Geographic Oversampling for Race/Ethnicity Using Data from the 2010 U.S. Population Census

Graham Kalton and Sixia Chen
Westat, 1600 Research Blvd, Rockville, MD 20850

Abstract

Using 1980 and 1990 U.S. Census data, Waksberg, Judkins, and Massey (1997) examined the effectiveness of disproportionate stratification to oversample areas with greater concentrations of a minority population in order to achieve a specified effective sample size for a national survey of that minority. The areas considered were Census block and block groups. The effectiveness of this oversampling depends on the degree of the minority's geographical concentration and on the relative cost of the full data collection to the screening cost. This paper updates the Waksberg et al. findings using 2010 U.S. Census data and compares these findings with those from the 1990 Census. In addition, the paper extends the application to subnational surveys that are concerned with a single census region or with Core Based Statistical Areas (CBSAs) or non-CBSAs, and with use of these variables for prior stratification for a national survey. It also examines the effect of using different cutpoints for defining the density strata.

Key Words: disproportionate stratification, rare populations, sampling minority populations.

1. Introduction

Many demographic surveys of the U.S. population are designed to oversample certain racial/ethnic domains in order that the domain sample sizes are adequate to support separate analyses by domain. A widely-used method for achieving such oversampling is to increase the sampling fractions in geographic areas where members of the domain of interest are more prevalent. Waksberg, Judkins, and Massey (1997) examined the effectiveness of such oversampling when the prevalence of the domains was based on the 1980 and 1990 U.S. Censuses. The geographical areas considered were Census blocks and block groups. The race/ethnicity domains studied were the Black, Hispanic, Asian and Pacific Islander, and the American Indian, Eskimo, and Aleut populations. The aim of this paper is to update the Waksberg et al. (1997) results using data from the 2010 U.S. Census and to add some subnational results.

At the outset, it should be made clear that any differences found between the 1990 and 2010 results cannot be clearly attributed to changes in the degree of racial or ethnic segregation during this twenty year period. Any differences are in fact the result of a combination of four factors: changes in the degree of segregation, changes in the way race was treated in the two censuses, changes in the definitions of blocks and block groups, and changes in the national prevalence of the minority. In particular, the 1990 census respondents were instructed to "Fill ONE circle for the race that the person considers himself/herself to be.", whereas in the 2010 Census the question and instructions were "What is this person's race? Mark X one or more boxes." The number

of blocks was about 25 percent larger in 2010 than in 1990 (increasing from just under 5 million to over 6 million) whereas the number of block groups declined slightly (from 224,000 to 217,000). The Hispanic and Asian minorities are far more prevalent in 2010 than in 1990. In view of these factors, comparisons of results between the two censuses are therefore to be interpreted in terms of the effectiveness of oversampling using the two census databases and not attributed to any one cause.

The first part of the paper replicates some of the Waksberg et al. (1997) results for 1990 using the 2010 Census data. Their main results address efficient sample design for a national survey of a given minority by oversampling the U.S. population in strata where the minority is more concentrated. It is assumed that simple random sampling is used within each stratum. The basic theoretical results for this simple design are summarized in Section 2. Section 3 then presents comparisons of the effectiveness of the geographical oversampling of different minority populations in 1990 and 2010 based on the census data for those years. Section 4 examines the effectiveness of the oversampling obtained by using different cutpoints to form the strata. Section 5 presents results for regions of the country and for urban and rural areas separately. Concluding remarks are given in Section 6.

2. Theoretical Results

This section presents theoretical results for the general situation when disproportionate stratification is used to oversample strata where the rare domain of interest is more prevalent. These results, taken from Kalton and Anderson (1986) and Waksberg et al. (1997), assume that:

a. The survey will collect data only for sample members in the rare domain, screening out all non-domain members;
b. Simple random sampling (SRS) is used within each stratum;
c. The parameter of interest is the domain mean of some variable, *Y*; and
d. The strata population standard deviations of *Y* are the same for all strata.

Under these conditions, the effectiveness of oversampling strata where the domain of interest is more prevalent depends on the cost of screening out a non-member of the domain to the cost of data collection for a member of the domain: the lower the screening cost, the less effective the oversampling of the strata where the rare domain is concentrated. Let c denote the ratio of the cost of a full interview to the cost of a screening interview. Then, the optimum sampling fraction for stratum h is

$$f_h \propto \sqrt{\frac{P_h}{P_h(c-1)+1}} \tag{1}$$

where P_h denotes the proportion of the population in stratum h who are members of the rare domain. In the case where the cost of data collection for a member of the rare domain is the same as the cost of screening out a nonmember, that is $c=1$, then $f_h \propto \sqrt{P_h}$.

Under the listed assumptions, the variance reduction (*VR*) due to oversampling strata where the rare domain is more prevalent, using the sampling fractions given in equation

(1), as compared to proportionate stratified sampling (or simple random sampling under the additional assumption that the stratum means are equal) is

$$VR = 1 - \frac{[\Sigma_h A_h \sqrt{c - 1 + P_h^{-1}}][\Sigma_h W_h P_h \sqrt{c - 1 + P_h^{-1}}]}{P(c-1)+1} \tag{2}$$

where P is the overall prevalence of the rare domain, W_h is the proportion of the total population in stratum h, and A_h is the proportion of the rare domain population in stratum h. In the special case where $c = 1$, VR reduces to

$$VR_1 = 1 - [\Sigma_h \sqrt{(A_h W_h)}]^2 \tag{3}$$

(Kalton, 2003).

In this paper, we focus on the values of f_h and VR or VR_1 for different race and ethic domains. The value of VR_1 represents the maximum variance reduction that can be achieved and serves as a guide to the utility of the oversampling approach. In practice, as a rule, c will be greater than 1 so that smaller gains will be obtained than indicated by VR_1. The gains are in fact generally much smaller as shown in Table 6.

As an extension, suppose that the population is first divided into major strata with proportionate stratification based on the total population across these strata. In Section 5, the examples of major strata are the four regions of the country and the Core Based Statistical Areas (CBSAs) vs. the non-CBSAs. The oversampling of the minority population is then carried out by oversampling the minority in density strata within each of the major strata. Note that the density strata can be defined differently (i.e., with different cutpoints) within the major strata. The variance reduction in this case is then

$$VR_{ps} = 1 - \Sigma_k A_k (1 - VR_k)$$

where the subscript ps denotes proportionate stratification for the major strata, A_k is the proportion of the rare domain population in major stratum k, and VR_k is given by equation (2) within stratum k. In the special case where $c = 1$, VR_{ps} reduces to

$$VR_{ps1} = 1 - \Sigma_k A_k [\Sigma_h \sqrt{(A_{h|k} W_{h|k})}]^2 \tag{4}$$

where $A_{h|k}$ and $W_{h|k}$ are the proportions of the rare domain population and of the total population in density stratum h in major stratum k, respectively.

An alternative approach is to employ disproportionate stratification for the KH combinations of major stratum ($k = 1, 2, ..., K$) and density stratum ($h = 1, 2, ..., H$), allocating the sample using the optimum sampling fractions given by equation (1). In this case, formulas (1) and (2) can applied with this full set of KH strata. The variance reduction with $c = 1$ is then and then given by

$$VR_{KH1} = 1 - [\Sigma_{(kh)} \sqrt{(A_{(kh)} W_{(kh)})}]^2 \quad (5)$$

where (*kh*) indexes density stratum *h* in major stratum *k*.

3. Comparisons of the Effectiveness of Oversampling in 1990 and 2010

This section compares the effectiveness of oversampling for various race categories and by ethnicity based on data obtained for blocks and block groups from the 1990 and 2010 Population Censuses, with the 1990 results taken from Waksberg et al. (1997). The process involves grouping blocks or block groups into strata based on their densities of the race or ethnic population of interest. For consistency across censuses, we retained the same definitions of the density strata for 2010 as used by Waksberg et al. (1997) for the 1990 Census. The 2010 results reported for the Black population are for Blacks defined as "Blacks alone" for all ages, rather than "Blacks alone or in combination with other races." The results reported for the other racial categories are also based on the "alone" definition. In all cases, very similar results were found when the corresponding, more inclusive, definitions of the race categories were used.

The key condition for oversampling by density strata to be effective is that the distribution of the rare population across the density strata A_h must be different from that of the total population W_h. This is readily seen from equation (3): the maximum value of the second term in that equation occurs when $A_h = W_h$, leading to no gains in precision from the use of this technique. Tables 1 to 4 display the distributions of A_h and W_h for the various race/ethnicity categories based on the 1990 and 2010 Censuses with density strata based on both blocks and block groups (BG).

A visual inspection of the data in Table 1 shows that the Black population is appreciably less segregated in 2010 than in 1990. For example, the percentage of the Black population in the highest density stratum based on blocks has fallen by almost 15 percent, from 61.4 percent to 46.8 percent, with a similar decline for the highest density stratum based on block groups. The Black population as a proportion of the total population has increased by only a small amount, and the distribution of the total population across the strata shifted somewhat from the two outside strata into the middle two strata.

Unlike Blacks, there have been major increases in the Hispanic population and the Asian, Native Hawaiian, and Pacific Islander population (the category was described as "Asians and Pacific Islanders" in 1990; henceforth simply Asians) as percentages of the total population, as shown in Tables 2 and 3. As a result, in each case the lowest density stratum is much smaller in that it contains a smaller percentage of the total population in 2010 than in 1990. The changes in stratum sizes make it hard to predict the relative efficiency of oversampling in 1990 and 2010 based on a visual inspection of these tables.

Table 1: Residential clustering of Blacks in 1990 and Blacks alone in 2010 (all ages)

Density stratum (Blacks as a percent of the stratification unit)	*Percentage of Blacks living in the stratum in the indicated year* (A_h)				*Percentage of the total population living in the stratum in the indicated year* (W_h)			
Stratification unit	*BG*		*Block*		*BG*		*Block*	
Measurement year	*1990*	*2010*	*1990*	*2010*	*1990*	*2010*	*1990*	*2010*
<10%	12.0	15.0	8.5	10.7	75.7	70.6	77.5	71.9
10%-30%	16.8	23.7	13.9	20.8	11.4	16.8	9.6	14.8
30%-60%	20.3	23.1	16.2	21.7	5.7	6.8	4.5	6.4
60%-100%	51.0	38.2	61.4	46.8	7.2	5.8	8.4	6.9
Total population (mn)	30.0	38.0	30.0	38.0	248.7	305.3	248.7	305.3
Blacks as a percent of total population	12.1	12.5	12.1	12.5				

Table 2: Residential clustering of Hispanics in 1990 and 2010 (all ages)

Density stratum (Hispanics as a percent of the stratification unit)	*Percentage of Hispanics living in the stratum in the indicated year* (A_h)				*Percentage of the total population living in the stratum in the indicated year* (W_h)			
Stratification unit	*BG*		*Block*		*BG*		*Block*	
Measurement year	*1990*	*2010*	*1990*	*2010*	*1990*	*2010*	*1990*	*2010*
<5%	10.6	6.1	6.6	3.6	68.4	43.4	68.9	47.6
5%-10%	8.7	7.5	8.1	6.1	10.9	17.2	10.3	13.8
10%-30%	22.8	23.4	22.1	21.5	11.8	21.7	11.5	19.9
30%-60%	24.1	26.5	23.3	25.9	5.1	10.1	4.9	9.9
60%-100%	33.9	36.5	39.8	42.9	3.8	7.6	4.4	8.7
Total population (mn)	22.4	50.0	22.4	50.0	248.7	305.3	248.7	305.3
Hispanics as a percent of total population	9.0	16.4	9.0	16.4				

Table 3: Residential clustering of Asian and Pacific Islanders in 1990 and Asians, Native Hawaiians, and Other Pacific Islanders in 2010 (all ages)

Density stratum (Asians as a percent of the stratification unit)	*Percentage of Asians living in the stratum in the indicated year* (A_h)				*Percentage of the total population living in the stratum in the indicated year* (W_h)			
Stratification unit	*BG*		*Block*		*BG*		*Block*	
Measurement year	*1990*	*2010*	*1990*	*2010*	*1990*	*2010*	*1990*	*2010*
<5%	30.5	21.8	19.4	12.7	86.4	75.4	85.2	75.2
5%-10%	17.2	16.7	17.7	15.2	7.2	11.7	7.4	10.6
10%-30%	27.8	33.2	32.1	35.7	5.0	9.9	5.7	10.5
30%-60%	14.6	20.1	18.0	24.1	1.0	2.4	1.3	2.9
60%-100%	9.8	8.3	13.0	12.3	0.4	0.6	0.5	0.8
Total population (mn)	6.97	15.09	6.97	15.09	248.7	305.3	248.7	305.3
Asians as a percent of total population	2.8	4.9	2.8	4.9				

Finally, Table 4 presents the 1990 and 2010 distributions for the American Indian, Native Hawaiian, and Alaska Native population, described as the "American Indian, Eskimo, and Aleut" population in 1990, and abbreviated here to AI/AN. The results show that the proportion of this population living in the lowest density stratum has increased between the two censuses, and also that that stratum includes all but a small percentage of the total population. These data suggest that even with any benefits from geographic oversampling for this population, the amount of screening needed will be very large, as noted by Waksberg et al. (1997).

Table 4: Residential clustering of American Indians, Eskimos, and Aleuts in 1990 and American Indians and Alaska Natives in 2010 (all ages)

Density stratum (AI/AN, as a percent of the stratification unit)	*Percentage of the AI/AN population living in the stratum in the indicated year* (A_h)				*Percentage of the total population living in the stratum in the indicated year* (W_h)			
Stratification unit	*BG*		*Block*		*BG*		*Block*	
Measurement year	*1990*	*2010*	*1990*	*2010*	*1990*	*2010*	*1990*	*2010*
<5%	50.3	59.5	34.6	39.3	98.3	98.1	97.4	96.6
5%-10%	7.4	6.7	12.1	14.4	0.8	0.9	1.4	2
10%-30%	12.4	10.9	15.9	16.4	0.6	0.6	0.8	1
30%-60%	6.0	5.4	7.7	7.0	0.1	0.1	0.1	0.2
60%-100%	23.8	17.6	29.6	22.9	0.2	0.2	0.2	0.2
Total population (mn)	1.79	2.89	1.79	2.89	248.7	305.3	248.7	305.3
AI/AN as a percent of total population	0.7	0.9	0.7	0.9				

While Tables 1 to 4 provide some insight into the effectiveness of geographic oversampling for the various race/ethnicity populations, it is useful to examine the values of the variance reduction that might be achieved based on the 1990 and 2010 Census data. Table 5 presents the results based on VR_1 in equation (3) for the simple case when the ratio of the cost of a full interview to the cost of a screening interview is $c = 1$. The variance reductions in Table 5 represent the maximum variance reductions that could be achieved: the variance reductions are smaller the larger the value of *c* and also they will be smaller at other times than the census year because of changes in the composition of the block groups and blocks over time (see Waksberg et al., 1997, for an examination of this issue). The results in Table 5 show declines of around 10 percentage points in the value of VR_1 from 1990 to 2010 for all the populations except for the Asians and Pacific Islanders, and for both block groups and blocks.

Table 5: Percentage variance reduction achieved by geographic oversampling with optimum sampling fractions and the cost of a full interview equal to the cost of a screening interview (VR_1, $c = 1$)

Minority	*1990 BG*	*2010 BG*	*1990 Block*	*2010 Block*
Blacks	45.4	36.1	53.1	43.7
Hispanics	43.0	30.8	50.6	38.7
Asians	35.9	33.0	46.7	44.6
AI/AN	38.5	29.4	52.3	44.7

Another issue related to the changes in the distributions of the various race/ethnicity distributions between the two censuses are changes in the sampling rates for the different density strata. For example, with $c = 1,$ from equation (1) the optimum sampling rates for the strata are given by $f_h \propto \sqrt{P_h}$. Thus, for Blacks, the optimum oversampling rate for the highest density stratum based on blocks relative to the rate for lowest density stratum in 1990 is 8.16, whereas in 2010, the corresponding oversampling rate is 6.75. The corresponding results for Hispanics are an oversampling rate of 9.72 for the highest density stratum in 1990, as compared with 8.07 in 2010.

The variance reductions given in Table 5 assume a cost ratio of $c = 1,$ whereas in practice *c* will almost always exceed 1, often by a large amount. Table 6 presents the variance reductions associated with values of *c* greater than 1 in 2010 based on blocks, computed according to equation (2). As can be seen from Table 6, for Blacks and Hispanics the variance reductions achieved by geographic oversampling drop off rapidly as *c* increases. When *c* is 20 or so, the variance reductions are modest, and those reductions are based on the assumption that the optimum sampling fractions are used. In practice, the variance reductions will be less because of changes in the compositions of the blocks or block groups between the time of the Census and the survey data collection, and indeed the oversampling could lead to a loss in precision. As also reported by Waksberg et al. (1997), the decline in the variance reductions for Asian and Pacific Islanders and for American Indians and Alaska Natives with increasing *c* is much slower than for Blacks and Hispanics; even for high values of *c* there remains some useful gain in precision from the oversampling.

Table 6: Percentage variance reductions achieved by geographic oversampling with optimum sampling fractions based on block data in 2010 $(VR_1,\ c = 1)$.

Cost ratio c	*Blacks*	*Hispanics*	*Asians*	*AI/AN*
1	43.7	38.7	44.6	44.7
3	29.1	24.4	36.7	40.9
5	21.4	17.4	31.2	38.2
10	12.1	9.4	22.2	32.9
20	5.7	4.4	13.2	25.9
30	3.5	2.6	8.7	21.1

In addition to basic demographics, the population censuses also collect data on housing ownership. Geographic oversampling for, say, rented accommodation can yield some limited variance reduction at the block level. With $c = 1,$ the variance reduction is $VR_1 = 22$ percent; with *c* much greater than 1, there will be only modest reductions at best.

4. Comparisons of the Effectiveness of Oversampling in 2010 by using Different Density Strata

This section examines the effect of the use of different cutpoints than those used in Waksberg et al. (1997) with the concern that, particularly for the Hispanic and Asian populations which have grown in a major way since 1990, the Waksberg et al. cutpoints might not work well in 2010. Two alternative ways of establishing the cutpoints were

used: the cumulative (cum.) $\sqrt{f}$ rule described in Cochran (1977); and an optimal procedure that established the cutpoints from a program that considered all possible cutpoints, working through the possibilities in increments of one percent. The number of strata formed was held constant.

This investigation of the effect of alternative cutpoints was carried with target populations of persons aged 18 and over with any response of a given minority to the multiple race question. The results are presented in Table 7. The most striking finding from this analysis is that, within the range of options considered, the variance reductions achieved (VR_1) show little variability. Only in the case of Blacks do the cumulative $\sqrt{f}$ rule and the optimal procedure show noticeable improvements in the variance reduction over that obtained by the original procedure.

Table 7: Percentage variance reductions achieved by geographic oversampling with optimum sampling fractions using the original strata, strata formed using the cumulative $\sqrt{f}$ rule, and strata formed using the optimal procedure $(VR_1,\ c=1)$.

Minority	*2010 Census BG*			*2010 Census block*		
	Original	*Cum.* $\sqrt{f}$	*Optimal*	*Original*	*Cum.* $\sqrt{f}$	*Optimal*
Blacks	35.0	37.9	38.0	42.3	46.7	46.8
Hispanics	32.4	32.4	32.5	39.8	39.9	39.9
Asians	31.6	31.4	31.8	41.9	41.8	42.1
AI/AN	17.0	16.7	17.0	31.7	31.4	31.7
Rented Housing		11.6	11.8		22.0	23.2

5. Applications to Census Regions, CBSAs, and non-CBSAs

This section first examines the effectiveness of geographic oversampling when the target population is a subdivision of the country. It then investigates the effectiveness of geographic oversampling when the national target population is first divided into major strata using these subdivisions.

Table 8 presents results for subdivisions of the country defined by Census region and by CBSAs or non-CBSAs. The table shows appreciable variability in the variance reductions across the regions for all the race/ethnic groups. For example, the variance reductions in the West are lower, and those in the Midwest are higher, than those for other regions, with the exception of the AI/AN population.

Geographic oversampling is particularly effective for all race/ethnicity groups in non-CBSAs. Over 90 percent of the Blacks in non-CBSAs live in the South, and 71 percent of them live in the highest density stratum comprising blocks in which 60 percent or more are Black, whereas 66 percent of the total population in non-CBSAs in the South live in the lowest density stratum comprising blocks in which under 10 percent are Black. Over 80 percent of the Hispanics in non-CBSAs live in the South or West, and 54 percent of them live in the highest density stratum in which 60 percent or more are Hispanic, whereas 77 percent of the total population in the South or West live in the lowest density stratum comprising blocks in which under 5 percent are Hispanic. These examples illustrate why geographic oversampling is so effective in non-CBSAs.

Table 8: Percentage variance reduction achieved in Census regions and CBSAs and non-CBSAs achieved by geographic oversampling with optimum sampling fractions using 2010 block data (VR_1, $c = 1$).

Subdivision	*Blacks*	*Hispanics*	*Asians*	*AI/AN*
National	46.8	39.9	42.1	31.7
Northeast region	47.2	39.7	39.5	16.7
Midwest region	55.2	44.8	41.2	34.6
South region	40.4	41.2	37.1	32.2
West region	35.4	24.9	34.0	31.0
CBSA	45.4	38.5	40.7	27.3
Non-CBSA	71.4	60.9	48.8	64.3

The effectiveness of geographical oversampling in combination with stratification by either region or CBSA/non-CBSA is examined in Table 9. For comparison purposes, the first row of the table repeats the values of VR_1 from Table 8, with only density stratification. The second two rows give results when stratification is first carried out by major strata, region. and CBSA/non-CBSA respectively, with proportionate stratification based on the total population counts. The density stratification is then carried out within each major stratum separately, using the optimal procedure described in Section 4, thus potentially producing different cutpoints by major stratum. The variance reduction for this design is then given by VR_{ps1} in equation (4). As is to be expected, under the assumptions made in deriving VR_{ps1}, the second two rows of the table show that this design is somewhat less effective than the design without major stratification. However, this form of stratification may serve other purposes.

An alternative way of combining major strata and density strata is to create *KH* strata by forming *H* density strata within each of the *K* major strata, using the optimum cutpoints within each major stratum, as outlined in Section 2, and then using the optimum sampling rates given by equation (1) in the resultant strata. The variance reduction for this design is then given by VR_{KH1} in equation (5). As can be seen by comparing the last two rows of Table 9 with the first row of the table, the gains in variance reduction from the addition of either type of major stratification with this design versus those for the design with only the basic density stratification are very small.

In passing, it may be noted that basing the geographical oversampling on regional or on CBSA/non-CBSA strata alone yields very little benefit. With oversampling just by region, the variance reductions are around 4 percent to 8 percent, and with oversampling just by CBSA/non-CBSA, they are from under 1 percent to 3 percent.

Table 9: Percentage variance reductions without major strata, with region or CBSA/non-CBSA used for initial proportionate stratification, and with the combination of major and density stratification, using 2010 block data and the cost of a full interview equal to the cost of a screening interview $(c = 1)$.

Stratification	*Blacks*	*Hispanics*	*Asians*	*AI/AN*
No major stratification: VR_1	46.8	39.9	42.1	31.7
Density within region: VR_{ps1}	43.7	34.7	36.5	30.6
Density within CBSA/non-CBSA: VR_{ps1}	46.4	39.0	40.8	32.0
Region × density: VR_{KH1}	46.8	40.0	42.3	33.0
CBSA/non-CBSA × density: VR_{KH1}	47.0	40.1	42.6	32.0

6. Concluding Remarks

The results presented in this paper show that, provided that the relative cost of the full data collection to the screening cost is not great, geographic oversampling remains an effective means of sampling minority populations in national surveys. However, the benefits of the approach are noticeably weaker than they were in 1990. Also, as demonstrated by Waksberg et al. (1997), the gains from the geographic oversampling decline later in the decade as the Census data become more dated. The variance reductions do vary by region and are particularly large for all minorities in non-CBSAs.

The results on the effect of different choices of cutpoints are reassuring. The variance reductions seem to be fairly robust for modest departures from the optimum cut-points. The simple cumulative $\sqrt{f}$ rule appears to work well. Initial stratification by region and by CBSA/non-CBSA does not add much benefit for oversampling minorities.

Two limitations of this research should be noted. First, the basic theory assumes a single stage sample with SRS within the density strata, whereas in practice multistage sampling is generally needed. See Clark (2009) for an approach using two-stage designs. Second, the research focuses on sampling a single minority population, whereas in practice surveys are often designed to estimate parameters for several minorities as well as the total population. These issues will be taken up in future research.

Acknowledgements

We thank Yuki Nakamoto and Baifan Li for their programming support of this research.

References

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

Cochran, W.G. 1977. *Sampling Techniques*, 3rd ed. Wiley, New York.

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

Kalton, G. 2003. Practical methods for sampling rare and mobile populations. *Statistics in Transition*, 6, 491-501.

Waksberg, J., Judkins, D. and Massey, J.T. 1997. Geographic-based oversampling in demographic surveys of the United States. *Survey Methodology,* 23, 61-71.