A Report To:
The National Institutes
of Allergy and Infectious Disease

SAMPLING METHODS AND WAVE 1 FIELD RESULTS
OF THE
SAN FRANCISCO MEN'S HEALTH STUDY

by

Thomas Piazza

Survey Research Center
University of California, Berkeley
April 1986
1. OVERVIEW

1.1 Background of the Study

The San Francisco Men's Health Study is a prospective epidemiological study designed to learn more about the natural history of Acquired Immune Deficiency Syndrome (AIDS). The study was funded by the National Institutes of Allergy and Infectious Disease (NIAID) under a contract with the School of Public Health of the University of California, Berkeley. Similar studies are being conducted in four other metropolitan areas in other states.

Several organizations are cooperating in carrying out the San Francisco study. The School of Public Health at the University of California is providing the overall administration for the project. The Survey Research Center, also at the University, is responsible for questionnaire design, sampling, recruitment of participants, interviewing, coding and data entry. Children's Hospital of San Francisco is the location for the interviewing and provides the facilities for the physical examinations, the various clinical tests performed, and the collection of specimens, some of which are sent to NIAID. The Irwin Memorial Blood Bank of San Francisco performs most of the blood analysis and also stores frozen lymphocytes. Research on the specimens retained locally is being carried out at the University of California, San Francisco.

Work on the project began in September, 1983, when the contract was signed. The first year was devoted to the design of the project and the first wave of interviews. The study plan calls for reinterviewing par-
participants five additional times at six-month intervals.

Sampling work began in October, 1983, although some preliminary
work was carried out during the preparation of the proposal. Final
selection of households was completed in January, 1984. Thomas Piazza,
of the Survey Research Center, was responsible for the overall design
and execution of the sample. Yu-Teh Cheng provided technical assis-
tance. Valuable consultation was provided by John Hall.

1.2 Definition of the Target Population

Our overall goal was to secure the participation of a sample of
individuals that would be representative of the San Francisco population
most at risk of contracting AIDS. Since most of the persons who have
contracted AIDS so far are male homosexuals, it was clear from the
outset that our study should concentrate on that group. On the other
hand, since it was considered possible that AIDS would eventually be
transmitted outside the gay community (even aside from drug users and
recipients of blood transfusions), we wanted to obtain at least some
non-gay participation in the study.

The rather specialized nature of our desired target population
posed a serious sampling problem. The ideal method to obtain such a
sample would be to select a large probability sample of the general
population and then screen on the basis of the desired characteristics.
That method, however, would have been very costly, and in any case we
were not sure that it would be possible to screen effectively for sexual
orientation in a short screening interview. One alternative, adopted by
bability sample of unmarried men in that age group would yield approximately the desired proportions of gays and non-gays for our study. This meant that we could draw a sample of the population in that area and screen only on the basis of gender, age, and marital status. No screening on sexual preference would be necessary -- a definite advantage. Although we were uncertain of the exact proportion of gays that this method would yield, we devised a reserve sample strategy (described below) that would allow us to exercise some control over that proportion.

Our final specification of the target population, therefore, is as follows:

1. Residents of 19 census tracts in San Francisco. The tracts are 162-171, 204-207, and 210-214.
3. English speaking -- to avoid the cost of questionnaires in other languages for very few cases.

Although our target population is limited to a part of the city containing only about 12 percent of the total population, we estimate that the area contains between one third and two thirds of the gay population of San Francisco. In general our sample should provide reasonably good estimates for the city's gay population.
the other study sites, would have been to advertise for volunteers and base the study on them. We determined at an early stage, however, that we did not want to abandon completely a population-based approach, which would require probability sampling.

There were some specific reasons for our decision to base our study on a probability sample of a known population. First of all, we are interested in estimating the prevalence of AIDS and other conditions among gays in the population. Basing such estimates on volunteers is certainly not reliable, and a probability sample is preferable for that purpose. Secondly, we are interested in estimating the magnitude of the relationships between the variables we are measuring. We suspected that volunteers could be atypical not only in the distribution of individual characteristics (sexual activity, drug use, and other suspected or known risk factors) but also in the relative impact of such characteristics on health. Since we were not in a position to control important variables experimentally, we wanted the distributions both of variables and of relationships to have as wide a representation as possible.

For all of these reasons, we decided that it was important to base our study on a probability sample of a defined population. Since it was not feasible to screen a large sample of the whole area, or even of the whole city of San Francisco, we decided to focus on one specific part of San Francisco. This part of the city was defined as nineteen contiguous census tracts that were known to include the residences of virtually all of the people with AIDS at that time.

Furthermore, we estimated that most of the unmarried men in those census tracts between the ages of 25 and 54 would be gay and that a pro-
1.3 General Design of the Sample

Once the target population was defined, we proceeded to develop a strategy to sample that population. The details of the sampling procedures are described below in Section 2. The general features of the design are summarized here.

The sample we drew was a stratified two-stage sample of all households within the target area. Each of the nineteen census tracts was a separate stratum. From each tract we selected several blocks with probability proportionate to the estimated number of housing units on each block (based on the 1980 Census). We then listed all housing units on each selected block. From each resulting list we selected a certain proportion of housing units by systematic random sampling. All eligible males in each selected housing unit were taken into the sample.

In regard to this last point, note that we did not restrict our sample to one person per household, as we usually do in samples of the general population. We reasoned that it might be important to be able to compare eligible males within the same household. There was also the practical difficulty of restricting the benefits of participation in the study to only one member of a household.

One additional feature of our sampling procedure was the development of a reserve sample. Since we did not have precise estimates of the number of eligible males we could expect to find per sampled household nor of the proportion of eligible males that would be gay, we needed to devise a sampling strategy that would allow us to release the sample in increments. The method adopted was to select more blocks than were needed, randomize their order within strata, release the minimum
number expected to be adequate, and then draw on the reserve blocks as necessary. Details of this procedure are given in Section 2.3 below. We did in fact need to use some of the reserve blocks, so the design served us well.

Our overall sampling fraction was approximately 1/13. Of the estimated 43,624 housing units in the nineteen census tracts, we selected 3,343 units. The sampling fraction varies somewhat, however, between strata, depending primarily on the number of reserve blocks used in each stratum. A weight inversely proportional to the probability of selection in each stratum should, therefore, be used when analyzing the data. The appropriate weight for each case is included in the data file. Weights are discussed below in Section 4.
2. SAMPLING PROCEDURES

2.1 Constructing the Sampling Frame

The statistics and maps of the 1980 U.S. Census provided the sampling frame for the selection of blocks. We obtained for this purpose a computer tape from the State Census Data Center in Sacramento, California. The tape contained for each block in our target area an identifier and the number of housing units enumerated by the Census. At the time this work began (late 1983), the Census data were already more than three years old. Nevertheless, we knew that there had been little new construction in our target area during those years, and we considered the Census data adequate for our purposes.

The frame for block selection was constructed by generating a list of the census blocks in each of the nineteen tracts. For this study we wanted each block to have at least fifty housing units, since we expected to select up to twenty-five housing units on each selected block. Blocks that were smaller than the minimum were linked with other blocks to create units of at least the desired minimum size. When we refer below to "blocks," therefore, we mean those units of at least fifty households, comprising one or more physical census blocks.

2.2 Selection of Blocks

Once the list of blocks within each tract (stratum) was complete, we proceeded to select several blocks from each tract with probability
proportionate to the estimated number of housing units on each block. The number of blocks selected from each tract was set to be roughly proportional to the size of the tract -- approximately 1/125 times the number of housing units. For example, if a tract had 2500 housing units, we selected 2500/125 = 20 blocks. Although this number of selections was a large proportion of the total number of blocks in each tract, we did not expect to use all of them. Most of them were put into reserve, as described next.

2.3 Creation of the Reserve Sample

In order to decide what proportion of the households we should sample, certain assumptions had to be made. First of all, we had to estimate what proportion of households would be vacant and what proportion of eligible persons would participate in the study -- assumptions that need to be made for every survey. In addition, however, we had to estimate what proportion of households would contain an eligible male, how many eligible males per household we would find, and what proportion of the eligible males would be gay. The best we could do to estimate these figures was to set a high and low range for each, based on inferences we drew from an examination of the Census data for each tract in the sample.

If all of our most optimistic assumptions were met, we estimated that a sample of about 1/15 of the housing units in our target area would be sufficient to reach our goal of completing Wave 1 interviews and tests with 1000 gay men and approximately 200 non-gay men. Our most
pessimistic estimates, on the other hand, indicated that a sample of almost 1/3 of all housing units in the area would be required. This difference between optimistic and pessimistic estimates was too large to handle by simply adopting a compromise estimate and hoping for the best. Our uncertainty on many points was genuine and could not be ignored. We therefore decided to design the sample in such a way that we could release part of it at the beginning of field work and then release additional segments as needed.

A reserve sample can consist of additional housing units on a selected block, or it can consist of entire reserve blocks. The former method maximizes the geographic spread of the sample, whereas the latter is easier to administer. In our case we opted for the administrative advantages of having entire blocks designated as reserve blocks, especially since our geographic coverage of the target area would be substantial anyhow. (Our final sampled included an average of more than seven blocks per tract.)

Our method of creating the reserve sample, consequently, was as follows: We selected enough blocks to meet our needs under the relatively pessimistic assumption that we would have to select 1/5 of all households in each tract of the target area. The selected blocks in each tract (stratum) were then put into a random order. Approximately the first 1/3 of the blocks (going down each randomized list) were released immediately for field work; the sampling fraction to begin with, consequently, was about 1/3 \times 1/5 = 1/15, corresponding to our optimistic assumptions. Additional blocks could be taken from all or some of the tracts, so long as the randomized order within each tract
was observed. We eventually used eleven of the reserve blocks, in seven
different tracts. Since we listed the households on a few reserve
blocks in each tract at the same time that the initial blocks were
listed, the reserve blocks could be put into the field on short notice.

2.4 Selection of Housing Units

Field workers were sent to each of the blocks in the initial sample
(plus a few reserve blocks in each tract) with instructions to list all
dwelling units. Each house was listed by its street address. Each
apartment was listed by its address and also by its apartment number
(when available) or by a description of its location within the struc-
ture. The outcome of this procedure was a list of housing units for
each block. This list constituted the sampling frame from which individu-
dual housing units were selected.

The actual selection of housing units was carried out by systematic
random sampling. As a first step, a target sample size for each
selected block was calculated. This target sample size was obtained by
applying the overall sampling fraction to the estimated number of hous-
ing units in each stratum and then dividing by the total number of
blocks selected (including reserve blocks) from that stratum. If \( M_h \) is
the estimated number of housing units in stratum \( h \), and \( a_h \) is the number
of blocks selected from stratum \( h \), and \( F \) is the inverse of the sampling
fraction, then the target sample size for a block, \( b_{hi} \), is given by:

\[
b_{hi} = \frac{M_h}{a_h F}
\]

(See the summary of notation in Exhibit 2.1.) For example, if a stratum
Exhibit 2-1
SUMMARY OF NOTATION

Subscripts

\[ h \] Stratum

\[ i \] Block within a stratum

\[ j \] Housing unit within a block

Numbers of Housing Units

\[ M_h, M_{hi} \] Estimated from Census block statistics

\[ N_h, N_{hi} \] Actually found during block listing

Sampling Fraction and Sample Sizes

\[ f = \frac{1}{F} \] Initial sampling fraction (1/5)

\[ a_h \] Number of blocks selected in stratum \( h \)

\[ B_h \] Proportion of selected blocks in stratum \( h \) actually used in the study

\[ B_h/F \] Final sampling fraction in stratum \( h \)

\[ b_{hi} \] Target number of housing units to be selected from block \( i \) of stratum \( h \)

\[ b_{hi} = \frac{M_h}{a_hF} \]

\[ n_{hi} \] Actual number of housing units selected from block \( i \) in stratum \( h \)

\[ n_{hi} = \left( \frac{N_{hi}}{M_{hi}} \right) \ast b_{hi} \]
(tract) was estimated to have 2500 housing units and 20 blocks were selected, then the target sample size (with Fr5) on each of the selected blocks in that stratum would be 2500/(20 x 5) = 25.

Once the target sample size for each block \((b_{hi})\) was calculated, the spacing interval for systematic selection was obtained by dividing the estimated number of housing units on each block by the target sample size for that block. If \(M_{hi}\) is the estimated number of housing units in block \(i\) of stratum \(h\), then the interval, \(I_{hi}\), is given as:

\[
I_{hi} = \frac{M_{hi}}{b_{hi}}
\]

This interval is the inverse of the probability of selecting any one housing unit on this block. For example, if a block is estimated to have 100 housing units, and if the target sample size is 25, then the spacing interval is 100/25 = 4. This means that we would take every fourth housing unit on that block, after a random start. In practice the target sample size and the spacing interval are not usually whole numbers, but that does not cause any problems.

The overall probability of selection for each household is

\[
P_{hij} = \frac{a_h M_{hi}}{M_h} \cdot B_h \cdot \frac{b_{hi}}{b_{hi}}
\]

The first term after the equal sign is the probability of selecting block \(i\); \(M_{hi}\) and \(M_h\) are the estimated numbers of housing units in the block and in the entire stratum (tract), respectively, and \(a_h\) is the number of blocks selected. The second term, \(B_h\), is the proportion of selected blocks in the stratum that were actually released for interviewing, as described above in the discussion of the reserve sample. The third term is the probability of selecting housing unit \(j\) on
block i; this term is the inverse of the selection interval for each block, as given in Equation 2.

If we substitute the value of $b_{hi}$ from Equation 1 into the third term of Equation 3, we have:

$$P_{hij} = \frac{b_{hi}}{M_{h}} \cdot b_{h} \cdot \frac{(M_{h} - b_{h}F)}{M_{h1}} = B_{h}$$

All of the values specific to a block or stratum cancel out, except for $B_{h}$. If we had used all blocks in the reserve sample, $B_{h}$ would equal 1.0 in all strata, and all housing units would have been selected with the same probability.

The actual probability of selection within each stratum is $B_{h}/5$ (since $F = 5$). For example, if 24 blocks were originally selected and 8 of those were used in the study, $B_{h}$ would equal $8/24 = 1/3$, and the probability of selection in that stratum would be $1/3 \times 1/5 = 1/15$. The final values of $B_{h}$ across strata vary from $1/3.33$ to $1/1.83$, and the corresponding overall selection probabilities ($B_{h}/5$) range from $1/15.7$ to $1/9.2$, respectively. Weights inversely proportional to the selection probabilities should be used in analyzing the data; such weights are provided in the data file. Although the precision of estimates can decrease as a result of weighting, the effect of weighting in our case is very small. We calculate that the standard error of estimates based on our sample will only be about 0.9% larger than for a design in which the same total number of cases was selected with a constant sampling fraction. ¹

¹This computation is based on equation 11.7.6 in Leslie Kish, *Survey Sampling* (New York: John Wiley, 1965), p.430.
2.5 Method of Recruiting Respondents

A letter was sent to each selected housing unit to introduce the study in general terms and to advise residents that an interviewer would be coming. Shortly thereafter an interviewer went to the household to determine whether any eligible males resided there. If one or more eligible males were found, they were given information packets on the study and, if possible, were immediately given appointments for a visit to the clinic, where the interview and examination would be administered. Several visits were made to each household, if needed, in order to locate eligible males and to schedule them for clinic appointments. Prior to each scheduled clinic appointment a reminder postcard was sent, and an attempt was also made to confirm the appointment by telephone on the day before.
3. FIELD OUTCOME AND RESPONSE RATES

The disposition of each sampled housing unit and each eligible male is summarized in this section. We will account first for the selected housing units. We will then account for the eligible males that were identified. Finally, we will summarize the results for individual census tracts.

3.1 Household Results

The household results for the sample as a whole are summarized in Exhibit 3-1. As shown there, a total of 3,343 housing units were selected by the sampling procedures described above in Section 2. Of the total selections, 6.9 percent were excluded because they were vacant or otherwise ineligible, leaving 3,112 housing units for the sample. Of these housing units, 95.1 percent were successfully enumerated -- that is, an interviewer was able to determine whether any eligible males resided there and, if so, how many there were. We found 1,245 housing units with one or more eligible males.

3.2 Wave 1 Results for Enumerated Males

The Wave 1 results for enumerated eligible males are shown in Exhibit 3-2. As we see there, a total of 1,759 eligible males were enumerated. Of that number, 40.7 percent either refused or were unable to participate. The relatively high refusal rate was undoubtedly a consequence of the nature of the study and the long-term commitment
### Exhibit 3-1

**HOUSEHOLD FIELD RESULTS**

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>% of Total</th>
<th>% of Valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Selections</td>
<td>3343</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Ineligible for the Sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacant</td>
<td>139</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>Not a residence</td>
<td>70</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>No one English speaking</td>
<td>22</td>
<td>.7</td>
<td></td>
</tr>
<tr>
<td>Total ineligible</td>
<td>231</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>Valid Sample Units</td>
<td>3112</td>
<td>93.1</td>
<td>100.0</td>
</tr>
<tr>
<td>Non-response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refused enumeration</td>
<td>129</td>
<td>3.9</td>
<td>4.1</td>
</tr>
<tr>
<td>Never at home</td>
<td>17</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>No access</td>
<td>5</td>
<td>.1</td>
<td>.2</td>
</tr>
<tr>
<td>Total non-response</td>
<td>151</td>
<td>4.5</td>
<td>4.9</td>
</tr>
<tr>
<td>Enumerated households</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No eligible males</td>
<td>1716</td>
<td>51.3</td>
<td>55.1</td>
</tr>
<tr>
<td>One or more eligibles</td>
<td>1245</td>
<td>37.2</td>
<td>40.0</td>
</tr>
<tr>
<td>Total enumerated</td>
<td>2961</td>
<td>88.6</td>
<td>95.1</td>
</tr>
</tbody>
</table>
### Exhibit 3-2

**WAVE 1 FIELD RESULTS FOR ENUMERATED MALES**

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Enumerated Males</strong></td>
<td>1759</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Non-response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refusals</td>
<td>545</td>
<td>31.7</td>
</tr>
<tr>
<td>Unable to participate</td>
<td>27</td>
<td>1.5</td>
</tr>
<tr>
<td>Moved out of area</td>
<td>24</td>
<td>1.4</td>
</tr>
<tr>
<td>Never home</td>
<td>13</td>
<td>.7</td>
</tr>
<tr>
<td>Participants in other studies</td>
<td>7</td>
<td>.4</td>
</tr>
<tr>
<td><strong>Total non-response</strong></td>
<td>716</td>
<td>40.7</td>
</tr>
<tr>
<td><strong>Completed Wave 1 Interviews</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homosexuals</td>
<td>649</td>
<td>36.9</td>
</tr>
<tr>
<td>Bisexuals</td>
<td>174</td>
<td>9.3</td>
</tr>
<tr>
<td>Heterosexuals</td>
<td>212</td>
<td>12.1</td>
</tr>
<tr>
<td>Unknown</td>
<td>8</td>
<td>.5</td>
</tr>
<tr>
<td><strong>Total completed</strong></td>
<td>1043</td>
<td>59.3</td>
</tr>
</tbody>
</table>
required of participants. The remaining 59.3 percent of enumerated males completed the Wave 1 interview and physical exams.

3.3 Wave 1 Results for Each Census Tract

A summary of the response rates for each census tract is given in Exhibit 3-3. The response rate in the column labeled "Households" corresponds to the last row of Exhibit 3-1 for the sample as a whole -- the percent of valid housing units that were successfully enumerated. The response rate in the column labeled "Eligibles" corresponds to Exhibit 3-2 -- the percent of enumerated eligible males that were interviewed in Wave 1. The final column, labeled "Combined," gives the product of the other two columns; this combined response rate was used in calculating weights for analysis, as described next.
<table>
<thead>
<tr>
<th>Tract</th>
<th>Response Rates for:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Households</td>
<td>Eligibles</td>
<td>Combined</td>
</tr>
<tr>
<td>162</td>
<td>91.6</td>
<td>60.0</td>
<td>54.9</td>
</tr>
<tr>
<td>163</td>
<td>94.8</td>
<td>50.0</td>
<td>47.4</td>
</tr>
<tr>
<td>164</td>
<td>95.1</td>
<td>50.8</td>
<td>48.3</td>
</tr>
<tr>
<td>165</td>
<td>98.0</td>
<td>50.0</td>
<td>49.0</td>
</tr>
<tr>
<td>166</td>
<td>93.2</td>
<td>53.0</td>
<td>49.4</td>
</tr>
<tr>
<td>167</td>
<td>96.7</td>
<td>79.3</td>
<td>77.4</td>
</tr>
<tr>
<td>168</td>
<td>91.2</td>
<td>60.0</td>
<td>54.7</td>
</tr>
<tr>
<td>169</td>
<td>93.9</td>
<td>56.1</td>
<td>52.7</td>
</tr>
<tr>
<td>170</td>
<td>96.2</td>
<td>61.0</td>
<td>58.7</td>
</tr>
<tr>
<td>171</td>
<td>97.8</td>
<td>62.8</td>
<td>61.4</td>
</tr>
<tr>
<td>204</td>
<td>94.0</td>
<td>58.0</td>
<td>54.5</td>
</tr>
<tr>
<td>205</td>
<td>95.7</td>
<td>64.8</td>
<td>62.1</td>
</tr>
<tr>
<td>206</td>
<td>93.3</td>
<td>70.8</td>
<td>56.0</td>
</tr>
<tr>
<td>207</td>
<td>98.3</td>
<td>70.3</td>
<td>69.1</td>
</tr>
<tr>
<td>210</td>
<td>96.5</td>
<td>46.4</td>
<td>44.8</td>
</tr>
<tr>
<td>211</td>
<td>96.3</td>
<td>50.9</td>
<td>49.0</td>
</tr>
<tr>
<td>212</td>
<td>95.3</td>
<td>70.8</td>
<td>67.5</td>
</tr>
<tr>
<td>213</td>
<td>93.3</td>
<td>60.0</td>
<td>56.0</td>
</tr>
<tr>
<td>214</td>
<td>95.7</td>
<td>60.8</td>
<td>58.2</td>
</tr>
<tr>
<td>Overall</td>
<td>95.1</td>
<td>59.3</td>
<td>56.4</td>
</tr>
</tbody>
</table>
4. CALCULATION OF WEIGHTS

Two kinds of weights were computed for this survey. The first weight was to compensate for unequal probabilities of selection. The second weight incorporated an adjustment for different rates of non-response. Each of the steps involved in the weighting will now be described.

4.1 Selection Probability

The probability that a specific housing unit would fall into the sample was not exactly the same in every census tract (stratum). As described above in Section 2.4, the probability of selection was a function of the proportion of selected blocks in each tract actually used in the study; probabilities ranged from 1/9.2 to 1/16.7. To compensate for unequal probabilities of selection, a weight inversely proportional to those probabilities was created:

Lowest weight: \( \hat{w} = k \times 9.2 = 0.73 \)

Highest weight: \( \hat{w} = k \times 16.7 = 1.33 \)

For this weight the constant, \( k \), was set equal to 0.079, which resulted in weights that summed to 1034 -- the number of completed cases (excluding nine cases that came into the study in an irregular manner and should be excluded when making estimates for the population from which the sample was drawn; those nine cases are assigned weights equal to zero). Note that this weight varies around 1.00. Those respondents that had a smaller than average chance of being selected were given weights greater than 1.00; those with a greater than average chance of
being selected were given weights that were less than 1.00.

4.2 Differential Non-response

The response rate varied in different census tracts, as was shown above in Exhibit 3-3. In order to adjust for this differential response rate, we assigned each case a weight which was inversely proportional to the overall response rate in its census tract. For example, cases from a tract with a 70 percent response rate would be weighted \(1/0.70 = 1.43\), whereas cases from a tract with a 60 percent response rate would be weighted \(1/0.60 = 1.67\). This adjustment for differential non-response was not used by itself; rather, it was combined with the adjustment for selection probability to create a combined weight, as described next.

4.3 Combined Weights

The adjustments for selection probability and non-response were combined (multiplied) into a single combined weight. All cases in the same tract received the same weight. The range of resulting combined weights is as follows:

- Lowest weight = \(k \times 14.77\)
- Highest weight = \(k \times 31.62\)

Each case was actually assigned two variants of this weight, corresponding to two different values for \(k\). Under the first variant, \(k\) was calculated so that the sum of the weights across cases was 1034, as for the weight to adjust for selection probability. Under the second variant, \(k\) was set equal to one, and the resulting sum of weights of 23,137 is
our estimate of the total number of eligible males residing in the area
from which the sample was drawn. This latter weight is the expansion
factor for estimating the number of eligible men that have some charac-
teristic measured in the survey.

4.4 Which Weight to Use

On the Wave 1 data file three weights are available, from which the
analyst may choose:

- **SAMFWT**: Adjustment for selection probability alone
- **COMBWT**: Adjustment for both selection and non-response
- **EXPAND**: Same as COMBWT, but summing to population

The first two weights sum to 1034, the number of completed cases in the
regular sample. The third weight sums to 23,137.

For most purposes, the combined weight COMBWT will be the appropri-
ate one to use. It adjusts for both selection probability and for non-
response (at the tract level), and the number of weighted cases
corresponds to the actual number of completed interviews.

Occasionally the expansion factor EXPAND will be useful. Both
COMBWT and EXPAND will generate the same percentages, regression coeffi-
cients, and other statistics; however, EXPAND will also provide a rea-
sonable estimate of the number of people in the population that have
some characteristic. Note that such estimates are not necessarily the
best possible ones, but at least they do not require auxiliary sources
of information such as Census data. When using the EXPAND weight with a
standard statistical package, the analyst should completely ignore any
printed standard error results; they are computed as if there were
23,137 actual cases in the sample. Printed standard errors will be
underestimates even with COMBWT, but with EXPAND they will not even be
close to reality.

The adjustment for selection probability alone (SAMPWT) could be
used if the analyst wishes either to ignore non-response bias or to deal
with it in a different way. The assumption made when using COMBWT or
EXPAND is that non-respondents in a particular census tract are like the
average respondent in the same tract. If SAMPWT is used without any
other non-response adjustment, the implicit assumption is that all non-
respondents are like the average respondent in the sample as a whole.

There may be occasions when it is appropriate not to use any
weights at all. If the focus is only on the selected persons (for some
kind of case control, for instance) and not on estimating statistics for
the population as a whole, then there is no reason to use weights. It
should be noted in any event that the effect of using weights with this
sample will generally be relative small, because most cases have proba-
bilities of selection that are close to equal. Most estimates, conse-
quently, will not differ by more than a few percentage points regardless
of weighting.