



Scottish Centre *for*  
Social Research

Survey design and  
methodology: sampling,  
response and weighting

Paul Bradshaw

Scotland's leading independent social research institute



# Outline

- Study design
  - Research objectives
  - Study design overview
  - Sample design
  - Data collection
- Response rates
- Weighting
  - Overview
  - Weighting for GUS



Scottish Centre *for*  
Social Research

# Study Design

Scotland's leading independent social research institute



## GUS: The ‘purpose’

“To generate, through robust methods, specifically Scottish data about outcomes throughout childhood and into adulthood for children growing up in Scotland across a range of key domains:

- Cognitive, social, emotional and behavioural development
- Physical and mental health and wellbeing
- Childcare, education and employment
- Home, family, community and social networks
- Involvement in offending and risky behaviour

Such data will encompass, in particular, topics where Scottish evidence is lacking and policy areas where Scotland differs from the rest of the UK.”



## Study Design: Outline

- National sample capable of analysis by urban/rural, deprived/non-deprived and other sub-groups of interest
- Sample drawn from Child Benefit Records
  - Good coverage
  - Some limitations
- Three cohorts:
  - Birth cohort 1: 5217 children aged 10.5 months at the 1<sup>st</sup> interview
  - Child cohort: 2859 children aged 34.5 months at the 1<sup>st</sup> interview
  - *Birth cohort 2: c6000 children aged 10.5 months at the 1<sup>st</sup> interview*



# Study design: ages and stages

Year Sweep	Age at interview							
	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8
2005 1	BC1 (5217)		CC1 (2858)					
2006 2		BC1 (4512)		CC1 (2500)				
2007 3			BC1 (4191)		CC1 (2331)			
2008 4				BC1 (3994)		CC1 (2200)		
2009 5					BC1 (3880)			-
2010 6						BC1 (3657)		-
2011 BC2 1		BC2						-
2012 7		-					BC1	



# Sample Design (1)

- Random/Probability sample
  - Every person in sample frame has a known (and non-zero) probability of selection
  - Statistical theory applies
    - You can calculate error
    - You can estimate non-response bias
  - Key concepts
    - Sampling error - extent to which estimates based on random samples vary from true value in the population
    - Confidence interval - estimate of the range in which actual value in the population will fall (+/-)
    - Confidence level - how confident you are about your estimates  
E.g. 40% (+/- 3% at the 95% confidence level)

## Sample Design (2)

### Clustering - area level sampling

- Areas made up by aggregating Data Zones
- Data Zones merged into larger areas
- Each merged area had an average of 57 births per year
- List of areas sorted by Local Authority and then by SIMD
- 130 areas selected at random

### Individual level

- Within each area, ALL babies and 3/5 of toddlers who met the date of birth criteria were selected
- Sampling undertaken monthly
- Multiple child households



## Sample Design (3)

Sample Number	Dates of Birth required	
	Birth Cohort	Child Cohort
1	01-June-2004 - 30-Jun-2004	01-June-2002 - 30-Jun-2002
2	01-Jul-2004 - 31-Jul-2004	01-Jul-2002 - 31-Jul-2002
3	01-Aug-2004 - 31-Aug-2004	01-Aug-2002 - 31-Aug-2002
4	01-Sep-2004 - 30-Sep-2004	01-Sep-2002 - 30-Sep-2002
5	01-Oct-2004 - 31-Oct-2004	01-Oct-2002 - 31-Oct-2002
6	01-Nov-2004 - 30-Nov-2004	01-Nov-2002 - 30-Nov-2002
7	01-Dec-2004 - 31-Dec-2004	01-Dec-2002 - 31-Dec-2002
8	01-Jan-2005 - 31-Jan-2005	01-Jan-2003 - 31-Jan-2003
9	01-Feb-2005 - 28-Feb-2005	01-Feb-2003 - 28-Feb-2003
10	01-Mar-2005 - 31 Mar-2005	01-Mar-2003 - 31 Mar-2003
11	01-Apr-2005 - 30-Apr-2005	01-Apr-2003 - 30-Apr-2003
12	01-May-2005 - 31-May-2005	01-May-2003 - 31-May-2003

## Data collection

- Face-to-face CAPI interview with self-complete (CASI) section – just over 60 minutes
- Respondent to be child's 'main carer' but aimed to get mother as far as possible (and did so in upwards of 99% of cases at sweep 1)
- At subsequent sweeps aim is to interview, where possible, respondent from previous sweep

## Timing of fieldwork

- Monthly 'waves'
- Target interview dates



Scottish Centre *for*  
Social Research

# Response rates

Scotland's leading independent social research institute



## Response and attrition rates

	No. cases achieved	Response rate	As % of sw1 achieved
<b>Birth cohort</b>			
Sweep 1	5217	80%	100%
Sweep 2	4512	88%	86%
Sweep 3	4193	90%	80%
Sweep 4	3994	91%	77%
Sweep 5	3833	92%	74%
<b>Child cohort</b>			
Sweep 1	2859	79%	100%
Sweep 2	2500	89%	87%
Sweep 3	2332	90%	82%
Sweep 4	2200	90%	77%

# Non-response

- Why is this an issue?
- After sweep 1, survey data and area-level variables are used to model non-response
- Factors affecting non-response are similar at each sweep
- Analysis indicates that non-response more likely amongst
  - Lower income families
  - Lone parents
  - Families living in more deprived areas
  - Mothers who had not breastfed
  - Parents who did not attend parent and child groups
  - Younger mothers



Scottish Centre *for*  
Social Research

# Weighting

Scotland's leading independent social research institute



# Weighting: Overview

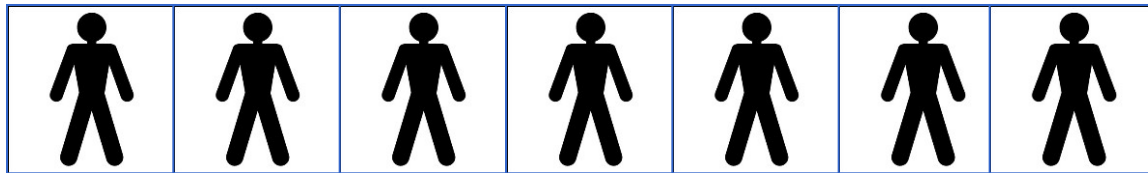
- Why do we need weights?
  - To make the achieved sample look as much like the population as possible
    - Selection weights – correcting for unequal selection probabilities
    - Non-response weights – to correct for any bias in achieved sample
- Advantages:
  - Correct for selection and non-response bias
  - Allow inferences about national population, not the sample
- Disadvantages
  - Reduce sample efficiency



# How weights work

Unweighted sample

70% male



30% female



$\times 0.71 =$

Weighted sample

$\times 1.67 =$

50% male



50% female







# Weighting: Overview

- Why do we need weights?
  - To make the achieved sample look as much like the population as possible
    - Selection weights – correcting for unequal selection probabilities
    - Non-response weights – to correct for any bias in achieved sample
- Advantages:
  - Correct for selection and non-response bias
  - Allow inferences about national population, not the sample
- Disadvantages
  - Reduce sample efficiency

## Sample efficiency (longitudinal)

Cohort	Actual sample size	Effective sample size	Sample efficiency	95% CI for an estimate of ...		
				10%	30%	50%
<b>Birth</b>						
Sweep 1	5217	5061	97%	0.8%	1.3%	1.4%
Sweep 2	4512	4294	95%	0.9%	1.4%	1.5%
Sweep 3	4120	3829	93%	1.0%	1.5%	1.6%
Sweep 4	3844	3484	91%	1.0%	1.5%	1.7%
Sweep 5	3621	3221	89%	1.1%	1.6%	1.8%
<b>Child</b>						
Sweep 1	2859	2777	97%	1.1%	1.7%	1.9%
Sweep 2	2500	2389	96%	1.2%	1.8%	2.0%
Sweep 3	2280	2146	94%	1.3%	1.9%	2.1%
Sweep 4	2100	2048	93%	1.3%	2.0%	2.2%



## GUS Weights

- Sweep 1
  - Single weight corrects for selection and non-response bias
- Sweep 2:
  - Two weights:
    - Main interview weight
    - Partner weight
  - Each correct for non-response at sweep 2
  - The main interview weight includes the weight from sweep 1, the partner weight includes the sw2 main interview weight
- Sweeps 3, 4 and 5
  - Two weights because two ‘samples’:
    - Those who responded at all sweeps
    - Those who responded at the individual sweep but missed an intervening sweep
  - Longitudinal weights
    - Same method as used at sw2 – combined non-response & sw2 weight
  - Cross-sectional weights
    - Calibration method

**At all relevant sweeps there are separate final weights for each cohort. The cohorts must be analysed separately.**



Scottish Centre *for*  
Social Research



Scotland's leading independent social research institute