MIDUS Refresher 1 Technical Report

on Sampling and Weighting

Charles D. Palit, Barry T. Radler, and Vicki Lein

Updated April 2025

Overview

The MIDUS Refresher baseline (MR1) survey data collection was conducted with two separate grants from the National Institute on Aging, which were lagged in time. Thus, sample data comes from two independent samples, referred to throughout this document as such:

1. Younger decades (MRY): Adults between the ages of 25 and 54 living in residential housing units in the USA. These cases were fielded and completed phone data collection between November, 2011 and October, 2012.

2. Older decades (MRO): Adults between the ages of 55 and 74 living in residential housing units in the USA. These cases were fielded and completed phone data collection between July, 2013 and May, 2014.

Total fielding period for the MIDUS Refresher 1 phone survey spanned 2 years 6 months, with an 8-month gap between the fielding of the MRY and MRO samples. The distinct fielding periods for these two samples were the result of NIA funding exigencies.

While there may be a small overlap in the frames used to draw these samples (caused by the time difference in their selection) the intent of the design is to produce two samples coming from two non-overlapping frames covering the entire population.

Sample Selection

The MRY Sample

Three Frames were used to select the MRY Sample.

- An RDD telephone sampling frame

- List frame targeted to reach each age bracket, 25 to 34 year olds, 35 to 44 year olds, and 45 to 54 year olds
- Cell phone sample frame made up of all possible cellphone phone numbers.

The RDD frame consisted of all possible blocks of 100 sequential and contiguous numbers generated from the set of all eligible area-code prefixes in the USA, which contained at least one listed phone number.

The list frame was drawn from a list of phone number/address combinations for which rough household information on age (often imputed) was available. This allowed the stratification of the pool into three strata, 25 to 34 year olds, 35 to 44 year olds, and 45 to 54 year olds.

The cell frame consisted of all possible phone numbers from the pure cell phone area code-prefix combinations in the USA. Cell phones from the landline-phone frames were covered by the sample from the landline phone frames.

Probability samples of phone numbers were selected independently from each frame. This sample of phone numbers was drawn as a number of statistically identical samples, which we call replicates. The number of replicates issued was used to control sample size. A replicate, once issued, became part of the sample. The release rule DID NOT allow the release of partial replicates.

In the RDD and List frames the sample of phone numbers designated a sample of housing units. A within-household respondent selection procedure was used to select a respondent from the pool of eligible respondents present in the household. The procedure allowed the within-household selection probability of potential respondents to vary from replicate to replicate. This latter device, together with the number of replicates released, was used to exercise stochastic control on the demographic distribution of the final sample of completes so that the released sample produced approximately the target sample sizes required by the study design. Note that these targets were achieved by controlling the sample selection probabilities, not the sample size directly.

In the case of the cell phone frame, the sample phone numbers pointed directly to sample individuals. In the cell phone sample the only potential respondent was the person answering the telephone.

MRY Eligibility

The eligible respondents for the MRY survey were English-speaking adults living in residential units in the USA between the ages of 25 through 54. Eligibility was determined at the time of contact. An additional eligibility requirement for the CELL PHONE sample excluded all sample elements who had a landline telephone through which they could be reached and qualify for the survey under the study rules for landline respondents and households. As a result, the CELL phone frame (and sample) has no overlap with the RDD or LIST phone frames.

MRY Frame Specific Probabilities and Weighting Procedures

In the CELL phone sample the person answering the cell phone was the respondent if they met the age qualification, so the frame-specific person selection probability is the selection probability for the CELL phone number, π_i.

π_i = (The size of the cell phone sample/The size of the cell phone frame).

By size we mean the number of population units (phone numbers) in the group (sample or frame).

The design weight (pre-post stratification record level weight) for the ith cell phone sample respondent, which is $(1/(\pi_i r_i)$, where π_i is the selection probability of the ith observation, and r_i is the response rate appropriate to the ith observation divided by 100.

The RDD and LIST sample phone numbers were drawn from landline telephone numbers which point at households. These households may contain more than one eligible respondent. Upon reaching a household using these sample phone numbers, a respondent was selected from the household with known probability. As a result, the frame-specific person selection probabilities for respondents from these two frames is the product of the probability of selecting the household and the probability of selecting the person, given that the person's HU was selected into the sample.

Since the RDD sample frame and the LIST frame have overlap, persons selected in the RDD frame have a chance of being selected in the LIST frame and vice versa. The sample design contained a rule that HU's actually selected as part of the RDD sample from the RDD frame could not be selected as part of the LIST sample. This rule makes sense operationally and simplifies the computation of the sample selection probabilities. For the persons selected in the RDD and LIST frames, the probability of being in the unified sample is the sum of the frame specific probability of the person being selected in the RDD sample and the frame specific probability of the person being selected in the LIST sample. (Lohr, S. and Rao, J.N.K. 2006), (Palit C.D. 2006).

The design weights for the survey are computed using the person selection probability for the unified frame and an estimated propensity to respond. The propensity to respond used in each case was the estimated response rate for the sample point for the frame through which the phone number/housing unit entered the sample during the selection process.

The design weights for j^{th} person selected from either the RDD or the LIST frame is calculated as

$W_j = (1/(\prod_{1j} + \prod_{2j}))$, where

$\prod_{1j} = ((\pi_1 \ r_1)p_j)$ if $x_j \in$ Frame 1 and 0 otherwise,

$\prod_{2j} = ((\pi_2 \ r_2)p_j)$ if $x_j \in$ Frame 2 and 0 otherwise,

p_j = the probability with which the jth respondent was selected from its housing unit, and

r_i = the response rate divided by 100.

Final weights used in the study are produced from these design weights after the data was post stratified by education, age, and race.

The MRO Sample

Two frames were used to produce the MRO Sample

- An RDD landline telephone sampling frame similar to the one used in the MRY sample.
- Cell phone sample frame made up of all possible cellphone phone numbers.

Probability samples of phone numbers were selected independently from each frame. The cell phone sample was a simple random sample of cell phone-numbers drawn from the cell phone frame. In practice the frame was built from the phone numbers in the set of 1000 series blocks of numbers reserved for cell phones. The landline phone sample was also a random sample drawn from the set of phone numbers built all blocks of 100 contiguous phone numbers with one or more listed phone numbers.

The MRO CELL Phone Sample

As before, in the CELL phone sample the person answering the cell phone was the respondent if they met the age qualification, so the frame-specific person selection probability is the selection probability for the CELL phone number, π_i.

π_i = (The size of the cell phone sample/The size of the cell phone frame).

By size we mean the number of population units (phone numbers) in the group (sample or frame).

The design weight (pre-post stratification record level weight) for the ith cell phone sample respondent, which is ($1/(\pi_i r_i)$), where π_i is the selection probability of the ith observation, and r_i is the response rate appropriate to the ith observation divided by 100

The MRO RDD Sample

These RDD frame phone numbers designated a sample of housing units from which a respondent was selected with known probability. For some of the phone numbers in the RDD sample it was possible to obtain information indicating the possible demographic composition of the household which might be attached to that phone number. Using this information the RDD sample was divided into three strata as follows:

1. Phone numbers with high probability of pointing to households containing persons aged 66 thru 75

2. Phone numbers with high probability of pointing to households containing persons aged 56 thru 65 and no persons over 65.
3. Other sample phone numbers

A phone number was judged to have a high probability of containing a person between the ages of 66 and 75 if one or more persons in that age group were reported as being in the attached household. Similarly a phone number was judged to have a high probability of containing a person between the ages of 56 and 65if at least one person in that age group were reported as being in the household attached to the phone numbers

Each stratum was divided into replicates which were statistically identical or exchangeable. The initial number of replicates in each stratum was the same, but the number of replicates released from each stratum varied by stratum. Replicate release at the stratum level and the within household selection probabilities assigned to the sampling units therein were used to control the sample size and demographic distribution. Once a replicate was issued all of the sample points in any released replicate were counted as part of the sample.

Household selection probabilities for the RDD sample varied with the stratum, 1, 2, or 3 from which the sample was drawn. For the kth stratum, this is P_k, k=1, 2, or 3.

P_k = ((No. of replicates released in stratum)/(No. of replicates originally generated in the stratum))

$P_k = (gk/G) /I,$

Where I = the sampling interval for the original sample,

G = The number of replicates generated for the original sample, and

g = The number of replicates released in the stratum

The personal selection probability is given by

$\quad P_k p_j r_k$, where

p_j = the probability with which the jth respondent was selected from its housing unit, and

r_k = the response rate for stratum k divided by 100.

The weight for the jth respondent is $1/(P_k p_j r_k)$.

MRO Eligibility

The eligible respondents for the MRO survey were English speaking adults living in residential units in the USA between the ages of 56 through 75. Eligibility was determined at the time of contact.

MRO Frame Specific Probabilities and Weighting Procedures

The cell phone frame and RDD frames are as defined for the MRY sample, except the eligible age was defined to be 56 through 75. As in the MRY sample the cell phone sample and the RDD sample come from non-overlapping frames.

The need to define the ages of the elements in the MRO sample as being between 56 and 75 is an effort to compensate for the fact that the MRO sample fielding took place one year later than the original MRY sample. Because the MRO sample was fielded approximately one year later, 54-year-olds in the younger sample would have turned 55 by the time the MRO sample was fielded. Left unadjusted, this would have resulted in the accumulation of more 55-year-old completes than necessary or desired in the MRO sample. Put another way, if the MRO sample was fielded without adjustment (as 55 to 74), then the probability of selection for 54 year olds would be inflated, i.e. 54 year olds would have 2 ways of "entering" the Refresher sample; through the first (MRY) and second (MRO) fielding periods.

In the RDD sample, respondents were drawn only from housing units with a landline telephone. Upon reaching a landline household, a respondent was selected from the household with known probability. As a result, the frame-specific person selection probabilities for the RDD frame is the product of the probability of selecting the household and the probability of selecting the person, given that the person's HU was selected into the sample.

In the CELL phone sample, the person answering the cell phone was the respondent. The result is that the frame-specific person selection probability is the selection probability for the CELL phone number.

The design weights for the survey are computed using the person selection probability for the unified sample and an estimated propensity to respond. Final weights were produced from these design weights after the data was post stratified by education, age, and race.

Combining and Weighting the Two Samples

Since the MRY Sample and the MRO Sample are essentially disjointed, but contiguous, the two samples can be pooled with the existing weights. Once the MRY and MRO datasets were combined, the full MR dataset was compared to the Census Current Population Survey on a variety of demographic strata, and final weight adjustments were made. Table 1 shows the CPS comparison with MR using unweighted and the final weighted data.

Table 1. Comparison of MIDUS Refresher and the Current Population Survey

MIDUS Refresher	CPS (October 2012) [1]	Unweighted MR data[2]	Weighted MR data
	(N=81,379)	(N=3,577)	(N=3,577)
	%	%	%
REGION			
Northeast	20.5	18.8	20.5
Midwest	23.0	28.9	23.0
South	31.4	33.9	31.4
West	25.1	18.3	25.1
SEX			
Male	47.9	48.1	47.9
Female	52.1	51.9	52.1
RACE[3]			
White	82.5	81.8	82.5
Black	9.7	6.3	9.7
Others	7.8	11.8	7.8
AGE[4]			
25-34	20.9	19.4	20.9
35-44	20.9	20.3	20.9
45-54	23.3	20.4	23.3
56-65	21.3	19.9	21.3
66-75	13.6	20.0	13.6
EDUCATION			
12 years or less	39.2	23.2	39.1
13 – 15 years	27.6	30.7	27.7
16 years or more	33.2	46.1	33.2
MARITAL			
Married	62.5	64.0	62.6
Separated	2.3	2.1	2.3
Divorced	12.9	13.4	12.9
Widowed	3.7	5.6	3.7
Never Married	18.6	15.0	18.6

---

[1] CPS data filtered by age: >/= 25 & </=74;sampled age of MR data is reported in Table 1.

[2] MR data: MR_CombinedSample_Charlie_20160616.sav; reporting valid percent when there is missing data.

[3] Race was trichotomized and Other Specify responses were included in coding Race.

[4] The sampled age (rather than the computed age) was used to compute the MRO weights. Further, the post-stratified weights were computed by adjusting the MRO sample (56-75) to match the proportion of 55-74 year olds in the CPS.

References

United States Census Bureau. (2012). Oct12pub.dat dataset. Retrieved 11/19/2012. Available from: http://thedataweb.rm.census.gov/ftp/cps_ftp.html

Hartley, H.O. (1962). Multiple Frame Surveys, Proceedings of The Social Statistics Section, American Statistical Association, 203 ff.

Lohr, S. and Rao, J.N.K. "Estimation in Multiple-Frame Surveys" JASA September 2006, Vol.101. No.475

Palit, C. D. (2006). "An Alternative Estimator for Multi-Frame Sample Designs", Proceedings of the Survey Research Methods Section, ASA.