2013

Statistical methods for analysis of aggregate health performance data

Luise Lago
University of Wollongong, lago@uow.edu.au

Statistical methods for analysis of aggregate health performance data

Abstract

Presentation to the COAG Reform Council.

Keywords

Publication Details

Statistical methods for analysis of aggregate health performance data

Presentation to the COAG Reform Council
Nov 2013

Luise Lago
Senior Research Fellow (Statistics)

1

Introduction – CHSD

- **CHSD** – Centre for Health Service Development. Health services research for 20+ years. Historically included other research 'centres' e.g. Australasian Rehabilitation Outcomes Centre (AROC), National Casemix and Classification Centre (NCCC).

- **AHSRI** – Australian Health Services Research Institute. Formed in 2011. Over-arching institute headed by Kathy Eagar. Includes CHSD + other centres, including:

- **CASiH** - Centre for Applied Statistics in Health. Formed in 2011. Currently 11 staff.

- **COAG Reform Council** worked with CHSD in 2011 comparing ED performance by state and territory (Kathy Eagar, Janette Green are two of the authors).

CHSD
centre for health service development

Project Team

- **Janette Green**
  - Director of CASiH
  - Inaugural CHSD statistician
- **Megan Blanchard**
  - Statistician in CASiH
  - Previously ABS demographer and census
- **Luise Lago**
  - Senior Statistician in CASiH
  - Previously ABS survey methodologist
- **AHSRI staff**
  - Additional input from others in CHSD including Kathy Eagar and Sam Allingham (previously ABS time series).

CHSD
centre for health service development

Questions/suggestions for workshop

- Present ideas in **health context** but make more broadly applicable
- **Practical significance** - Is it possible to develop a general rule or consistent approach across different kinds of data - survey, admin, census, NAPLAN? (Issue in Education and skills reports)
- What can we do with a **time series** of 5-10 data points?
- What are the minimum number of time points needed before using **time series** methods?

CHSD
centre for health service development

Questions/suggestions for workshop

- Any others?

CHSD
centre for health service
development

Structure

- **Topic 1: Concepts and Cross-sectional methods**
  - including data types, errors
  - significance testing for survey and administrative data
- **Topic 2: Time series methods**
  - short span time series, 'formal' time series analysis, baselines
- **Topic 3: Practical significance and health context**

Each topic: ≈20m presentation of report then 20m discussion

Examples – cross sectional survey data

Example A: Risk of long-term harm from alcohol

Figure 2.8   Adults at risk of long-term harm from alcohol by sex, 2011–12
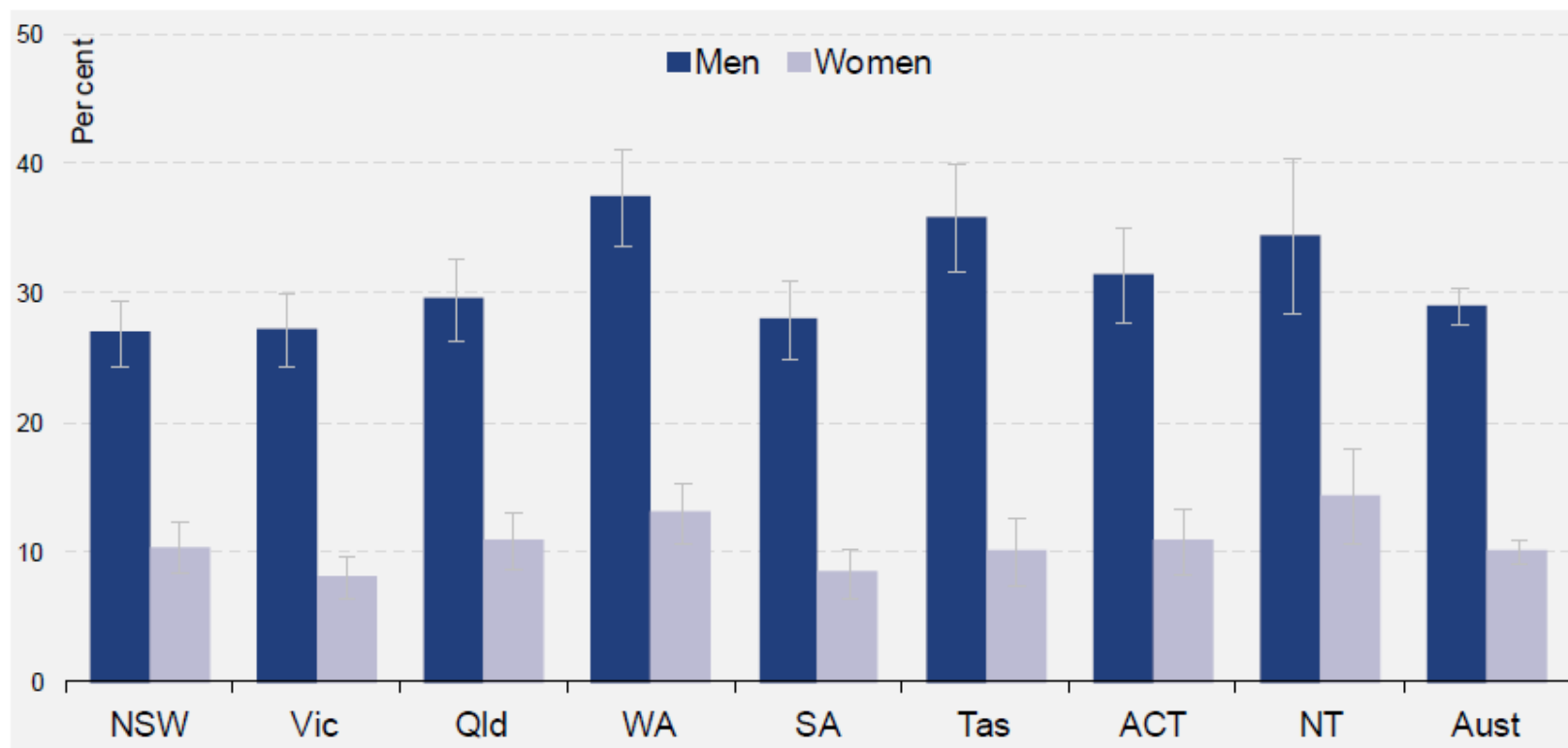
Source: COAG Reform Council HEALTHCARE 2011–12: COMPARING PERFORMANCE ACROSS AUSTRALIA (p35)
Data from ABS Australian Health Survey

Examples – time series administrative data

Example B: Waiting times for elective surgery

Figure 3.4   Waiting times for elective surgery at the 50th percentile (median)

Source: COAG Reform Council HEALTHCARE 2011–12: COMPARING PERFORMANCE ACROSS AUSTRALIA (p44)
Data from AIHW National Elective Surgery Waiting Times Data Collection

Examples – time series survey data

Example C: Waiting times for GPs

Figure 3.2 The proportion waiting 24 hours or longer for an 'urgent' appointment with a GP

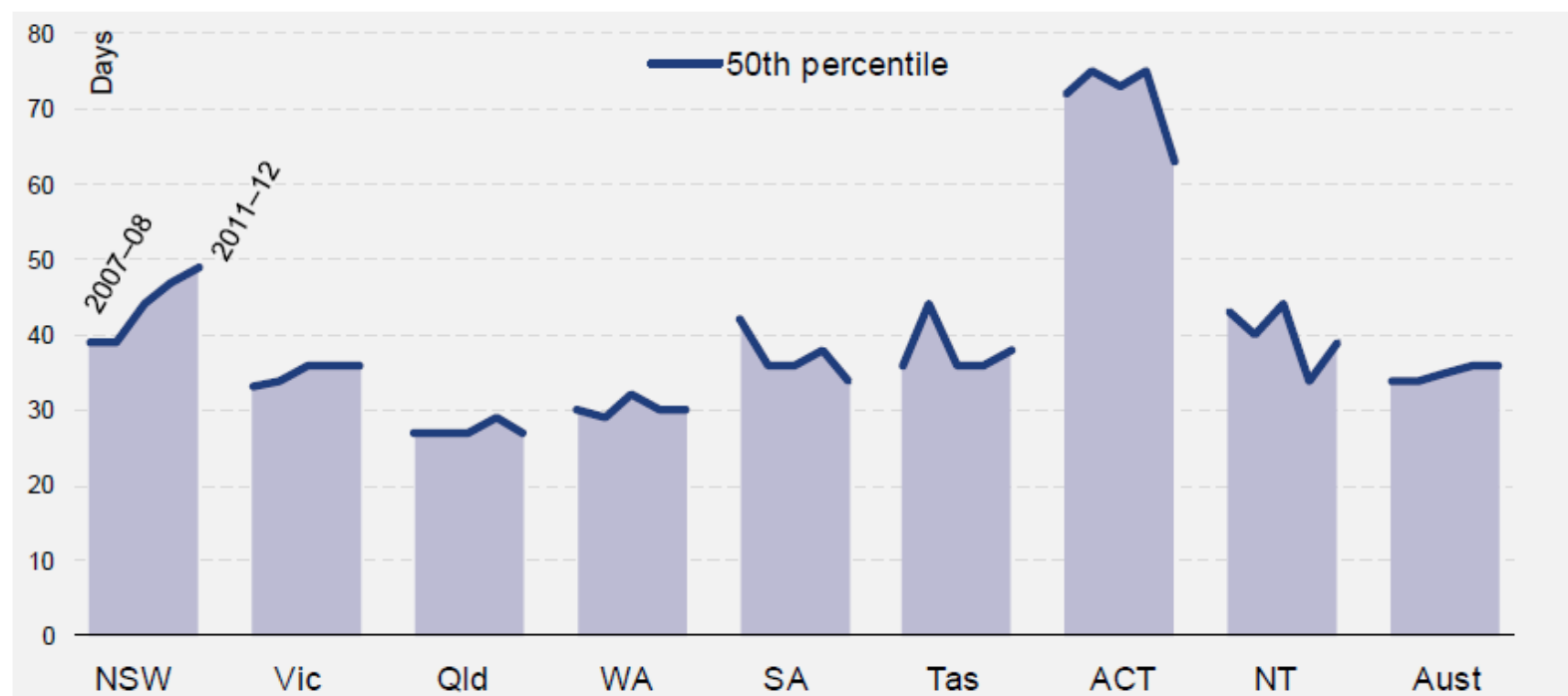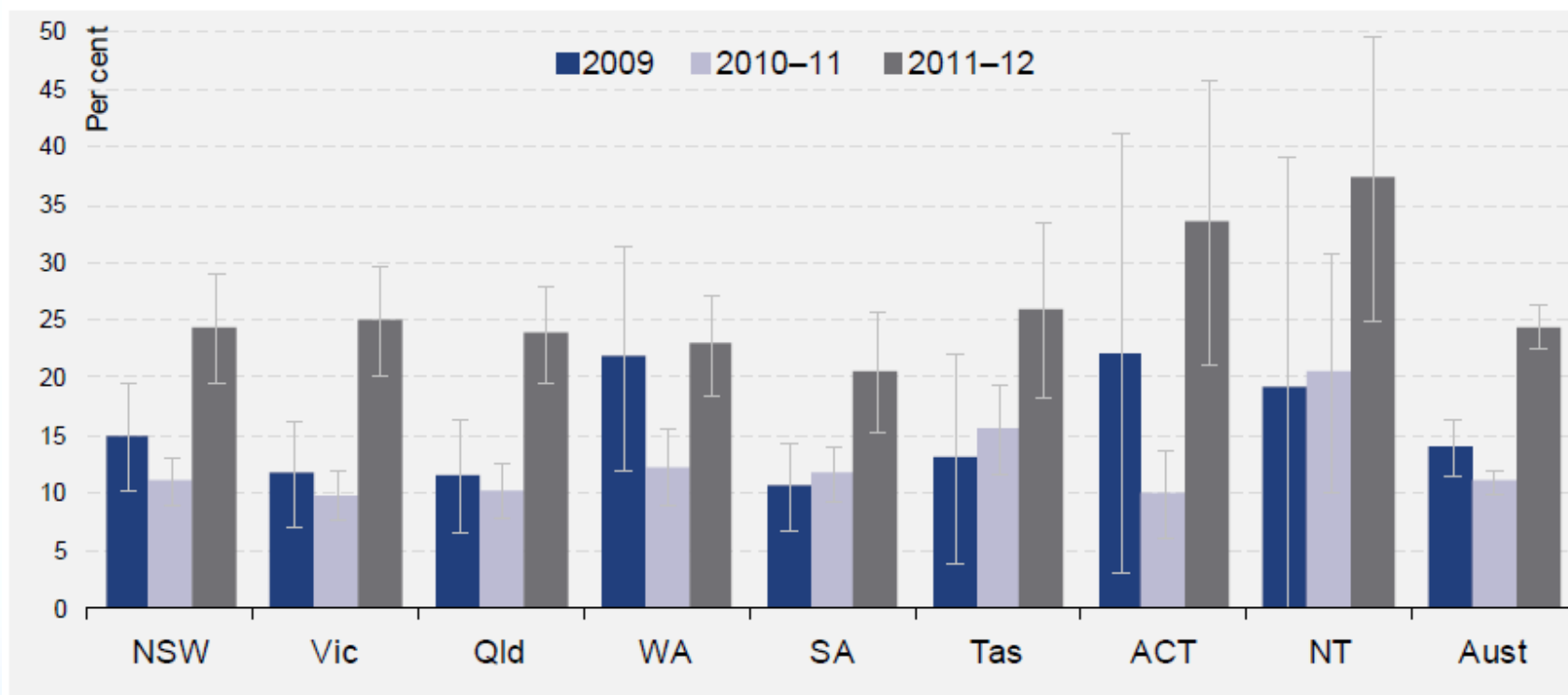

Source: COAG Reform Council HEALTHCARE 2011–12: COMPARING PERFORMANCE ACROSS AUSTRALIA (p43)

Data from ABS Patient Experience Survey

Topic 1: Concepts and Cross-sectional methods

CHSD
centre for health service
development

Concepts – types of data

Purpose
of data collection

Administrative

Often records or transactional data

e.g. births, deaths, hospital records

Statistical

Main purpose for collecting data is statistical

Population
for data collection

Entire population (census)

Subset of population (sample)

CHSD
centre for health service development

Concepts – types of data

- Example A: Risk of long-term harm from alcohol
  - Purpose: Statistical (Australian Health Survey)
  - Population: Sample survey
- Example B: Waiting times for elective surgery
  - Purpose: Administrative
  - Population: Census

Concepts – types of error
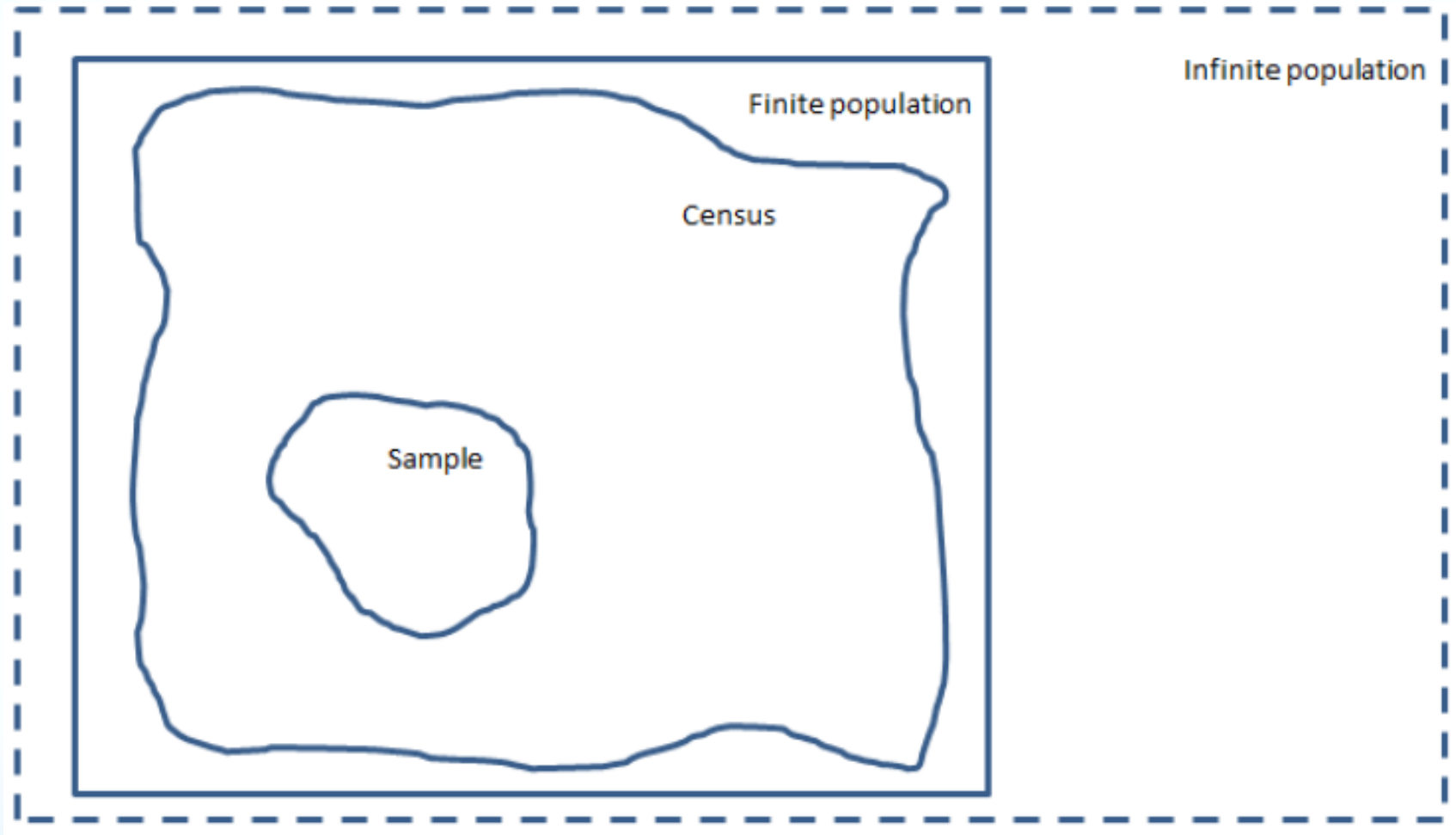
Sampling error

- Subset of population selected for survey

- Eg a sampling fraction of 1 in 100

Non-sampling error

- Coverage error

- Nonresponse error

- Reporting error

- Transcription error

Concepts – types of error

Concepts – types of error

- Example A: Risk of long-term harm from alcohol
  - Sampling error (subset of population)
  - Nonresponse bias
- Example B: Waiting times for elective surgery
  - Transcription error eg. dates entered incorrectly
  - Reporting errors such as difference in business rules across jurisdictions
- Example C: Waiting times for GPs
  - Sampling error

Cross-sectional data

- Cross-sectional data refer to observations on different individuals at a given time.

Figure 2.8   Adults at risk of long-term harm from alcohol by sex, 2011–12
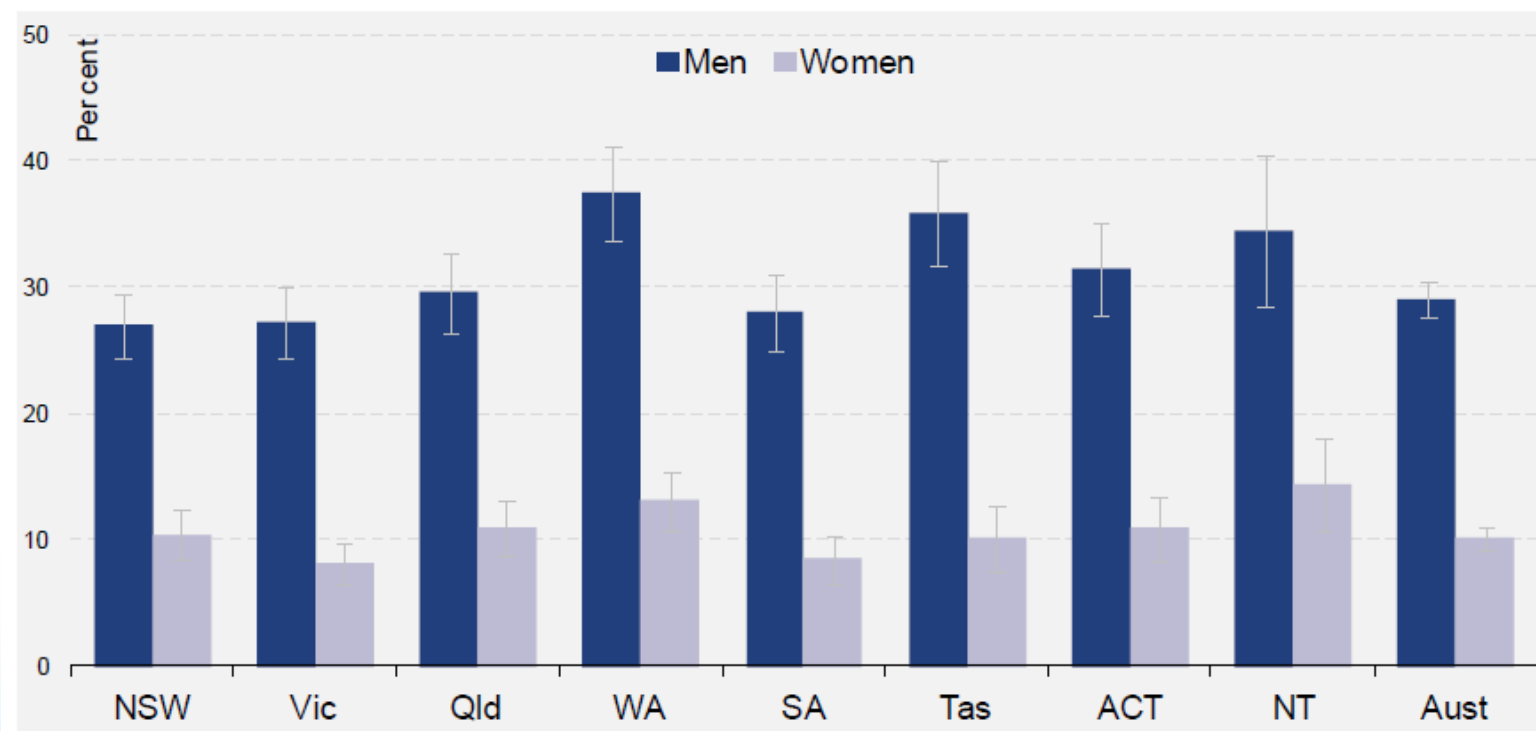
Statistical Significance

- Statistical significance measures *the probability that an observed effect occurred because of chance.*

- Carrying out a test of statistical significance requires the assumption that the outcome is a random variable, with a *probability distribution.*

  e.g. $$95\% \; CI = \bar{x} \pm 1.96 \times SE_{\bar{x}}$$

- i.e. a model needs to be specified describing the assumed distribution of the random variable

Statistical Significance

- Statistical significance can be presented in various ways including confidence intervals, hypothesis tests and p-values.

- Confidence intervals give a measure of the precision of an estimate. They account for natural variation (the same underlying conditions but a different set of data), and non-sampling error (for samples) but do not account for errors such as non-response bias.

- A hypothesis test results in a yes/no conclusion about the statistical evidence for a specific research question.

Statistical analysis steps

1. Type of analysis (cross-sectional, time series)
2. Type of data and measure of interest
3. Data presentation
4. Special features of data
5. Appropriate analysis method
6. Results assessed for statistical and practical significance

Being clear on the research question (or questions) is fundamental to each of these steps.

CHSD
centre for health service development

Research question

- The research question may be "What is the value of this health indicator and does it differ across subpopulations?" or may just be to present nationally consistent statistics for others to interrogate.

- Eg . Adults at risk of long-term harm from alcohol consumption
  - What is the national rate?
  - Is the rate higher for men than women?
  - Are there differences in the rate across states and territories?

Example – Adults at risk of long-term harm from alcohol consumption

Figure 2.8   Adults at risk of long-term harm from alcohol by sex, 2011–12
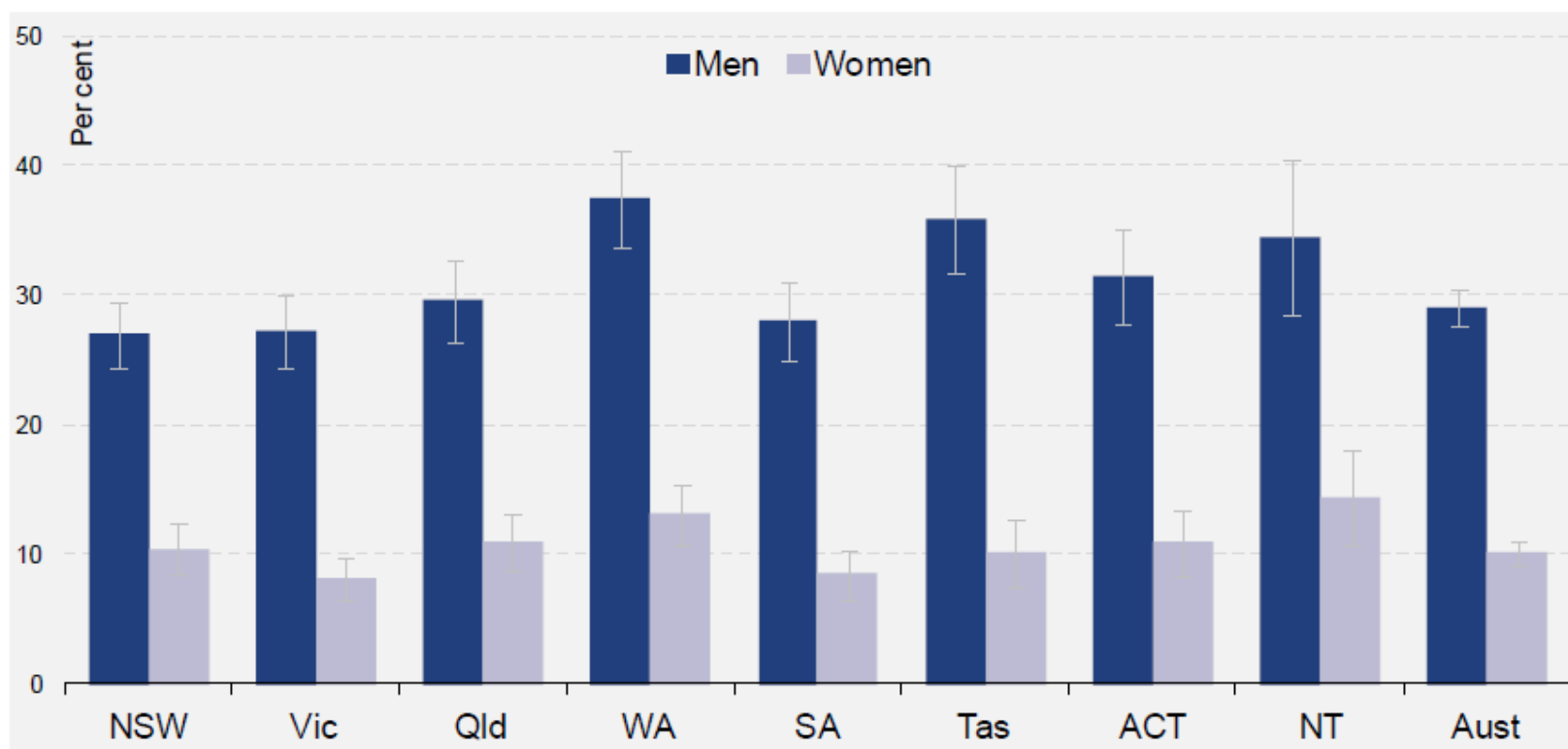
Statistical analysis steps long-term harm from alcohol

1. Type of analysis (cross-sectional, time series)
   o Questions are of a cross-sectional nature
2. Type of data and measure of interest
   o Survey data – proportion of adults who have consumed more than 2 drinks per day
3. Data presentation
   o Tabular or graphical suit a multi-faceted questions
4. Special features of data
   o Not for cross-sectional analysis. Issue with time series for NT
5. Appropriate analysis method
   o Confidence intervals suit a multi-faceted question
6. Results assessed for statistical and practical significance
   o Conclusions about differences across states can be interpreted in terms of rates and counts (eg. an additional 'n' men than women)

Statistical Significance

- Sample data should always include CI's, but they can also be used for administrative data[1].

- CI's are used when the uncertainty in an estimated is required, for example when a comparison is performed. The outcome variable is assumed to be a random variable with a probability distribution that needs to be specified appropriately for the context.

- Note: Standardised mortality ratios or other age-standardised rates which use different 'standard populations' cannot be compared using CIs[2]

1 Brilllinger, D. (1986) The Natural Variability of Vital Rates and Associated Statistics *Biometrics* 42 p693-734

2 Washington State Department of Health (2012) Guidelines for Using Confidence Intervals for Public Health Assessment

CHSD
centre for health service
development

Statistical Significance

- There are a large number of situations where a CI may be required. The two papers referenced on the previous slide give recommendations for the type of model and how to calculate a CI for commonly used health statistics e.g. age-standardised rates.

- An example follows from a recent publication where the data source is administrative, a hospital discharge database and the outcome variable is median waiting times for selected elective surgeries.

Statistical Significance

From Petrelli et al (2012)[1]

Table 4 Median waiting times and 95% CI for variables used in the study

	CABG	Angioplasty	Coronarography	Endarterectomy
Education				
None/low Education	12.0 [11.5,12.5]	8.0 [7.6,8.4]	8.0 [7.8,8.2]	28.0 [26.3,29.7]
Middle Education	12.0 [11.4,12.6]	7.0 [6.5,7.5]	7.0 [6.7,7.3]	27.0 [24.1,29.9]
High Education	10.0 [9.5,10.5]	6.0 [5.7,6.3]	7.0 [6.8,7.2]	20.0 [17.3,22.7]
Gender				
Male	11.0 [[10.7,11.4]	7.0 [6.7,7.3]	7.0 [6.8,7.2]	26.0 [24.5,27.5]
Female	11.0 [10.4,11.6]	7.0 [6.6,7.4]	7.0 [6.8,7.2]	28.0 [25.7,30.3]

1. Socioeconomic differences in waiting times for elective surgery: a population-based retrospective study *BMC Health Services Research* 2012 **12**:265 - Excerpt from Table 4

CHSD
centre for health service
development

Statistical and Practical Significance

"Significant" in a statistical sense means "not likely to happen just by chance". It does not mean "important" (Moore 1995). Statistical significance is usually expressed in terms of a significance level which is a percentage, but no matter what that percentage is, significance does not equate to importance. In a study in which the sample size is small, an important effect might not show up as being statistically significant. Conversely, an effect that is not at all important might be statistically significant in a large study. Furthermore, the significance of an effect (whether or not an important effect) can be increased by increasing the sample size of the study or through a more effective design. (ABS)

Statistical and Practical Significance

- It is important to be aware of the limitations of statistical significance:
  - Depends on sample size
  - Probability distribution is unknown, but one needs to be assumed
  - Analysis may be overly-simplistic (doesn't adjust for confounding variables)
  - What is an appropriate statistical significance level?
  - It doesn't necessarily imply practical significance.

- Further discussion of practical significance in Topic 3

Concepts - outliers

- Outliers in cross-sectional data
  - Define: eg less than (Q_1 - 1.5 * *IQR*) or greater than (Q_3 + 1.5 * *IQR*)
  - Resolve: eg remove (consider carefully for sample data as impacts on survey weights) or winsorize (set to a cut-point value)
  - If unit record data not available the data custodian should be able to describe the treatment of outlier values (if any) or interrogate any suspicious data.
- Further discussion on outliers for time series in Topic 2

CHSD
centre for health service development

Questions/Discussion

- Any questions or discussion points on concepts or cross-sectional data?

CHSD
centre for health service development

Topic 2: Time series methods

Time-series data

- Time series data are repeated, regular observations over time.

Figure 3.4   Waiting times for elective surgery at the 50th percentile (median)
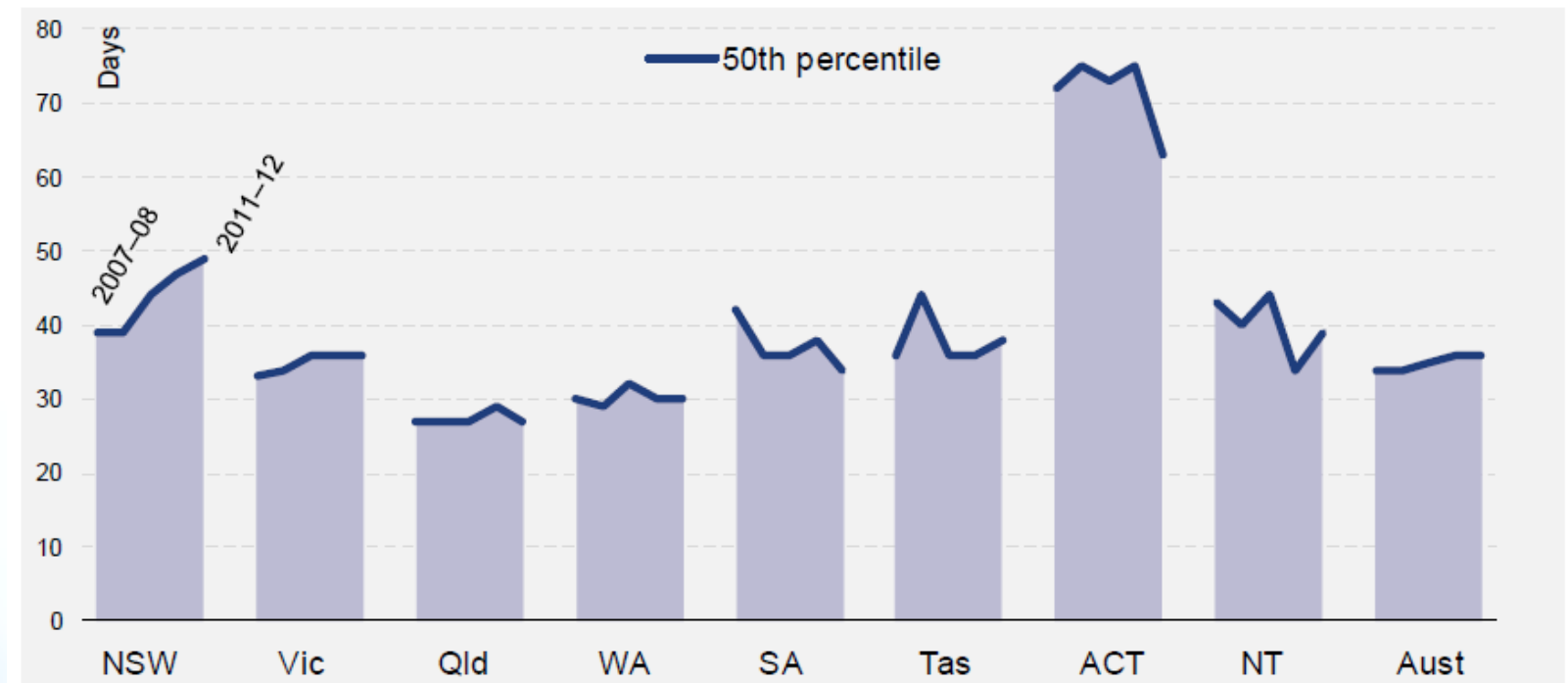
Time series types

- COAG Reform Council reports include:
  - Simple baseline comparisons (two time periods)
  - Short-span time series
  - Longer time series
  - Projections
- Frequency
  - Mostly annual, or less frequent

CHSD
centre for health service
development

Simple baseline comparisons

'Simple' baseline comparisons occur when a data item is compared over two time periods, a baseline year and a recent year:

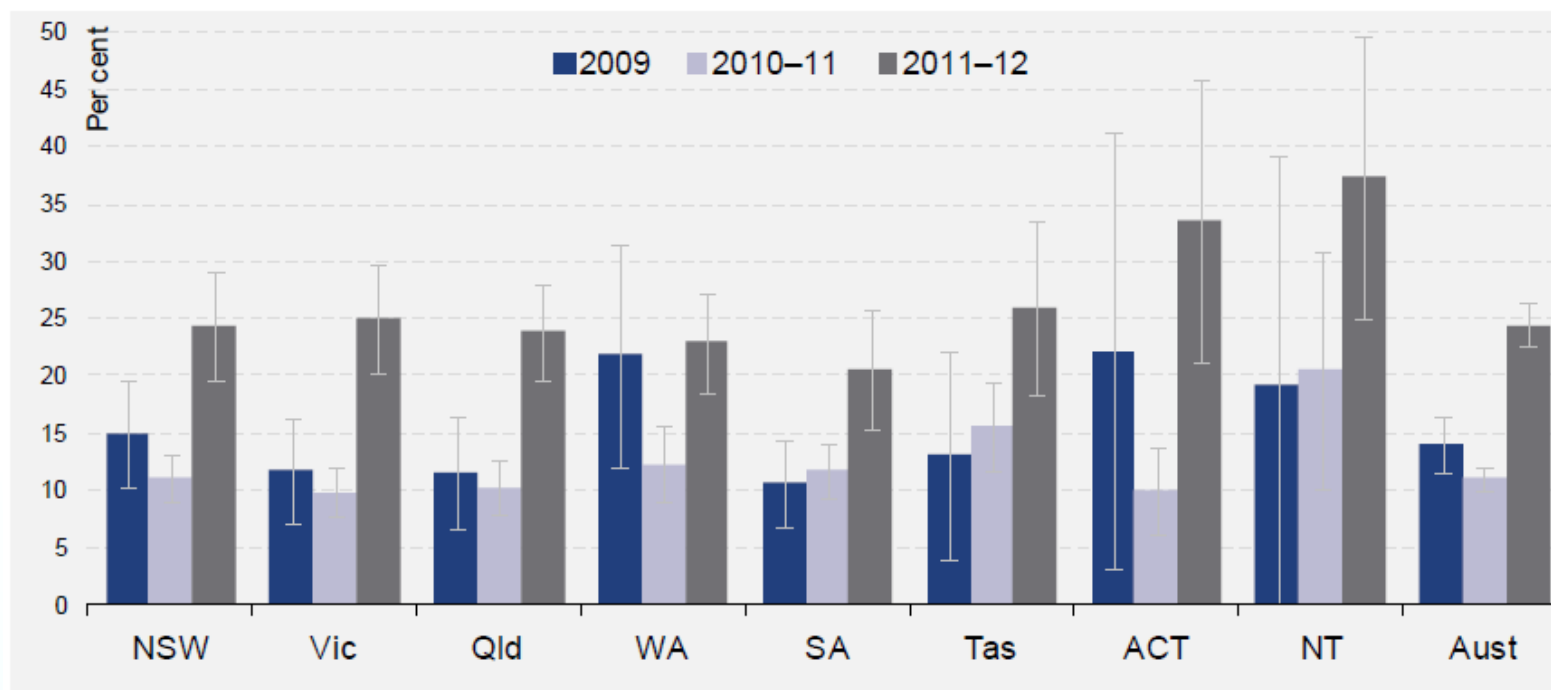Figure 2.4   Smoking rates by State and Territory, 2007–08 to 2011–12
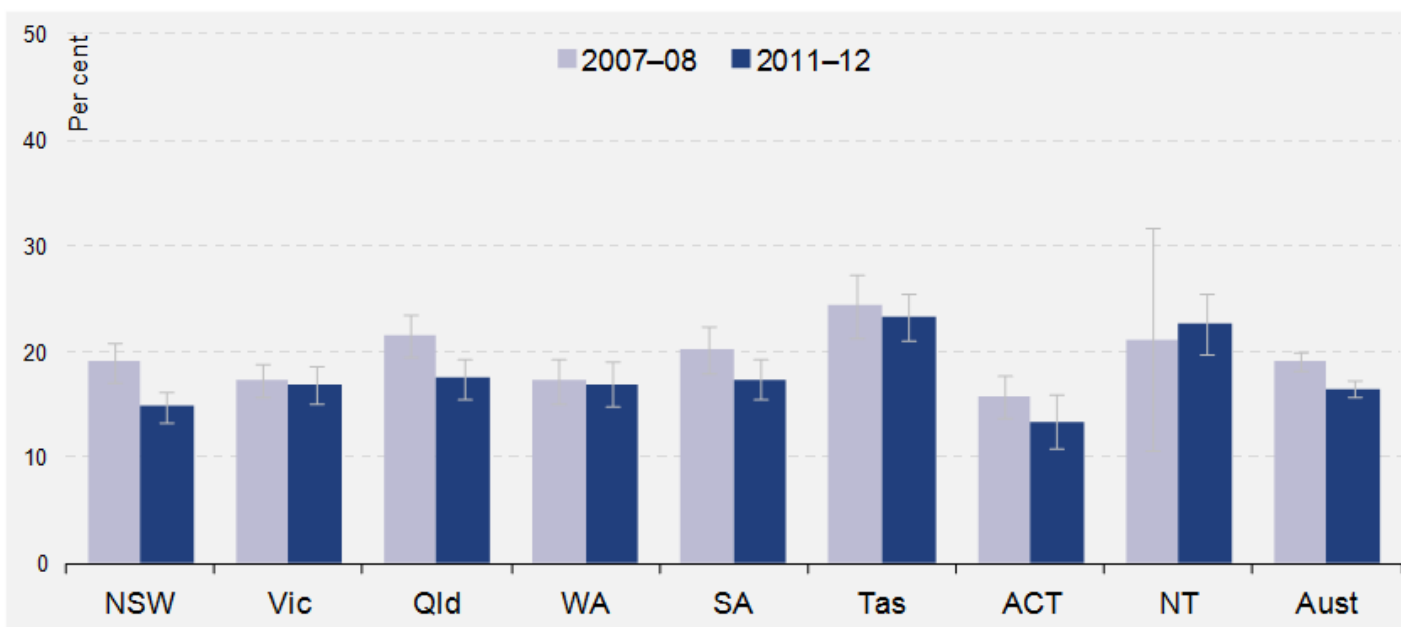
Notes:
1.   Data for the Northern Territory are not comparable over time.

Short-span time series

Short-span time series cover changes in an indicator over a small number of years:
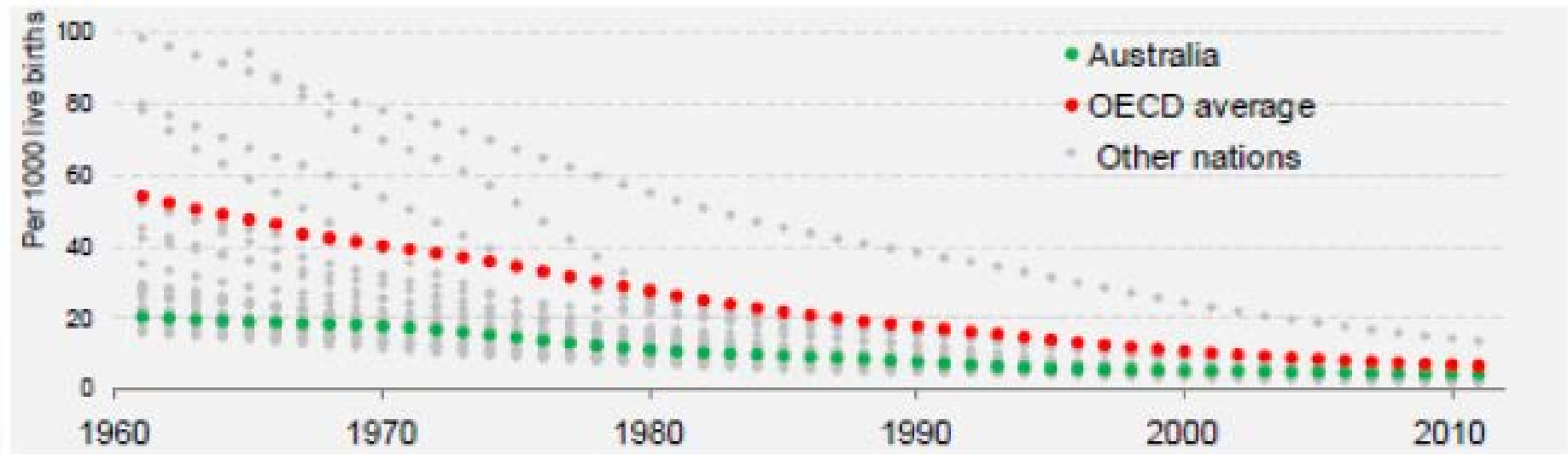

Figure 3.2   The proportion waiting 24 hours or longer for an 'urgent' appointment with a GP

Longer time series

The Council also present a small number of **longer time series** showing the value of an indicator over several years:

Figure 1.8   Infant death rates, Australia and other OECD countries since 1960

CHSD
centre for health service
development

Time series example

Projections can present future predictions based on a statistical model, or present future targets for a performance indicator:

Figure 1.4   Northern Territory death rates per 100 000 standard population 1998–2011, projections 2006–2031
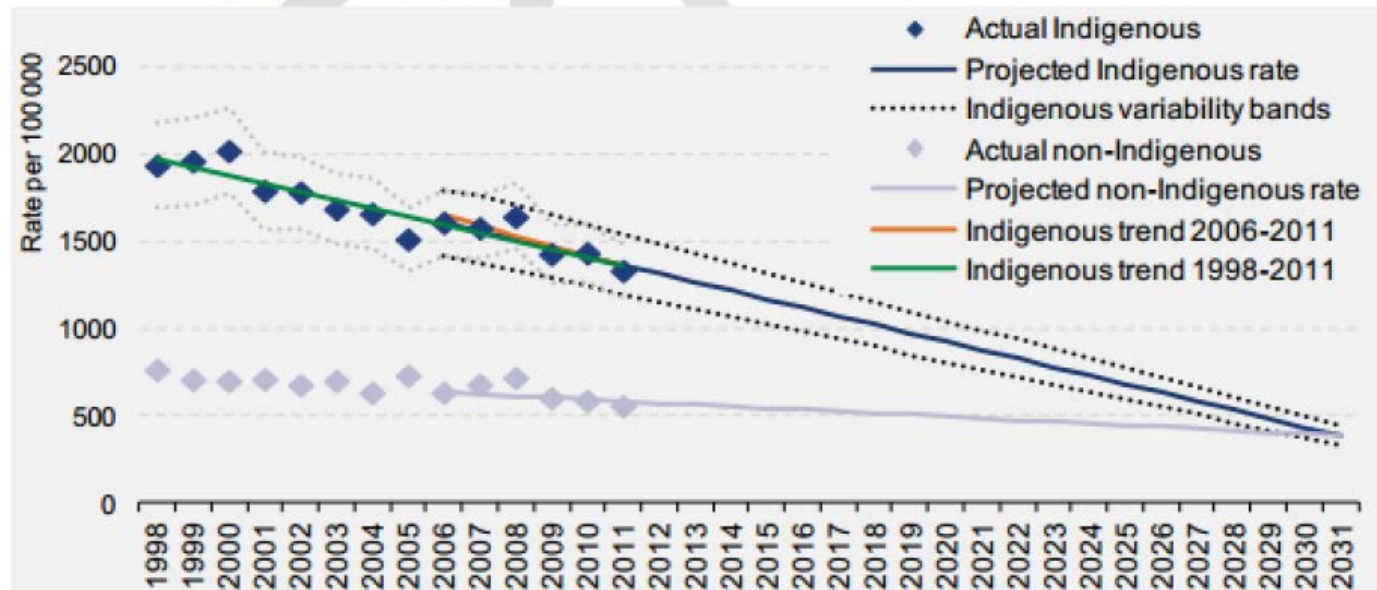
Time series research questions

- Time series analysis is primarily concerned with whether and by what magnitude a data item such as a performance indicator has changed over time.

- Possible specific research questions include:
  - Has the indicator changed since the baseline?
  - How has the indicator changed over time?
  - What is the trend over the time series?
  - What are the predicted future values of the series?

CHSD
centre for health service
development

Time series methods

- The specific method for time series analysis depends on several factors:
  - Research question
  - Time series length
  - Available data (unit record or aggregate)
  - Frequency of the data

Time series methods - Decomposition

- An *original* time series can be modelled in separate components, the *seasonal,* the *trend* and the *irregular (*decomposition).
  - Seasonal and cyclical effects: Calendar related effects (eg. winter peaks in ED presentations) and regular fluctuations that may be daily, weekly, monthly, annual (eg business cycle).
  - Irregular: one-off events unrelated to a season or cycle
  - Trend: overall 'trend' of the data over time

    (ABS Statistical Language – Time Series Data)

Time series methods - Decomposition

- Seasonally adjusted series
  - Seasonality is estimated first and removed from the series.
- Trend series
  - The irregular component is 'smoothed' out of the seasonally adjusted series leaving an estimate of trend.
- This method requires unit record data and is typically used for data with a strong seasonal pattern, usually monthly or quarterly where the outcome variable depends on factors such as number of days per month or season, and shows a predictable pattern.

Time series example

Time series example

42
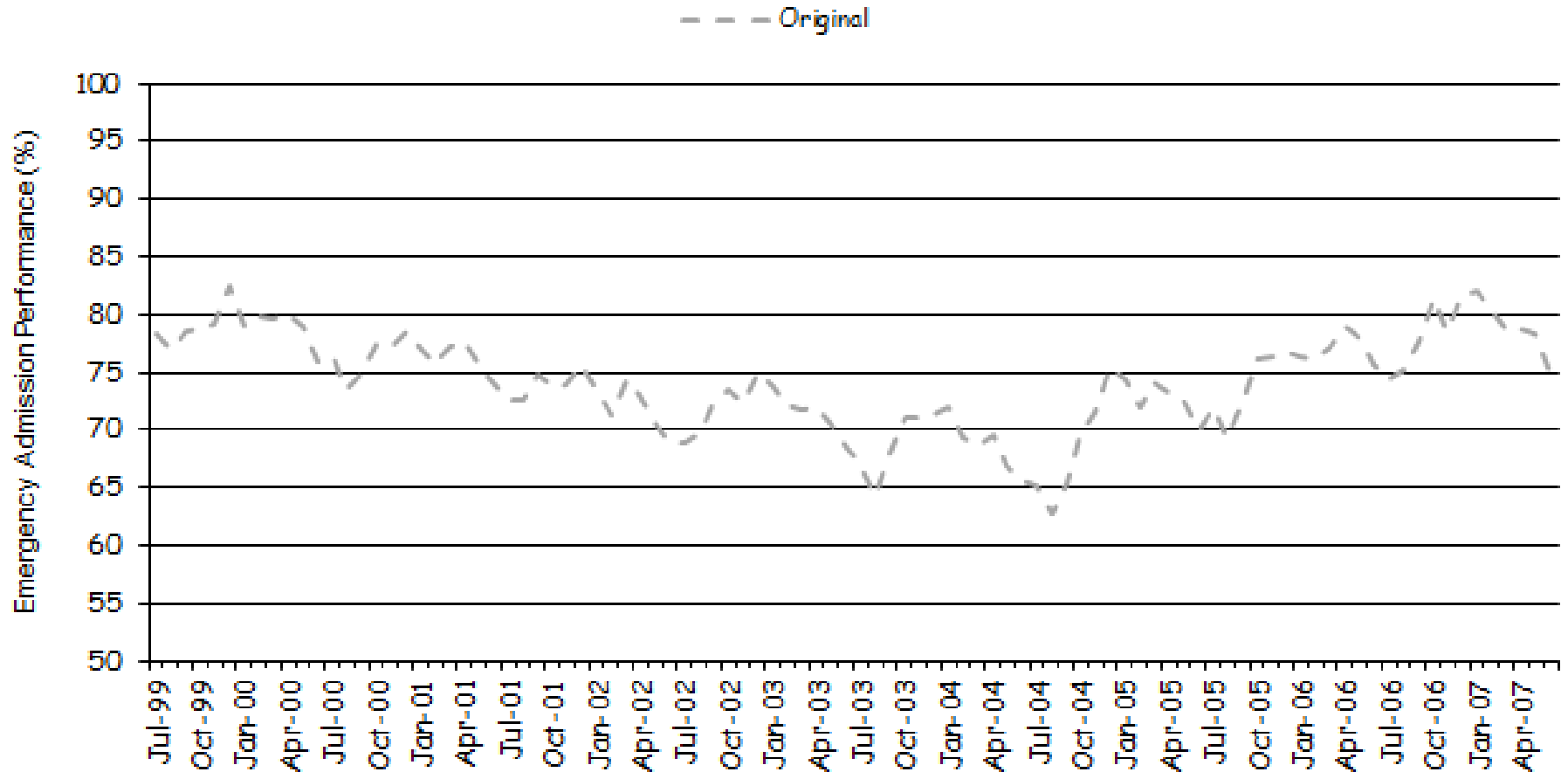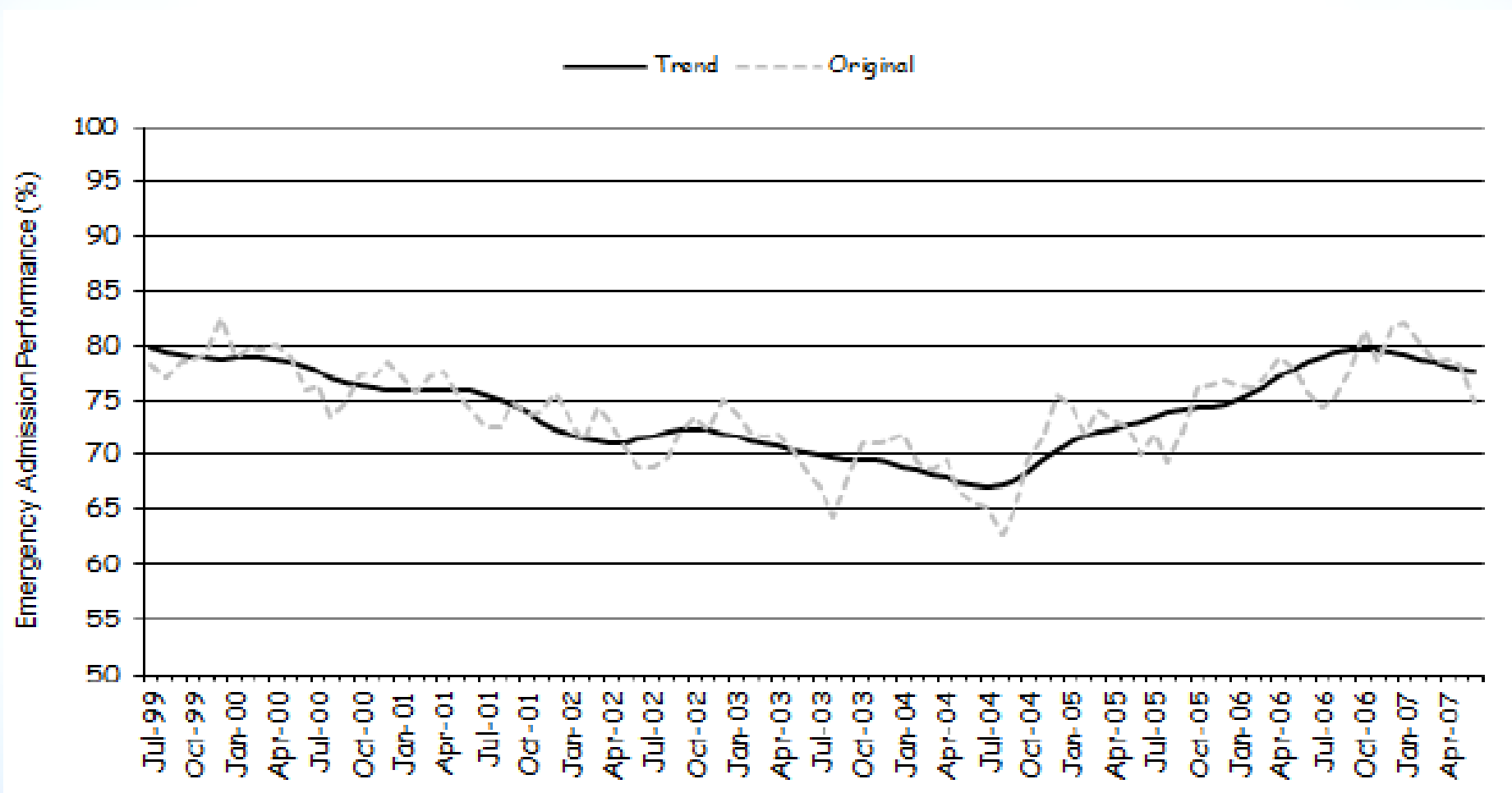
Time series example
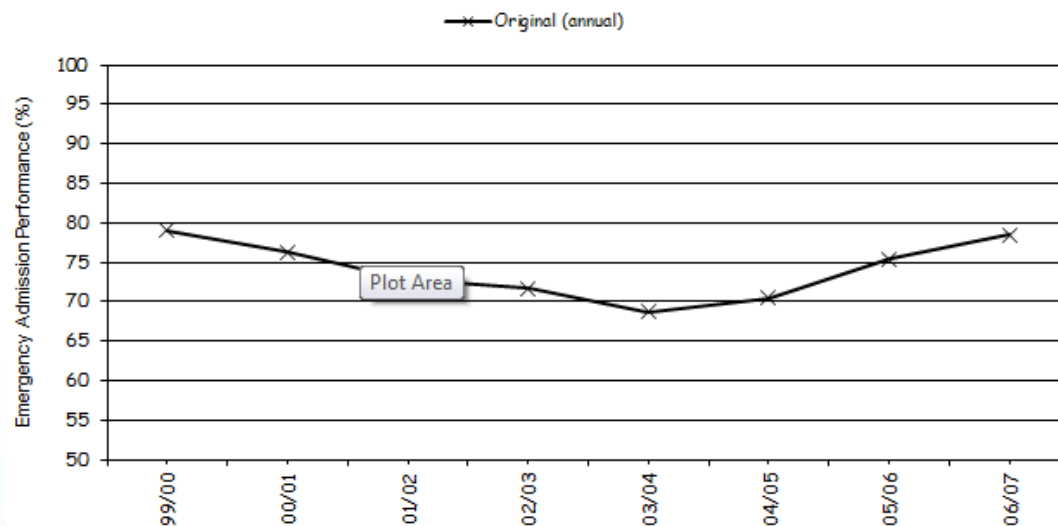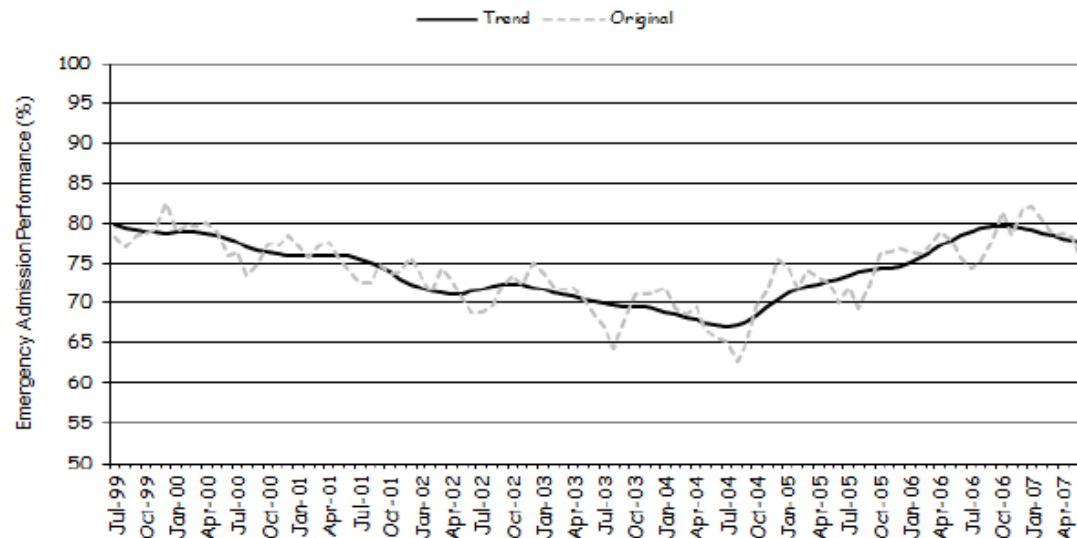
Time series example
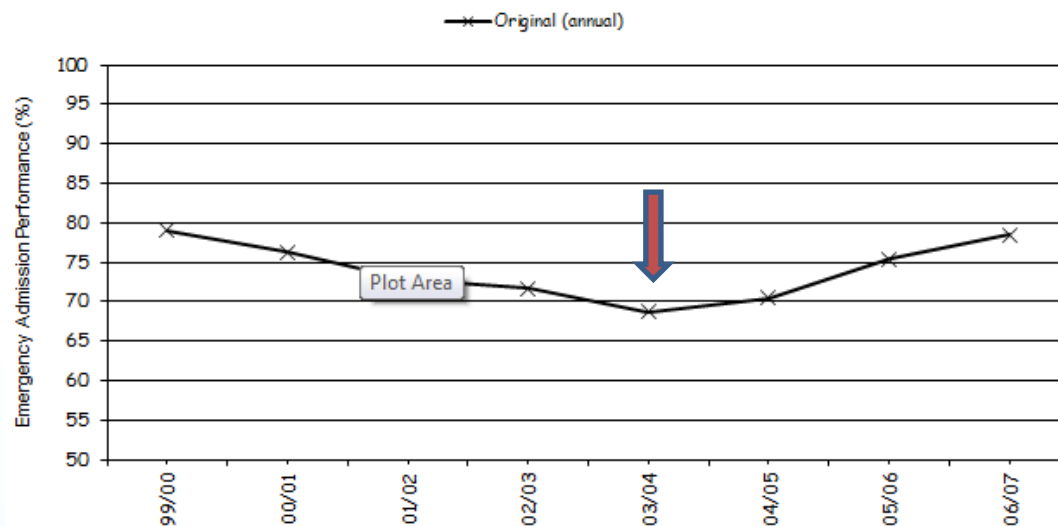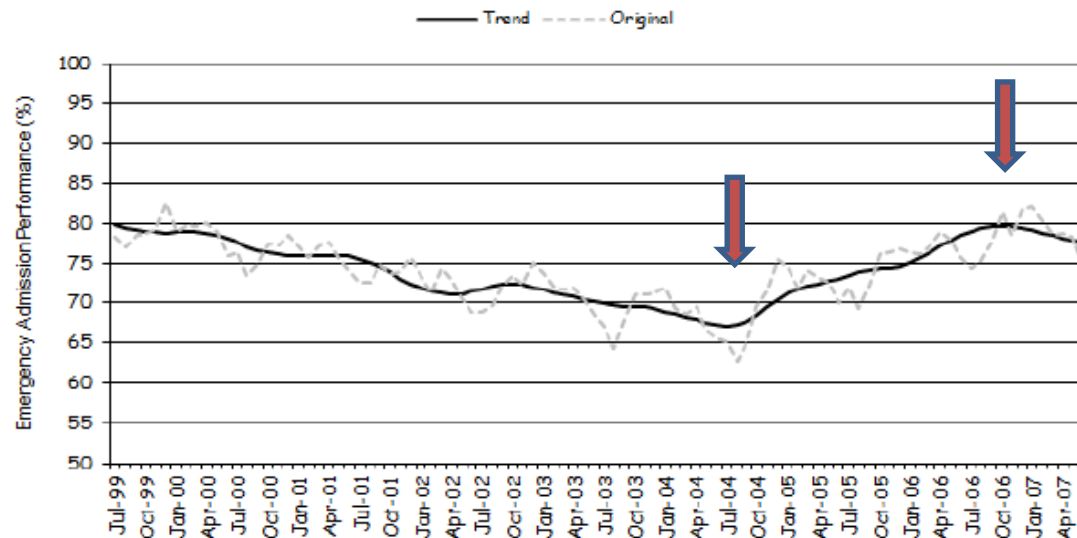
Time series methods - Decomposition

- The research question focuses on the underlying direction of series unobscured by seasonal or irregular effects.

- " The ABS recommends that at least seven years of data be used to ensure that the results of the **seasonal adjustment process** are reliable, as it can take some time for seasonal patterns to evolve. Experimental estimates are possible with fewer observations, although a minimum of five years of data is preferable." (ABS Interpreting Time Series Data)

  5 years x 4 quarters = 20 data points
  5 years x 12 months = 60 data points

CHSD
centre for health service
development

Time series methods - smoothing

- Smoothing is a method of analysis when the trend is of interest but seasonality is not strong.

- Moving averages are one method of smoothing that can be used as an alternative to linear trend lines to provide a summary of the direction of the series with the irregular component's contribution reduced.

- Example for annual mortality rates in Alberta:

Time series example

Figure 14-3: Age-Standardized Incidence Rates (ASIRs)[†][*] and Mortality Rates (ASMRs)[*][†][‡] and 95% Confidence Intervals (CI) for Childhood Cancer, Ages 0-14, Both Sexes Combined, Alberta, 1990-2010
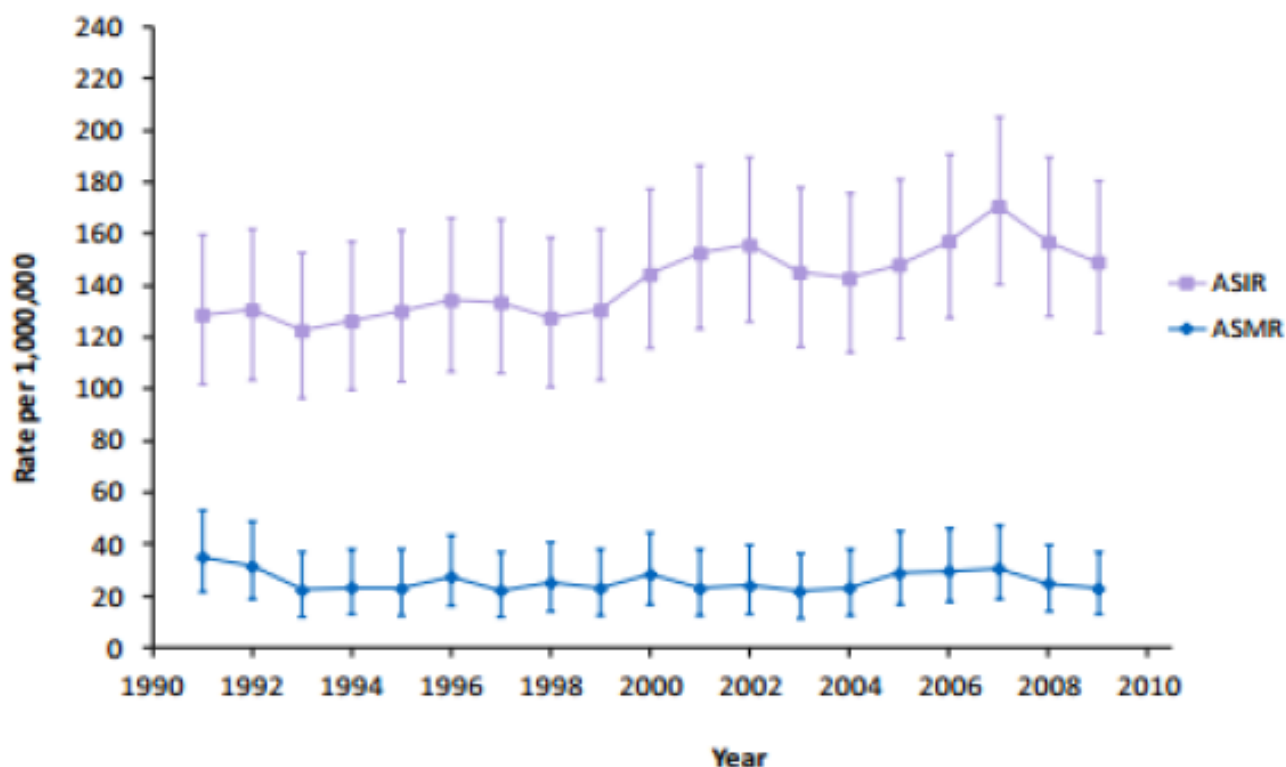
[*] Three year moving averages
[†] Standardized to 1991 Canadian Population;
[‡] ASIRs and ASMRs are rates per 1,000,000

CHSD
centre for health service
development

47

Time series – outliers

- Time series
  - Define: If the series is of an adequate length calculate the IQR of the aggregate series, and determine the distance of any possible outliers outside the IQR as a factor.
  - Resolve: eg correct value, or make note of in interpretation of trend

Time series – selecting a baseline measure

- A baseline measure is selected as a point of comparison for future values, and can be used to set targets for performance improvement.

- A baseline value may be set at 0, be a current value, a historical value, an arbitrary value or an average of historical values.

CHSD
centre for health service
development

Guidelines for selecting a baseline measure

- Look at historical data for the time series to determine if a potential baseline is typical of the trend.

- When unit record data are available, measure the interquartile range over recent periods to a proposed baseline, to assess it's likelihood of being an atypical value, or outlier.

Guidelines for selecting a baseline measure (cont)

- Consider using an average of estimates covering a small number of time points, such as a moving average as described in section 3.4.5. This is particularly important for time series with high noise to trend ratios, which may be seen for example in series with small counts.

- Scrutinise data quality statements associated with a potential baseline year to ensure that there are no concerns with the data contributing to estimates for this time point

- Consider using a historical value, for example 2 years prior, as the additional year of recent data can provide support for the baseline being a typical value.

CHSD
centre for health service
development

Time series issues – What to do with 5-10 obs?

- What is the research question?
  - Understanding trend?
    - 3 point moving average would be more meaningful at closer to 10 time points (end point problem).
  - Assessing differences from one time to another?
    - CI or hypothesis test
- What are the characteristics of the data
  - Seasonality present?
- What are data are available?
  - Unit record data can allow for more sophisticated methods such as simulation to remove irregular component.

Discussion / questions

- Any questions or discussion on time series methods?

CHSD
centre for health service
development

Topic 3: Practical significance and health context

Practical significance

- The goal of statistical inference is to discover if there is a **real** difference between the populations underlying two data sets being compared.

- Its aim is not only to find if the two data sets produce a different result, as they almost always would.

Practical significance and health data

- Health data sets
  - generally large
  - differences are often statistically significant, even when small.

- Statistically significant differences
  - are unlikely to be different due to random chance.
  - should also be tested for practical significance, to decide if they are important.

CHSD
centre for health service
development

Practically significant differences

- A difference that is practically significant is a difference that is meaningful or important enough to cause you to change your behaviour or thinking, **or to take action**.

- Example:
  - Rate of potentially preventable hospitalisations may be statistically significantly different between NSW and Victoria.
  - Because states have large populations, this result can occur when the difference is very small.
  - There may be no justification for implementing large scale strategies to address this difference.

CHSD
centre for health service development

Practical significance

- There is a danger in applying strict rules to a decision about practical significance, the same applies to statistical significance.

- Vaske (2002) "Practical significance must always involve a value judgement made by both the researcher and the consumer of research about the theoretical and applied implications of a study's findings".

Practical significance methods - proportions

- First step - **gain an understanding of the size of the difference** that has been found, in terms of the underlying data.
- A difference in proportions can be converted to a difference in absolute numbers to assess the practical significance of the difference.
  - e.g. If the proportion of men at risk of long-term harm from alcohol is higher than the national proportion for one state
  - Difference in proportions can be calculated by subtracting one from the other.
  - This can be converted to an actual number, using the population of that state.
- Practical implications of this difference can be made with experts from the field.

Practical significance methods - proportions

- Another method[1] is to **first seek expert guidance** on what a non-acceptable difference in proportions would be.

- Example:
  - Expert advice was 3% is important.
  - Calculate a CI for the difference in proportions in the data.
  - If this interval lies is entirely above 3%, then the result is not practically significant.

Rowe (2007) *Essential statistics for the pharmaceutical sciences.* John Wiley & Sons

Practical significance methods - means

- Report **effect size** and statistical significance.
- Effect size can be interpreted as the standardised difference between the two groups.
  - A commonly used measures is Cohen's d statistic
  - Difference between the group means, divided by the pooled standard deviation.
- Guidelines (Cohen) for interpretation:
  - Less than 0.2 small effect,
  - Close to 0.5 medium effect
  - At least 0.8 is a large effect.
- Example of methods for rates in CHSD report (p36).

CHSD
centre for health service development

Practical significance: accuracy

- Differences need to be interpreted considering the accuracy in the data item.
- For example, waiting times in the ED:
  - Captured in whole minutes.
  - If a statistically significant difference found with absolute difference < 1 minute could not be considered to be practical significant.
- Such a difference could have no credibility when the data are not even captured at this level of accuracy.

Practical significance summary

- Expert advice can be sought to determine an unacceptable difference

- Context and expert opinion are crucial

- Statistical testing should incorporate an assessment of practical significance.

CHSD
centre for health service
development

Health context

1. Develop a list of possible contextual factors
2. Refine the list of possible contextual factors to a shortlist
3. Carry out analysis of the shortlisted contextual variables
4. Consult with key stakeholders
5. Report using contextual factors

CHSD
centre for health service development

1. Possible contextual factors

- **Develop a list of possible contextual factors** relevant to each indicator with consideration of the following issues:
  - Data must be available (or able to be requested) on the contextual factor to allow the Council to report on it (n*ecessary*)
  - Evidence is available demonstrating the contextual factors effects the performance indicator. This may be from previous reports, other sources such as literature (including grey literature) or other expert advice (*strongly desirable)*
  - The possible contextual factors is not a long list of factors of minor importance which would distract from the key messages (*strongly desirable)*

2. Refine contextual factors

Refine the list of possible contextual factors to a shortlist by:

- A literature review or discussion with experts on the contextual factors relevant for this indicator.

- Considering the implications of adding each contextual factor, such as its relevance and whether the factor is outside the control of the jurisdiction (such as remoteness).

- By carrying out statistical analysis, or request the data custodian to carry out statistical analysis, to assess the contribution of the factor to explaining variation in the indicator and the interaction between the factors.

- A combination of the above, such as by contracted research as in Emergency Department performance (Eagar et al 2011).

CHSD
centre for health service
development

3. Analysis

Carry out analysis of the shortlisted contextual variables via cross-tabulations or other exploratory analysis to assess the indicator on possible subgroups.

- Consider the statistical and practical significance of any differences

- Assess the relative importance of each of the contextual factors and whether their inclusion adds to the understanding of performance

- Determine whether the contextual factor can be incorporated via an adjustment (such as age-standardisation and casemix adjustment) or via sub-population analysis.

Consultation with key stakeholders such as data custodians and jurisdictions.

- Does the data quality support this level of analysis?

- Does the further information provided by the breakdown impact on performance targets?

5. Reporting on contextual factors

Report on contextual factors for this indicator.

- Determine whether the report requires separate analysis be presented by these contextual factors or whether a discussion will be sufficient to convey the findings of the analysis (eg. when there are no differences)

- If statistical analysis will be presented for the contextual factor, present using simple graphics to show clear 'distinctions' of the subgroups.

Examples of contextual factors

Examples of potentially influential contextual factors that should be considered when making comparisons across jurisdictions:

Table 4   *Examples of contextual factors for health system performance indicators*

Performance Indicator	Figure reference	Data Source	Examples of key contextual factors
Health System			
Waiting times for GPs	Fig 3.2	ABS Patient Experience Survey	Remoteness Socioeconomic Status Bulk billing rates No. GPs per capita
Elective Surgery waiting times	Fig 3.4	AIHW National Elective Surgery Waiting Times Data Collection	Elective surgery urgency profile Public/private mix (hospitals and patients) Casemix complexity Wait times by Service Related Group
Emergency Department waiting times	Fig 3.7	AIHW National Non-admitted Patient Emergency Department Care database	Triage category profile Hospital peer group

Excerpt from p31 of CHSD report

Questions / Discussion

- Any question or discussion for issues of practical significance or health system context?

CHSD
centre for health service development