Selecting a Sample from a Changing Frame of Program Beneficiaries: An Application of Adaptive Sampling

Eric Grau
Mathematica Policy Research, P.O. Box 2393, Princeton, NJ 08543-2393

Abstract
Probability samples are usually selected from a fixed sample frame, which is a close facsimile of the target population. In those situations, the probability of selection can be easily quantified. The sample design becomes more complicated when the sample frame is not fixed, but changes over time. In this paper, we describe a sample design for a target population that represents program beneficiaries with disabilities who had successful work experience. The highest priority in this design is to minimize the length of time between the interview date and the period of successful work. We describe a design that was intended to do this, while at the same time accommodating the changing frame. We describe issues with the sample design's implementation in the first round of sampling. We show how the entire sample frame was divided into seven segments that were revealed over time, and how the sample allocations changed as we moved through time. We revisit some of the assumptions made when designing the sample, and assess what changes would be required in the next round of data collection.

1. Introduction

The Social Security Administration (SSA) is interested in understanding the work interest and experiences of Supplemental Security Income (SSI) recipients and Social Security Disability Insurance (SSDI) beneficiaries. Since 2003, SSA, with assistance from its contractor Mathematica Policy Research, Inc., has conducted six rounds of the National Beneficiary Survey (NBS), with planning currently underway for the seventh round. The NBS seeks to uncover important information about the factors that promote beneficiary self-sufficiency and, conversely, the factors that impede beneficiary efforts to maintain employment. Rounds 6 and 7 of the NBS include surveys from two samples, one selected from among all SSI and SSDI beneficiaries (the "representative beneficiary sample," or RBS), and the other among those SSI and SSDI beneficiaries called "successful workers," who were able to sustain a minimum level of earnings over a period of time (the "successful worker sample," or SWS). These surveys use the same survey instrument, but were conducted independently and simultaneously. The RBS has consisted of cross-sectional samples of about 4,000 individuals in each round, which in Round 6 we drew from a single frame of 14 million SSI and SSDI beneficiaries. Among those 14 million, there were about 200,000 successful workers in Round 6 who sustained monthly earnings above the non-blind Substantial Gainful Activity (SGA) level[1] for at least 3 consecutive months. For the SWS, the target number of completed interviews was 4,500, and in Round 7 the target number of completed interviews will be 3,000. The remainder of this paper will focus on the SWS. Our objective for the SWS was to obtain samples of individuals from this subpopulation. The difficulty which this paper seeks to address is how to sample these cases when the administrative data used to identify the earnings are not immediately available, but become available over an extended period of time.

[1] SGA was defined as $1,130 per month for 2016 and $1,180 per month in 2018.

This type of sampling is relevant any time a short-lived aspect of the target population requires updated information over the course of the field period. An example of such a short-cycle event includes hospital stays, recent job searches, or large purchases. In such cases, you would want to talk to respondents shortly after the event, and this becomes difficult if the field period is long and is compounded by reporting delays regarding the event of interest, which often occurs.

2. Target Population

For the purposes of the NBS, successful workers are defined as SSI or SSDI beneficiaries as of June 30 of the target year, who are less than 62 years old on that date, and have earnings above SSA's non-blind SGA earnings level for a minimum of three consecutive calendar months. Earnings are determined using administrative data from SSA in the Disability Control File (DCF). For some successful workers, administrative earnings data are available on the DCF shortly after the work occurs but, for other beneficiaries, there is a gap in time of up to three years between the first three-month period of successful work and the appearance of the earnings data in the DCF that allows their identification as successful workers. Therefore, it was not possible to include in the target population all of the successful workers; those who possessed this long lag were missing. Another constraint that SSA placed on the target population was that the successful work had to have occurred within six months prior to the interview date, to ensure that they could recall characteristics of their successful work. From preliminary work that was done with simulated populations of successful workers using older data, we had estimated that there were about 150,000 to 200,000 successful workers in a given year, but we also had estimated that only about 15,000 successful workers both met the success criteria and could be identified within six months of doing so. SSA was concerned that this would not capture enough of the population they were interested in because: (1) we were missing too many successful workers that had the aforementioned lag and (2) the successful work was too concentrated in the time period closest to a single sample selection date.

To avoid this problem, we expanded the population so that the successful work could occur at any time between August 1 of the target year and July 31 of the data collection year a year later. (The target year for Round 6 was 2016 and the data collection year was 2017.) We also decided to create multiple frames, instead of just one (details are given below).

3. Sampling Frames

The sampling frame for the SWS must necessarily be a subset of the sampling frame used for the RBS sample: SSI and SSDI beneficiaries who were active or in suspense status as of June 30 of the target year. In order to identify successful workers whose successful work could occur at any time in the year and to ensure that the time between the successful work and the interview date did not exceed six months, we created seven successive sample frames in Round 6 instead of one. This addressed the issue of obtaining successful workers throughout the year, where the time between the successful work and interview date could be less than six months. It also partially addressed the lag problem, at least for successful workers whose successful work persisted for more than three months. These frames consist of those who can be identified as successful workers using administrative data that is available on or before September 1 of the data collection year.

We obtained these frames by requesting extracts of administrative data from SSA (approximately) every six weeks, to ensure that enough new successful workers could be identified in each new extract. These extracts were identified as those beneficiaries who met the successful work criteria. We constructed a total of seven successive extracts, the first of which is pulled on the first Monday after December 1 of the target year. For the next five of the successive frames, we extracted data on the Monday or Tuesday after the following dates in the data collection year: January 15, March 1, April 15, June 1, and July 15. Due to the short data collection window available for successful workers in the final extract, we performed the extraction for the final frame on the Tuesday before September 1.

For those who met these criteria to be included in the extract, sample members were asked in the questionnaire if they had worked in the past six months. If they answered negatively, they were screened out. The period between the last month of successful work and the interview date was limited to six months to avoid issues of recall about the sample member's successful work period. The issue of recall could affect the accuracy of the responses to the screening question to ensure the successful work occurred recently. To mitigate this risk, we defined the extracts so that the potential elapsed time period between the final identified month of the successful work period and the interview date did not exceed six months. This means that each extract had to be limited to successful workers whose successful work ended late enough to satisfy this requirement. Table 1 summarizes the earliest acceptable final month of successful work for a successful worker to be included in each extract in Round 6, given the time period when data collection began.[2] Also included in this table is the first month of ineligibility for those whose successful work actually ended on the earliest acceptable final month shown.

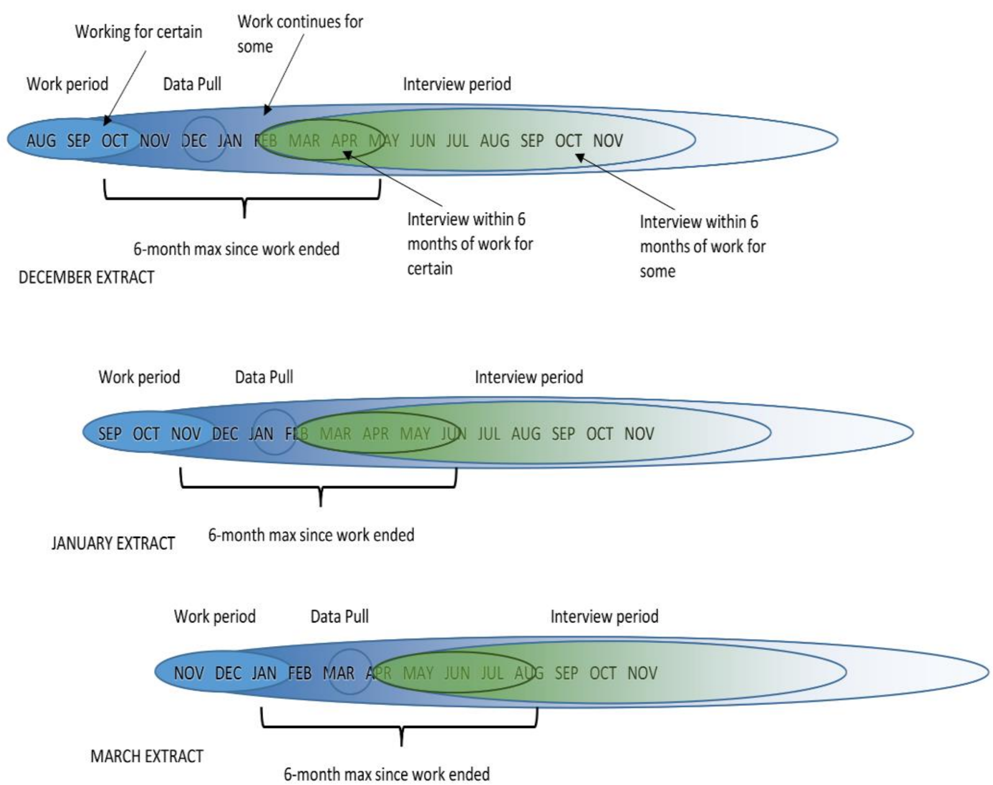
Table 1: Earliest acceptable final identified month of successful work for each extract, and resulting first month of ineligibility for the Round 6 SWS

Extract	*Earliest acceptable final month of successful work*	*Dates when data collection period began*	*First month of ineligibility for those with earliest acceptable final month of successful work*
Dec. 1, 2016	October, 2016	February 1, 2017	May, 2017
Jan. 15, 2017	November, 2016	February 23, 2017	June, 2017
Mar. 1, 2017	December, 2016	April 17, 2017	July, 2017
Apr. 15, 2017	February, 2017	June 12, 2017	September, 2017
June 1, 2017	March, 2017	July 28, 2017	October, 2017
July 15, 2017	May, 2017	September 2, 2017	December, 2017
Sept. 1, 2017	June, 2017	September 29, 2017	January, 2018

The window of time that a successful worker could be identified for inclusion in an extract, selected for the sample, and have an attempted interview, is illustrated in Figure 1 for three of the seven extracts. The figure shows the length of time between the successful work and the interview, and how this elapsed time must not exceed six months. The first oval corresponds to the first sample extract, which is limited to those whose successful work either ended in October or November of the target year, or continued at the time of the

[2] Given that data collection in the first extract began on February 1, we would expect that most interviews in the sample from this extract would have been completed by May, 2017. The gap between October 31, 2016 and May 1, 2017 just exceeds six months, but most of those successful workers would have been interviewed in February, March, or April, 2017.

extract creation in early December. It excludes those whose three consecutive months of successful work ended earlier than October of the target year. This is because, for the December extract, we estimated that the successful workers' interview date could be as late as April of the data collection year. For someone whose successful work ended in September, this would be more than six months of recall. It is possible that the interview date would be sooner than April of the data collection year, in which case we would be excluding someone from the frame whose successful work ended fewer than six months beforehand. By the same token, if the interview was in May, someone whose successful work ended on October 31 would have more than a six-month gap until the interview date (and would be screened out from the screener question in the questionnaire). However, constructing the frames in this way ensures that most will have a gap that is less than six months, and that few cases would be screened out based on the response to the screening question in the questionnaire.



Note: Solid ovals identify the "for certain" periods, and gradients represent the decline in certainty over time.

Figure 1. Timeline for extracts in Successful Worker Sample, including work period, data pull dates, and admissible data collection period for each extract.

Using these constraints to define the target population for the sample in this round, we created seven sample frames with a total of 89,939 successful workers in Round 6. However, we believe there were approximately 200,000 individuals who were successful workers in Round 6, which means that over half could not be identified in time to be included in the sample frames.[3]

4. Sample Design

For all survey rounds, the NBS has used a multistage sampling design. In Round 6, we used such a design for both the RBS and SWS, with an independently drawn supplemental single-stage sample for the SWS. We drew the SWS and RBS independently, from separate frames, although the SWS frame was a subset of the RBS frame. The RBS and the main sample of the SWS involved selecting individuals within selected clusters of geographic areas, and is therefore referred to as a "clustered sample." The supplemental sample (for the SWS only) was selected across the entire population of successful workers and was therefore not limited to those residing in selected clusters. It is therefore referred to as an "unclustered sample." We created the clustered and unclustered samples because of concerns about the number of successful workers within strata and their distribution across PSUs within each extract. Even though the clustered and unclustered samples are independently drawn, we combine the two samples when the data in the SWS are analyzed using composite weights which account for the probabilities of selection in the two samples. Therefore, the clustered and unclustered samples are referred to as "components" of the SWS as a whole. The combination of the two samples into a single sample is called a "dual-sample design," which is discussed in detail in Touzani et al (2002).

Construction of the PSUs began with county-level counts of beneficiaries in the four age strata that were the strata used for the RBS. We formed 1,330 PSUs that each included one or more counties. For sampling purposes, we used a size measure (Folsom et al. 1987) that incorporates the count of beneficiaries and the desired sampling rate of beneficiaries in each age stratum. This measure of size, referred to as a composite size measure, presents a "population" for each PSU that is essentially a weighted average of the population sizes within each age group, where the weight is the sampling rate. Due to time constraints, we used the same PSUs for the SWS even though the strata differed.

We selected a sample of PSUs from the set of 1,330 PSUs with probability proportional to size and with minimal replacement using Chromy's procedure (1979). We classified two PSUs as certainty selections (Los Angeles County and Cook County[4]), based on the selection frequencies for the PSUs computed using the composite size measure. We allocated the Los Angeles County PSU twice the sample size allocated to the other PSUs due to its population size relative to the other PSUs. To complete the sample of 80 PSUs, we selected 77 PSUs with probability proportional to size (PPS), where the size was defined by the composite size measure.

We used two types of sampling strata for the sample selection in the NBS—explicit strata and implicit strata. Explicit strata are required in cases where oversampling or undersampling are used or in other instances where it is necessary to directly control the size of the sample by certain characteristics. In the clustered component of the SWS, stratification was used in multiple stages. In the first stage of selection, the PSUs were not initially defined within explicit strata; however, with two certainty selections these PSUs became explicit strata. For the clustered component of the SWS, second stage strata (strata

[3] We conducted a study based on simulated data and concluded that, based on administrative data, observable differences between the population with and without the lag were minimal.

[4] Los Angeles County includes the city of Los Angeles; Cook County includes the city of Chicago.

within the PSUs) were defined by beneficiary type (SSDI-only and SSI, the latter of which includes both SSI-only and beneficiaries of both SSI and SSDI) and extract. The unclustered component did not have stages of selection; explicit strata were defined by beneficiary type (SSDI-only and SSI), extract, and whether the beneficiary lived in a PSU. In both the clustered and unclustered components of the SWS, implicit strata were used to ensure the distribution of variables in the samples matched that of the population. Implicit stratification variables included disability diagnosis, beneficiary type (three separate categories: SSDI-only, SSI-only, and beneficiaries of both SSI and SSDI), race and ethnicity, gender, and ZIP code.

5. Sample Allocation of Targeted Complete Interviews

We did not know the size of each extract before sample selection or what the overall proportion will be in the clustered sample or residing in the PSUs for the unclustered sample. The initial sample size allocation[5] to the samples in each extract was based on simulated successful worker populations from prior years. The proportion of the sample that was allocated to the clustered and unclustered samples in each extract was designed to minimize bias and cost. After the release of each extract, we adjusted the allocation of sample sizes to the samples from the remaining extracts to make the allocation as proportional as possible to the population of successful workers over time within each of the two beneficiary-type strata (SSDI-only and SSI). We did not complete sample selection until after the release of the last extract. The allocations that were calculated after the release of selected extracts are given in Table 2.[6]

Table 2: Round 6 SWS Changes in Sample Allocations after the Release of Each Extract

		Allocation						
Sample extract	*Original projected popula-tion*	*Orig-inal*	*After 1st extract*	*After 3rd extract*	*After 5th extract*	*After 6th extract* [a]	*Final popula-tion counts*	*Final completed interviews*
12/1/16	10,500	631	860	860	860	955	17,059	982
1/15/17	12,500	737	697	662	662	696	13,006	723
3/1/17	12,900	773	681	694[b]	694	694	17,595	740
4/15/17	10,500	627	584	589	578	578	11,341	606
6/1/17	11,500	657	614	620	663	664	13,476	582
7/15/17	9,600	573	593	599	581	489	10,109	442
9/1/17	8,400	502	471	476	462	424	7,353	512
Total	75,200	4,500	4,500	4,500	4,500	4,500	89,939	4,587

Source: NBS Round 6 (the second round of NBS–General Waves).
[a] The counts for the first few extracts in this column show the actual number of completed interviews rather than the target. We released too many cases in the earlier extracts to obtain the desired target because we underestimated the unclustered yield rate in the first two extracts.

[5] "Sample size allocation" refers to the target number of completed interviews.
[6] The allocations were revised after every new extract was downloaded. However, to save space, Table 2 only shows the original allocations and how they changed after the first, third, fifth, and sixth extracts.

6. Size of Released Sample

The size of the released sample was difficult to know beforehand since this population had never been sampled before. The number of cases to release depended upon

1. The projected yield rate (YR)

   YR = Number of targeted complete interviews/Number of released cases

2. The proportion of the sample allocated to the clustered and unclustered components.

The last point is important because, in the clustered component of the SWS, phone interview nonrespondents are followed up in the field, whereas all cases in the unclustered component of the SWS must be resolved by phone. We therefore would need to release more cases in the unclustered component than the clustered component for an equivalent number of targeted completed interviews.

6.1 Estimating the Yield Rate

Although this subpopulation had never been sampled before, we have had a long history of surveying general beneficiaries of SSI and SSDI through the RBS. We felt that three things would negatively affect the yield rate of the SWS, when compared to the RBS: (1) all sampled successful workers that indicate that they had not worked in the past 6 months would be screened out,[7] (2) those who are working may be busier and thus more difficult to reach, and (3) later extracts would have less time for data collection due to a compressed schedule. Item (1) improves as we move through the extracts; item (2) stays the same from extract to extract, and item (3) gets worse as we move through the extracts.

The unweighted yield rate in the all-clustered RBS in Round 5 was 52.9 percent. This included four releases with yield rates that decreased with each release. The decrease was particularly apparent for the third and fourth release, mainly due to a shorter data collection period. As a result of this, we decided that most of the sample cases in the Round 6 RBS will come from the first two releases. In this spirit, and also because we did not anticipate more than one release for each extract of the SWS, we used the yield rate for the first two releases (57.3 percent) of the Round 5 RBS as our starting point to project the yield rate for the SWS clustered sample. In Table 3, we provide a summary of the reduction in yield rate from 57.3 percent due to the following three sources: (1) proportion screened out, (2) proportion reduced due to working population, and (3) proportion reduction due to a shorter field period. The final column, the total yield rate reduction, is the sum of the three sources of yield rate reduction. The yield rate reductions range from 16 to 20 percent in the first six extracts, and jumps to 32 percent for the seventh extract, due to the very short data collection period for this extract. The assumed yield rate for the clustered sample is shown in the last column of Table 3, ranging from approximately 41 percent (57 – 16) for the first extract, to 25 percent for the final extract (57 – 32).

[7] We had already limited the frames to exclude cases that had not worked in the prior six months. However, we felt that, for the earlier extracts, some would be screened out based on the questionnaire screener question because the data collection period was ten months long.

Because there is no in-person nonresponse follow-up in the unclustered sample, the yield rate is substantially lower in the unclustered sample than in the clustered sample. Based upon our experience from a prior study with a dual sample design, the yield rate in the unclustered sample was expected to be 10 to 20 percentage points lower than for the clustered sample. Given the difficulty getting telephone interviews in Round 5 of the NBS, the SWS unclustered sample was expected to have a yield rate closer to (or perhaps exceeding) the maximum of that range. Therefore, we estimated that the yield rate in the unclustered sample would be 20 percent for the first six extracts, and 10 percent for the final extract.

Table 3: Total yield rate reductions, by extract

	Yield rate reduction due to				
Extract	*Proportion screened out*	*Successful workers working*	*Shorter data collection period*	*Total yield rate reduction*	*Final projected clustered yield rate*
Dec. 2016	14	2	0	16	41
Jan. 2017	13	2	2	17	40
Mar. 2017	13	2	3	18	39
Apr. 2017	11	2	7	20	37
Jun. 2017	5	2	13	20	37
Jul. 2017	0	2	17	19	38
Sep. 2017	0	2	30	32	25

Table 4 shows how the actual yield rates compared with the initial projected yield rates. As is apparent, our assumptions for the clustered sample were not far off. However, we severely underestimated the yield rate in the unclustered component; the field nonresponse follow-up had very little impact on the overall yield rate in the later extracts due to the short data collection period. We revised our assumed yield rates for the unclustered component after the release of the third extract, assuming unclustered yield rates closer to 30 percent instead of 20 percent.

Table 4: Projected yield rates compared with actual yield rates in clustered and unclustered SWS

	Clustered sample		*Unclustered sample*	
Extract	*Initial projected yield rate*	*Actual yield rate*	*Initial projected yield rate*	*Actual yield rate*
December 1, 2016	41	40	20	32
January 15, 2017	40	37	20	30
March 1, 2017	39	41	20	32
April 15, 2017	37	42	20	31
June 1, 2017	37	32	20	31
July 15, 2017	38	33	20	31
September 1, 2017	25	30	10	28
Total		37		30

Within each stratum (where the extract was part of the defined stratum), we selected an equal probability sample of successful workers by using a sequential selection algorithm with the sampling frame sorted by disability diagnosis, beneficiary title, race and ethnicity, gender, and ZIP code to form the sample. These sorting factors ensured an approximate proportional allocation of the sample across levels of these factors and therefore enhanced the face validity of the sample across these factors. We released a sample of 13,272 successful workers across the seven SWS extracts.[8]

6.2 Determining the Proportion of the SWS in the Clustered and Unclustered Components

Determining the appropriate value for the percentage of the completed sample to allocate to the clustered sample component is somewhat subjective. It requires consideration of the balance among five considerations, three of which push toward the 0 percent clustered end of the scale and two of which push toward the 100 percent clustered end of the scale. The three that push toward 0 percent clustered are:

1. Minimizing the design effect due to clustering[9]
2. Accommodating small populations where the existing population is clumped in a few PSUs, resulting in a large unequal weighting effect, or where the total number of cases available for sampling within PSUs is too small to meet analytic objectives[10]
3. Minimizing cost (though this is mitigated to some degree by the fact that the low yield rates in the unclustered sample will require releasing many more sample cases, decreasing the savings)[11]

The two that push toward 100 percent clustered are:

4. Minimizing bias due to nonresponse
5. Maximizing yield rates[12]

The way we measure the negative impact of clustering and unequal weighting on the variance of estimates is by the effective sample size (ESS). The ESS is the number of

[8] For fielding purposes in the SWS, we selected a larger sample than needed (called the augmented sample) in each extract to ensure that reserve sample cases would be available if we found that the response and eligibility rates during data collection differed from our initial assumptions. We selected 18,400 successful workers in the augmented sample across extracts, of which 13,272 were released for data collection, and 5,128 were held in reserve. The 5,128 were never released due to the short time window of eligibility for some successful workers.
[9] The design effect due to clustering is defined as $deffc = 1 + \rho(m-1)$, where $\rho = 0.01$ is the intraclass correlation coefficient (ICC) and m is the average cluster size. The ICC value is based upon estimated ICC values observed in prior rounds in the NBS.
[10] The design effect due to unequal weighting, deffw, is calculated by taking the product of the sample size and the sum of squared weights and dividing by the sum of weights squared.
[11] For this goal, the lower cost with an unclustered sample assumes no field follow-up. In fact, we know that cases that included field follow-up cost over five times as much as those that did not in the Round 6 data collection effort. Obviously, if we had field follow-up in the unclustered sample (reducing nonresponse bias), the costs would be extraordinarily high.
[12] Clearly these last two goals are interrelated. Higher yield rates often (but not always) lead to lower nonresponse bias.

completed interviews divided by the product of the design effects due to clustering and unequal weighting. The ESS takes only goals 1 and 2 into account.

Given these tradeoffs and the limited information we had at our disposal, it was not possible to find a single "right" answer for the question of what percent clustered to use, but we could try to get close. Our goal was to determine a range of percentage clustered over which the gains from moving toward a larger share in the clustered sample (the value of coming closer to goals 4 and 5) were neither extremely high nor extremely low relative to the costs in terms of variance and expense (of moving away from goals 1, 2, or 3). Having an entirely clustered sample meant we would have minimized bias through a higher response rate, but we may not have had enough observations to choose from within PSUs (or the distribution is such that unequal weighting will be a factor, increasing the variance), and we would pay a big price in ESS. Having an entirely unclustered sample meant we would have the highest possible ESS with the lowest cost, given the total sample size,[13] but we would pay a price in bias and response rate and it would be necessary to release a lot of sample cases.

In Table 5, we provide an illustration of how this decision was made in the Round 6 SWS. The table shows the ESS relative to the proportion of the completed sample that is in the clustered component for the first extract among SSDI-only cases in the first (December) extract of the Round 6 SWS. The table is based upon 372 completed interviews, the target number of completed interviews for the SSDI-only stratum. We used this table to determine the allocation to the clustered and unclustered sample in the December extract of the Round 6 SWS. It shows how the ESS decreases with each increase in the proportion of interviews obtained from the clustered sample. For samples in which the clustered samples account for 20 percent or less of the completed interviews, the effective sample size is 338. As the proportion of the interviews from the clustered sample increases past 30 percent, the decline in the effective sample accelerates. The ESS is calculated separately for the clustered and unclustered sample. For the clustered portion:

ESS_C = (clustered completes)/(deffc*deffw*deffnr),

where deffc is the design effect due to clustering, deffw is the design effect due to unequal sampling weights, and deffnr is the design effect due to post-data collection adjustments such as nonresponse weights. For the unclustered portion:

ESS_U = (unclustered completes)/(deffw*deffnr).[14]

Adding the two together gives us a crude measure of the total ESS, called ESS_E.[15]

Table 5 provides various hypothetical percentages of completed interviews that were clustered, ranging from 0 to 100, plus two interpolated percentages used to obtain optimal values. The first column shows the percentage of respondents from the clustered sample and the second column shows how many sample cases would need to be selected to get a

[13] "Total sample size" includes the entire selected sample, not just completed cases.
[14] Because we are using proportional allocation within extracts, we would expect that the deffw for the unclustered sample will be close to one.
[15] The actual ESS for the combined sample would also include an effect due to compositing the two samples. That measure, however, requires calculation of a shrinkage estimator of the weights that combines weights from the two samples, which can only be done after the percent clustered is determined.

total of 372 completed cases, given the respective yield rates for clustered and unclustered sample members. The third column shows the ESS_E accounting for the design effects due to clustering (for clustered sample) and weighting. The fourth column of Table 5 shows how each ESS_E compares to the maximum ESS_E of 338. The fifth column shows the ratio of the ESS_E to the selected sample, indicating the impact of sample yield and clustering on the ESS_E. In that column, a higher number is better, indicating that the design jointly maximizes yield (minimizing bias) and precision (minimizing variance). The optimal value in the column is associated with a clustered percentage of 56. The final column (which we will refer to as the ratio of ratios) shows how each ratio of ESS_E to selected sample compares to the optimal value of 0.226 for 56 percent clustered.

Table 5: SWS effective sample size for the SSDI-only stratum by percentage clustered, with statistics by level of percentage clustered, Round 6 December extract, for 372 target completed cases

Percent clustered among respondents	*Total selected sample*	*Effective sample size (ESS_E)[a]*	*Percent of maximum ESS_E[b]*	*Ratio of ESS_E to selected sample*	*Percent of maximum ratio[c]*
0	1,860	338	100	0.182	80.5
10	1,764	338	100	0.192	84.8
20	1,669	338	100	0.203	89.6
30	1,574	336	99.3	0.213	94.4
40	1,479	328	96.9	0.221	98.0
50	1,383	312	92.2	0.226	99.8
56	**1,330**	**301**	**89.3**	**0.226**	**100.0**
60	1,288	291	86.0	0.226	99.8
70	1,193	265	78.4	0.222	98.4
78	**1,120**	**243**	**72.0**	**0.217**	**96.2**
80	1,098	237	70.0	0.216	95.4
90	1,002	206	60.9	0.205	90.8
100	908	174	51.4	0.191	84.7

[a]The formula for ESS_E includes design effects due to clustering (deffc), unequal sampling weights (deffw), and post-data collection adjustments (deffnr). The value for deffc varies according to the size of the clusters and is calculated under the assumption that ICC = 0.01, ranging in value between 1.00 and 1.07. The value for deffw varies according to the percentage clustered and ranges between 1.00 and 1.82. The value for deffnr is set at 1.1.
[b]The ratio of the ESS_E over the maximum value of the ESS_E.
[c]This is calculated as the "ratio of ESS_E to selected sample" column divided by the maximum value in that column. We sometimes refer to this as the ratio of ratios.

To save space in Table 5, we did not specify the number of cases in the selected sample for the clustered and unclustered samples separately, only the combination. For each allocation of "percent clustered," we multiplied the percentage by 372 to determine the number of clustered completes and unclustered completes. We then determined the number of sample cases required by dividing the clustered completes by 0.41 (representing the yield rate we had assumed for the Round 6 December extract) and unclustered completes by 0.20

(representing the assumed 20 percent yield rate in the Round 6 December extract).[16] For example, in the SSDI-only stratum for the December extract shown in Table 5, the number of sample cases required to get 372 targeted completed cases is determined by the percentage clustered. For zero percent clustered, all 372 completed cases would be obtained entirely from the unclustered sample, so we can obtain the total selected sample by dividing 372 by 0.20, giving us 1,860 cases (372/0.20 = 1,860). In another example, with 40 percent clustered and 60 percent unclustered, the 372 completed cases comprise 149 clustered completes (372 x 0.4 = 149) and 223 unclustered completes (372 x 0.6 = 223). The 223 unclustered cases can be obtained by dividing 223 by 0.20, giving us 1,115 sampled cases and the 149 clustered cases can be obtained by dividing 148 by 0.41 giving us 364 sampled cases. This results in a total of 1,479 sampled cases (1,115 + 364 = 1,479) in the total selected sample.

If we chose the percentage clustered simply to maximize the ESS_E, we would choose 0 to 20 percent clustered for the December extract in the SSDI-only stratum because, according to the table and plot, that value has the highest value for ESS_E. Because a lower yield rate results in a higher total selected sample size, a statistic that accounts for both ESS_E and the yield rate is the ratio between ESS_E and the selected sample size. This is maximized when the percentage clustered is 56 percent. However, because there is no nonresponse follow-up in the unclustered sample, resulting in a lower yield rate, the potential for nonresponse bias could be a major problem with a large proportion of unclustered cases. Therefore, we believed that this method of determining the percentage clustered does not sufficiently account for low yield rates in the unclustered sample and the resulting potential for nonresponse bias, and recommended using a higher percentage clustered than 56 percent. However, we also did not want to get too far away from this value. In the December extract for SSDI-only in Round 6, we decided to allocate 77.8 percent of the sample to the clustered sample. With 77.8 percent, the ESS_E was 72 percent of the maximum possible ESS_E value and the ratio of ratios was 96 percent. Our strategy, as we proceeded through the extracts, was to choose a value for percentage clustered where the ESS_E was relatively close to the maximum value, but also ensure that the percentage of the maximum ratio of ESS_E to the selected sample (the ratio of ratios) was close to its maximum value. The methodology was ad hoc, reflecting the subjectivity of the exercise. In the first three extracts, we chose a percentage clustered that resulted in an ESS_E that was 72 percent of the maximum ESS_E possible. However, from the fourth extract onwards, the percentage of the maximum ratio of ratios began dropping below 90 percent when the percentage of the maximum ESS_E was set to 72 percent, so we changed our strategy to ensure that the maximum ratio of ratios was at least 91 percent. The percentage clustered for each of the extracts in the Round 6 SWS is provided in Table 6. The last two columns provide the values for the percentage of maximum ESS_E and percentage of the maximum ratio.

Notice the large drop in the percentage clustered in the September (final) extract. This is likely due to the short data collection period; so little time is available for the nonresponse follow-up in the clustered sample that the advantage of the clustered sample over the unclustered sample is muted.

[16] These were the assumed yield rates for the December extract prior to selecting the Round 6 SWS sample for that extract. In fact, the clustered yield rate that we achieved in Round 6, 40.3 percent, was very close to the assumed rate of 41.0 percent for December, as shown in Table 3. However, we underestimated the unclustered yield rate, which was 32.3 percent (also in Table 3).

Table 6: Percentage clustered for the SWS in Round 6, by strata

Data extraction date	*Beneficiary type*	*Percent clustered*	*Percent of maximum ESS_E*[a]	*Percent of maximum ratio*[b]
12/1/16	SSDI-only	77.6	72.0	96.2
12/1/16	SSI	73.3	72.0	95.4
1/15/17	SSDI-only	74.5	72.0	95.2
1/15/17	SSI	72.6	72.0	94.9
3/1/17	SSDI-only	85.8	72.0	90.9
3/1/17	SSI	84.1	72.0	90.3
4/15/17	SSDI-only	70.9	78.1	91.0
4/15/17	SSI	68.6	77.9	91.0
6/1/17	SSDI-only	70.1	78.2	91.0
6/1/17	SSI	68.6	77.9	91.0
7/15/17	SSDI-only	61.3	81.3	91.0
7/15/17	SSI	78.5	78.8	91.0
9/1/17	SSDI-only	39.9	85.8	91.0
9/1/17	SSI	44.2	85.4	91.0

[a]The percent of maximum ESS is a ratio of the ESS over the maximum value of the ESS.
[b]The percent of maximum ratio is a ratio of the value in the "ratio of ESS to selected sample" column divided by the maximum value in this column. We sometimes refer to this as the ratio of ratios.

7. Conclusion

In this paper, we have shown that using rolling sample frame is an innovative way to sample from as much of the target population as possible when the target population is not revealed at one point in time. In the case of the NBS, the SWS was a sample of successful workers, a population with this attribute. We showed how we sampled from this population using a succession of seven sampling frames, called extracts, where it was necessary to adjust the sample allocation after each extract was downloaded. This application had an added complication in that we used a dual sample design, with clustered and unclustered components, each with different yield rates. We needed to see how the distributions of data changed over time, adjusting the allocations and the proportion assigned to the two components. Not only did the allocations change, but the assumed yield rates also had to adapt to changes from the initial assumptions. We discovered that nonresponse follow-up had little effect on the overall yield rate for later extracts, since the data collection periods are much shorter. Our assumptions for the next round of data collection will be revised accordingly.

Acknowledgments

I would like to give thanks to the Social Security Administration (SSA), who partnered with Mathematica to produce the National Beneficiary Survey. In particular, I would like to thank Paul O'Leary of SSA, who provided valuable input for this paper, including the graphic in Figure 1 which he produced. I would also like to thank Barbara Carlson of Mathematica, who also provided valuable comments, and Frank Potter and Elaine Gilby, who along with Paul O'Leary were instrumental in the development of the design.

References

Chromy, J. R. (1979). "Sequential Sample Selection Methods." *Proceedings of the Section on Survey Research Methods,* American Statistical Association: Alexandria, VA. 401-406.

Folsom R. E., F. J. Potter and S.R. Williams (1987). "Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains." *Proceedings of the Section on Survey Research Methods,* American Statistical Association: Alexandria, VA. 792-796.

Touzani, M., F. Potter, and A. Zambrowski (2002). "Sampling Children for a Study of the State Children's Health Insurance Program." *Proceedings of the Section on Survey Research Methods,* American Statistical Association: Alexandria, VA. pp. 3484-3489.