SAMPLING METHODS AND FIELD RESULTS
OF THE
SAN FRANCISCO DRUG USE STUDY

by

Thomas Piazza
and
Yu-Teh Cheng

Survey Research Center
University of California, Berkeley
November 1999
Table of Contents

1. Overview
   1.1 Background of the Study .............................................. 1
   1.2 Definition of the Target Population ................................. 2
   1.3 General Design of the Sample ....................................... 3

2. Sampling Procedures
   2.1 Constructing the Sampling Frame .................................... 4
   2.2 Selection of Blocks ................................................... 4
   2.3 Selection of Housing Units .......................................... 5
   2.4 Creation of the Reserve Sample ...................................... 7
   2.5 Selection of Respondents ............................................ 9

3. Field Outcome and Response Rates
   3.1 Field Results for Prevalence Interviews .......................... 11
   3.2 Field Results for Career Marijuana Users ....................... 11
   3.3 Field Results by Geographic Area ................................. 14

4. Calculation of Weights
   4.1 Selection Probability ................................................. 16
   4.2 Non-response Adjustment ............................................. 17
   4.3 Combined Weights ................................................... 18
   4.4 Which Weight to Use ................................................ 18

Exhibits
   2-1 Summary of Notation ................................................ 6
   3-1 Household Field Results ........................................... 12
   3-2 Field Results for Career Marijuana Users ...................... 13
   3-3 Field Results by Geographic Area ................................ 15
1. OVERVIEW

1.1 Background of the Study

The San Francisco Drug Use Study is part of an international investigation designed to compare marijuana use patterns and practices, and their social and health consequences. The San Francisco study is identical to ones being conducted in Amsterdam (The Netherlands), and Bremen (Germany). Amsterdam was chosen because it is legal to purchase small amounts of marijuana there. Bremen was chosen because it is a city similar in character to Amsterdam but with more restrictive drug laws.

The U.S. study is funded by the National Institute on Drug Abuse (NIDA), which is part of the National Institutes of Health. NIDA is interested in studying the long-term patterns of marijuana use and its consequences, in order to provide better information for ongoing drug policy debates.

The San Francisco study employs a rigorous study design in which a random sample of household addresses was selected from among all households in the City. Survey Research Center interviewers contacted each sampled household to conduct an enumeration of the household’s residents and then randomly selected one adult resident for the Prevalence Interview. The Prevalence Interview was designed to assess the general prevalence of past and current drug use of all types -- not only marijuana, but also legal drugs like cigarettes and alcohol, and illegal drugs like cocaine and hallucinogens. The Prevalence Interview’s second purpose was to identify “career” users of marijuana (defined as those who had used marijuana 25 days or more in their lifetime) and to recruit those people to participate in a longer In-Depth Interview about their personal history of marijuana use and its consequences. The In-Depth Interview was
designed to investigate how many of various types of people, with what psychological expectations and in what social settings, have increased their use of marijuana over time, experienced negative consequences, become dependent, and gone on to use other illicit drugs. The study seeks to identify the predictors of the various career patterns -- including infrequent use and cessation -- that are extant in the general population.

The San Francisco survey was carried out by the Survey Research Center at the University of California at Berkeley. The Survey Research Center designed the sample, conducted the interviews, and prepared the data files. Sampling work begin in January, 1998. Interviewing began in August, 1998, and finished in August, 1999.

1.2 Definition of the Target Population

The target population for the Prevalence Survey was all English- or Spanish-speaking adults residing in housing units within the city of San Francisco; for the longer In-Depth Interviews, however, the target population was limited to English speakers. Non-transient residents of hotels, motels, and rooming houses were also included. Excluded from the study population were residents of student dormitories, cooperative housing, nursing homes, hospitals, jails, half-way houses and other group quarters.

According to the 1990 Census, approximately 17% of San Francisco residents are of Chinese origin, and almost half of them reported that they did not speak English well. However, the cost of translating the questionnaire into Chinese and of hiring interviewers to conduct interviews in Cantonese or Mandarin would have been prohibitive for the few additional cases to be obtained, especially for the In-Depth Interview, and such an effort was beyond the scope of the present study. The study was therefore limited to English speakers and, for the Prevalence Interview, Spanish speakers.
To summarize then, our target population was defined as follows:

1. Household residents of San Francisco
2. Persons 18 years of age or older
3. English or (for the Prevalence Interview) Spanish speaking.

1.3 General Design of the Sample

Once the target population was defined, we proceeded to develop a strategy to sample that population. Our desired sample size was approximately 200 completed In-Depth Interviews, plus the number of Prevalence Interviews necessary to attain that number. The details of the sampling procedures are described below in Section 2. The general features of the design are summarized here.

The sample was a two-stage area sample of all households within the city of San Francisco. City blocks served as the primary sampling units, and housing units on selected blocks were the second-stage units. After geographically sorting the 1990 Census list of blocks, we selected 100 blocks from the sorted list with probability proportionate to the estimated number of housing units on each block (based on the 1990 Census). We then listed all housing units on each selected block. From each resulting block list we selected a certain proportion of the housing units by systematic random sampling.

In each selected housing unit one randomly selected eligible adult was requested to complete the Prevalence Interview. All completed Prevalence Interviews were sent back to SRC for review, and respondents meeting the criterion for "career marijuana users" (used marijuana 25 days or more in their lifetime) were contacted again and requested to complete the In-Depth Interview. The field results of these efforts are described below in Section 3.
2. SAMPLING PROCEDURES

2.1 Constructing the Sampling Frame

The statistics of the 1990 U.S. Census provided the sampling frame for the selection of blocks. We obtained for this purpose a computer file from the UCDATA Program at the University of California in Berkeley. The file contained for each block in San Francisco an identifier and the number of housing units enumerated by the Census. At the time this work began, the Census data were already more than eight years old. Nevertheless, we knew that there had been little new construction in our target area during those years and the Census data were adequate for our purposes.

For this study we wanted each block to have at least forty-five housing units, since we originally expected to select approximately thirty housing units on each selected block. Blocks that had fewer than forty-five housing units 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 forty-five households, comprising one or more physical Census blocks.

2.2 Selection of Blocks

Once the list of blocks within each tract was complete, the full list of blocks was sorted by tract and block number, in order to provide implicit geographic stratification by neighborhood of San Francisco. We then proceeded to select 150 blocks from the sorted block list with probability proportionate to the estimated number of housing units on each block. The selection was carried out by using systematic random sampling, with a random start, in order to preserve the implicit geographic stratification of the blocks.
2.3 Selection of Housing Units

Field workers were sent to each of the selected 150 blocks with instructions to list all housing units. The selected blocks were located with the help of block maps prepared for this project by the Survey Research Center at the University of Michigan. 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 structure. The outcome of this procedure was a list of housing units for each block. This list constituted the sampling frame from which individual 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. Since we originally estimated that we would need to select 4,500 households to complete enough Prevalence Interviews to generate the desired number of In-Depth Interviews with career users, we designed the sample to select 30 households on each of the 150 selected blocks.

Once the target sample size for each block \((b_i = 30)\) was set, the selection 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 \(M_i\) is the estimated (from the 1990 Census) number of housing units in block \(i\), then the interval, \(I_i\), is given as:

\[
I_i = \frac{M_i}{b_i}
\]  
(1)

(See the summary of notation in Exhibit 2.1.) 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 90 housing units, and if the target sample size is 30, then the selec-
Exhibit 2-1
SUMMARY OF NOTATION

Subscripts

i  Block
j  Housing unit within a block

Numbers of Housing Units

\( M_i, \ M \)  Estimated from Census block statistics
\( N_i, \ N \)  Actually found during block listing

Sampling Fraction and Sample Sizes

\( f = 1/F \)  Sampling fraction
\( a \)  Number of blocks selected
\( b_i \)  Target number of housing units to be selected from a block \( i \)
\( n_i \)  Actual number of housing units selected from block \( i \)
\( n_i = (N_i/M_i) \ast b_i \)
tion interval is $90/30 = 3$. This means that we would take every third housing unit on that block, after a random start. In practice the target sample size and selection interval are not usually whole numbers, but that does not cause any problems.

Note that the target sample sizes on the various blocks are used to calculate the selection interval for each block, but the actual number of housing units selected depends on how close the estimated number of housing units ($n_i$) is to the actual number ($N_i$) found when listing all the housing units on the block. For example, if a block is estimated to have 90 housing units, and the target sample size is 30, the selection interval of $1/3$ is applied to the entire list. If the block actually contained only 75 housing units, we would select $75/3 = 25$ housing units.

2.4 Creation of the Reserve Sample

The total number of selected blocks (150) and households (targeted at 30 per block) was based on worst-case assumptions. A random part of the selections were put into a reserve sample, to be used only as needed. Two methods were used to set part of the sample aside as a reserve sample.

The first type of reserve sample was a random subset of entire blocks. To ensure that both the main sample and the reserve samples were spread maximally over the entire city of San Francisco, we first sorted the 150 blocks into 31 neighborhood groups. Then we drew three systematic random samples (with separate random starts) of 25 blocks each from the sorted list to create three sets of reserve blocks. Using this procedure, we ensured that the main sample and each of the three reserve samples were distributed throughout the city. (We eventually used 25 of the 75 reserve blocks.)
The second type of reserve sample was a random subset of housing units on the selected blocks. The original target of 30 housing units was reduced to 15 on about half of the blocks, by retaining every other selection after a random start. In order to maximize the geographic spread of the sample, the number of selected housing units was reduced on some blocks to a random quarter of the original 30 selections, to allow more blocks to be included in the sample.

In the end we found that the prevalence of career marijuana users was higher than anticipated. As a result, we did not need to use most of the reserve sample. In total, we used 100 blocks. On 35 of those blocks we selected the original target of 30 housing units. On 48 of the blocks we sampled at half that rate, and on the remaining 17 blocks we sampled at a quarter of the original rate.

Accordingly the overall probability of selection for each household is

$$P_{ij} = \frac{a \cdot M_i}{M} \cdot \frac{b_i}{M_i}$$

(2)

The first term after the equal sign is the probability of selecting block \(i\); \(M_i\) is the estimated number of housing units on the block, \(M\) is the estimated number of housing units in the entire list of blocks in San Francisco (328,472), and \(a\) is the number of blocks selected for use in the sample (100). The second term after the equal sign 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 1.

If we substitute the known values of \(a\) and \(M\) into the first term after the equal sign in Equation 2, we have:

$$P_{ij} = \frac{100 \cdot M_i}{328,472} \cdot \frac{b_i}{M_i} = \frac{b_i}{328472}$$

(3)

There are three main probabilities of selection for this sample based on the three target subsample sizes on the blocks (\(b_i = 30\), or 15, or 7.5). Corresponding to each of these
are overall sampling fractions (1/F) of 1/109.5, 1/219, or 1/438.

There are some additional differences in probabilities of selection on a few large blocks that were segmented prior to selecting households. On those large blocks, one segment was selected at random with probability proportional to the estimated size of each segment. The selection interval on that block was then calculated as the ratio $s_j$ divided by $b_j$, where $s_j$ is the estimated size of the selected segment $j$ in block $i$. The probability of selection for a household in such a segment is:

$$P_j = \frac{100 \cdot \frac{M_i}{328,472} \cdot \frac{s_j}{\sum_j s_j} \cdot \frac{b_j}{\sum_j b_j}}{328,472}$$  \hspace{1cm} (4)

There were some small differences between the original estimated size of each large block ($M_i$) and the sum of the segments ($\sum_j s_j$). The resulting differences in probabilities of selection were compensated for by one of the components of the sampling weight, as described below in Section 4.

2.5 Selection of Respondents

At each sampled household, an enumeration of all household members by age and gender was first completed with any responsible adult to determine whether any eligible persons resided there. When the household contained more than one eligible person aged 18 or over, the eligible persons were assigned sequential respondent numbers in a standard sequence by age within gender, beginning with the oldest male and continuing to the youngest female. A respondent selection table affixed to each enumeration form was then consulted to determine which adult to interview in a household with that number of eligible adults. The selection tables were taken from Leslie Kish, *Survey Sampling*, page 399, and are designed to closely approximate equal probability selection of respondents in households of up to six eligible members.
Once an adult was selected, an attempt was made to administer the Prevalence Interview to that person; no substitutions were allowed. If the selected person turned out to qualify as a career marijuana user, he or she was contacted later, and an attempt was made to administer the In-Depth Interview to that person. The outcome of these attempts to obtain interviews is reported next.
3. FIELD OUTCOME AND RESPONSE RATES

The disposition of each sampled housing unit and each identified career marijuana user is summarized in this section. We will account first for the Prevalence Interviews in the selected housing units. We will then account for the career marijuana users that were identified by the study criteria. Finally, we will summarize the results for individual geographic areas.

3.1 Field Results for Prevalence Interviews

The prevalence results for the sample as a whole are summarized in Exhibit 3-1. As shown there, a total of 1,951 housing units were selected by the sampling procedures described above in Section 2. Of the total selections, 13.3 percent were excluded because they were vacant or otherwise ineligible, leaving 1,691 housing units for the sample. Of these housing units, 962 (56.9 percent) were successfully enumerated -- that is, an interviewer was able to determine whether any eligible persons resided there and, if so, to list the eligible persons and select one at random. In the enumerated housing units, 71 selected adults either were unable to participate or refused to complete the Prevalence Interview. A final total of 891 randomly selected adults completed the Prevalence Interview and 349 of those 891 (39.2 percent) were later identified as career marijuana users.

3.2 Field Results for Career Marijuana Users

The results for career marijuana users are shown in Exhibit 3-2. A total of 349 career marijuana users were identified, of whom three did not speak English adequately for the In-Depth Interview and therefore were treated as ineligible. Of the 346 eligible career marijuana users, 23.1 percent either refused or were never found at
# Exhibit 3-1

## FIELD RESULTS FOR PREVALENCE INTERVIEWS

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>% of Total</th>
<th>% of Valid</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Selections</strong></td>
<td>1951</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td><strong>Ineligible for the sample</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacant</td>
<td>85</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>Not a residence</td>
<td>46</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>No one English/Spanish speaking</td>
<td>129</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td><strong>Total ineligible</strong></td>
<td>260</td>
<td>13.3</td>
<td></td>
</tr>
<tr>
<td><strong>Eligible households</strong></td>
<td>1691</td>
<td>86.7</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Non-response: Enumeration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refused enumeration</td>
<td>595</td>
<td>30.5</td>
<td>35.2</td>
</tr>
<tr>
<td>Never at home</td>
<td>95</td>
<td>4.9</td>
<td>5.6</td>
</tr>
<tr>
<td>Not-accessible</td>
<td>39</td>
<td>2.0</td>
<td>2.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>729</td>
<td>37.4</td>
<td>43.1</td>
</tr>
<tr>
<td><strong>Enumerated household</strong></td>
<td>962</td>
<td>49.3</td>
<td>56.9</td>
</tr>
<tr>
<td><strong>Non-response: Prevalence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unable to participate</td>
<td>38</td>
<td>2.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Refused</td>
<td>33</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>71</td>
<td>3.6</td>
<td>4.2</td>
</tr>
<tr>
<td><strong>Completed Prevalence Interviews</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non career users</td>
<td>542</td>
<td>27.8</td>
<td>32.1</td>
</tr>
<tr>
<td>Career users</td>
<td>349</td>
<td>17.9</td>
<td>20.6</td>
</tr>
<tr>
<td><strong>Total completed</strong></td>
<td>891</td>
<td>45.7</td>
<td>52.7</td>
</tr>
</tbody>
</table>
## Exhibit 3-2

**FIELD RESULTS FOR CAREER MARIJUANA USERS**

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>% of Total</th>
<th>% of Valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Career Users</td>
<td>349</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Ineligible for the study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not English speaking</td>
<td>3</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Eligible persons</td>
<td>346</td>
<td>99.1</td>
<td>100.0</td>
</tr>
<tr>
<td>Non-response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refusals</td>
<td>68</td>
<td>19.5</td>
<td>19.7</td>
</tr>
<tr>
<td>Never at home</td>
<td>12</td>
<td>3.4</td>
<td>3.5</td>
</tr>
<tr>
<td>Total non-response</td>
<td>80</td>
<td>22.9</td>
<td>23.1</td>
</tr>
<tr>
<td>Completed in-depth Interviews</td>
<td>266</td>
<td>76.2</td>
<td>76.9</td>
</tr>
</tbody>
</table>
home to complete the In-Depth Interview. The remaining 76.9 percent (266 persons) completed the In-Depth Interview.

3.3 Field Results for Each Geographic Area

A summary of the response rates and proportions of career marijuana users for each geographic area is given in Exhibit 3-3. The column labeled "Prevalence Response Rate" corresponds to the last row of Exhibit 3-1 for the sample as a whole -- the percent of valid housing units at which a Prevalence Interview was successfully completed. The column labeled "Proportion Career User" shows the unweighted proportion of career marijuana users among those who completed the Prevalence Interview. The column labeled "Career Users Response Rate" corresponds to the last row of Exhibit 3-2 -- the percent of eligible career marijuana users that completed the In-Depth Interview. The final column, labeled "Combined Response Rate", gives the product of the Prevalence and the Career Users response rates. The Prevalence Response Rate was used in calculating weights for estimating the prevalence of drug use in San Francisco. The Combined Response Rate was used in calculating weights for the analysis of data on career marijuana users. The procedures for calculating those weights are described next.
### Exhibit 3-3

**FIELD RESULTS BY GEOGRAPHIC AREA**

<table>
<thead>
<tr>
<th>Geographic Area</th>
<th>Prevalence Resp. Rate</th>
<th>Proportion Career User (unweighted)</th>
<th>Career Users Resp. Rate</th>
<th>Combined Resp. Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinatown/N. Beach</td>
<td>40.0</td>
<td>29.0</td>
<td>77.8</td>
<td>31.1</td>
</tr>
<tr>
<td>Nob Hill/Marina</td>
<td>42.7</td>
<td>36.4</td>
<td>87.5</td>
<td>37.4</td>
</tr>
<tr>
<td>Pacific Hts</td>
<td>39.5</td>
<td>42.9</td>
<td>66.7</td>
<td>26.3</td>
</tr>
<tr>
<td>Univ SF</td>
<td>70.0</td>
<td>45.2</td>
<td>63.2</td>
<td>44.2</td>
</tr>
<tr>
<td>Haight-Ashbury</td>
<td>55.9</td>
<td>60.7</td>
<td>79.5</td>
<td>44.5</td>
</tr>
<tr>
<td>Western Addition</td>
<td>52.1</td>
<td>48.6</td>
<td>83.3</td>
<td>43.4</td>
</tr>
<tr>
<td>Tenderloin/S Market</td>
<td>66.2</td>
<td>48.9</td>
<td>78.3</td>
<td>51.8</td>
</tr>
<tr>
<td>Mission/Potrero</td>
<td>69.2</td>
<td>50.0</td>
<td>83.3</td>
<td>57.7</td>
</tr>
<tr>
<td>Mission</td>
<td>63.7</td>
<td>34.5</td>
<td>73.7</td>
<td>47.0</td>
</tr>
<tr>
<td>Castro</td>
<td>71.2</td>
<td>61.9</td>
<td>84.6</td>
<td>60.2</td>
</tr>
<tr>
<td>Noe Val/Twn Pk/Brln Ht</td>
<td>51.3</td>
<td>48.3</td>
<td>75.9</td>
<td>38.9</td>
</tr>
<tr>
<td>Outer Mission</td>
<td>48.0</td>
<td>30.6</td>
<td>80.0</td>
<td>38.4</td>
</tr>
<tr>
<td>Bayview/Hunters Pt</td>
<td>55.4</td>
<td>19.6</td>
<td>82.8</td>
<td>45.4</td>
</tr>
<tr>
<td>GG Hc/W Portal/Inglesd</td>
<td>56.0</td>
<td>22.5</td>
<td>68.4</td>
<td>38.3</td>
</tr>
<tr>
<td>Sunset/Merced</td>
<td>44.3</td>
<td>24.2</td>
<td>68.8</td>
<td>30.5</td>
</tr>
<tr>
<td>Richmond #1</td>
<td>52.2</td>
<td>41.7</td>
<td>66.7</td>
<td>34.8</td>
</tr>
<tr>
<td>Richmond #2</td>
<td>55.6</td>
<td>36.9</td>
<td>83.3</td>
<td>46.3</td>
</tr>
<tr>
<td>Overall</td>
<td>52.7</td>
<td>39.2</td>
<td>76.9</td>
<td>40.5</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

There are three components to the weight that compensates for unequal selection probabilities, referred to as the sampling weight. First, the probability that a specific housing unit would fall into the sample was not the same in every Census tract and block. As described above in Section 2.4, there were three main probabilities for selecting households: 1/109.5, 1/219, and 1/438. The corresponding weight is 1/H, where H is the probability for selecting households. The second component of the sampling weight is the adjustment for the segmented blocks. The probability of selecting housing units in segmented blocks (S) varied by a factor in the range of .8729 to 1.0420. The relative weight to compensate for this factor is 1/S. The third component of the sampling weight is an adjustment for the number of eligible adults in each selected household. Since only one eligible adult was selected to be interviewed, persons residing in households with more eligible persons were less likely to be selected than persons residing in households with fewer persons. The relative weight to compensate for this factor is P, where P is the number of eligible persons in the selected household (1-4 = actual number, and 5 = 5 or more).

The sampling weight (sampwt) for each case in the Prevalence data file is the product of these three adjustments: sampwt = k*P/(H*S), where k is a constant to scale the weight so that the weighted number of cases in each of the data files equals
the unweighted number of cases.

4.2 Non-response Adjustment

Without a non-response adjustment we implicitly assume that the non-responders are like the average respondent. However, since the individuals in each geographic area are somewhat homogeneous ethnically and perhaps in other ways as well, it is preferable to assume that the non-responders are like the average respondent in the same geographic areas rather than the average respondent overall. Since different geographic areas had different response rates, it is worth making an adjustment for that differential rate of non-response.

In order to adjust for this differential response rate, we assigned each case a factor which was inversely proportional to the Prevalence response rate in its geographic area. For example, cases from a geographic area with a 65 percent Prevalence response rate would be weighted 1/.65 = 1.54, whereas cases from a geographic area with a 55 percent Prevalence response rate would be weighted 1/.55 = 1.82. The largest Prevalence non-response adjustment is 1/.395 = 2.53; the smallest is 1/.712 = 1.40; the ratio of largest to smallest is 2.53/1.40 = 1.81.

There are also differential response rates for the In-Depth Interviews among the career marijuana users. Therefore, besides the Prevalence response rate adjustment, we also assigned each career user another adjustment which was inversely proportional to the Career User response rate in its geographic area. The largest Career User non-response adjustment is 1.58; the smallest is 1.14; the ratio of largest to smallest is 1.58/1.14 = 1.38. The adjustments for differential Prevalence non-response and for Career User non-response were not used by themselves; rather, they were combined with the sampling weight to create two combined weights, as described next.
4.3 Combined Weights

The sampling weight and the two non-response adjustments were combined into two combined weights. For the Prevalence data, the combined weight is basically the ratio of the sampling weight \( sampwt \) divided by the Prevalence response rate \( (prev\_rr) \):

\[
combwt(prev\_prevalence) = k_1 * sampwt / prev\_rr
\]

For the In-Depth Interview data for the career users, the combined weight is basically the ratio of the sampling weight divided by the product of the Prevalence response rate and the Career User response rate:

\[
combwt(career\_user) = k_2 * sampwt / (prev\_rr * user\_rr)
\]

In creating the two combined weights, \( k_1 \) and \( k_2 \) are two constants to scale each weight so that the weighted number of cases both in the Prevalence data file and in the Career User data file equals the unweighted number of cases in each data file.

4.4 Which Weight to Use

For most purposes, the combined weight \( combwt \) will be the appropriate one to use in both data files. It adjusts both for selection probability and for non-response (geographically), and the number of weighted cases equals the actual number of completed interviews in each of the two data files.

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 \( combwt \) is that non-respondents in a particular geographic area are like the average respondent in the same geographic area. 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. Nevertheless, we should note that although the adjustment for differential non-response affects some estimates, the sampling weight component seems to contribute the main impact of the weights on the survey estimates.