

APPENDIX 2: STATISTICAL SIGNIFICANCE TESTING AND ITS USE IN THIS REPORT

Understanding Statistical Significance Testing

Inferential statistics are used when a researcher wishes to use a sample to draw conclusions about the population as a whole (e.g. weighing a class of 10 year old boys, in order to estimate the average weight of all 10 year old boys in New Zealand). Any measurements based on a sample however, even if drawn at random, will always differ from that of the population as a whole, simply because of chance. Similarly, when a researcher wishes to determine whether the risk of a particular condition (e.g. lung cancer) is truly different between two groups (smokers and non-smokers), they must also consider the possibility that the differences observed arose from chance variations in the populations sampled.

Over time, statisticians have developed a range of measures to quantify the uncertainty associated with random sampling error (e.g. to quantify the level of confidence we can have that the average weight of boys in our sample reflects the true weight of all 10 year old boys, or that the rates of lung cancer in smokers are really different to those in non-smokers). Of these measures, two of the most frequently used are:

P values: The p value from a statistical test tells us the probability that we would have seen a difference at least as large as the one observed, if there were no real differences between the groups studied (e.g. if statistical testing of the difference in lung cancer rates between smokers and non-smokers resulted in a p value of 0.01, this tells us that the probability of such a difference occurring if the two groups were identical is 0.01 or 1%. Traditionally, results are considered to be statistically significant (i.e. unlikely to be due to chance) if the probability is <0.05 (i.e. less than 5%) [1].

Confidence Intervals: A 95% Confidence Interval suggests that if you were to repeat the sampling process 100 times, 95 times out of 100 the confidence interval would include the true value. In general terms, if the 95% confidence intervals of two samples overlap, there is no significant difference between them (i.e. the p value would be ≥ 0.05), whereas if they do not overlap, they can be assumed to be statistically different at the 95% confidence level (i.e. the p value would be <0.05) [1].

The Use of Statistical Significance Testing in this Report

In the preparation of this report a large range of data sources were used. For the purposes of statistical significance testing however, these data sources can be considered as belonging to one of two groups: Population Surveys and Routine Administrative Datasets. The relevance of statistical testing to each of these data sources is described separately below:

Population Surveys: A number of indicators in this report utilise data derived from national surveys (e.g. 2011/12 New Zealand Health Survey), where information from a sample has been used to make inferences about the population as a whole. In this context statistical significance testing is appropriate, and where such information is available in published reports, it has been incorporated into the text accompanying each graph or table (i.e. the words *significant*, or *not significant* in italics are used to imply that a test of statistical significance has been applied to the data and that the significance of the associations are as indicated). In a small number of cases however information on statistical significance was not available in published reports, and in such cases any associations described do not imply statistical significance.

Numbers and Rates Derived from Routine Administrative Data: A large number of the indicators in this report are based on data derived from New Zealand's administrative datasets (e.g. National Minimum Dataset, National Mortality Collection), which capture



information on all of the events occurring in a particular category. Such datasets can thus be viewed as providing information on the entire population, rather than a sample and as a consequence, 95% confidence intervals are not required to quantify the precision of the estimate (e.g. the number of leukaemia deaths in 2003–2007 although small, is not an estimate, but rather reflects the total number of deaths during this period). As a consequence, 95% confidence intervals have not been provided for any of the descriptive data (numbers, proportions, rates) presented in this report, on the basis that the numbers presented are derived from the total population under study.

Rate Ratios Derived from Routine Administrative Data: In considering whether statistical significance testing is ever required when using total population data Rothman [2] notes that if one wishes only to consider descriptive information (e.g. rates) relating to the population in question (e.g. New Zealand), then statistical significance testing is probably not required (as per the argument above). If, however, one wishes to use total population data to explore biological phenomena more generally, then the same population can also be considered to be a sample of a larger super-population, for which statistical significance testing may be required (e.g. the fact that SIDS in New Zealand is 10 times higher in the most deprived NZDep areas might be used to make inferences about the impact of the socioeconomic environment on SIDS mortality more generally (i.e. outside of New Zealand, or the 5 year period concerned)). Similarly, in the local context the strength of observed associations is likely to vary with the time period under study (e.g. in updating 5-year asthma admission data from 2004–2008 to 2005–2009, rate ratios for Pacific children are likely to change due to random fluctuations in annual rates, even though the data utilised includes all admissions recorded for that particular 5-year period). Thus in this report, whenever measures of association (i.e. rate ratios) are presented, 95% confidence intervals have been provided on the assumption that the reader may wish to use such measures to infer wider relationships between the variables under study [2].

The Signalling of Statistical Significance in this Report

In order to assist the reader to identify whether tests of statistical significance have been applied in a particular section, the significance of the associations presented has been signalled in the text with the words *significant*, or *not significant* in italics. Where the words *significant* or *not significant* do not appear in the text, then the associations described do not imply statistical significance or non-significance.

References

1. Webb P, Pirozzo S. 2005. Essential Epidemiology: An Introduction for Students and Health Professionals. Cambridge: Cambridge University Press.
2. Rothman K. 2002. Epidemiology: An Introduction. New York: Oxford University Press.

