

The Use of External Data Sources to Assist in Sampling

Nick Moon

NOP World

Two types of sampling

- Random
- Quota

Two types of interviewing

- Face to face
- Telephone

Random Face to Face

- Postcode Address File (PAF)
- Electoral Register
- Other official records

Postcode Address File

- THE sample frame of choice for major surveys
 - ★ Almost all government surveys
 - Labour Force
 - British Crime Survey
 - General Household Survey
 - ★ BARB
 - ★ National Readership Survey

Postcode Address File

- Updated quarterly
- Easily available
- Almost universal coverage
- Even includes properties not yet built

Postcode Address File

- Includes non-residential addresses
- Separate Small User File
 - ★ Removes large businesses
 - ★ Many smaller businesses and organisations remain
 - ★ Typically around 10% deadwood

Postcode Address File

- Crucially, it is a sample of addresses, not individuals (or even households)
- Have to deal with multi-household addresses
- More importantly, have to deal with selection of individuals within households

Postcode Address File

- Selecting Individuals
 - ★ Kish Grid
 - ★ Next birthday
 - ★ Head of household or spouse

Postcode Address File

- Over-samples people in smaller households
 - ★ Someone in single-adult household bound to be selected
 - ★ Each person in a six-adult household has only a one in six chance of selection
- Good for samples of households, bad for samples of individuals

Postcode Address File

- Oversampling of small households needs to be corrected by weighting
- But this weighting increases the sampling error
 - ★ Weight = no of people in household, so range of weights is high
- The effective sampling size is reduced by c15%
 - ★ you are paying for 1,000 interviews and only getting 850 interviews' worth of data

Postcode Address File

- Oversampling of small households can be avoided
 - ★ Only interview at every other single-adult household found
 - ★ Attempted once on NRS, but unpopular with interviewers
 - ★ Could be worth trying again

Electoral Register

- Plus

- ★ Sample of individuals, not addresses
 - More suitable for most surveys

- Minus

- ★ Severe coverage problems in wake of Poll Tax
 - Now largely overcome
- ★ Only compiled once a year
 - Three months out of date when released, fifteen months out of date at end of its life
 - New rolling register will alleviate this

Electoral Register

- Now wholly unusable for research
 - ★ People given right not to appear on the publicly available version
 - ★ As many as 25% in some areas chose to do this

Local Authority Records

- Council Tax Register is usually up to date, and can separate domestic from residential property
BUT
- Not publicly available, so useful only for the authority itself

Telephone Interviewing

- High and growing ex-directory rate
- UK 36% overall London 50%+
- Epsem samples impossible by normal methods

Telephone Interviewing

Effects of ex-directory

Under-represented

urban

younger

female

Over-represented

rural

older

male

Telephone Interviewing

- Directory plus n not a solution
 - ★ Originally Directory +1, then became more sophisticated
- Unlisted numbers tend to be clustered
 - ★ If 10 consecutive numbers are unlisted then directory + any number from the first one will yield no listed number
 - ★ the chance of any number being selected depends on the proportion of the previous nine numbers that are in the directory

Telephone Interviewing

01727 891079	yes	01582 891079	yes
01727 891080	no	01582 891080	yes
01727 891081	no	01582 891081	yes
01727 891082	yes	01582 891082	yes
01727 891083	no	01582 891083	no
01727 891084	no	01582 891084	yes
01727 891085	yes	01582 891085	yes
01727 891086	yes	01582 891086	yes
01727 891087	no	01582 891087	no
01727 891088	no	01582 891088	yes
01727 891089	no	01582 891089	yes

Telephone Interviewing

Directory plus 1, plus 2

01727 891080	yes	01582 891080	yes
01727 891081	no	01582 891081	yes
01727 891082	yes	01582 891082	yes
01727 891083	no	01582 891083	no
01727 891084	no	01582 891084	yes
01727 891085	yes	01582 891085	yes
01727 891086	yes	01582 891086	yes
01727 891087	no	01582 891087	no
01727 891088	no	01582 891088	yes
01727 891089	no	01582 891089	yes

4 in 9 chance

7 in 9 chance

Telephone Interviewing

- Random Digit Dialling (RDD) used widely in the US
 - ★ Mitofsky-Waksberg
 - ★ List assisted
- Several times attempted without success in the UK
 - ★ Cost of dealing with non-working numbers
 - ★ Lack of knowledge about the numbering system

Telephone Interviewing

- A ten digit telephone numbering system allows 1 billion numbers
 - ★ only 20 million households
 - ★ vast majority of possible numbers are not in use
 - ★ how do you find out which these are, short of ringing them?

Telephone Interviewing

- The 01727 exchange allows 1 million numbers, for a town of 50,000 people
 - ★ But all start with either 8 or 7
 - This immediately reduces total possible to 200,000
- More information now available from Ofstel over and above exchange listings
 - ★ Can identify blocks of 100,000 numbers that you know exist
 - Working blocks still contain huge numbers of non-allocated numbers that must be winnowed out

Telephone Interviewing

- **The Noble-Moon experiment**
- Empirical mapping of number system using large scale tele-marketing databases
 - ★ 15 million numbers, so can map nearly the whole system
- Identifies nearly all 100 banks currently in use
- Random sample of 100 banks
- Where working line found. fixed no of working lines drawn in 100-bank
- Yields 66% working lines at second stage

Telephone Interviewing

- Specialist agencies now sell RDD samples
 - ★ Epsem
 - ★ List-assisted
- Pre-dialled with auto-dialler to weed out non-working numbers

Quota Sampling

- Choosing the variables
 - ★ Correlate with key variables
 - ★ Correlate with availability

Quota sampling

- Setting the targets
 - ★ Census the most reliable
 - ★ Available at very local level
 - ★ Can be very out of date
 - Quotas set today would still need to be based on 1991 Census

Quota Sampling

- Constituency-level quotas can be set on basis of national figures
 - ★ tweaked to reflect known differences from norm, eg social class
- Enumeration district (ed) level quotas need to reflect the nature of that ed
 - ★ Interviewers' task must remain possible