

Appendix 2: Data quality and its effect on interpretation

This appendix provides a discussion of data quality for each data collection. Data quality problems arose for various reasons, particularly from changes in question structure over time, a high proportion of missing responses for some data items and associated difficulties in matching records between data collections. Data repair was undertaken to resolve some of these issues.

A2.1 Child Health Check data collection

Records of CHCs were sent to the AIHW as paper forms in almost all cases, and data from these forms were manually entered into the CHC database. The AIHW developed a system that allows for the receipt of electronically transmitted CHC forms from clinics within the Northern Territory that operate an electronic patient record system. This system was operational from 12 June 2008, but only one valid electronic record was processed over the duration of the CHCI.

Form versions

Over the 2-year period during which the NTER Child Health Checks were conducted (July 2007 to June 2009), six main versions of the health check form have been used by the medical teams conducting the checks. Eight communities in the Katherine East region provided information on CHCs using non-standard forms (referred to as Version 7). Data from 1,700 non-standard forms were received and cannot be analysed at present.

The changes to the forms over time rectified deficiencies of earlier versions; in particular, reducing the high proportions of missing responses for many data items. Changes involved adding several questions and improving the structure of a number of existing questions, as well as adding a response option of 'unsure' to many questions, which allowed health teams who did not have the relevant information to provide an answer to the question, rather than leave it blank. These changes improved quality of information provided using the new forms, but created some difficulties comparing information across versions.

Missing data

Forms received by the AIHW often contained missing or illegible information in one or more fields, resulting in an incomplete CHC database. The proportion of missing responses for a particular data item is an indicator of data quality; high levels pose a problem for analysis and interpretation because they reduce reliability of the results. A high level of missing responses for a data item was often related to the format of the question and/or the available response options. For instance, when a response was not provided, it was often unclear whether this meant: no testing was undertaken; testing was undertaken but no abnormality was detected or referral given; the health team was unsure

of the correct response; or the question was inadvertently skipped. This ambiguity was one of the main reasons for improvements between form versions, and modifications to question format resulted in a substantial decrease in the proportion of missing responses over the course of the collection of the CHCI data.

The quality of the CHC data for different items in relation to missing responses varied, ranging from 0% to 79%. The extent of missing data must be taken into account when using and interpreting data for each item. Analysis methods were adjusted where appropriate (for instance, by excluding children who were known not to have been screened for a particular condition in calculating prevalence of that condition) to account for high levels of missing responses. Levels of missing data must also be considered when making comparisons across regions.

Data repair

High levels of missing responses were of particular concern among the identification variables (including HRN, date of birth, sex, community ID, and so on) because these are used to identify children and have the potential to create significant problems in CHC analysis and in matching records between data sets.

In an attempt to repair these records, the AIHW performed both internal and external data repair processes. Data records were repaired where possible during data entry based on the information available to the AIHW, and these repairs were documented. Where a response was missing, there was sometimes enough information available to arrive at an approximation. For instance, although the exact age could not be determined for 2.6% of children, sufficient information was available (such as the completion of age specific questions) to determine the age group for all children resulting in no missing data in relation to age group. Missing fields could also be repaired internally where a child had completed two or more checks and identifying information was missing on one but provided on the other.

Where it was not possible for the AIHW to repair data internally, missing information was requested from NT DHF. The AIHW had been receiving data files from the NT DHF since March 2008, which provided missing information on date of birth (DOB), sex and, in some cases, HRN for many records that were originally incomplete. Before the most recent NT DHF external data repair process in September 2009, there were 114 records where the date of birth of the child was missing or wrong, 21 cases where sex of the child was missing or wrong and 216 cases where HRNs was wrong or missing were repaired. Other improvements included changes in the sex or date of birth so that these variables were consistent across the four data collections.

These data repair processes resulted in a CHC data collection with minimal missing responses. Once this clean CHC data set was obtained, this was used as the standard data set to clean and repair the three follow-up data collections through further internal AIHW data repair processes.

Other issues to consider in interpreting data

The data included in the CHC collection were a by-product of a clinical process where health professionals providing the CHCs and follow-up services documented the results of those checks and services on standard data collection forms. The aim of the checks was to detect, treat or refer children for clinically significant problems rather than establish a