JACOBS FOUNDATION

Our Promise to Youth

Methodological issues: Sampling and data management

End goal

A cleaned data set of at least 1,000 cases per age group which is broadly representative of children of that age in mainstream schools in the country

Sampling strategy

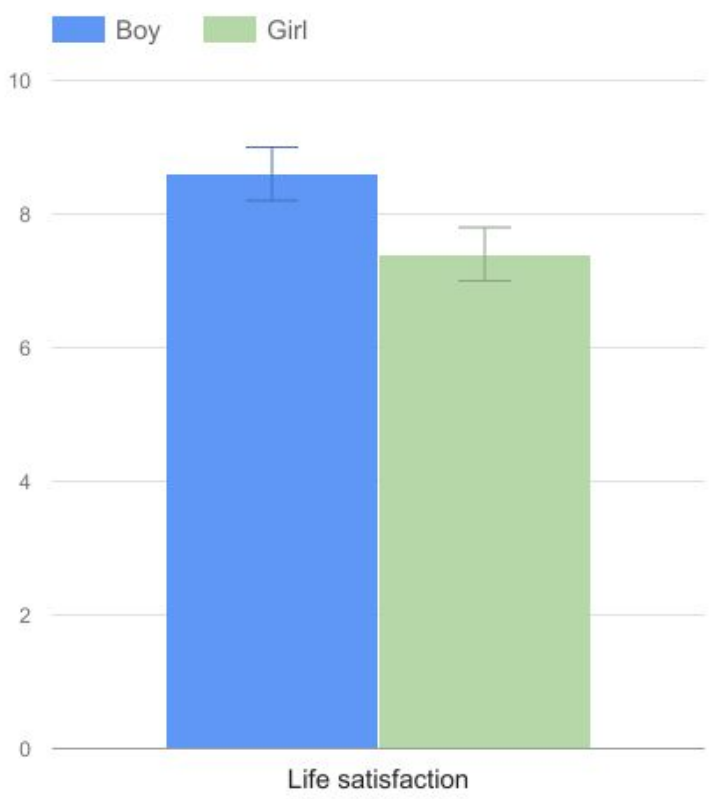
- Country or region?
- Sampling frame - Any exclusions?
- Stratification
- Number of schools (clusters) (note: need at least 2 per stratum)
- Selection method for schools - e.g. uniform or proportional to size
- Selection method for classes within schools
- Selection method for children within classes

Sampling strategy - England

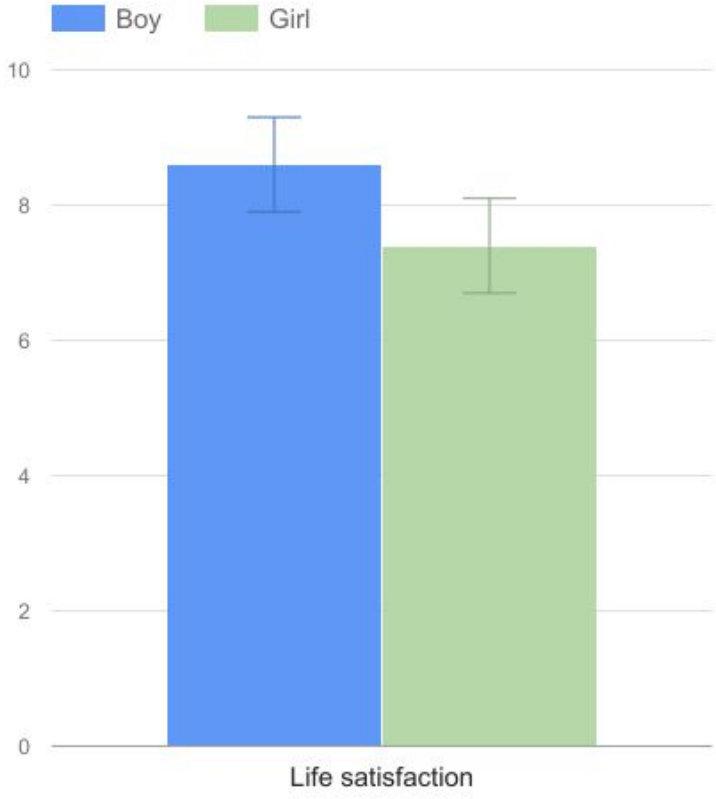
- Whole country
- Some private fee-paying schools excluded (< 1% children)
- Five strata based on proportion of children in school entitled to free school meals. Strata equally sized based on number of pupils
- 8 schools per stratum
- Schools selected with probabilities proportional to number of pupils
- One class group (not based on ability) randomly selected in each school
- All children within selected class invited to take part

Problems of clustering

Simple random sample



Clustered sample



Sampling method

Uniform probability

- 10,000 children in 100 schools
- Select 2 schools
- School A (200 children, 5 classes)
  - School probability = 1/50 = 0.02
  - Class probability = 1/5=0.2
  - Child probability = 0.2*0.02 = 0.004
  - 4 in 1,000 chance
- School B (40 children, 1 class)
  - School probability = 1/50 = 0.02
  - Class probability = 1/1 = 1
  - Child probability = 0.02*1= 0.02
  - 20 in 1,000 chance

Requires more weighting

Probability proportional to size

- 10,000 children in 100 schools
- Select 2 schools
- School A (200 children, 5 classes)
  - School probability = 200/10000 = 0.02
  - Class probability = 1/5=0.2
  - Child probability = 0.2*0.02 = 0.004
  - 4 in 1,000 chance
- School B (40 children, 1 class)
  - School probability = 40/10000 = 0.004
  - Class probability = 1/1 = 1
  - Child probability = 0.004*1= 0.004
  - 4 in 1,000 chance

Requires less weighting

Record of data administration

For each school, important to keep record of:

- Strata membership
- Number of children in school for each age group
- Number of classes in school for each age group
- Number of classes selected in each age group
- Number (total) of children in classes selected
- Number of children participating in survey (completing questionnaires)
- Any other notes

A summary of this will be included in a sampling report after data collection is ended

Data preparation

- Data templates provided in SPSS, Stata and Excel format
- Each research team inputs data and checks inputing
- Research team e-mails data sets to University of York
- 2 weeks: data checked and sent back to research team with queries
- Research team responds to queries
- 2 weeks: data finalised
- Research team sends sampling report to University of York
- 2 weeks: proposal for weighting sent back to research team
- Research team responds to proposals
- 2 weeks: final weighted data sent back to research team

Data checking

- Coding
- Age range
- Variables with high levels of missing data
- Cases with high levels of missing data
- Logic of answers
- Systematic responding