Oversampling Minorities in the National Alcohol Survey using the Zip Code Tabulation Area File

Pedro Saavedra[1], Shelley Osborn,[1] Naomi Freedner-Maguire[1], Thomas Greenfield[2], Katherine Karriker-Jaffe[2]

[1] ICF International
[2] Public Health Institute

3/20/2014 – BLS Conference Center

The National Alcohol Survey and its sample

- Funded by NIH
- Nationally Representative Sample
- Calls for sample of 5400
- Random Digit Dialing telephone survey
- 60% Landline – 40% cell
- Oversample of Hispanics and African Americans (over 1,000 of each)

Oversampling by Racial Minorities

- It is not unusual to get a request for an address-based sample or an RDD sample using addresses to identify racial/ethnic composition of a set of telephone numbers.
- Today this can be complicated because strategies for different racial or ethnic groups can be at odds with each other.
- Some groups have a preference for cell phones while others tend to use landlines.

Obstacles to Oversampling

- If screening potential respondents took no time or resources, the oversampling task would not present a problem. One could simply create a very large self-weighting sample and drop a certain percent of non-minorities.
- If one did not care about effective sample size or confidence intervals, concentrating in a few predominantly minority areas would do the trick.

Zip Code Tabulation Areas

- **ZCTAs are developed by the Census**
- **Correspond to Zip Codes, but not all are included. Among exclusions are:**
  - Zip Codes for P.O. boxes only
  - Some Zip Codes used in primarily commercial areas
- **Crosswalks link excluded Zip Codes to ZCTA**
- **Census files have demographic and economic data for ZCTA**

Sampling with ZCTAs

- **Requires a database which links telephone exchanges to ZCTAs**
- **Easier with landlines than with cell phones since a billing code does not necessarily correspond to a residence**
- **Must be prepared for mismatches in data bases**
  - Data bases usually match by Zip Code and not ZCTA
  - Some zip codes are not present in database
  - Some exchanges do not have a database attached
  - Data for some ZCTA is contradictory in some sources
- **Several sampling strategies are possible**
  - Stratification of ZCTAs
  - Using a PPS sample, with greater probabilities for ZCTAs with larger minority population
  - Screening out non-minorities in minority areas

Data

- 2010 Census Zip Code Tabulation Area File
- # of residents for each minority group are calculated for each zip code
- For purposes of the study, groups other than African-American or Hispanics will be classified with whites
- Percentages are adjusted at the national level to correspond with recent distribution of ethnicities in telephone surveys

Stratification used for NAS

- **Similar to strata used in previous studies.**
- **Definitions**
  - 1) At least 60% white
  - 2) At least 40% African-American and more African-Americans than Hispanics
  - 3) At least 40% Hispanics and more Hispanics than African Americans
  - 4) Less than 60% white, but no other group reaches 40%

Two Similar Designs

- **One way of using the stratification is to create two frames and design a dual-frame study**
  - One yielding a proportionally distributed sample
  - One yielding a sample drawn from the minority strata
  - Since the entire population is covered by the first, non-minority cases may be screened out in the second, since they have a chance of selection
- **A second way is a stratified sample with greater allocations for minority strata**
  - Minority strata respondents were screened and where white respondents were identified in minority strata, a certain proportion of them were dropped from the sample.
  - The second approach was used in the NAS

NAS Cell Phone Sample

- It is difficult to identify a ZCTA for each exchange
- Consequently, exchanges were linked to counties and a stratification analogous to the one for landlines, but using counties, was implemented.
- The national adjustment was implemented using previous cell phone survey results
- A county was defined by the ZCTAs that were primarily in the county.

Objectives of the Design

- The objective was to maximize the actual sample size for the two minorities, while optimizing the confidence interval for the total sample and the racial/ethnic groups.
- We recall that the effective sample size is the size of a simple random sample that would yield the same confidence interval as the sample being considered.
- The design does not in theory do any clustering, except at the household level, so the effective sample size need only consider weighting.
- The design tries to rely on screening as little as possible.

Allocations by way of Spreadsheet

- The use of spreadsheets allows for experimentation and the estimation of results at once.
- It allows for iterations and can provide a sense of what changes would be more effective.
- Usually a few parameters can define the allocations.
- The final parameters are presented here. The preliminary ones were modified to achieve the desired sample sizes.

Steps to Derivation of Allocation

- Calculate population for Strata by ZCAT and Strata by county
- Select a starting allocation parameter, a little higher than half the desired total sample (2898 was chosen in this instance.
- Distribute 60% of this number proportionately across landline strata
- Distribute 40% of this number proportionately across cell strata

Steps to Derivation of Allocation (continued)

- Create a second set of allocations, but only for the minority strata
- Add the two sets of allocations
- Screen out half the non-minorities in the African-American and Mixed Strata, and two thirds of the non-minorities in the Hispanic strata
- Modify the starting allocation parameter until the total allocation gives the desired number

Adjusted Population Totals by Stratum

Strata	Total	White	Black	Hispanic
Non-Minority– Landline	**213,482,176**	192,774,405	11,101,073	9,606,698
Black – Landline	**10,249,435**	3,587,302	6,354,650	307,483
Hispanic – Landline	**5,340,637**	2,141,595	117,494	3,086,888
Mixed – Landline	**4,441,850**	2,411,925	994,974	1,034,951
Non-Minority– Cell	**189,128,053**	154,895,875	17,967,165	16,265,013
Black – Cell	**13,802,717**	7,329,243	5,755,733	717,741
Hispanic – Cell	**6,986,775**	3,144,049	586,889	3,255,837
Mixed – Cell	**23,564,515**	14,397,919	3,723,193	5,443,403

Initial Proportional Sample Allocation

Strata	Total	White	Black	Hispanic
Non-Minority– Landline	1,491	1,347	78	67
Black – Landline	72	25	44	2
Hispanic – Landline	37	15	1	22
Mixed – Landline	31	17	7	7
Non-Minority– Cell	1,026	840	97	88
Black – Cell	75	40	31	4
Hispanic – Cell	38	17	3	18
Mixed – Cell	128	78	20	30

Initial Oversample Allocation

Strata	Total	White	Black	Hispanic
Non-Minority– Landline	0	0	0	0
Black – Landline	834	292	517	25
Hispanic – Landline	435	174	10	251
Mixed – Landline	362	196	81	84
Non-Minority– Cell	0	0	0	0
Black – Cell	394	209	164	20
Hispanic – Cell	199	90	17	93
Mixed – Cell	673	411	106	155

Pre-Screening Sample

Strata	Total	White	Black	Hispanic
Non-Minority– Landline	1,491	1,347	78	67
Black – Landline	906	317	562	27
Hispanic – Landline	908	364	20	524
Mixed – Landline	393	213	88	91
Non-Minority– Cell	1,026	840	97	88
Black – Cell	469	249	196	24
Hispanic – Cell	437	197	37	204
Mixed – Cell	801	489	126	185

After-Screening Sample

Strata	Total	White	Black	Hispanic
Non-Minority– Landline	**1,491**	1,347	78	67
Black – Landline	**747**	159	562	27
Hispanic – Landline	**664**	120	20	524
Mixed – Landline	**286**	107	88	91
Non-Minority– Cell	**1,026**	840	97	88
Black – Cell	**344**	124	196	24
Hispanic – Cell	**305**	65	37	204
Mixed – Cell	**556**	245	126	185
Total	5,420	3,006	1,203	1,211

Sample Size, Design Effects and CI

			95% Confidence Interval for Several Estimates						
			50%	40%	30%	20%	10%	5%	1%
	N	DEFF							
Total	5,420	1.7	1.73 %	1.69 %	1.58 %	1.38 %	1.04 %	0.75 %	0.34 %
Black	1,203	2.9	4.83 %	4.73 %	4.43 %	3.86 %	2.90 %	2.10 %	0.96 %
Hispanic	1,211	3.6	5.31 %	5.21 %	4.87 %	4.25 %	3.19 %	2.32 %	1.06 %

Contact Information

- Pedro.Saavedra@icfi.com
- Shelley.Osborn@icfi.com
- Naomi.Freedner@icfi.com
- kkarrikerjaffe@arg.org
- tgreenfield@arg.org


ICF
INTERNATIONAL