

Sampling Subpopulations

Robert Clark ¹ Robert Templeton ²

¹University of Wollongong

²formerly New Zealand Ministry of Health

Frontiers in Social Statistics Methodology
8 February 2017

Outline

- Features of hard-to-sample subpopulations
- Overview of the NZ Health Survey
- Possible Alternative Methods for Oversampling Māori
 - disproportionate sampling by area
 - proxy screening
 - electoral roll
- Combining the Various Methods in the NZ Health Survey
- Non-Probability Samples

Features of hard-to-sample subpopulations I

- There is no reliable frame (e.g. name and address list) of the subpopulation (may be a partial list). So, we need to sample from a broader population.
- Subpopulation is small. No clear cutoff for small. Even if subpopulation is quite large, some of the issues of screening still apply. But wouldn't consider special sample design approaches in this case.
- Unreliability of identification: most populations of interest are self-identification. Short / quick / cheap screeners often inaccurate.
- Geographically dispersed; mobile; or more likely to be in remote areas.

Features of hard-to-sample subpopulations II

- Over-surveyed, which may contribute to low response rates.
- Prevalences (e.g. of health risk factors or conditions) may be higher than for the general population.
- Language or cultural differences may affect survey accuracy unless appropriately recognised.

Overview of the NZ Health Survey 2011-2013

- A rolling quarterly cross-sectional survey, with approximately 14,000 adults and 5,000 children per year.
- Monitor health indicators and prevalences, and service access and usage of NZ adults and children.
- Estimates of minorities are important, particularly Māori, but Pacific and Asian populations are also of interest. National statistics also required.

Disproportionate Sampling by Area I

- Assign households a higher probability in regions with greater proportions of subpopulations of interest according to the last Census.
- Unequal probability sampling results in more subpopulation members in sample, but greater variation in their selection probabilities and hence weights.

Disproportionate Sampling by Area II

- Suppose we screen for subpopulations, and only proceed to full interview if pass the screen. Optimal design for fixed cost has final probability of selection proportional to

$$\sqrt{\text{density}/(\text{screening-cost} + \text{density})} \propto \sqrt{\text{density}}$$

where density is the proportion of the region's population who are in the subpopulations. See Kalton and Anderson (1986), JRSS-A, 65-82 for stratified sampling, and Clark (2009), Statistics in Medicine vol.28, 3697-3717 for two-stage sampling.

Disproportionate Sampling by Area III

- Smaller regions lead to more effective targeting if census data is correct, but changes over time detract from efficiency. Larger regions are more stable. Assuming no change and relative screening cost of 0.4, efficiency gains of 4%, 8% and 15% from targeting Māori at the district health board level (about 100,000 dwellings), area unit level (about 1000 dwellings) or mesh block level (about 40 dwellings).

Proxy Screening

- Household form in 06/07 survey completed by any adult includes ethnicity and age of all residents.
- In core households, one adult and one child selected. In booster households, one eligible (Māori, Pacific or Asian) adult and one eligible child selected.
- Dropped in 2011-2013 survey, as it may give a poor first impression of the survey. The screener did improve the efficiency of the design, but not as much as might be expected, particularly for Māori, of whom 20% are missed by screener (even in single person households!)

Use of Electoral Roll in NZ Health Survey 2011-2013 I

- Sample consists of two components:
 - sample of addresses from the electoral roll where at least one member indicated Māori descent; and
 - a general population area-based sample.
- Both components were self-weighting samples of households, with meshblocks as primary sampling units. One adult and one child selected from each household.
- 52% of adults eligible for the roll component are Māori (according to quarter 2 of survey). 68% of Māori (according to survey) are eligible.

Use of Electoral Roll in NZ Health Survey 2011-2013 II

- To give a rough idea of the value of the Roll, suppose only Māori statistics are of interest, and we can stratify and optimally allocate all NZ adults into those in Māori households vs others. The relative efficiency compared to equal probability sampling would be 0.73.

Voting Enrolment Form



Enrolling to vote: Application

NO POST
 USE ONLY

DATE
 ENTER

YOU MUST ENROL
 if you are qualified to do so.

When you enrol to vote in parliamentary elections, your details are also made available to your local authority for the purpose of including you on the rolls for local elections.

If you have any questions about enrolment -

- See the information attached to this form.
 - Ask your Registrar of Elections, or the staff of any New Zealand PostShop or agency, or call 0800 ENROL NOW (0800 54 54 54).
 - See the elections website at www.elections.org.nz
- If you are physically disabled or outside New Zealand, another person may be able to fill out this form and sign it for you.

SECTION A Please print using black or blue ink pen

My details

My surname or family name is:

My given or first names are:

My title is:

Mr. Mrs. Miss Ms. Other

My residential address is:

Flat/Room

Street/Place

Suburb/Town

City/Town/Village

Have you resided for at least the last month at this address?

Yes No

My postal address is:

My date of birth is:

Day / Month / Year

My occupation is:

Do you want to be able to update your details electronically in future?

Yes No

My contact telephone numbers are:

Home Work Other

Are you a New Zealand Māori or a descendant of a New Zealand Māori?

No Yes

To find out if you can choose to enrol for a Māori electorate or a General electorate, first read the information attached to this form.

This is the address where you choose to make your home. If your house or flat does not have a street or road number, please give extra details in Section E on the next page.

If you answer 'No' or you live outside New Zealand, please fill in Section C on the next page.

Please give your postal address if different from your residential address

If you answer 'Yes' please fill in Section D on the next page

You must enrol for a General electorate. Please sign in the General electorate box in Section B.

SECTION B Declaration Sign in one of the boxes below. You must sign and date this declaration yourself, unless you are physically disabled or outside New Zealand. See the information attached to this form.

General electorate

- I believe I am qualified to enrol as a voter.
- My details are given correctly on this form.
- I apply to enrol for a General electorate.

Signature _____ Date _____

Māori electorate

- I believe I am qualified to enrol as a voter.
- My details are given correctly on this form.
- I am a New Zealand Māori or a descendant of a New Zealand Māori.
- I apply to enrol for a Māori electorate.

Signature _____ Date _____

*You must enrol for a
General electorate.
Please sign in the General
electorate box in Section B.*

Mobile Work Home
Are you a New Zealand Māori or a descendant of a New Zealand Māori?
← No Yes →
*To find out if you can choose to enrol for a Māori
electorate or a General electorate, first
read the information attached to this form.*

SECTION B

Declaration

Sign in one of the boxes below

You must **sign** and date this declaration **yourself**, unless you are physically disabled or outside New Zealand.
See the information attached to this form.

General electorate

1. I believe I am qualified to enrol as a voter.
2. My details are given correctly on this form.
3. I apply to enrol for a General electorate.

<input type="text"/>	/	/
----------------------	---	---

Signature

Date

Māori electorate

1. I believe I am qualified to enrol as a voter.
2. My details are given correctly on this form.
3. I am a New Zealand Māori or a descendant of a New Zealand Māori.
4. I apply to enrol for a Māori electorate.

<input type="text"/>	/	/
----------------------	---	---

Signature

Date

Many Design Choices to be Made

- Aim was to minimise the sum of the three subpopulation SEs.
- Household probabilities of selection in the general population component of the sample were defined to be proportional to;

$$f_i = w_1\sqrt{\text{Māori MB Density}} + w_2\sqrt{\text{Māori AU Density}} + w_3\sqrt{\text{Māori DHB Density}} \\ + w_4\sqrt{\text{Pacific MB Density}} + w_5\sqrt{\text{Pacific AU Density}} + w_6\sqrt{\text{Pacific DHB Density}} \\ + w_7\sqrt{\text{Asian MB Density}} + w_8\sqrt{\text{Asian AU Density}} + w_9\sqrt{\text{Asian DHB Density}} + w_{10}$$

- Let p_{scrn} be the proportion of households where a proxy screener is applied, and only eligible ethnicities are selected.
- Let p_{list} be the proportion of the sample selected via the Roll.
- Might as well impose $\sum_{j=1}^{10} w_j = 1$, so 11 parameters (θ) need to be set.

Optimal Allocation allowing for Imprecise Design Data

- The design depends on both θ and census and electoral roll design datasets. The final design will be based on 2006 Census data and 2011 Roll data.
- To choose θ , we evaluated the design based on 2001 Census data and 2006 Roll data. We then estimated the objective $F(\theta)$ using 2006/2007 sample survey data.
- To do this, we calculated the probability of selection from the postulated design for every respondent in the 06/07 sample, and then applied a simple approximation for the design effect, adjusting for the 06/07 complex design.
- We then numerically minimised $\hat{F}(\theta)$ with respect to θ .

Resulting Design

- Household probabilities of selection in general area component of sample proportional to
$$0.31\sqrt{\text{Pacific MB Density}} + 0.37\sqrt{\text{Pacific AU Density}} + 0.09\sqrt{\text{Asian MB Density}} + 0.20\sqrt{\text{Asian AU Density}} + 0.03$$
(no weight attached to Māori densities!)
- Limited use of the proxy screener (we set to 0).
- 14% of the household sample to come from the Roll.
- If we calculated $\hat{F}(\theta)$ using 06/07 sample data but assuming a design based on 2006 Census data, we would have ended up using MB densities primarily (or only).

Snowball Sampling

- Select sample of people;
- Ask subpopulation members to identify others among their acquaintance;
- Advantage: don't need to contact as many people to achieve the same number of subpopulation members in sample.
- Issues:
 - Can be biased towards people with more friends?
 - unbiased estimation possible provided that everyone is linked to others;
 - Subpopulation members need to know each other;
 - Image problem for government?
 - RDS is a related method, which is less biased.

Intercept Point Surveys

- Example: sample homeless population by selecting individuals visiting selected soup kitchens (?aggregation points?) at selected times.
- At each location, individuals are asked how often they visit this and other aggregation points. Weights are then inversely proportional to reported frequency.
- Very cost-efficient, but biased, possibly extremely so.

An Empirical Comparison

- McKenzie and Mistaien, "Surveying Migrant Populations: A Comparison of Census-Based, Snowball and Intercept Point Surveys" (Jnl of the Royal Statistical Society Series A 2008)
- Group of interest: Japanese-Brazilian families (0.9%)
- Snowball Survey: poor response rate; most respondents did not want to provide referrals
- Snowball and intercept point surveys selected individuals more tied to the Nikkei community.
- Intercept point can be useful for exploratory investigation but snowball not much cheaper than probability sampling.

Table 8: Comparison of Migration and Remittances of Nikkei Across Different Sampling Methods

	Sao Paulo			
	Stratified Survey	Unweighted Intercept	Weighted Intercept	Snowball
<i>% of households which:</i>				
Migration				
Have a Member currently in Japan	19	35***	33**	30
Have a Member who has returned from work/study in Japan	41	65***	59***	51
Remittances				
Receive remittances from Japan	14	10	10	10
Refuse to say if they receive remittances	3	2	3	7
Refuse to say how much they receive if receiving	31	73***	82***	50 ⁺
Amount (Reales)				
Annual Amount Received Conditional on Receiving Remittances and Reporting Amount				
Mean	5404	1483*	1429*	9400
Median	2500	1792	1792	2400
Sample Size	204	516	516	100

Summary

- The NZ Electoral Roll is useful for oversampling the Māori population.
- A combination of tools can and should be used to oversample subpopulations, but you need a strategy to make many interacting design choices, while reflecting the imperfections of the various design data.
- Non-probability samples are much cheaper but do not reliably represent the wider population, at least for some variables.

Acknowledgements

- NZ Ministry of Health, particularly Robert Templeton
- Statistics New Zealand who sponsored an Official Statistics Research Fund project on Sampling Subpopulations in Two-Stage Surveys

Related Papers

Clark, R.G. and Templeton, R. (2014). Sampling the Māori Population using Proxy Screening, the Electoral Roll and Disproportionate Sampling in the New Zealand Health Survey. Chapter 22 of Hard-to-Survey Populations. Cambridge University Press.

Clark, R.G. (2013). Sample design using imperfect design data. *Journal of Survey Statistics and Methodology*, 1(1), pp.6-23.

Clark, R.G. (2009). Sampling of subpopulations in two-stage surveys. *Statistics in Medicine*, 28(29), 3697-3717.

Kalton, G. and Anderson, D.W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society Series A*, 149(1), 65-82.