APPENDIX G

APPENDIX G

RAKING

The raking procedure was conducted for the combined national and Puerto Rico RDD Samples. The purpose of the raking procedure was to correct for undercoverage due to households without telephones and households with unlisted telephones in the "zero-listed telephone banks." Moreover, the raking results in improvement in the precision of the survey estimates. We used a two dimensional raking procedure in which the two raking dimensions were formed from the cross classification of veterans according to the demographic/education/region characteristics of the veterans. The first dimension was formed from the cross classification of three age categories (under 50, 50-64, over 64) with four education levels (no high school diploma, high school diploma, some college, bachelor's degree or higher) and four race categories (Hispanic, Black, Other, and White), resulting in 48 cells. The second dimension was formed from the cross classification of gender (male, female) and the four census regions (Northeast, Midwest, South, West), resulting in 8 cells. By using a set of cross classified variables for each raking dimension, the internal correlation structure of the data could be better preserved.

We use the indices r (for row) and c (for column) to denote the cells in the two dimensional table for the raking procedure. We also denote by N_{rc} the number of veterans in the population in the cell defined by row r and column c. In our case, the population counts N_{rc} are unknown, but both sets of marginal veteran population totals are known. The two sets of marginal veteran population totals are defined as

$$N_{r.} = \sum_{c=1}^{8} N_{rc}; r = 1, 2, \ldots 48,$$

$$N_{.c} = \sum_{r=1}^{48} N_{rc}; c = 1, 2 \ldots 8. \qquad \text{(G-1)}$$

The cell population counts N_{rc} can be estimated from the sample. We start with the nonresponse adjusted survey weights $W_i^{(0)}$ and use the two sets of known marginal population counts to further adjust the weights by applying the sequence of weight adjustments as follows.

$$W_i^{(1,1)} = \left(\frac{N_{r.}}{\sum_{c=1}^{8} \tilde{N}_{rc}^{(0)}} \right) \times W_i^{(0)}; \quad i \in r,$$

$$W_i^{(1,2)} = \left(\frac{N_{.c}}{\sum_{r=1}^{48} \tilde{N}_{rc}^{(1,1)}} \right) \times W_i^{(1,1)}; \quad i \in c. \tag{G-2}$$

In the above equations $\tilde{N}_{rc}^{(0)}$ is an estimate of the cell population count N_{rc} using the initial weights $W_i^{(0)}$. The updated weights $W_i^{(1,1)}$ are used to compute $\tilde{N}_{rc}^{(1,1)}$, which is also an estimate of the cell population count N_{rc}. At the end of the first iteration we obtain the adjusted weights given by $W_i^{(1,2)}$, which is used as the input weight for the second iteration. The iterative process is continued until the specified convergence criteria described later in this Appendix are satisfied.

In general, let $W_i^{(t,1)}$ and $W_i^{(t,2)}$ be, respectively, the weights during t^{th} iteration after adjusting to the first marginal and second marginal totals, respectively. Then the weights during the t^{th} iteration can be expressed as

$$W_i^{(t,1)} = \left(\frac{N_{r.}}{\sum_{c=1}^{8} \tilde{N}_{rc}^{(t-1,2)}} \right) \times W_i^{(t-1,2)}; \quad i \in r,$$

$$W_i^{(t,2)} = \left(\frac{N_{.c}}{\sum_{r=1}^{48} \tilde{N}_{rc}^{(t,1)}} \right) \times W_i^{(t,1)}; \quad i \in c, \tag{G-3}$$

$$t = 1, 2, \ldots, T.$$

In the above equation, the weight $W_i^{(0,2)} = W_i^{(0)}$.

As before, $\tilde{N}_{rc}^{(t-1,2)}$ is an estimate of the cell population count N_{rc} based on the weights $W_i^{(t-1,2)}$, and the estimate $\tilde{N}_{rc}^{(t,1)}$ is computed in a similar fashion from the updated weights $W_i^{(t,1)}$. If T is the total number of iterations then $W_i^{(T,2)}$ are the raked survey weights. For the sake of simplicity, we will denote the raked weights $W_i^{(T,2)}$ by W_i.

The iterative process stops when the prespecified number of iterations is completed or one of the stopping rules (given below) is satisfied. Two options are available to define convergence or stopping rules. The percent relative difference value that each relation satisfies, $e\%$, can be specified and the following convergence criterion is applied.

$$100 \times \left| 1 - \frac{\sum_{c=1}^{8} \tilde{N}_{rc}}{N_{r.}} \right| < e\%, \quad r = 1, 2, \ldots, 48,$$

$$100 \times \left| 1 - \frac{\sum_{r=1}^{48} \tilde{N}_{rc}}{N_{.c}} \right| < e\%, \quad c = 1, 2, \ldots, 8. \tag{G-4}$$

Alternatively, an absolute value of the difference that each relation satisfies, d can be specified and the following convergence checks are applied.

$$\left| N_{r.} - \sum_{c=1}^{8} \tilde{N}_{rc} \right| < d, \quad r = 1, 2, \ldots, 48,$$

$$\left| N_{.c} - \sum_{r=1}^{48} \tilde{N}_{rc} \right| < d, \quad c = 1, 2, \ldots, 8. \tag{G-5}$$

We applied the convergence criteria in terms of percent absolute relative difference. We applied the raking procedure for both the full sample weights and the replicate weights. For the full sample weights, the raking procedure was stopped when the percent absolute relative difference was less than 0.01 percent for all marginal population counts. For the (raked) replicate weights the iterative procedure was stopped when the percent absolute relative difference was less than 0.1 percent for all marginal population counts.