

Overview of process

Stage 1	Data input	Middle of May
Stage 2	Data checking	Middle of May
Stage 3	Responses to queries	End of May
Stage 4	Finalise data sets Create unweighted data set	First week of June
Stage 5	Agree weightings	Middle of June
Stage 6	Create initial weighted data set	End of June
Stage 7	Analysis	June onwards
Stage 8	Agree final data set	By June 2015?

Total data received so far

	8 years old	10 years old	12 years old
Algeria	1385	1216	1359
Nepal	1000	1000	1000
Estonia	1131	1034	1033
Spain	1066	1082	1767
Colombia	1000	1000	1000
Turkey	1045	1079	1029
Ethiopia	1000	1000	1000
Korea	2437	2446	2623
Germany	1096	1075	866
England			1318
Israel	976	1005	927
Romania	1422	1424	
Norway			
Total	13558	13361	13922

Data checking process

1. Check of variable and value specifications (so that data sets are compatible for merging) and some basic data scanning.
2. Levels of missing data
 - (a) within cases and
 - (b) for different variables across the data set
3. Some basic logical checks
4. Evidence of response sets which may indicate systematic patterns in responses for particular children.

General comments

- Excellent quality of data files
- Only minor issues to resolve for individual countries
- Queries in a few countries about children who ticked more than one box for a question

Children who ticked more than one box

Options (if boxes are adjacent):

1. Use mid-point (e.g. 8 and 9 = 8.5)
2. Use lower value
3. Use higher value
4. Use value closest to mid-point
5. Select value randomly from two options
6. Treat as missing data

If boxes are not adjacent, treat as missing data.

Variables with more than 10% of values missing in more than one country

	8 years old	10 years old	12 years old
Frequency of pocket money	n/a	n/a	3
Whether has access to internet		2	2
Frequency of worrying about family money	6	4	2
Whether town council asks views	n/a	n/a	6
Frequency of being excluded by peers	5		
Frequency of classes outside school	5		
Frequency of reading for fun	3		
Frequency of using a computer	3	2	2
Whether lives with same adults as one year ago	n/a	n/a	3

Cases with missing data

- 8 years old survey:
1.6% of cases had more than 20 missing values
- 10 years old survey:
1.3% of cases had more than 25 missing values
- 12 years old survey:
0.9% of cases had more than 30 missing values

Age ranges

- Three different patterns in different countries:
 - Some only include children of the target age
 - Some include children one year either side of target age
 - Some include a wider age range of children
- If we use a narrow age range (e.g. one year either side) this will exclude quite a lot of children (2% to 3%) particularly in Algeria (7% to 20%) and Ethiopia (11% to 14%).
- If we use an age range of two years either side of target age) this will exclude much fewer children (less than 1% in each age group)

Change in children's lives

(12 years old data)

	Moved house	Changed area	Changed school	Other country	Not same adults
Algeria	18%	12%	35%	9%	14%
Colombia	33%	25%	26%	12%	13%
England	26%	17%	14%	10%	6%
Estonia	14%	9%	6%	6%	5%
Ethiopia	17%	12%	13%	3%	11%
Germany	14%	8%	11%	7%	6%
Israel	21%	15%	16%	7%	9%
Nepal	34%	24%	44%	14%	19%
S Korea	19%	6%	6%	3%	6%
Spain	23%	19%	40%	16%	4%
Turkey	25%	n/a	13%	7%	21%

Response sets - process

Looking for same responses to five types of questions in the same format throughout the questionnaire.

1. Agreement questions (5 point scale)
2. Satisfaction / Happiness questions (11 point scale)
3. Time use questions (family, friends, general) (4 point scale)
4. Agreement questions (11 point scale)*
5. Questions about qualities aspired to (11 point scale)*

* 10 and 12 years old questionnaires only

Identifying cases with more than one of the above

Response sets - outcomes

- High levels of responses sets for Agreement (11 point scale) and Qualities (11 point scale), especially in 10 years old survey (up to 40% of more).
- So just focusing on first three types of questions and identifying cases with more than one response set. This affects:
 - 1.3% of 8 years old cases
 - 3.2% of 10 years old cases
 - 1.1% of 12 years old cases

In summary

Excluding cases on the basis of:

- High levels of missing data
- Outside the age range
- More than one response set

will exclude:

- 3.4% of 8 years old cases (maximum of 7.7% in one country)
- 5.2% of 10 years old cases (maximum of 10.1% in one country)
- 2.5% of 12 years old cases (maximum of 5.9% in one country)