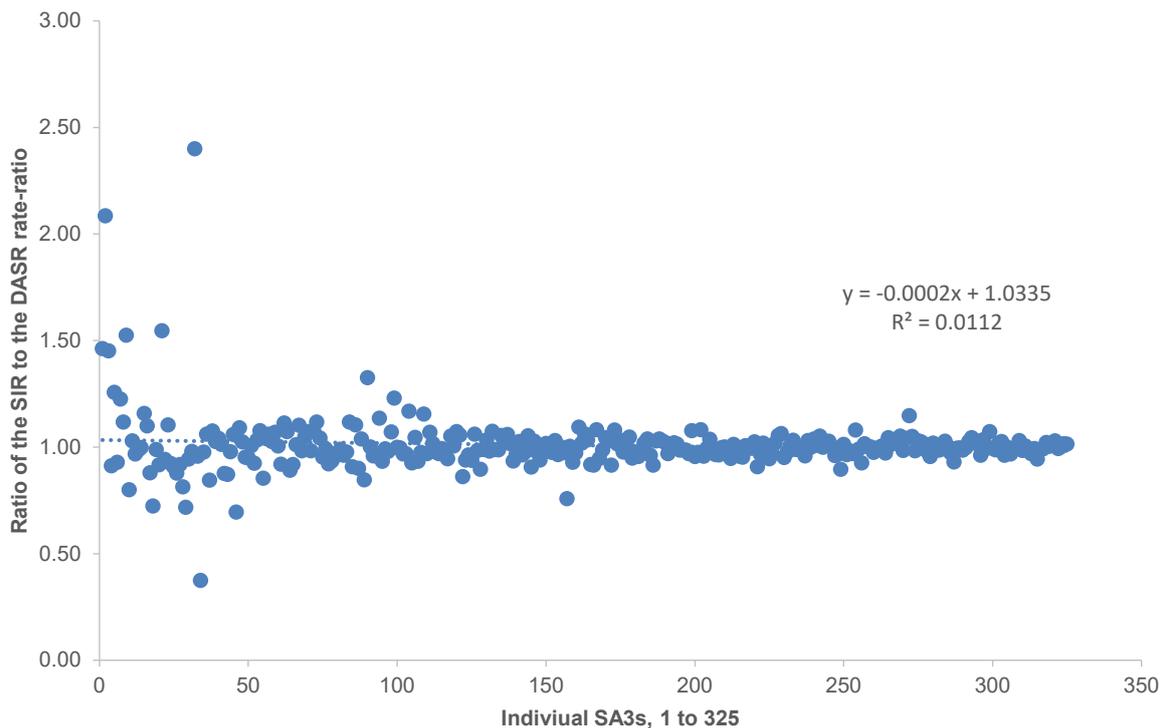
On the right hand side of the figure, where relatively large numbers of cases annually (typically hundreds (to a maximum of seven hundred)) contribute to the calculation of both the DASR and the SIR, there appears to be strong agreement between the two methods, typically within 1%, sometimes within 2%, and rarely 3% different.

Of ten SA3s with cancer incidence of more than 100 new cases per year, where the difference between the SIR and the DASR rate-ratio was greater than 1.5%, five had SIRs and DASRs that were both statistically significantly different from the national rates, one had an SIR that was not (and a DASR that was) statistically significantly different from the national rate, while all of the rest had both SIRs and DASRs that were not statistically significantly different from the national rate.

Repeating this work with one of the less common cancers, a similar pattern emerges, but with less clear congruence between the SIR and the DASR rate-ratios, because of the smaller number of cancers typically diagnosed (in this example, the maximum number of new cases in any PHN was 18 per year, with most PHNs typically having fewer than 10 new cases per year).

Figure 3 describes the ratio of the SIR to the DASR rate-ratio, for males in each SA3 diagnosed with stomach cancer in the period 2010–14.

Figure 3: Comparison of the SIR with the direct age-standardised incidence rate-ratio, for males resident in each of 325 SA3s, diagnosed with stomach cancer in the period 2010–14, sorted by incidence



Note: Rates that are based on fewer cases are on the left hand side of the figure, while rates that are based on the more cases are on the right hand side of the figure. Eight SA3s in which no males were diagnosed with stomach cancer in the period have been omitted.

Source: AIHW analysis of the AIHW Australian Cancer Database.

Figure 3 illustrates the substantial differences between SIRs and directly age-standardised rate-ratios when the number of cases reported are relatively small. In this case, with such small numbers of cases contributing to the calculation of the rates, it is likely that DASRs will be less reliable than SIRs, with much of the differences between the two being a consequence of the inadequacy of the direct method to cope with small numbers, much more than with any error associated with the indirect method.

The second method

Another way of reviewing the ‘comparability’ of SIRs, or of assessing the drift associated with the method, is to compare the calculated SIRs for a multitude of small areas, which have been allocated identical age-specific cancer incidence rates.

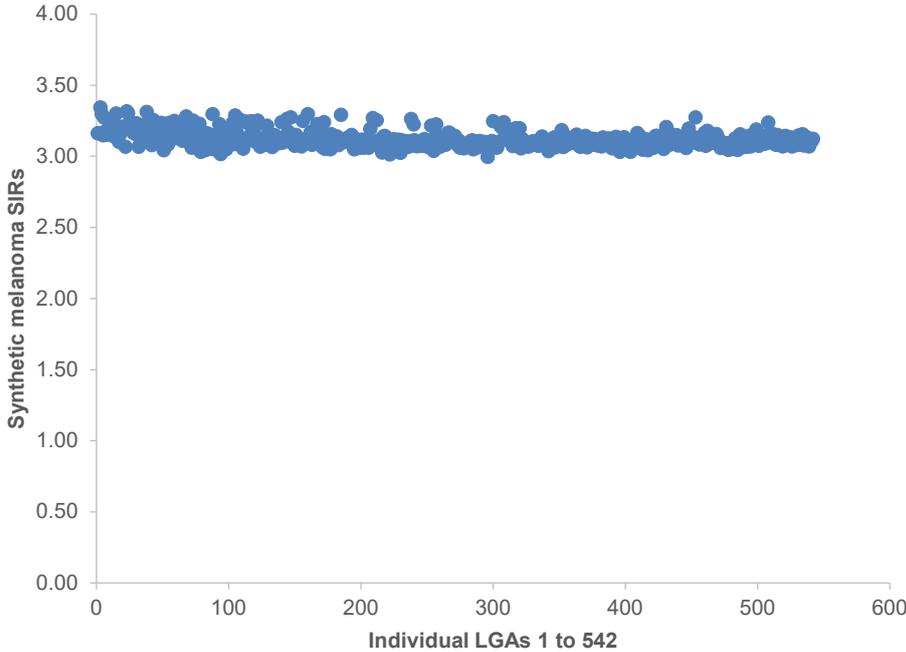
Age-specific melanoma incidence rates for Northwest Melbourne PHN were applied to the populations of each of 544 LGAs so as to calculate an 'expected' number of new cases of melanoma. These expected cases are the number expected if rates in each LGA were identical to those in Northwest Melbourne PHN (which has a low melanoma incidence rate) at that time.

Age-specific melanoma incidence rates for Gold Coast PHN (which has a high melanoma incidence rate) were applied to the populations of each of 542 LGAs so as to calculate a synthetic 'observed' number of new cases of melanoma. These synthetic observed cases are the number of new cases that would be observed in each LGA if rates in each LGA were identical to those in the Gold Coast PHN.

The synthetic 'observed' number of new cases for each LGA was divided by the expected number to calculate a synthetic SIR.

Ideally, the calculated synthetic SIRs for all of the LGAs would be identical (because the underlying age-specific rates are identical). Any differences will be due to the slightly different age structure of the 544 LGA populations, and be indicative of the error associated with the indirect method when comparing against other study populations.

Figure 4: Calculated synthetic SIRs for 542 LGAs, having the same age-specific melanoma incidence rates as the Gold Coast PHN, standardised to rates in the North West Melbourne PHN, 2009–13, sorted by number of cases diagnosed



Source: AIHW analysis of the AIHW Australian Cancer Database

Figure 4 above describes the distribution of SIRs for the 542 LGAs. The median SIR is 3.10, with values ranging from 3.00 to 3.34, with 80% lying between 3.07 and 3.22. The furthest any LGA is from the median is ~7% and all show a relatively high SIR, approximately three times higher than the standard.

Conclusion

For the purpose of reporting cancer rates in PHNs and SA3s, Indirect age-standardisation appears to generally work well and reliably, with the following observations:

- For areas with smaller numbers of cancer cases per year, direct age-standardisation is unlikely to work well, and SIRs and SMRs are more likely to be meaningful.
- In areas with large numbers of (typically more than 50 or 100) cancer cases per year, DASRs would be expected to work well. In such areas, SIRs and SMRs appear to align well with DASRs in so far as their relationship to national rates are concerned, with differences where they occur (relative to national rates) of typically up to one or two percent.
- The use of confidence intervals appears to protect well against questionable comparisons based on small numbers.
- SIRs and SMRs appear to be reliable means of comparing local with national rates. DASRs appear to be the more reliable means of comparing between local areas (for example, between PHNs). Given the level of agreement between SIRs or SMRs and DASR rate-ratios, SIRs and SMRs may also have a role in comparing between areas if used carefully and conservatively.

The use of indirect age-standardisation in reporting small area cancer incidence and mortality trends

The indirect method is not recommended for trend analysis because of the potential for statistical drift associated with shifts in age structure of the population.

However, the indirect method is better at reporting age-standardised rates of disease in small populations than the direct method, and so the drift associated with the indirect method has been assessed for reporting cancer incidence and mortality over the short eight-year period 2006 to 2014.

Background

The basis of this assessment is that, if the underlying age-specific rates in a population did not change over time and, the age structure of the population remained the same, then SIRs and SMRs would remain unchanged over time.

However, if either the age specific rates or the age structure change, then the SIR and SMR will change.

If the age structure changes but the underlying age-specific rates remain unchanged, then differences in calculated SIRs and SMRs will be a direct consequence of changes in the age structure of populations. The difference can be used to assess the error that might be introduced when using the indirect method to describe underlying trends (that is, trends that take account of changes in the age structure of the population).

Method

For the purposes of the assessment, the 2014 population for each PHN is assumed to have the same age-specific cancer incidence rates as the 2006 population in each of those PHNs. The only difference between the two (2006 and 2014) populations is the age-structure.

The age-specific rates evident in the 2006 population were applied to each of the age groups in the 2014 population to calculate an artificial 'observed' number of cases (that is, the

number that would have occurred in the 2014 population if age specific rates hadn't changed since 2006).

The standard rates (those for Australia in the period 2006–14) were applied to both the 2006 and 2014 population of each PHN, to calculate the expected number of cases in the population in each year.

For both the 2006 and the 2014 population of each PHN, the total observed number of cases is divided by the total expected number of cases to calculate two SIRs for the same population, with the same age specific rates but slightly different population structures.

Any difference between the two SIRs will be due to the slightly different age structure of the 2 populations separated by 8 years, and be indicative of the 'drift' or error that could be introduced in assessing the trend over the period 2006–2014.

SIRs were calculated for 2006 and 2014 assuming no change in age-specific cancer incidence rates, for 'all cancers combined', cancer of the pancreas, melanoma, lung cancer and colorectal cancer.

For those PHNs for which the trend was statistically significantly different from zero (that is, for which cancer incidence rates appear to have changed over the period), the size of the 'drift' was compared with a number of pertinent statistics, including the actual 2014 SIR, the predicted 2014 SIR (applying the slope of the trend to the 2006 SIR), and the slope itself.

Results

Table 1 describes the 2006 SIR, the 2014 (2006 adjusted) SIR, the actual 2014 SIR, the drift apparent in the method, the slope of the trend, and the drift expressed as a percentage of the slope, for 5 cancer categories (all cancers combined, pancreas, melanoma, lung and colorectal), for those PHNs where the trend was statistically significantly different from zero. (The 2014 (2006 adjusted) SIR is the SIR as it would be in 2014, if age-specific rates had remained unchanged from 2006.)

The table illustrates the drift inherent in the indirect method, and also indicates the size of the effect. Drift was typically within 1%, and at most within 4% (although drift in one outlier was 10% (pancreatic cancer in Western Queensland—but the trend was so large that the drift is very small in comparison). Similar analysis for all PHNs and all cancers assessed, irrespective of whether trends are statistically significantly different from zero, show a similar range of drift.

From Table 1, drift can constitute up to 20% of the trend. In that particular case (melanoma in Northern Queensland), the trend is for the SIR to increase by 0.18 over the eight years, while the drift is minus 0.04 (–0.04), indicating that the true trend may be closer to 0.22 over the 8 year period. Nonetheless, the trend of an increasing SIR of around 0.18 points over that period appears unlikely to be very misleading. Relative to the 2006–2014 national rates of melanoma, the SIR for this PHN increased from 1.21 in 2006 to 1.39 in 2014 (with calculated drift—or error associated with the method—sitting at around –0.04).

For all the other comparisons in Table 1, drift as a percentage of trend exceeded ten percent in about 10% of cases, and otherwise exceeded five percent in about 25% of cases, being five percent or less in 65% of cases.

Drift can be positive or negative (that is, trends based on SIRs and SMRs will either slightly overstate or slightly understate true underlying trends (that is, trends that account for changes in the age structure of the population).

It is possible that the value of the drift could be used to 'adjust' trends based on SIRs or SMRs, but this approach would need to be validated.

Table 1: Statistical drift as a percentage of trend

PHN	2006 SIR	2014 (2006 adjusted) SIR	2014 SIR (actual)	Drift (2014 SIR–2006 SIR)	Trend 2006–14	Drift as a percentage of 2006–2014 trend %
All cancers combined						
South Eastern Melbourne	1.01	1.00	0.94	–0.006	–0.09	7
Gippsland	1.00	1.00	0.97	0.001	–0.07	–1
Western Victoria	0.99	0.98	0.95	–0.004	–0.10	4
Gold Coast	0.99	0.99	1.06	0.001	0.07	2
Tasmania	1.04	1.04	0.97	–0.004	–0.10	4
Pancreas						
Nepean Blue Mountains	1.29	1.26	0.92	–0.031	–0.34	9
Hunter New England and Central Coast	0.89	0.90	1.12	0.004	0.19	2
Eastern Melbourne	0.95	0.95	1.23	0.001	0.20	1
Gold Coast	0.89	0.92	1.05	0.026	0.26	10
Western Queensland	2.16	2.26	0.40	0.104	–1.25	–8
Adelaide	0.95	0.95	1.03	0.004	0.16	2
Australian Capital Territory	0.69	0.68	1.03	–0.006	0.55	–1
Melanoma						
Northern Sydney	1.04	1.03	1.14	–0.002	0.10	–2
North Coast	1.47	1.44	1.73	–0.024	0.32	–7
Murray	1.14	1.12	0.91	–0.020	–0.20	10
Brisbane South	1.14	1.13	1.33	–0.011	0.18	–6
Gold Coast	1.19	1.18	1.62	–0.013	0.42	–3
Central Queensland and Sunshine Coast	1.27	1.25	1.56	–0.014	0.27	–5
Northern Queensland	1.21	1.18	1.39	–0.037	0.18	–20
Country SA	0.68	0.67	0.85	–0.013	0.18	–8
Northern Territory	0.62	0.60	0.82	–0.021	0.19	–11
Lung						
Central and Eastern Sydney	1.05	1.04	0.88	–0.007	–0.12	5
North Coast	1.14	1.13	1.01	–0.009	–0.13	7
Northern Queensland	1.23	1.22	1.08	–0.009	–0.12	7
Country SA	0.90	0.88	0.82	–0.019	–0.12	16
Colorectal						
Central and Eastern Sydney	0.96	0.96	0.79	–0.002	–0.14	1
Northern Sydney	1.07	1.07	0.84	0.000	–0.19	0
Western Sydney	1.10	1.10	0.87	–0.007	–0.22	3
South Eastern NSW	1.09	1.09	0.87	–0.003	–0.21	1
Western NSW	1.07	1.06	1.03	–0.010	–0.12	8
Hunter New England and Central Coast	1.11	1.12	1.02	0.004	–0.12	–3

(continued)

Table 1 (continued): Statistical drift as a percentage of trend

PHN	2006 SIR	2014 (2006 adjusted) SIR	2014 SIR (actual)	Drift (2014 SIR–2006 SIR)	Trend 2006–14	Drift as a percentage of 2006–2014 trend %
North Western Melbourne	0.99	0.99	0.86	–0.001	–0.16	0
Eastern Melbourne	1.02	1.02	0.89	0.000	–0.11	0
South Eastern Melbourne	1.05	1.04	0.83	–0.005	–0.17	3
Gippsland	1.15	1.17	0.87	0.013	–0.32	–4
Brisbane North	1.16	1.16	0.97	0.003	–0.19	–1
Brisbane South	1.09	1.10	1.00	0.011	–0.17	–7
Gold Coast	1.15	1.16	0.96	0.009	–0.19	–5
Northern Queensland	1.13	1.13	0.97	0.001	–0.15	–1
Country SA	1.07	1.07	1.05	0.002	–0.14	–1
Perth South	0.91	0.91	0.84	0.003	–0.11	–3
Tasmania	1.18	1.17	0.93	–0.005	–0.30	2
Australian Capital Territory	1.06	1.06	0.79	–0.006	–0.21	3

Source: AIHW analysis of the AIHW Australian Cancer Database.

Conclusion

Using SIRs and SMRs to describe trends in cancer incidence and mortality is less reliable than using DASRs. But in small area analysis, the reliability of DASR suffers, particularly for those less common cancers in smaller populations.

While SIRs and SMRs do not appear to be able to provide exact estimates of trend, they do seem capable of providing approximate measures of trend, which appear unlikely to be misleading, if interpreted as such (that is, as approximate).

Statistical drift associated with the method is typically around (or less than) 1% and rarely (~1 in 40) 4%—but in one case drift was 10% in a PHN with an especially high incidence (SIR>200).

In comparison with the SIR trends, the magnitude of the drift is typically less than 5% of the size of the trend, and rarely approaching 20% the size of the trend. Even at the higher end, drift appears to have little material effect on the overall conclusions that could be drawn from the presented trends.

Trends based on SIRs and SMRs should be used and interpreted carefully. They should be interpreted as approximate and indicative of general movement in the underlying age-specific rates of cancer incidence and death.

Appendix A: Data quality statements

The data quality statement for the Australian Cancer Database can be found at [Australian Cancer Database, 2014; Quality Statement](#).

The data quality statements underpinning the AIHW National Mortality database can be found on the following ABS web pages:

- ABS quality declaration summary for Deaths, Australia (ABS cat. no. 3302.0)
<http://www.abs.gov.au/ausstats/abs%40.nsf/mf/3302.0/>
- ABS quality declaration summary for Causes of death, Australia (ABS cat. no. 3303.0)
<http://www.abs.gov.au/ausstats/abs%40.nsf/mf/3303.0/>

For more information on the AIHW NMD see Deaths data at AIHW
<http://www.aihw.gov.au/deaths/aihw-deaths-data/>.

Acknowledgments

This report was prepared by Andrew Phillips, Brett Davis, Henry Wong, Justin Harvey, Mark Short and Brent Bufton, under the direction of Richard Jukes. Substantial contributions were made by David Whitelaw.

The authors would like to thank all colleagues for their contribution towards the development of the associated small area cancer data, including members of the Australasian Association of Cancer Registries and the Cancer Monitoring Advisory Group who provided expert advice and assistance.

The support of the Australasian Association of Cancer Registries through providing cancer incidence data to the AIHW's Australian Cancer Database is gratefully acknowledged.

Abbreviations

ABS	Australian Bureau of Statistics
ACD	Australian Cancer Database
AIHW	Australian Institute of Health and Welfare
CI	Confidence interval
DASR	Direct age-standardised rate
ERP	Estimated resident population
IARC	International Agency for Research on Cancer
LGA	Local government area
NMD	National Mortality Database
PHN	Primary Health Network
SA2	Statistical Areas level 2
SA3	Statistical Areas level 3
SIR	Standardised incidence ratio
SMR	Standardised mortality ratio

References

- AIHW (Australian Institute of Health and Welfare) 2017. Health risk factors in 2014–15. Cat. no. HPF 13. Canberra: AIHW. Viewed 01 October 2019, <https://www.aihw.gov.au/reports/risk-factors/health-risk-factors-in-2014-15>
- AIHW 2018. Cancer in Aboriginal and Torres Strait Islander people of Australia, selecting for males and liver cancer. Viewed 11 September 2019, <https://www.aihw.gov.au/reports/cancer/cancer-in-indigenous-australians/contents/cancer-type/liver-cancer-c22>
- AIHW 2019. Cancer in Australia 2019. Cancer series no.119. Cat. no. CAN 123. Canberra: AIHW
- Australian Government 2012. Gen Aged Care Data, population projections 2012 to 2027. Australian Government and Australian Institute of Health and Welfare. Viewed 11 September 2019, <https://webarchive.nla.gov.au/awa/20170818144806/http://www.aihw.gov.au/nacdc/population-projections/>
- Dobson AJ, Kuulasmaa K, Eberle E & Scherer J 1991. Confidence intervals for weighted sums of Poisson parameters. *Statistics in Medicine* 10:457–62.
- IARC (International Agency for Research on Cancer) 2019. Agents Classified by the IARC Monographs, Volumes 1–124. Viewed 11 September 2019, <https://monographs.iarc.fr/agents-classified-by-the-iarc/>
- Mendenhall W 1983. *Introduction to Probability and Statistics* (sixth edition). Duxbury Press, Boston.

List of tables

Table 1: Statistical drift as a percentage of trend 34

List of figures

Figure 1: Year to year variability in liver cancer incidence rates, Aboriginal and Torres Strait Islander males, 1998 to 2013..... 5

Figure 2: Comparison of the SIR with the direct age-standardised incidence rate-ratio, for males resident in each of 331 SA3s, diagnosed with any cancer in the period 2010–14, sorted by incidence 29

Figure 3: Comparison of the SIR with the direct age-standardised incidence rate-ratio, for males resident in each of 325 SA3s, diagnosed with stomach cancer in the period 2010–14, sorted by incidence..... 30

Figure 4: Calculated synthetic SIRs for 542 LGAs, having the same age-specific melanoma incidence rates as the Gold Coast PHN, standardised to rates in the North West Melbourne PHN, 2009–13, sorted by number of cases diagnosed..... 31

Related publications

This user guide and methods manual supports the release of cancer statistics for small geographic areas—specifically cancer incidence, mortality and prevalence data at <https://www.aihw.gov.au/reports/cancer/cancer-incidence-mortality-small-geographic-areas/data>.

Other related resources include:

- Cancer in Australia 2019
<https://www.aihw.gov.au/reports/cancer/cancer-in-australia-2019/contents/table-of-contents>
- Cancer data in Australia 2019
<https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/summary>
- CIMAR books
<https://www.aihw.gov.au/reports/cancer/cimar-books/contents/cimar-books>
- Cancer incidence and mortality in Australia by small geographic areas
<https://www.aihw.gov.au/reports/cancer/cancer-incidence-mortality-small-geographic-areas/data>
- Healthy Community Indicators
<https://www.aihw.gov.au/reports-data/indicators/healthy-community-indicators/national/all-australia/can/cancer>
- Cancer screening by small area
<https://www.aihw.gov.au/reports/cancer-screening/national-cancer-screening-programs-participation/contents/summary>



aihw.gov.au

AIHW

Stronger evidence,
better decisions,
improved health and welfare

