

APPENDIX 1 : STATISTICAL SIGNIFICANCE TESTING AND ITS USE IN THESE INDICATORS

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 (i.e. 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 these Indicators

In the development of these indicators, a range of data sources were used. For the purposes of statistical significance testing, 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 is described separately below:

Population Surveys: A number of indicators utilise data derived from national surveys (e.g. the 2004 Living Standards 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 number of indicators are based on data derived from New Zealand's administrative data sets (e.g. Hospital Admissions, Mortality), 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 drowning deaths in 2002-2006, 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, 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 SUDI 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 SUDI 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 2004-2008, 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 these indicators, 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 order to assist the reader to identify whether tests of statistical significance have been applied in a particular section, the *Data Sources and Methods* text box accompanying each indicator includes a small paragraph entitled *Statistical Significance Testing* (see examples below). It is suggested the reader briefly reviews this information before considering the analyses presented in the sections which follow.

Data Sources and Methods

Statistical Significance Testing Example

Note: Tests of statistical significance (in the form of 95% confidence intervals) have been applied to some of the data in this section. Where relevant, the significance of these associations has been signalled in the text (with the words *significant*, or *not significant* in italics being used to denote the statistical significance of the observed association). Where the words *significant* or *non-significant* do not appear in the text, then the associations described do not imply statistical significance or non-significance.



APPENDIX 2: NATIONAL MORTALITY COLLECTION

Mode of Data Collection

The Mortality Collection is a dataset managed by the New Zealand Health Information Service (NZHIS), which classifies the underlying cause, for all deaths registered in NZ since 1988. Fetal and infant data is a subset of the Mortality Collection and contains extra information on factors such as birth weight and gestational age [3].

Each month Births, Deaths and Marriages send NZHIS electronic death registration information, Medical Certificates of Cause of Death and Coroner's reports. Additional information on the cause of death is obtained from the National Minimum Dataset (NMDS), private hospital discharge returns, the NZ Cancer Registry (NZCR), the Department of Courts, the Police, the Land Transport Authority, Water Safety NZ, Media Search and from writing letters to certifying doctors, coroners and medical records officers in public hospitals. Using information from these data sources, an underlying cause of death (ICD-9 and ICD-10) is assigned by NZHIS staff according to the World Health Organisation's rules and guidelines for mortality coding [3].

Data Quality Issues Relating to the Mortality Collection

Unlike the NMDS, where information on the principal diagnosis is coded at the hospital level and then forwarded electronically to the NZHIS, for the Mortality Collection each of the approximately 28,000 deaths occurring in NZ each year is coded manually within NZHIS. For most deaths the Medical Certificate of Cause of Death provides the information required, although coders also have access to the information contained in the NMDS, NZ Cancer Registry, LSTA, Police, Water Safety NZ and ESR [4]. As a consequence, while coding is still reliant on the accuracy of the death certificate and other supporting information, there remains the capacity for a uniform approach to the coding which is not possible for hospital admission data.

While there are few published accounts of the quality of coding information contained in the Mortality Collection, the dataset lacks some of the inconsistencies associated with the NMDS, as the process of death registration is mandated by law and there are few ambiguities as to the inclusion of cases over time. As a consequence, time series analyses derived from this dataset are likely to be more reliable than that provided by the NMDS.



APPENDIX 4: THE NATIONAL MINIMUM DATASET

Mode of Data Collection

The National Minimum Dataset (NMDS) is New Zealand's national hospital discharge data collection and is maintained by the New Zealand Health Information Service (NZHIS). The information contained in the dataset has been submitted by public hospitals in a pre-agreed electronic format since 1993. Private hospital discharges for publicly funded events (e.g. births, geriatric care) have been submitted since 1997. The original NMDS was implemented in 1993, with public hospital information back loaded to 1988 [5]. Information contained in the NMDS includes principal and additional diagnoses, procedures, external causes of injury, length of stay and sub-specialty code and demographic information such as age, ethnicity and usual area of residence.

Dataset Quality and Changes in Coding Over Time

There are a number of key issues which must be taken into account when interpreting information from the NMDS. Many of these issues arise as a result of regional differences in the way in which data is coded and uploaded to the NMDS. These include

1. Inconsistencies in the way in which different providers upload day cases to the NMDS, and how this has changed over time.
2. The changeover from the ICD-9 to ICD-10 coding system, and irregularities in the way in which diagnoses and procedures are allocated ICD codes.
3. Changes in the way in which ethnicity information has been collected over time and across regions.

The following sections discuss the first two of these issues.

1. Inconsistencies in the Uploading of Day-Cases to the NMDS

One of the key issues with time series analysis using hospital discharge data is the variability with which different providers upload day cases to the NMDS. Day cases are defined as cases that are admitted and discharged on the same day, with the "three hour rule" (treatment time >3 hours) traditionally being utilised to define an admission event. In contrast patients who spend at least one (mid)night in hospital are classified as inpatients irrespective of their length of stay [6].

In the past, there have been significant regional variations in the way in which different providers have uploaded their day cases to the NMDS, leading to problems with both time series analysis and regional comparisons. These inconsistencies have included:

1. During the mid 1990's, a number of providers began to include A&E events as day cases if the total time in the Emergency Department (including waiting time) exceeded 3 hours, rather than uploading only those whose actual treatment time exceeded 3 hours [6]. NZHIS provided feedback which rectified this anomaly and since January 1995 the correct procedure has been used (these additional cases were coded using medical and surgical sub-specialty codes and are thus difficult to filter out using traditional Emergency sub-specialty filters).
2. Over time, a number of providers have become more efficient at recording the time of first treatment within the Emergency Department (rather than time of attendance) and thus during the late 1990s and early 2000s have become more efficient in identifying emergency department cases which meet the 3-hour treatment rule and are thus eligible to be uploaded to the NMDS. This has resulted in a large number of additional cases being uploaded to the NMDS, particularly in the upper North Island.



3. In addition, some providers admit cases to their short stay observation units while other providers do not, leading to regional variations in the appearance of day cases in the NMDS [7].

Previous Attempts to Address Inconsistent Uploading at the Analytical Stage

When producing their annual Hospital Throughput reports, the Ministry of Health has adopted the following filter to ensure regional and time series comparability with respect to day patient admissions [7]. In its analyses it excludes all cases where:

1. the admission and discharge date are the same (length of stay = 0)
2. and the patient was discharged alive
3. and the health specialty code on discharge is that of Emergency Medicine (M05, M06, M07, and M08).

While this coding filter succeeds in ensuring a degree of comparability between regions and across time (although it fails to correct the anomalies occurring during the mid 1990s when A&E cases were uploaded using medical sub-specialty codes), the exclusion of emergency day cases from time series analysis has a number of limitations including:

1. Exclusion of only those with a length of stay of 0 days means that those emergency cases who begin their treatment late at night and are discharged in the early hours of the following morning (up $\frac{1}{4}$ of emergency cases have a length of stay of 1 day in some DHBs) are included as genuine hospital admissions, whereas those who begin their treatment early in the morning and are discharged late in the afternoon or the evening of the same day are excluded.
2. With a move towards the development of specialist paediatric emergency departments in larger urban centres (e.g. Auckland), there remains the possibility that some larger DHBs are now seeing and treating a number of acute medical patients within the emergency setting, while in regional centres similar patients continue to be assessed on the paediatric medical ward / assessment unit and thus receive a paediatric medical specialty code. The exclusion of all emergency presentations from time series and sub-regional analysis may thus differentially exclude a large portion of the workload occurring in large urban centres where access to specialist advice and treatment is available within the Emergency Department setting.

The potential impact of inconsistent uploading of day cases to the NMDS is likely to be greatest for those conditions most commonly treated in the emergency department setting. Analysis of 2001-2003 hospital admission data suggests that $>1/3$ of NMDS emergency department discharges for those 0-24 years were due to injury, with another $1/3$ were due to ambulatory sensitive conditions (e.g. asthma, gastroenteritis, respiratory infections). In contrast, only 2% of those presenting with bacterial meningitis and 4% of those with septic arthritis were discharged with an emergency sub-specialty code.

Further sub-analysis of these two admission categories however demonstrated that inclusion / exclusion of emergency department admissions had quite different effects depending on the category of admission under study (injury vs. ambulatory sensitive admissions) and whether the region had access to a specialist Paediatric Emergency Department. In this analysis the Wider Auckland Region, (comprising $1/3$ of the NZ population and whose residents have access to specialist Paediatric Emergency Departments) was compared to the rest of NZ. For ambulatory sensitive admissions, exclusion of emergency department cases resulted in Auckland's admission rates being consistently lower than in the rest of New Zealand. It was only when emergency cases were included in this analysis that Auckland's admission rates began to approximate those of the rest of New Zealand. In contrast for injuries, inclusion of emergency department cases resulted in hospital admissions in the Auckland Region consistently exceeding the rest of New Zealand. It was only when emergency cases were excluded from the analysis that Auckland's injury admission rates began to approximate those of the rest of New Zealand.



Zealand. (These findings occurred despite Auckland having a similar proportion of children living in the most deprived NZDep small areas as the rest of NZ).

Loosely interpreted, the findings of this analysis suggest that the workload of large specialist paediatric emergency departments must not be discounted when examining trends in ambulatory sensitive or other medical admissions, as it is only when emergency cases are included in the analysis that the admission rates of the Wider Auckland Region (with its access to Specialist Paediatric Emergency care) begin to approximate the rest of NZ. In contrast, it is possible that specialist paediatric emergency departments have much less of an influence on admission thresholds for injury, with these being handled in a similar manner by different emergency departments across the country. Thus for injury data, the greater tendency for some emergency departments to upload their cases to the NMDS must be taken into account in any analysis.

Implications for Interpreting Time Series Analyses in these Indicators

For the indicators on this website, analysis of time series and other information has been undertaken using unfiltered hospital admission data, with the exception of the injury and assault sections. Here emergency department discharges have been filtered out of the dataset, in an attempt to address some of the inconsistencies discussed above. Despite such an approach, there remains the potential for the inconsistent uploading of day cases to significantly influence the time series analyses presented in this report. In particular, such practices may lead to an over estimate of the number of medical admissions commonly treated in the emergency department setting (e.g. asthma, skin infections, respiratory tract infections), while at the same time the filtering out of injury/poisoning emergency cases may lead to undercounting for a number of more minor types of injury. Nevertheless, the filtering process utilised in this report are thought to provide the best balance when considering hospital admissions amongst those 0-24 years. Despite this, the reader must bear in mind that a potential for significant residual bias remains, when interpreting the time series analyses presented on this website.

2. Data Quality and Coding Changes over Time (ICD-9 and ICD-10)

Change Over from ICD-9 to ICD-10 Coding

From 1988 until June 1999, clinical information in the NMDS was coded using variants of the ICD-9 classification system (ICD-9 CM until June 1995, then ICD-9-CM-A until June 1999). From July 1999 onwards, the ICD-10 classification system has been used, although for time series analysis, back and forward mapping between the two classification systems is possible using pre-defined algorithms [5].

The introduction of ICD-10 represents the most significant change in the International Classification of Diseases (ICD) in over 50 years and uses an alphanumeric coding system for diseases in which the first character of the code is always a letter followed by several numbers. This has allowed for the expansion of the number of codes to provide for recently recognised conditions and to provide greater specificity about common diseases (there are about 8,000 categories in ICD-10 as compared to 5,000 in ICD-9). While for most conditions there is a reasonable 1:1 correspondence between ICD-9 and ICD-10 codes, for some this may lead to some irregularities in time series analysis [8]. Where possible such irregularities will be highlighted in the text, although care should still be taken when interpreting time series analysis across the 1999-2000 period as some conditions may not be directly comparable between the two coding systems.

Accuracy of ICD Coding

In recent years the NZHIS has undertaken a number of reviews of the quality of ICD coding in the NMDS. In the latest audit 2708 events were audited over 10 sites during a 3 month period during 2001/2002. Overall the audit found that 22% of events required a change in coding, although this also included changes at the fourth and fifth character level. The average ICD code change was 16%, with changes to the principal diagnosis being 11%, to additional diagnoses being 23% and to procedure coding being 11%. There were 1625 external causes of injury codes, of which 15% were re-coded differently [9]. These findings were similar to an audit undertaken a year previously.



While the potential for such coding errors must be taken into consideration when interpreting the findings for each indicator, it may be that the 16% error rate is an overestimate, as in the majority of the analyses undertaken in this report, only the principal diagnosis (with an error rate of 11%) is used to describe the reason for admission. In addition, for most admissions the diagnostic category (e.g. lower respiratory tract infections) is assigned using information at the 3 digit level (with the 16% error rate also including issues with coding at the 4th or 5th digit level).

Conclusion

In general the inconsistencies outlined above tend to make time series and (regional) comparative analyses based on the NMDS less reliable than those based on Mortality or Birth Registration data (where legislation dictates inclusion criteria and the type of information collected). While hospital discharge data still remains a valuable and reasonably reliable proxy for measuring the health outcomes of children and young people in this country, the reader is cautioned to take into consideration the biases discussed above, when interpreting the findings outlined for each indicator.



APPENDIX 9. METHODS USED TO DEVELOP THE CHILDREN'S SOCIAL HEALTH MONITOR

Introduction

In response to deteriorating economic conditions in New Zealand and Australia over the past 18 months, a Working Group of health professionals from a range of Organisations¹ with an interest in child health was formed in early 2009. Over the course of the year, this Working Group discussed the conceptualisation of an indicator suite to monitor the impact of the recession on child wellbeing, the range of indicators which might be included, and the criteria by which indicators should be selected. Following these discussions, it was proposed a Children's Social Health Monitor be developed, which comprised the following:

1. *A Basket of Indicators to Monitor Prevailing Economic Conditions:* Ideally, indicators would capture different facets of economic wellbeing (e.g. in a recession several quarters of negative growth (GDP) may precede upswings in Unemployment Rates, which in turn will influence the number of Families with Children Reliant on Government Benefits).
2. *A Basket of Indicators to Monitor Children's Wellbeing;* Ideally indicators would responded relatively quickly (e.g. months-small number of years) to family's adaptations to deteriorating economic conditions (e.g. hospitalisations for poverty related conditions) and would provide an overview of family wellbeing from a variety of perspectives.

Indicator Selection Criteria

In selecting these indicators, it was decided that only routinely collected data sources which were of good quality, and which provided complete population coverage would be used, in order to ensure the indicator suite was methodologically robust and could be consistently monitored over time. In order to achieve this aim, the Working Group developed a set of selection criteria, against which candidate indicators were scored. These selection criteria included:

Conceptual Criteria

Criteria for Indicators to Monitor Prevailing Macroeconomic Conditions

1. Internationally Recognised and Reported Measure of Economic Performance / Wellbeing
2. Should impact on at least one facet of children's wellbeing (i.e. the pathway(s) via which it impacts on children's wellbeing should either be relatively well understood, or an association between the indicator and wellbeing at least documented in the literature).
3. Likely to change in response to a recession (i.e. months-small number of years)

Criteria for Indicators to Monitor Children's Health and Wellbeing

1. The condition is likely to be influenced by family's physical adaptations to worsening economic conditions (e.g. saving on heating to pay for food, moving in with family to save on rent).
2. The condition is likely to be influenced by family's psychological adaptations to worsening economic conditions (e.g. increased family conflict in response to financial stress).

¹ The Paediatric Society of New Zealand, the Population Child Health Special Interest Group of the Royal Australasian College of Physicians, the New Zealand Child and Youth Epidemiology Service, TAHA (the Well Pacific Mother and Infant Service), the Maori SIDS Program, the Kia Mataara Well Child Consortium, the New Zealand Council of Christian Social Services, and academics from the Universities of Auckland and Otago



3. The condition exhibits a socioeconomic gradient (e.g. rates are higher in poorer areas)
4. The condition is likely to respond to changing economic conditions in the short to medium term (e.g. months to 1-2 years)

Data Quality Criteria

Data Quality Criteria (for Either of the Above Indicator Categories)

1. Needs to be routinely collected
2. Available at the national level i.e. complete coverage of target population
3. Updated at least annually (although quarterly preferable)
4. Availability of consistent time series data going back several years (i.e. standard and stable method of data collection)
5. Distribution can be broken down by e.g. ethnicity, socioeconomic status, region

Selection of the Baseline Indicator Set

In mid-2009 a long list of candidate indicators (selected by means of a scan of the available literature, email consultation with child health networks, and the suggestions of Working Group members) were then scored against each of these criteria by a group of Working Group members and other health professionals (n=20). Those scoring the indicators were also asked to select a Top 5 Economic and Top 5 Health and Wellbeing Indicators for inclusion in the Children's Social Health Monitor. The resulting Top 5 Economic and Wellbeing indicators (as determined both by criteria scoring and priority ranking) were:

Economic Indicators:

- Gross Domestic Product
- Income Inequality
- Child Poverty
- Unemployment Rates
- The Number of Children Reliant on Benefit Recipients

Child Health and Wellbeing Indicators:

- Hospital Admissions with a Social Gradient
- Mortality with a Social Gradient
- Infant Mortality
- Hospital Admissions and Mortality from Non-Accidental Injury
- Ambulatory Sensitive Hospital Admissions

Methodology for Developing the Hospital Admissions and Mortality with a Social Gradient Indicators

While all of the Top 5 Economic Indicators, and a number of the Child Health and Wellbeing indicators already had established methodologies, the hospital admissions and mortality with a social gradient indicators had to be derived specially for the Children's Social Health Monitor. The methodology used to derive these indicators is outlined briefly below:

Hospital Admissions

In considering which conditions should be included in the analysis of hospital admissions with a social gradient, the 40 most frequent causes of hospital admission in children aged 0-14 years (excluding neonates) were reviewed, and those exhibiting a social gradient (a rate ratio of ≥ 1.8 for NZDep Decile 9-10 vs. Decile 1-2; or for Māori, Pacific or Asian vs. European children) were selected. A small number of conditions with rate ratios in the 1.5-1.8 range were also included, if they demonstrated a consistent social gradient (i.e. rates



increased in a stepwise manner with increasing NZDep deprivation) and the association was biologically plausible (the plausibility of the association was debated by Working Group members).

Inclusion and Exclusion Criteria

Neonatal hospital admissions (<29 days) were excluded on the basis that these admissions are more likely to reflect issues arising prior to / at the time of birth (e.g. preterm infants may register multiple admissions as they transition from intensive care (NICU) → special care nurseries (SCBU) → the postnatal ward), and respiratory infections / other medical conditions arising in these contexts are likely to differ in their aetiology from those arising in the community.

For medical conditions, only acute and arranged hospital admissions were included, as Waiting List admissions are likely to reflect service capacity, rather than the burden of health need (e.g. the inclusion of Waiting List admissions would result in a large number of children with otitis media and chronic tonsillitis (who were being admitted for grommets and tonsillectomies) being included, and the demographic profile of these children is very different from children attending hospital acutely for the same conditions).

For injury admissions, filtering by admission type was not possible, as a number of DHBs admitted injury cases under (now discontinued) ACC admission codes, making it difficult to distinguish between acute and waiting list admissions in this context. As with other NZCYES reports, all injury cases with an Emergency Department Specialty Code (M05-M08) on discharge were excluded as a result of inconsistent uploading of Emergency Department cases across DHBs (see Appendix 4 for further detail). This differential filtering however, means that it is not possible to accurately compare the magnitude of the social gradients between the medical condition and injury categories, as they were derived using different methodologies (and social differences in Emergency Department vs. primary care attendances for minor medical conditions may have accounted for some of the social gradients seen). No such differential filtering occurred for mortality data however (see below), and thus the magnitude of the social differences seen in this context is more readily comparable.

Mortality

In the case of mortality, because in many instances, the number of deaths from a particular condition was insufficient to calculate reliable rate ratios by NZDep and ethnicity, the rate ratios derived from the analysis of hospital admission data were used to denote category membership. The most frequent causes of mortality in those 0-14 years (excluding neonates) were reviewed however, in order to ensure that no additional conditions making a large contribution to mortality had been missed by the analysis of hospital admission data. This identified two further conditions (which by analysis of mortality of data met rate ratio criteria); deaths from drowning and Sudden Unexpected Death in Infancy, which were then included in the coding algorithms (for both hospital admissions and mortality data). A number of deaths were also identified, which were attributed to issues arising in the perinatal period (e.g. extreme prematurity, congenital anomalies), but in order to preserve consistency with previous exclusion criteria (i.e. the exclusion of conditions arising in the perinatal period), these were not included in coding algorithms.

In Conclusion

While it is hoped that over time this indicator set will be expanded and further refined, it is intended that the NZ Child and Youth Epidemiology Service will monitor this core minimum indicator set on an annual basis, until the economic position of New Zealand children improves appreciably. It is also hoped that further adaptations to this indicator set will be made, so that it can also be used in the Australian context.

