

Children's Worlds Survey 2013

Guidance for national research teams on data inputting and checking

1. Data inputting by national research teams

Data entry

Three SPSS template files have been created to gather the survey data from each country in a standard format. National research teams can use these files to input the data directly. Alternatively, if a team wishes to use other data entry software to input data then the same system of variable names, value codes and missing data values should be used for data entry and the data should be transferred to the SPSS template in order to be sent to the data checking team at University of York.

If you do not have access to SPSS software, please contact Gwyther Rees (contact details below) to arrange an alternative format for returning the data (e.g. Microsoft Excel).

Missing values

We have used values from 90 to 99 to indicate missing data values in the templates as follows:

90 = 'Not sure' or 'Don't know' where this one of the options to answer the question.

91 = 'Not applicable'. Mainly used for the questions about second homes if the child only lives in one home.

92 = 'Not asked in this country' where a question has been left out of a particular national survey.

99 = 'Missing' where the child gave no answer at all to a question.

The only exception to this is for the two variables for inputting amounts of pocket money received (per week or month). For these variables, to avoid confusion with non-missing values, the missing values are negative.

Dealing with national variations in questionnaires

Please do not change or edit the variable and value specifications in the template files. All possible values are specified in the template and only these values should be used.

If your survey includes additional questions not contained within the standard international questionnaire then additional variables can be added to the file, using a similar system of naming variables and values. Additional variables should be added according to their location in the paper questionnaire. Please note that these variables will not be included in the final combined international data set

If your survey excludes some of the questions contained within the standard international questionnaire then please enter the code of 92 for all cases in the data set.

Data inputting

Please input all survey questionnaires received, even if there is some missing data. A standard approach will be taking to excluding cases with high levels of missing data. This will be dealt with at the data cleaning stage so that there is consistency in how this issue is handled across all countries.

Please input the data as written by the child, even if there are some inconsistencies (for example if a child says they live in one home but then ticks boxes for people living in first and second homes). Again, a standard approach will be taken to resolving inconsistencies like this at the data cleaning stage.

It is important that national research teams ensure that the data is input accurately. It is recommended that checks are run regularly for selected cases - comparing the data on the computer with the answers on the paper questionnaire - to ensure the quality of data inputting and that there are no systematic inputting errors.

It is also vital that a unique reference number is used for each questionnaire in the data file (variable QRef) and that this number is written on the paper questionnaire at the time of data inputting in order to make it possible to check back if there are any queries about the data input.

Please also include a unique reference for each school in the survey (SchoolRef). This will be helpful for weighting purposes and also because researchers may wish to explore variation in responses between and within schools. However, please do not provide any identifying details such as names of schools.

Sending the data for checking and cleaning

Once all data has been input in line with the guidelines above, please send the data files directly to Gwyther Rees by e-mail, copying in Tamar Dinisman.

2. The data checking and cleaning process

Initial checking and cleaning

The data sets will be cleaned and checked using a standard process by staff at the University of York in England, which is acting as the central point for data cleaning, checking and collation.

The process was developed and tested during the pilot wave of the Children's Worlds survey in 2012. The following key checks will be undertaken:

1. An initial check will be run to ensure that all the variable names and value labels match the template.
2. A check on levels of missing data for each variable in the file.
3. A check on cases with high levels of missing data.
4. A check on logical consistency in the data set. This will particularly apply to the set of questions on family structure.
5. A check on cases with systematic responses suggesting perhaps that the child was not fully engaged in the survey. An example would be if the child selected the same value for each question about satisfaction with different aspects of life throughout the questionnaire.

Dealing with queries

Once these steps have been completed, a report on the process will be sent back to the national research team together with a version of the data set with additional variables to identify any queries. This will include:

- a. Queries about particular cases or values that might need to be checked with the original paper data.
- b. Suggestions and recommendations for changing in codings for particular values (e.g. where there are logical inconsistencies).
- c. Indications of cases that might need to be excluded from the final combined international data set due to levels of missing data or systematic response patterns.

At this stage the original data will remain unchanged and national research teams will have the opportunity to comment on the queries.

Finalising the data

Once the national research team has responded to the queries raised in the report, the team at the University of York will make any necessary changes and finalise the data set to be added to the combined international data set. At this stage, any weightings due to sampling strategy that are needed will be calculated for the data.

3. Time scales

Because work on the current wave of the survey is a two-year project funded by the Jacobs Foundation it is very important that the final combined international data set is ready as soon as possible. This will mean that analysis work can be done on the data in preparation for a report at the end of the funding period. Ideally we would like to finalise the combined data sets by April 2014.

In order to try to achieve this, we will be aiming to check, clean and finalise the data quickly once it has been sent in by the national research team.

It is also very important that, as part of the funded work, each national research team allocates a person who will have time to respond quickly to any queries about the data.

The time scales for the data cleaning process are as follows:

1. Within three days of the data being received by University of York, you will be sent an e-mail confirming this and letting you know when you will receive the report on queries with the data. This will usually be within two weeks.
2. Once you have received the queries, we would ask that you are able to respond to them within two weeks.
3. We will then finalise the data and allocate any weightings as appropriate within a further two weeks. At that point you will receive a final copy of the data from your country to be added to the international data set.

Further information

If you have any queries about data inputting or cleaning, please contact Gwyther Rees at University of York, England at the following e-mail address: gwyther.rees@york.ac.uk.

Please copy Tamar Dinisman (tdinisman@gmail.com), the project co-ordinator, into all e-mail correspondence.