

## Avoiding Sequential Sampling with Random Digit Dialing

J. MICHAEL BRICK and JOSEPH WAKSBERG<sup>1</sup>

### ABSTRACT

The Mitofsky-Waksberg procedure is an efficient method for selecting a self-weighting, random digit dialing (RDD) sample of households. The Mitofsky-Waksberg procedure is sequential, requiring a constant number of households be selected from each cluster. In this article, a modified Mitofsky-Waksberg procedure which is not self-weighting or sequential is described. The bias and variance for estimates derived from the modified procedure are investigated. Suggestions on circumstances which might favor the modified procedure over the standard Mitofsky-Waksberg procedure are provided.

**KEY WORDS:** Random digit dialing; Telephone sampling; Cluster sampling; Trimming.

### 1. INTRODUCTION

The Mitofsky-Waksberg procedure for selecting random digit dialing samples of households (Waksberg 1978) is frequently used for sample selection in telephone surveys. As described in the Waksberg paper, it is an efficient method of producing a self-weighting sample, that is, one in which all telephone households have the same probability of selection (except for households with more than one telephone number). The efficiency is due to the sharp reduction in the proportion of nonhousehold telephone numbers that have to be dialed in order to identify sample households.

The Mitofsky-Waksberg procedure is a two-stage sample design. In the first stage, a sample of clusters is chosen where the clusters consist of blocks of 100 telephone numbers, or multiples of such blocks. The clusters (or blocks of 100 telephone numbers) are first selected with equal probability. One telephone number is chosen at random in each cluster and dialed. If the number is that of a household, the cluster is retained. Otherwise, it is rejected. The second stage is the selection of households within the retained sample clusters. For the self-weighting feature of the sample to apply, a constant number of households per cluster is required. Some organizations (including Westat Inc.) generally go a little further and specify a constant number of interviewed households per cluster (or screened households if the first part of data collection is screening). The rationale is that substituting another randomly selected household within the same cluster for each nonrespondent is a reasonable way of reducing nonresponse bias.

There is an awkward operational feature to this system. It sometimes takes a fairly large number of callbacks to determine whether or not a telephone number is residential, particularly for numbers that repeatedly ring with no answer. Even more are needed to learn which households cooperate. Such determinations must be made for an initially selected sample to ascertain which clusters require more telephone numbers to achieve the desired cluster size and how many telephone numbers have to be added. In effect, a sequential scheme is necessary for each cluster, where all previous cases need to be cleared up before it is known whether the sample needs to be increased. This process is particularly inconvenient when there is a tight time schedule for data collection.

Several attempts to modify the Mitofsky-Waksberg method have been proposed which reduce or eliminate the sequential features of the plan. Potthoff (1987) developed a generalization

---

<sup>1</sup> J. Michael Brick and Joseph Waksberg, Westat Inc., 1650 Research Boulevard, Rockville, Maryland 20850, U.S.A.

of the Mitofsky-Waksberg technique in which  $c$  telephone numbers are chosen per cluster in determining whether to retain the cluster, whereas Mitofsky-Waksberg use only one. A self-weighting sample is achieved by having clusters in which only one of the  $c$  telephone numbers dialed continue with a sampling plan that includes having a fixed number of households per cluster, and the remaining clusters having a fixed number of telephone numbers. The latter group of clusters does not require a sequential approach. Potthoff reports that in practice most clusters will fall into the second class so that the sequential operations, although not eliminated, are sharply reduced.

Lepkowski and Groves (1986) describe a sampling method in which blocks of telephone numbers with more than a trivial number of telephone numbers listed in directories (and other sources, if available) are selected with probability proportionate to the number of listed numbers. Blocks of numbers which contain zero or very few telephone numbers are sampled through the Mitofsky-Waksberg procedure. Sudman (1973) had previously proposed sampling blocks of numbers with probability proportionate to the numbers listed in directories, but without making any provision for empty blocks (which could have unlisted numbers). Drew and Jaworski (1986) describe an RDD survey carried out in Canada in which purchased counts of residential numbers (both published in directories and nonpublished) were used as measures of size. Since the counts were considered as virtually complete, there was no need to sample empty blocks. As far as we are aware, there is no way of getting virtually complete counts of residential numbers in the U.S.

Neither the Potthoff nor the Lepkowski-Groves sample design completely eliminates the need for a sequential process, although both appear to reduce the portion of the sample which requires it. There are also some other disadvantages to the two procedures. The Potthoff technique appears to be rather complex – as far as we know it has not been used much for RDD surveys. For national surveys, the Lepkowski-Groves technique involves the purchase of a directory list covering the total U.S. and processing it to obtain measures of size. Such commercial lists are available, but they are expensive. Furthermore, a number of recent reports indicate the percentage of all residential numbers that are listed in directories is not very high, and is rapidly decreasing. Tucker (1989) describes an analysis of listed numbers in a group of U.S. counties and cities which shows listing rates varying from 48 to 62 percent. An article by Linda Piekarski (1990) states that if the rate of increase of unlisted numbers continues at the current level, “as many as 62% of the nation’s households may be unlisted by the year 2000.” The measures of size thus are probably only moderately correlated with the actual number of households in a working block.

Waksberg has suggested an alternative modification of the Mitofsky-Waksberg procedure (Waksberg 1984) which completely eliminates the need for sequential sampling. Westat has used this method in a large number of studies using RDD. Cummings (1979) had previously used the same procedures as a result of an error in implementing the Mitofsky-Waksberg procedure. Cummings did not recognize its usefulness in avoiding sequential sampling and did not explore its features for use in other surveys. We describe the method and its mathematical and statistical properties.

## **2. ALTERNATIVE METHOD OF ESTABLISHING CLUSTER SIZES WITH MITOFSKY-WAKSBERG TECHNIQUE**

As indicated earlier, the Mitofsky-Waksberg technique requires a constant number of sample residential numbers per cluster (or block of numbers) to produce a self-weighting sample. The alternative that is proposed is to use a constant number of telephone numbers per cluster for

the sample ( $K$ ). The first stage of selection is unchanged. (The first-stage selects clusters with probability proportionate to the number of households.) With a constant number of telephone numbers per cluster, the sample numbers can be designated in advance eliminating the sequential process. We note that followup effort is still necessary to determine which sample telephone numbers are residential, both in the first and in the second stages of sampling. However, this has to be done for a fixed set of telephone numbers. A sequential process is not involved.

The alternative procedure does not produce a self-weighting sample. Since the first stage is selected with PPS, the probability of a cluster being selected is  $r N_i/100$  where  $r$  is the sampling rate for selection of the clusters, that is, the first stage selection rate, and  $N_i$  is the number of residential numbers in the  $i$ th cluster. The weight should be proportional to  $N_i^{-1}$ , but since  $N_i$  is not known, it is taken to be proportional to  $n_i^{-1}$ , the number of sample households in the cluster.

This modification of the Mitofsky-Waksberg method has good features for survey operations. It is simple. The sample can be virtually preselected and no costly control operations are needed. Although weighting is required, the weights are directly available from the sample data, and they can be mechanically produced without any extensive professional oversight.

There are, however, some serious problems. First, there is a bias when  $N_i^{-1}$  is estimated by  $K/100n_i$  where  $K$  is the number of telephone numbers selected per cluster (a constant number in all clusters). The bias is fairly small, but it does exist. It cannot be eliminated or reduced by minor modifications of the weights, such as using  $1/(n_i + t)$  instead of  $n_i^{-1}$ , with “ $t$ ” denoting a fixed constant. Secondly, the introduction of variable weights increases the variances of the estimates substantially. (The increase is not so much caused by the weights as by the fact they reflect variable probabilities of selection.) Finally, the modification loses one of the useful features of the Mitofsky-Waksberg method – the ability to fix the exact sample size desired. The Mitofsky-Waksberg method’s use of a constant number of households per cluster means that any desired sample size can be obtained by selecting a sample with the appropriate number of clusters. With the modification, the sample size becomes a random variable, which generally will not be exactly equal to the desired sample size. Although the deviations are usually small, the ability to achieve exact target sizes is useful when contracts or budget commitments require the survey organization to satisfy exact target requirements. We discuss these issues in Sections 3 and 4.

Before going on to a discussion of the variances and biases, it is useful to examine the distribution of cluster sizes in the U.S. Tables 1 to 3 show estimates of such distributions prepared from data reported in two large national U.S. surveys conducted via RDD by Westat Inc. Both of these surveys used the modification of the Mitofsky-Waksberg procedure described above. The sample for the survey summarized in Table 1 was selected in 1986 and consisted of 2,427 clusters (retained after first-stage sampling) with 15 telephone numbers per cluster, or 36,405 total numbers. There were 18,756 completed screeners, 2,396 refusals, 1,727 nonresponse for other reasons, and 13,526 nonresidential or nonworking numbers, ring no answers, and cases that could not be classified. The analysis is restricted to the 18,756 completed cases. The data in Tables 2 and 3 are based on a 1989 sample of 1,000 clusters with 30 telephone numbers per cluster or 30,000 telephone numbers, of which 19,586 were residential with screeners completed. Table 2 shows the distribution of the 15,030 completed cases and Table 3 shows the distribution of the 19,586 residential numbers found in the 1,000 clusters. The cluster weights shown are expressed as  $\bar{n}/n_i$  where  $\bar{n}$  is the average number of households per cluster. It seems useful to express them in this form since they then show the deviations from a self-weighting sample. The design effects only account for the increased variances arising from variable sampling fractions. They do not include effects of other aspects of the sample design.

**Table 1**  
 Number of Completed Screeners per Cluster in 1986 Survey  
 (Based on sample of 2,427 clusters with 15 telephone numbers per cluster)

| Number of<br>Completes<br>per Cluster | Average<br>Cluster<br>Weight <sup>1</sup> | Household Distribution |         |                       | Cluster Distribution |         |                       |
|---------------------------------------|---|------------------------|---------|-----------------------|----------------------|---------|-----------------------|
|                                       |   | Frequency              | Percent | Cumulative<br>Percent | Frequency            | Percent | Cumulative<br>Percent |
| 0                                     | xx  | 0                      | 0       | 0.0                   | 62                   | 2.6     | 2.6                   |
| 1                                     | 7.93 <sup>2</sup>                         | 54                     | 0       | 0.3                   | 54                   | 2.2     | 4.8                   |
| 2                                     | 3.97 <sup>2</sup>                         | 106                    | 0.6     | 0.9                   | 53                   | 2.2     | 7.0                   |
| 3                                     | 2.64                                      | 258                    | 1.4     | 2.2                   | 86                   | 3.5     | 10.5                  |
| 4                                     | 1.98                                      | 440                    | 2.3     | 4.6                   | 110                  | 4.5     | 15.0                  |
| 5                                     | 1.59                                      | 810                    | 4.3     | 8.9                   | 162                  | 6.7     | 21.7                  |
| 6                                     | 1.32                                      | 1,290                  | 6.9     | 15.8                  | 215                  | 8.9     | 30.6                  |
| 7                                     | 1.13                                      | 1,960                  | 10.5    | 26.2                  | 280                  | 11.5    | 42.1                  |
| 8                                     | 0.99                                      | 2,656                  | 14.2    | 40.4                  | 332                  | 13.7    | 55.8                  |
| 9                                     | 0.88                                      | 2,862                  | 15.3    | 55.6                  | 318                  | 13.1    | 68.9                  |
| 10                                    | 0.79                                      | 2,990                  | 15.9    | 71.6                  | 299                  | 12.3    | 81.2                  |
| 11                                    | 0.72                                      | 2,717                  | 14.5    | 86.1                  | 247                  | 10.2    | 91.4                  |
| 12                                    | 0.66                                      | 1,548                  | 8.3     | 94.3                  | 129                  | 5.3     | 96.7                  |
| 13                                    | 0.61                                      | 780                    | 4.2     | 98.5                  | 60                   | 2.5     | 99.2                  |
| 14                                    | 0.57                                      | 210                    | 1       | 99.6                  | 15                   | 0.6     | 99.8                  |
| 15                                    | 0.53                                      | 75                     | 0       | 100.0                 | 5                    | 0.2     | 100.0                 |
| Total                                 |   | 18,756                 | 100.0   | xx                    | 2,427                | 100.0   | xx                    |
| Mean cluster size <sup>3</sup>        |   |                        | 7.93    |                       |                      |         |                       |
| Design effect <sup>4</sup>            |   |                        | 1.31    |                       |                      |         |                       |

<sup>1</sup> The cluster weight is the mean cluster size (*i.e.*, 7.93) divided by the number of completes in the *i*-th cluster.

<sup>2</sup> Trimming the weights would bring these weights down to 3.

<sup>3</sup> The mean cluster size is the average over the 2,365 clusters with one or more completed screeners.

<sup>4</sup> The design effect is reduced to 1.12 if the maximum weight is 3.

It should be noted that Table 1 is based on a sample of 15 telephone numbers per cluster and Tables 2 and 3 used 30 telephone numbers per cluster. Estimates of the percent residential in a cluster based on 15 telephone numbers will, of course, be subject to a higher sampling error than an estimate based on 30 telephone numbers. However, the number of clusters used in Table 1 was more than twice those in Tables 2 and 3 which should largely offset the effect of the different cluster sizes.

There are two differences between Tables 2 and 3. One is that Table 2 shows the distribution of completed screeners (as does Table 1) while Table 3 is based on all sample households. The use of only completed cases in Table 2 reduces the estimate of the average number of households per cluster and shifts the entire distribution. In addition, it introduces more variability to the estimates of the distribution shown because the distributions reflect sampling errors of both the distribution of households per cluster and the distribution of nonresponse rates per cluster. The second difference is that Table 2 (and Table 1) is expressed in terms of the number of cases per cluster and Table 3 shows the distributions by the percentage of residential numbers per cluster. It was convenient to express Table 3 in that form for analyses described later in this report.

One other feature of the percentages shown in Tables 1 to 3 should be noted. They reflect the size distributions of clusters which fell into the sample, not the distribution of clusters in the U.S. The use of probability proportionate to size sampling results in an oversampling of clusters with a high proportion of residential numbers and an underrepresentation of clusters with a small number. It is possible to convert the distribution from one that represents the sample to one that represents the population by multiplying each percentage by the cluster weights and computing the percentage distribution of the resulting figures. Since the weights are exactly proportional to the reciprocal of the number of completes per cluster, it turns out that converting the household distribution so that it represents the distribution in the population produces the percentages shown in the cluster distribution. The cluster distribution in the sample is thus the same as the household distribution in the population.

We show distributions of both all-sample households and completed cases because both are of interest to researchers. The Table 3 data have been used for the analyses in Sections 3 and 4.

**Table 2**  
Number of Completed Screeners per Cluster in 1989 Survey  
(Based on sample of 1,000 clusters with 30 telephone numbers per cluster)

| Number of<br>Completes<br>per Cluster | Average<br>Cluster<br>Weight <sup>1</sup> | Household Distribution |         |                       | Cluster Distribution |         |                       |
|---------------------------------------|---|------------------------|---------|-----------------------|----------------------|---------|-----------------------|
|                                       |   | Frequency              | Percent | Cumulative<br>Percent | Frequency            | Percent | Cumulative<br>Percent |
| 0                                     | xx  | 0                      | 0       | 0.0                   | 8                    | 0.8     | 0.8                   |
| 1 or 2                                | 7.57 <sup>2</sup>                         | 6                      | 0       | 0.0                   | 3                    | 0.3     | 1.1                   |
| 3 or 4                                | 4.33 <sup>2</sup>                         | 37                     | 0.2     | 0.3                   | 10                   | 1.0     | 2.1                   |
| 5 or 6                                | 2.75                                      | 126                    | 0.8     | 1.1                   | 22                   | 2.2     | 4.3                   |
| 7 or 8                                | 2.02                                      | 403                    | 2.7     | 3.8                   | 53                   | 5.3     | 9.6                   |
| 9 or 10                               | 1.59                                      | 688                    | 4.6     | 8.4                   | 72                   | 7.2     | 16.8                  |
| 11 or 12                              | 1.32                                      | 1,325                  | 8.8     | 17.2                  | 115                  | 11.5    | 28.3                  |
| 13 or 14                              | 1.12                                      | 1,987                  | 13.2    | 30.4                  | 147                  | 14.7    | 43.0                  |
| 15 or 16                              | 0.98                                      | 2,636                  | 17.5    | 50.0                  | 170                  | 17.0    | 60.0                  |
| 17 or 18                              | 0.85                                      | 2,692                  | 17.9    | 65.9                  | 154                  | 15.4    | 75.4                  |
| 19 or 20                              | 0.78                                      | 2,387                  | 15.9    | 81.8                  | 123                  | 12.3    | 87.7                  |
| 21 or 22                              | 0.70                                      | 1,673                  | 11.1    | 92.9                  | 78                   | 7.8     | 95.5                  |
| 23 or 24                              | 0.64                                      | 816                    | 5.4     | 98.3                  | 35                   | 3.5     | 99.0                  |
| 25 or 26                              | 0.55                                      | 254                    | 1.7     | 100.0                 | 10                   | 1.0     | 100.0                 |
| 27 or 28                              | xx  | 0                      | 0       | 100.0                 | 0                    | 0       | 100.0                 |
| 29 or 30                              | xx  | 0                      | 0       | 100.0                 | 0                    | 0       | 100.0                 |
| Total                                 | xx  | 15,030                 | xx      | xx                    | 1,000                | xx      | xx                    |
| Mean cluster size <sup>3</sup>        |   |                        |         | 15.11                 |                      |         |                       |
| Design effect <sup>4</sup>            |   |                        |         | 1.33                  |                      |         |                       |

<sup>1</sup> The cluster weight is the mean cluster size (*i.e.*, 15.15) divided by the number of completes in the *i*-th cluster.

<sup>2</sup> Trimming the weights would bring these weights down to 3.

<sup>3</sup> The mean cluster size is the average over the 992 clusters with one or more completed screeners.

<sup>4</sup> The design effect is reduced to 1.12 if the maximum weight is 3.

**Table 3**  
 Proportion of Residential Numbers per Cluster in 1989 Survey  
 (Based on sample of 1,000 clusters with 30 telephone numbers per cluster)

| Proportion of<br>Residential<br>nos. per Cluster | Average<br>Cluster<br>Weight <sup>1</sup> | Household Distribution |         |                       | Cluster Distribution |         |                       |
|--|---|------------------------|---------|-----------------------|----------------------|---------|-----------------------|
|  |   | Frequency              | Percent | Cumulative<br>Percent | Frequency            | Percent | Cumulative<br>Percent |
| 0  | xx  | 0                      | 0.0     | 0.0                   | 5                    | 0.5     | 0.5                   |
| .001 to .049                                     | 21.76 <sup>2</sup>                        | 5                      | 0.0     | 0.0                   | 3                    | 0.3     | 0.8                   |
| .05 to .099                                      | 8.70 <sup>2</sup>                         | 18                     | 0.1     | 0.1                   | 6                    | 0.6     | 1.4                   |
| .10 to .149                                      | 5.22 <sup>2</sup>                         | 41                     | 0.2     | 0.3                   | 9                    | 0.9     | 2.3                   |
| .15 to .199                                      | 3.73 <sup>2</sup>                         | 48                     | 0.2     | 0.6                   | 8                    | 0.8     | 3.1                   |
| .20 to .249                                      | 2.90                                      | 53                     | 0.3     | 0.8                   | 7                    | 0.7     | 3.8                   |
| .25 to .299                                      | 2.37                                      | 144                    | 0.7     | 1.6                   | 16                   | 1.6     | 5.4                   |
| .30 to .349                                      | 2.01                                      | 178                    | 0.9     | 2.5                   | 17                   | 1.7     | 7.1                   |
| .35 to .399                                      | 1.74                                      | 408                    | 2.1     | 4.6                   | 34                   | 3.4     | 10.5                  |
| .40 to .449                                      | 1.54                                      | 459                    | 2.3     | 6.9                   | 34                   | 3.4     | 13.9                  |
| .45 to .499                                      | 1.37                                      | 840                    | 4.3     | 11.2                  | 56                   | 5.6     | 19.5                  |
| .50 to .549                                      | 1.24                                      | 1,040                  | 5.3     | 16.5                  | 63                   | 6.3     | 25.8                  |
| .55 to .599                                      | 1.14                                      | 1,926                  | 9.8     | 26.3                  | 107                  | 10.7    | 36.5                  |
| .60 to .649                                      | 1.04                                      | 2,126                  | 10.9    | 37.2                  | 109                  | 10.9    | 47.4                  |
| .65 to .699                                      | 0.97                                      | 3,255                  | 16.6    | 53.8                  | 155                  | 15.5    | 62.9                  |
| .70 to .749                                      | 0.90                                      | 2,610                  | 13.3    | 67.1                  | 116                  | 11.6    | 74.5                  |
| .75 to .799                                      | 0.84                                      | 3,022                  | 15.4    | 82.6                  | 126                  | 12.6    | 87.1                  |
| .80 to .849                                      | 0.79                                      | 1,556                  | 7.9     | 90.5                  | 61                   | 6.1     | 93.2                  |
| .85 to .899                                      | 0.75                                      | 1,458                  | 7.4     | 98.0                  | 54                   | 5.4     | 98.6                  |
| .90 to .949                                      | 0.71                                      | 399                    | 2.0     | 100.0                 | 14                   | 1.4     | 100.0                 |
| .95 to .999                                      | xx  | 0                      | 0.0     | 100.0                 | 0                    | 0.0     | 100.0                 |
| Total  | xx  | 19,586                 | 100.0   | xx                    | 1000                 | 100.0   | xx                    |
| Mean cluster size <sup>3</sup>                   |   |                        | 19.68   |                       |                      |         |                       |
| Design effect <sup>4</sup>                       |   |                        | 1.28    |                       |                      |         |                       |

<sup>1</sup> The cluster weight is the mean proportion in a cluster (*i.e.*, 0.653) divided by the proportion of residential numbers in the *i*-th cluster.

<sup>2</sup> Trimming the weights would bring these weights down to 3.

<sup>3</sup> The mean cluster size is the average over the 995 clusters with one or more residential numbers.

<sup>4</sup> The design effect is reduced to 1.12 if the maximum weight is 3.

### 3. VARIANCE IMPLICATIONS OF THE MODIFIED MITOFSKY-WAKSBERG METHOD

In the standard Mitofsky-Waksberg method the variance of a sample estimate is dependent upon the number of households selected per cluster and the homogeneity of the households within and between clusters. The variance for a cluster sample can be written as the variance for a simple random sample multiplied by  $[1 + \rho(\bar{n} - 1)]$ , where  $\rho$  is intraclass correlation and  $\bar{n}$  is the average number of households per cluster. Since telephone clusters are often related to geographic areas and tend to be somewhat homogeneous, selecting a large number of households per cluster can be inefficient.

When the modified Mitofsky-Waksberg method is used, another source of variance is introduced because the number of households selected per cluster is allowed to vary from cluster to cluster. As pointed out in Section 2, the denominator of the second stage probability of selection does not cancel with the number of households in the cluster (which is proportional to the probabilities in the first stage) and the overall probabilities of selecting households vary from cluster to cluster.

The variability among clusters in the overall household sampling rates causes the variances of the estimates to be larger than those in the standard Mitofsky-Waksberg method where each household has the same probability of selection. Methods for estimating the increase in the variance of an estimate arising from unequal probabilities of selection are discussed by Kish (1965) and by Waksberg (1973). A simple approximation to the variance of an estimate under an unequal weighting scheme (where the weights do not reflect variable sampling rates in strata deliberately chosen to reduce sampling variances) is the sampling variance which would occur with a self-weighting sample multiplied by a variance inflation factor (VIF), given by  $VIF = \{1 + \text{Relvar(weights)}\}$ . We will use this approximation below to investigate the variance implications associated with the modified Mitofsky-Waksberg method.

The relative variance of the weights was computed by partitioning the process into two components. First, the mean and variance of the weights were computed conditioned upon sampling from a truncated (since zero households cannot be obtained if the cluster is sampled in the first stage) hypergeometric distribution, defined by the household density in the cluster and the cluster sample size. The unconditional mean and variance of the weights were then computed by integrating over the distribution of households in the sampled clusters shown in Table 3. The distribution of households in the sample is critical in the evaluation of the VIF.

The natural weight assigned to a household in the modified Mitofsky-Waksberg is proportional to  $n_i^{-1}$ , where  $n_i$  is the number of households observed in sample cluster  $i$ . This weight can vary by factors which range from as little as  $1/K$  to 1, where  $K$  is the number of telephone numbers selected in a cluster. The average weight is roughly  $1.5/K$ , since about 65 percent of numbers in the sampled clusters are residential.

If the number of telephone numbers sampled per cluster is between 5 and 30, then the increase in variance due to the weighting is about 30 percent. The VIF decreases slightly as the number sampled per cluster increases beyond 30, reaching approximately 17 percent when all the numbers in the cluster are sampled.

The VIF or the relative variance of the weights is a function of the distribution of the number of households across clusters and random sampling variability within the clusters. This decomposition is made explicit by expressing the variance of the weights as the sum of the mean of the conditional variance of the weights and the variance of the conditional mean of the weights, where the conditioning is with respect to the household density of the cluster.

When the cluster sample size is small, the mean of the conditional variance is the dominant component of the overall variance. As the cluster sample size increases, the variance of the conditional mean becomes more dominant. This is why the relative variance of the weights, shown in the first row of Table 4, is not a monotonic function of the cluster sample size.

**Table 4**  
Approximate Variance Inflation Factors (VIF) for Modified  
Mitofsky-Waksberg Random Digit Dialing Samples

| Weight         | Cluster Sample Size ( $K$ ) |      |      |      |      |
|----------------|-----------------------------|------|------|------|------|
|                | 5                           | 10   | 30   | 60   | 100  |
| $1/n_i$        | 1.31                        | 1.34 | 1.29 | 1.23 | 1.17 |
| $1/(n_i + .5)$ | 1.18                        | 1.21 | 1.20 | 1.18 | 1.16 |
| $1/(n_i + 1)$  | 1.12                        | 1.15 | 1.16 | 1.15 | 1.14 |
| $1/(n_i + 2)$  | 1.07                        | 1.09 | 1.11 | 1.12 | 1.13 |

### Variances Using Different Weights

Weights other than ones proportional to the inverse of the number of households were also examined to determine their impact on the bias and variance of the estimates. Many of the alternative weights studied were derived from variance stabilizing transformations suggested for binomial variables.

Of all the alternatives examined, the estimators with the best bias and variance properties involved simple adjustments of the natural weight. In particular, adding a small constant to the observed number of households (estimators of the form  $(n_i + t)^{-1}$  where  $t$  is .5, 1, or 2) resulted in reducing the increases in variance due to differential weighting. The addition of the constant reduces the range of the weights by cutting the values of the largest weights while only slightly modifying the weights for clusters where more households are found in the sample.

Table 4 shows the VIF for the estimators of the form  $(n_i + t)^{-1}$  for different numbers of telephone numbers sampled per cluster. The table also is based on the household and cluster distributions shown in Table 3. It is clear from the table that a substantial reduction in the variance due to unequal weighting can be achieved by using  $(n_i + 1)^{-1}$ , rather than the natural estimator. This is especially true for RDD designs which sample 30 or fewer telephone numbers per cluster. The increase in variance due to differential weighting for  $(n_i + 1)^{-1}$  is only 16 percent when 30 numbers are selected per cluster as opposed to a 29 percent increase when the natural estimator is used.

### Variances with Trimmed Weights

A practice that is often used to mitigate the variance inflation associated with varying weights is the truncation of very large weights. This truncation, or trimming of weights, is usually fixed at a weight above which relatively few observations are found. In many Westat RDD samples, weights that exceed two or three times the mean weight have been truncated. For this research, we have examined weights truncated at about 3 times the mean weight. For samples of 10 per cluster, the weights were truncated at 2 times the mean weight because so few observations are affected otherwise.

Table 5 shows the VIF for the estimators for different cluster sample sizes when the weights are trimmed at three times the mean weight for  $n_i^{-1}$ . The VIF's for samples of 5 per cluster are not given because the truncation point in samples of this size is nearly at unity, the largest possible weight.

**Table 5**  
Approximate Variance Inflation Factors (VIF) for Modified Mitofsky-Waksberg  
Random Digit Dial Samples with Trimmed\* Weights

| Weight         | Cluster Sample Size ( $K$ ) |      |      |      |
|----------------|-----------------------------|------|------|------|
|                | 10                          | 30   | 60   | 100  |
| $1/n_i$        | 1.12                        | 1.11 | 1.09 | 1.09 |
| $1/(n_i + .5)$ | 1.11                        | 1.10 | 1.09 | 1.09 |
| $1/(n_i + 1)$  | 1.09                        | 1.10 | 1.09 | 1.09 |
| $1/(n_i + 2)$  | 1.07                        | 1.09 | 1.09 | 1.08 |

\* All weights trimmed at 3 times the mean weight, except samples of 10 trimmed at 2 times the mean.



The tabled values show trimming substantially reduces adverse impact of the differential weights on the variance of the estimates. The most dramatic reduction is for the natural estimator; its VIF is reduced by over 50 percent by the use of trimming. The VIFs for the other estimators are improved somewhat, but the reductions are less striking since they already had smaller VIF's than the natural estimator. Trimming has the potential of introducing biases which may counteract the advantage in variance reduction. Biases are discussed in Section 4.

### Variances with Augmented Sampling

A third approach to reducing the variability of the weights is the use of augmented sampling. Large weights occur when the number of households identified in the cluster is small relative to the expected number of households per cluster. To reduce the chance for this happening, an augmented sampling procedure can be used. If the number of households in a cluster is smaller than a fixed number (say less than one third of the mean number per cluster), then the sample size in the cluster can be doubled or increased by some other amount.

This procedure could be iterated to insure that the number of households per cluster reaches a specified limit or until all numbers in the cluster are used. The obvious disadvantage of this iterative plan is that it requires monitoring sample yield by cluster and the very fact that it is sequential. Another disadvantage of the method is that it results in sampling more telephone numbers from clusters that have a lower household density (the ones most likely to need augmentation), hence reducing productivity.

Despite the operational shortcomings of the augmented sampling approach, we did a limited examination of the method. Since the results for the augmented sample approach was not better than trimming the weights, this method is not discussed further.

## 4. BIAS IMPLICATIONS OF THE MODIFIED MITOFSKY-WAKSBERG METHOD

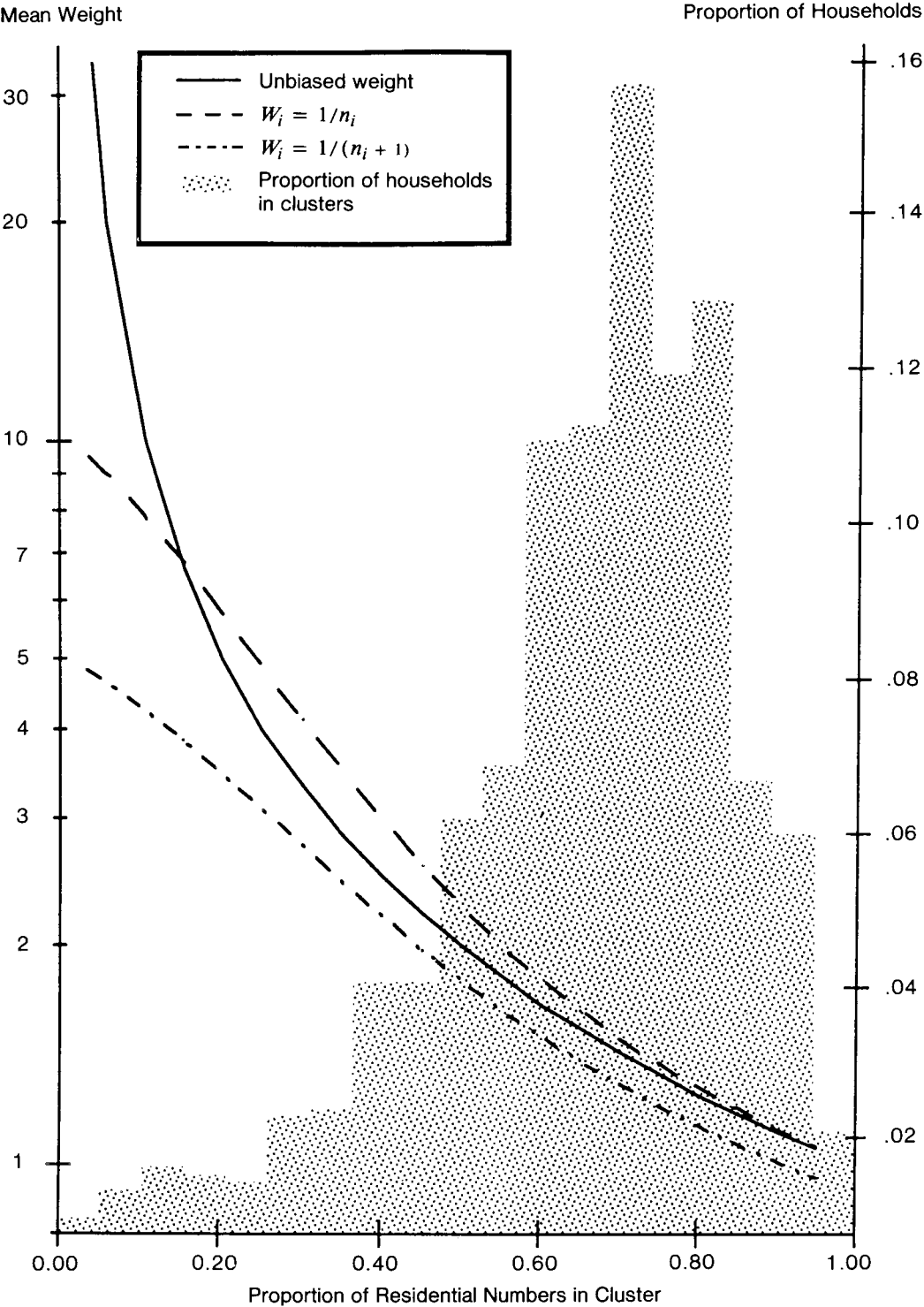
The increase in variance is just one of the consequences of using the modified Mitofsky-Waksberg method of sampling. Another important feature of the method is the bias in the resulting estimates. If a fixed sample size is selected in a cluster and no weight adjustment is made, the variance of the estimates are not increased but the bias has the potential of being very large.

The unbiased weight ( $W_u$ ) for the modified method is

$$W_u = \frac{100}{rN_i} \times \frac{100}{K},$$

where the terms are as defined above. The problem is that  $N_i$  is unknown and does not cancel with the second stage term, as it does in the standard Mitofsky-Waksberg method. Weights are therefore introduced in an effort to reduce the bias.

We refer to the estimator which uses a weight of  $n_i^{-1}$  as the natural estimator because  $n_i/K$  is an unbiased estimator of  $N_i/100$  in sampling from a binomial or hypergeometric distribution. (We use the weight of  $n_i^{-1}$  although the weight is actually  $K/100n_i$ . Since  $K/100$  is a constant, the relationship among the weights are not affected by using  $n_i^{-1}$ .) This weight appears to be the natural estimator despite the fact that  $n_i^{-1}$  is not unbiased for  $N_i^{-1}$  unless all 100 numbers are selected in a cluster. The bias of  $n_i^{-1}$  is discussed in literature; for example, see the discussion on stratification after sampling in Hansen, Hurwitz and Madow (1953). No simple unbiased estimator, certainly none of the form  $(n_i + t)^{-1}$ , is likely to exist for all possible cluster sample sizes.



**Figure 1.** Mean Weights of Estimators Conditional on the Proportion Residential with Shaded Histogram of Proportion of Households in Cluster

One of the ways to examine the potential bias is by comparing the expected value of the estimators (the mean weight using estimators of the form  $(n_i + t)^{-1}$ ) with the unbiased weight,  $W_u$ . Since both the unbiased weight and the expected value of the estimators are functions of  $N_i$ , we will begin by investigating these quantities conditioned on  $N_i$ .

Figure 1 shows the graph of the unbiased weight and the mean weights, using the estimators  $n_i^{-1}$  and  $(n_i + 1)^{-1}$ , when there are  $K = 10$  telephone numbers selected per cluster. The constant cluster sampling rate,  $r$ , has been omitted from all of the weights. A logarithmic scale has been used for the mean weights because of the range in  $W_u$ .

The graph clearly shows that the biggest differences between  $W_u$  and the mean weights for the two estimators are found when  $N_i/100$  is small. Once the residential density exceeds 20 percent when  $(n_i + 1)^{-1}$  is used, and 10 percent for  $n_i^{-1}$ , the differences are relatively minor. The graph shows that the weight  $(n_i + 1)^{-1}$  is always smaller than  $W_u$ , but this will not be true if poststratification is used. Poststratified weights are not used in the graph because poststratification really operates on the unconditional weights rather than the conditional weights shown here. The unconditional bias is addressed below.

The shaded histogram in the figure shows the distribution of households from Table 3. It has been overlaid to illustrate the fact that the large differences in weights occur in clusters which account for a very small fraction of the sampled households.

### Bias in Sample Size and Bias in Estimates

In nearly all RDD surveys, including those using the Mitofsky-Waksberg sample design, poststratification of the sample to known totals of persons or households is used. One of the prime reasons for using poststratification is to adjust the estimates to the levels associated with all persons, not just those in households with telephones. Massey and Botman discuss this and other benefits of poststratification in RDD surveys in Chapter 9 of Groves *et al.* (1988).

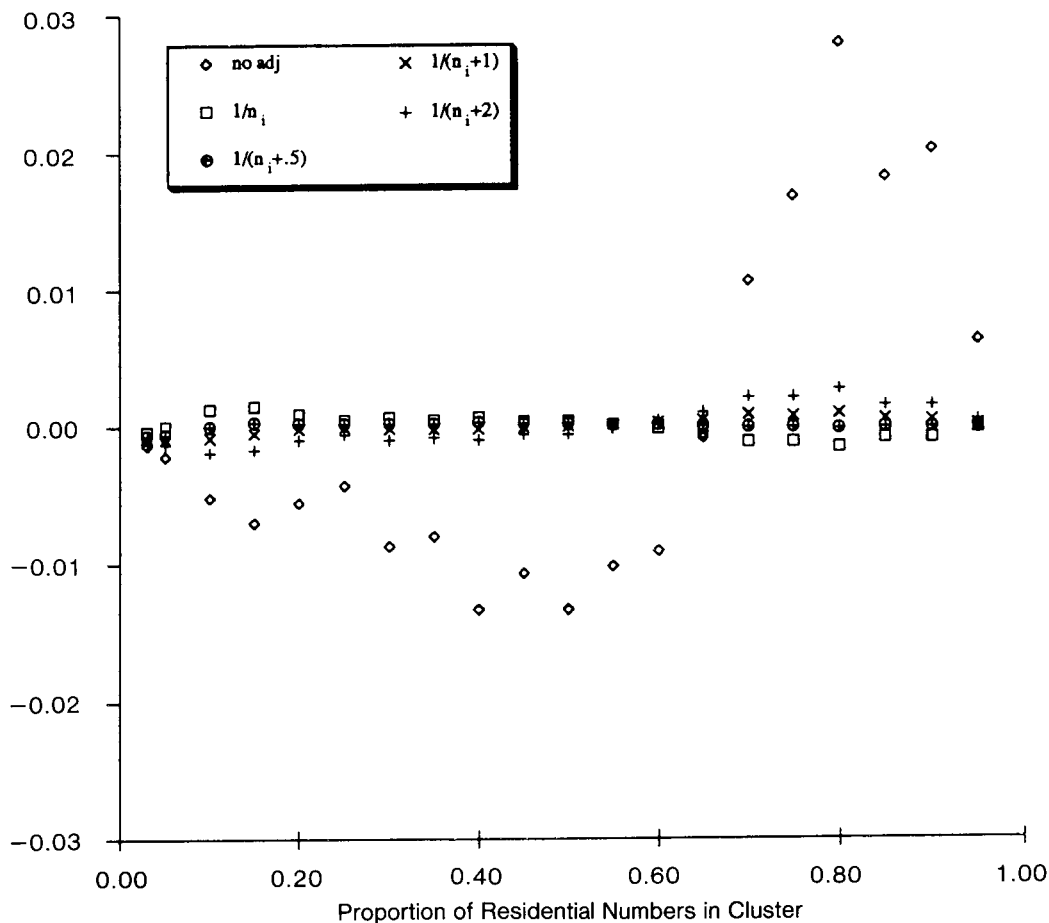
Regardless of the reasons for using it, poststratification results in estimates that are equal to known totals irrespective of the weights applied to the individual households. Since this bias, which can be considered as bias in sample size, is always zero, it is difficult to find a single statistic that measures unconditional bias directly. To attack this problem, we will examine the relative contribution to the bias in sample size over the range of household density values.

The following steps were taken to compute a measure of this contribution to bias in sample size. First, the different weighting functions or estimators were computed using the empirical household density shown in Table 3. Then, the estimates were poststratified to equal unity and the contribution to the total was computed for different values of  $N_i/100$ . Finally, the relative bias in sample size was defined as the difference between the contribution to the total from the particular estimator and the contribution from the total using  $W_u$  as the weight.

This measure thus takes into account both the difference in the weights for fixed values of  $N_i$  and the distribution of households across all the values of  $N_i$ . Thus, sampled households from clusters with values of  $N_i$  that are rare will not contribute heavily to the relative bias in sample size even if they are associated with large differences in weights.

To illustrate these computations, Figure 2 shows the relative bias in sample size for some estimators for samples of 30 numbers per cluster. One of the estimators uses the unadjusted weight, *i.e.*, the weight is a constant for all households regardless of the number of households identified in a cluster. The relative bias in sample size for the estimator with unadjusted weights is much larger than when other weights are used. The unadjusted weight has relative biases in sample size that range from about  $-2$  percent to  $+3$  percent.

Relative Bias in Sample Size

**Figure 2.** Relative Bias from Sample Size for Samples of 30 Cluster

The size of the bias in the estimate of a characteristic is bounded by the size of the bias in sample size. In other words, the relative bias in the estimate can be no larger than the relative bias in the sample size. For almost all characteristics, this upper bound will not be attained. The upper bound is only attained when the characteristic and the residential density are perfectly correlated. Very high correlations are not likely in national samples, but might be more feasible in samples in restricted geographic areas.

It can be seen that there are patterns in the biases; for example, the unadjusted estimator is uniformly too low in low proportion residential clusters and too high in clusters with a high proportion of households. When there are differences in the characteristics between low and high density clusters, the biases can be quite serious. The bias in estimates resulting from using unadjusted weights can be seen for some characteristics in Table 1 in Cummings (1979). The biases are not very large, but appropriate weighting will effectively eliminate them.

In general, the relationship between the estimate and the number of households in a cluster will be unknown and inconsistent across all the characteristics to be estimated. Therefore, a reasonable practice is to choose an estimator that has a relative bias in sample size that is small across the range of values of  $N_i$ . If the relative bias in sample size for a set of the estimators is small, then the choice of estimators can be dictated by variance considerations.

### Biases Using Different Weights

The relative bias in sample size were computed using different estimators for samples of 5, 10, 30, and 60 telephone numbers per cluster. The relative bias in sample size is negligible for the cluster sample sizes of 30 and 60 numbers, except when the unadjusted weights are used. Any of the adjusted estimators could be used for cluster samples sizes of this size without incurring biases in the estimates. When 10 numbers are selected per cluster, all of the weights except the unadjusted one again perform reasonably well. The bias performance of  $(n_i + .5)^{-1}$  is especially encouraging.

For the smallest cluster size studied, 5 numbers per cluster, the potential for bias is somewhat greater. The natural weight,  $n_i^{-1}$ , has a somewhat lower bias in sample size than the weight  $(n_i + .5)^{-1}$ , but the difference is not very large. The relative bias in sample size for both of these weights is always less than one percent. For residential densities between about 45 percent and 80 percent the bias is positive and elsewhere it is negative. This pattern might be problematic only for the few characteristics that are very highly correlated with residential density.

### Biases with Trimmed Weights

The introduction of trimming can produce significant biases, depending on the relationship between the characteristics being estimated and the weights which are being trimmed. In some applications, the bias associated with trimming may limit the amount of trimming that can be applied and, hence, its usefulness for variance reduction.

The relative bias in sample size was also computed for cluster samples of 10, 30 and 60 numbers and the weights trimmed at about 3 times the mean weight. The trimming for samples of 10 numbers per cluster was done at a factor of 2 rather than 3 as described previously.

The difference between the relative bias in sample size for the trimmed and untrimmed weights is largely inconsequential for all cluster sample sizes and most values of  $N_i/100$ . The only noticeable difference occurred when the residential density is under about 10 to 15 percent. There is a slightly greater potential for bias in these regions. However, the relative bias in sample size for the trimmed weights is still much less than one percent at all residential density values.

## 5. CONCLUSION

The standard Mitofsky-Waksberg method is an effective method of producing a self-weighting, RDD sample of fixed size. However, the sequential monitoring of the number of cases per cluster is an awkward operational feature of this method. One of the consequences of the sequential monitoring of caseloads by cluster is that it is difficult to complete data collection in a tight time frame. The data collection period has to be flexible enough to allow for obtaining the appropriate number of cases in each cluster. The more extensive data collection period and the monitoring of caseloads also result in increasing costs. Another problem with the sequential operations is that the requirement for frequent monitoring of the caseloads can lead to frustration arising from complications of combining sample selection and data collection operations.

The modified Mitofsky-Waksberg approach eliminates the sequential nature of the design and, with it, the need to monitor the work by cluster. A fixed number of telephone numbers are assigned to each sampled cluster in the modified method. Therefore, the costs associated with monitoring caseloads and a longer data collection period are not incurred. However, the modified Mitofsky-Waksberg method does introduce new components of bias and variance into the estimates. These statistical concerns should be addressed before the modified approach is used.

Specific recommendations on when the standard or modified Mitofsky-Waksberg method should be used are difficult to formulate since they depend upon circumstances which vary from survey to survey. Guidelines for choosing between the methods are suggested below.

A simple rule is that for surveys which require either very tight controls on sample size or a nearly self-weighting sample, then the standard Mitofsky-Waksberg approach is advisable. Even though the sample size in the modified method can be estimated relatively precisely, some variation, especially because of uncertainty of the nonresponse rates, can be expected. A self-weighting sample, which is not achieved when the modified Mitofsky-Waksberg method is used, also has some advantages in simplifying standard statistical analysis.

Since the costs for standard and modified methods are different, it would be very useful to have cost-variance models to help evaluate the two methods. Unfortunately, the differences in costs of the standard and modified methods are not easy to quantify. In fact, the lack of reasonable cost models is a major and pervasive problem that limits the ability to establish optimal survey design.

Because of lack of reasonable cost-variance models, we suggest some conditions in which one approach might be favored over another. One of the conditions that favors the modified approach is a relatively brief interview length. As the interview becomes longer, the cost savings associated with the modified method is likely to become smaller relative to the increases in variances of the estimates.

The length of the interview is particularly important for surveys which screen households to find those with particular characteristics. For example, some RDD surveys screen households and only interview if a member is in a particular target group. In these situations, the screening interview is often very brief. The modified Mitofsky-Waksberg approach may be very beneficial. Surveys in which households are screened also tend to have large cluster sample sizes, and this improves the performance of the modified procedure. When 10 or more numbers are selected per cluster (equivalent to about 6 households per cluster), the biases in the estimates under the modified Mitofsky-Waksberg approach are virtually inconsequential and the increases in variance with trimming are only about 10 percent. Samples of 10 or more numbers per cluster are frequently acceptable for screening purpose although such large cluster sizes are typically inefficient for the interview sample, even when the intraclass correlation is small.

Based on these factors, a general guideline is that the modified Mitofsky-Waksberg method can be recommended when households within the clusters must be screened. More specifically, the modified method with trimmed weights should be considered if the following conditions exist: (1) Ten or more numbers are sampled per cluster; and (2) the total cost for the modified method is at least 10 percent less than the standard method, or the data collection period is relatively short. If both of these conditions are not met, then the choice between methods must be made on evaluations of other survey requirements.

When the cluster sample size is less than 10, the bias and variance arising from the use of the modified Mitofsky-Waksberg method are more serious concerns. Any characteristics correlated with the proportion of residential numbers in a cluster could be affected with a cluster sample size this small. Also, the variance of the estimates with the modified method will be

20 to 30 percent larger than with the standard method since trimming is not very effective with small sample size. Therefore, in most surveys with sample sizes of less than 10 numbers per cluster, the problems of implementing the standard method should be quite serious before a decision is made to abandon it and use the modified method.

### ACKNOWLEDGEMENT

The author would like to thank the referees. Their comments were very helpful in improving the presentation of the paper.

### REFERENCES

- CUMMINGS, K.M. (1979). Random digit dialing: A sampling technique for telephone surveys. *Public Opinion Quarterly*, 233-244.
- DREW, J.D., DICK, P., and SWITZER, K. (1989). Development and testing of telephone survey methods for household surveys at Statistics Canada. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 120-127.
- GROVES, R.M., BIEMER, P.P., LYBERG, L.E., MASSEY, J.T., NICHOLLS, W.L., and WAKSBERG, J. (editors) (1988). *Telephone Survey Methodology*. New York: John Wiley and Sons.
- HANSEN, M.H., HURWITZ, W.N., and MADOW, W.G. (1953). *Sample Survey Methods and Theory*, 2. New York: John Wiley and Sons, 138-139.
- KISH, L.(1965). *Survey Sampling*. New York: John Wiley and Sons, 429-430.
- LEPKOWSKI, J.M., and GROVES, R.M. (1986). A two phase probability proportional to size design for telephone sampling. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 73-98.
- PIEKARSKI, L. (1990). Working block density declines. *The Frame*, a publication of Survey Sampling Inc.
- POTTHOFF, R.F. (1987). Generalizations of the Mitofsky-Waksberg technique for random digit dialing. *Journal of the American Statistical Association*, 82, 409-418.
- SUDMAN, S. (1973). The uses of telephone directories for survey sampling. *Journal of Marketing Research*, 10, 204-207.
- TUCKER, C. (1989). Characteristics of commercial residential telephone lists and dual frame designs. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 128-137.
- WAKSBERG, J. (1984). Efficiency of alternative methods of establishing cluster sizes in RDD sampling. Unpublished Westat Inc. memorandum.
- WAKSBERG, J. (1978). Sampling methods for random digit dialing. *Journal of the American Statistical Association*, 73, 40-46.
- WAKSBERG, J. (1973). The effect of stratification with differential sampling rates on attributes of subjects of the population. *Proceedings of the Social Statistics Section, American Statistical Association*.