Creating Design Weights for a Panel Survey With Multiple Refreshment Samples: A General Discussion With an Application to a Probability-Based Mixed-Mode Panel

Matthias Sand[1], Christian Bruch[1], Barbara Felderer[1], Ines Schaurer[2], Jan-Philipp Kolb[3] and Kai Weyandt[1]

[1] *GESIS – Leibniz Institute for the Social Sciences*
[2] *City of Mannheim*
[3] *Federal Statistical Office of Germany (DESTATIS)*

Abstract

Panel surveys suffer from attrition, where participants drop out over time. To maintain generalizability, refreshment samples are frequently employed, bringing in new individuals, increasing the number of panelists, and balancing sample composition. Although refreshment samples offer numerous advantages, the inclusion of new panel members may introduce bias into the analysis if the design weights are not appropriately tailored to these new members and adjusted to align with existing panel members. If not correctly accounted for, their inclusion may bias results. This paper addresses the issue of designing proper weights by applying the multiple-frame weighting approach proposed by Kalton and Anderson, which is generally used for cross-sectional surveys, to ongoing panel studies with refreshment samples. We demonstrate its application to a synthetic data set and a probability-based mixed-mode panel with an initial sample and two refreshment samples. We compare estimates obtained using multiple-frame weighting with those obtained using unweighted and naively weighted methods (where design weights are used as calculated for the respective samples without adjusting for the fact that some members of the population have a chance of being sampled more than once due to the refreshments). These comparisons showcase the potential for bias introduced by neglecting proper weighting and underscore the importance of both a multiple-frame weighting approach and meticulous sample documentation.

Keywords: panel surveys, GESIS Panel, refreshment samples, multiple-frame weighting, inclusion probabilities

To study social change, panel surveys of the same individuals over time are crucial. Ensuring the validity of the panel's findings requires that the panel members adequately represent the population. Panel surveys, which usually start with a random sample drawn from the population of interest, face attrition, as some panel members choose to discontinue their participation, can no longer be contacted, or die. Attrition introduces the risk of a panel being selective for certain population subgroups, especially if members of some subgroups drop out at higher rates than others. In addition to potential attrition bias, the reduced sample size decreases the precision of sample estimates.

To counteract the negative effects of attrition, panels such as the Longitudinal Internet Studies for the Social Sciences (LISS) panel (Scherpenzeel, 2011), the German Internet Panel (Blom, Gathmann, & Krieger, 2015), and the GESIS Panel (Bosnjak et al., 2018) are usually refreshed after some time by recruiting new panel members. In scientific research, both the initial recruitment sample and the refreshment sample(s) are usually drawn using a random sampling approach. It may be a simple task to determine for each recruitment sample an individual's propensity to be sampled (in the case of sampling designs that are not too complex). However, combining the initial recruitment sample and refreshment samples drawn at different points in time from a population of interest that is described in the same way (e.g., persons aged 18 years or older) is not a trivial task. This is due to the fact that each sample is drawn sequentially and independently of the previous samples. One key challenge is therefore to account for the fact that, in principle, some members of the population have a chance of being sampled more than once, whereas others do not, as they were not part of the population of interest when the previous samples were recruited. This results in very different probabilities of being included in the panel survey. Naive weighting strategies, such as directly adopting design weights using the design weights of the individual samples without adjusting for potential overlap between the sampling frames or the probability of being sampled several times, fail to yield valid inferences for panel surveys with refreshments, as they would lead to an overestimation of the population in cases where the population of interest of each recruitment overlaps (Gabler et al., 2012; Lohr, 2011; Sand & Gabler, 2018).

In this paper, we show how the multiple-frame weighting methodology proposed by Kalton and Anderson (1986) and Lohr and Rao (2006), which was originally developed for cross-sectional surveys with more than one sampling frame, can be used to create weights in a panel context. We demonstrate that using the initial design weights for the recruitment and refreshment samples

Direct correspondence to
Matthias Sand, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany
E-mail: Matthias.Sand@gesis.org

separately may result in significant biases, even when these samples are individually self-weighted. However, there is a limited body of literature on the correct calculation of design weights for panel surveys with refreshment samples. Therefore, using the GESIS Panel—a German probability-based mixed-mode panel—as an illustrative example, we showcase how the multiple-frame weighting approach provides more accurate estimates. We assess the weights by comparing unweighted, naively weighted, and multiple-frame-weighted estimates of age and region with their corresponding actual population values.

The remainder of the paper is structured as follows: In the next section (Multiple-Frame Approach), we introduce the multiple-frame weighting approach proposed by Kalton and Anderson (1986) and discuss how it can be applied and understood in the panel context rather than its original context of application, multiple cross-sectional samples. We further demonstrate its use under ideal (and controlled) conditions using a synthetic data set. In the third section (Applying the Multiple-Frame Approach to the GESIS Panel), we apply this approach to actual panel data. We conclude with a discussion of our findings (Discussion).

Multiple-Frame Approach

Sampling theory allows for the use of inclusion probabilities from the sampling design to estimate population values (e.g., totals) with the Horvitz-Thompson estimator (Horvitz & Thompson, 1952). However, when a survey is conducted using several sampling frames that partially cover the entire population, frames may intersect. Therefore, multiple frames, which are common in real-world surveys, require multiple-frame approaches to address the overlap of frames and ensure unbiased estimates and/or to calculate the inclusion probabilities. Approaches such as that proposed by Kalton and Anderson (1986) adjust the inclusion probabilities to account for individuals appearing in multiple frames by incorporating the overlap into the estimation process. Therefore, such an approach prevents overestimation of the accessible population (Brick et al., 2005; Lohr & Rao, 2006; Sand & Gabler, 2018; Skinner & Rao, 1996).

A common application of such an approach is a telephone survey of all residents in a country. In such surveys, two sampling frames are typically used: a list of all landline numbers and a list of all mobile phone numbers. Neither list contains all or most of the population members; for example, younger individuals may be missing from the landline list, and older individuals may be missing from the mobile phone list (Heckel & Wiese, 2011). When conducting surveys using two different sampling frames, two types of individuals can be identified: those who can participate in the survey via both frames and those who can participate only via one of the frames.

The challenge lies in the potential overlap of the different sampling frames, as some individuals may be accessible via both frames, thereby increasing their likelihood of being selected for the survey. This circumstance must be accounted for by using a multiple-frame approach when calculating design weights. Several methods can be used during the estimation process to account for individuals being part of multiple sampling frames. The most notable methods (e.g., the multiplicity approach or convex combinations) involve transforming or weighting the design weights of individuals belonging to both frames (Brick et al., 2005; Singh & Mecatti, 2011) or calculating a joint inclusion probability, as in the Kalton–Anderson approach (Kalton & Anderson, 1986; Lohr, 2007).

Multiple-frame approaches are commonly used in cross-sectional surveys. We propose to view the initial sample and the refreshment samples in panel surveys as multiple frames and to apply a multiple-frame approach to derive accurate design weights. In panel surveys, where the same group of individuals is surveyed repeatedly over time, fluctuations due, for example, to migration, births/deaths, or aging in the population may also pose a challenge. As the composition of the population changes, some individuals may become unreachable or no longer meet the survey's eligibility criteria. To address this issue, researchers may opt to use refreshment samples, which involve introducing new participants into the panel to replace those who have attrited or become ineligible. Additionally, the population from which the initial sample was drawn also ages. Hence, when initiating a refreshment, a compelling case can be made for treating the dynamic fluctuations within a population from one point in time to another as distinct frames, albeit with a substantial overlap. Recognizing these temporal shifts as separate frames is crucial to prevent biased estimates, particularly when there is a risk of overestimating subpopulations sampled at multiple time points. To address this concern, adopting a multiple-frame approach becomes imperative. It is noteworthy that the existing literature focuses predominantly on cross-sectional surveys and that there is a notable dearth of documented applications of multiple-frame approaches to panel surveys. This underscores the urgency of considering the temporal evolution within a population as distinct frames—especially when conducting refreshments—to ensure more accurate and unbiased estimates.

Panels With Refreshment Sample(s) as a Special Case of a Multiple-Frame Survey

Assuming a panel survey originates from a simple random sample design and incorporates refreshment sample(s) at a specific point in time to counteract the adverse effects of panel attrition, it is essential to consider changes in the inclusion probabilities for each element in the panel survey. If the sampling frame for the refreshment sample(s) has not been adjusted for the elements that were already part of the original (gross) sample, there is a theoretical possibility of an element drawn in the initial sample also entering the second sample. Furthermore, the inclusion probabilities of these specific elements increase with each subsequent refreshment. Additionally, due to the assumed time gap between the original sample and the refreshment sample, the latter may include elements that were not yet eligible at the time the original sample was drawn and that entered the panel exclusively via the refreshment sample. Consequently, regarding the estimation based on this "refreshed" panel survey, it can be viewed as a multiple-frame survey, given the disparity between the original and refreshment sampling frames, even though considerable overlap is presumed.

The refreshment of a panel survey may take place at several points in time. The different sampling frames may occur similarly to the Venn diagram in Figure 1.

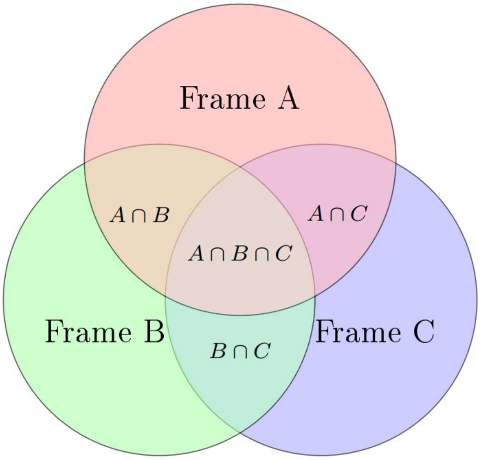

Figure 1 Schematics of a multiple-frame survey comprising three sampling frames

In the context of panel surveys, we have the original sample drawn from Frame A and two refreshment samples drawn from Frames B and C (see Figure 1). As the refreshment samples are generally drawn sequentially at two different points in time, we end up with three different frames, two or more of which overlap. In the case of a panel survey with several refreshments, one might assume the intersections between the frames to be considerable.

This implies that individuals who are in two or three sampling frames (and therefore form the intersection) could enter the panel at multiple time points and thus have a higher probability of being sampled for the panel compared with those who are part of only one of the frames. For individuals within the intersection of two or more frames, it is crucial to accurately calculate the inclusion probability, accounting for the possibility of being sampled from more than one frame, to prevent estimation bias.

Simply using the inclusion probability of each of the three samples based on each of the sampling frames may lead to an overestimation of the number of elements within the intersections and an underestimation of those that can be sampled from only one frame. However, simply adding up the probabilities of inclusion for those elements within each intersection would lead to an overestimation of their inclusion probability. Therefore, it is crucial to accurately calculate an individual's overall inclusion probability by appropriately adding and subtracting their corresponding joint inclusion probabilities of each frame. This mechanism includes and excludes particular overlaps of the respective frames when calculating the inclusion probability. In this particular example, the design weights must be generated as follows: For each of the three samples, the probability of being included in the corresponding sample s_h, with $h \in A, B, C$ is given by

$$\pi_i^{s_h} = \frac{n^{s_h}}{N^{s_h}}, \tag{1}$$

where n refers to the (gross) sample size of a sample and N to the number of elements within each sampling frame.

To adjust for the multiple-frame sampling design, three groups of individuals can be distinguished: (a) individuals who can enter the survey via all three sampling frames, (b) individuals who can enter the survey via two sample frames, and (c) individuals who can enter the survey via only one of the three sampling frames.

For individuals in Group (a), the (adjusted) inclusion probabilities are given by

$$\pi_i = (\pi_i^{s_A} + \pi_i^{s_B} + \pi_i^{s_C}) - (\pi_i^{s_A} * \pi_i^{s_B}) - (\pi_i^{s_A} * \pi_i^{s_C}) - (\pi_i^{s_B} * \pi_i^{s_C}) + (\pi_i^{s_A} * \pi_i^{s_B} * \pi_i^{s_C}). \tag{2}$$

Group (b) consists of three subgroups: (1) individuals who can enter the survey via Samples A and B, (2) individuals who can enter it via Samples A and C, and (3) individuals who can enter it via Samples B and C. For the first subgroup, π_i could be derived by setting the inclusion probability of the frame of which the individuals are not part in equation (2) to zero. The inclusion probabilities are then given by

$$\pi_i = (\pi_i^{s_A} + \pi_i^{s_B}) - (\pi_i^{s_A} * \pi_i^{s_B}) \tag{3}$$

Inclusion probabilities for the other two subgroups are generated accordingly. For individuals in Group (c), the multiple-frame inclusion probabilities are equal to the inclusion probabilities of the corresponding sample.

In the present paper, the importance of using the correct design weights to perform inference from a multiple-frame survey will be illustrated in a synthetic data example and the GESIS Panel (Bosnjak et al., 2018). There are other methods for adjusting design weights in multiple-frame designs, for example, fixed weight adjustment by Hartley (1962), the multiplicity approach by Mecatti (2007), and the pseudo-maximum likelihood method by Lohr and Rao (2006). Sand (2018) showed that using a composite approach can lead to more precise estimates, while point estimates are almost identical to the Kalton-Anderson approach. Composite weighting approaches adjust the design weight of an element in the overlap population by a factor between 0 and 1. However, these approaches require further information on the sampling frames, their respective sizes and overlap, or knowledge about the original frame that was used to sample a particular element. In our case, using only the data set of the GESIS Panel that was provided to us, we could not identify the original sampling frame. However, as we worked only on a reduced GESIS panel data set that included only age and region, it was easy to determine whether an individual belonged to the overlap population. Using the full GESIS Panel data set, one could nevertheless also employ a composite approach similar to that suggested by Brick et al. (2005).

Illustration of the Multiple-Frame Approach Based on a Synthetic Data Set

To demonstrate the workings of a multiple-frame weighting approach for panels with refreshment samples under controllable conditions, we initially generated a synthetic data set, mimicking the sampling approach and the related sampling frames of the GESIS Panel. The synthetic population was constructed in accordance with official statistics.

In our example, we assumed that the recruitment of the original sample started with a population aged between 18 and 69 years (Frame 1). Two years later, a first refreshment sample was drawn. Hence, each member of the synthetic population who was at least 16 years old when the initial panel recruitment started (and 18 years old at the time of the first refreshment) could be part of that refreshment sample (Frame 2). Three years after that, the second refreshment sample was drawn in accordance with the design of the first refreshment (Frame 3). For that particular sample, each individual who was at least 13 years old when the initial panel recruitment took place could be part of the second refreshment (Frame 3). All three frames together cover a population of 68 million elements (100%). Frames 1 and 2 jointly cover 65.7 million elements (96.6%), and Frame 1 contains only 53.4 million elements (78.5%). Table 1 illustrates the varying target populations of the underlying sampling frames based on the age of the persons at the time of recruitment of the original sample.

Table 1 Target populations of the underlying sampling frames based on the age of the persons at the time of recruitment of the original sample

Age category [1]	Frame 1	Frame 2	Frame 3
13–15 years	✗	✗	✓
16–17 years	✗	✓	✓
18–69 years	✓	✓	✓
70+ years	✗	✓	✓

[1] The age category refers to an individual's age at the time of recruitment of the original sample.

As can be seen in Table 1, there is an overlap of all three sampling frames for those elements aged 18–69 years when the original sample was recruited, and there is an overlap of Frames 2 and 3 for those aged 16–17 years and 70 years and over when the original sample was recruited. However, those aged 13–15 years when the original sample was drawn could come only from Frame 3. As this simulation study mimics the approach of the GESIS Panel, the recruitment of the first and second refreshments (Frames 2 and 3) differs from that of the initial sample (Frame 1). Whereas the initial sample was restricted by a maximum age of 69, the first and second refreshments were not. Hence, persons who were at least 70 years old when the original sample was recruited could be sampled only from Frames 2 and 3.

Similar to sampling designs often used in Germany, we further divided the synthetic population into two strata, "east" and "west," in accordance with the distribution of the population across eastern and western German federal states. We did so to achieve a close approximation of the GESIS Panel, which will be discussed in the next section.

From the synthetic population divided into the strata "east" and "west," we drew the three samples using an approach similar to that used by the GESIS Panel. We also used the GESIS Panel's gross sample sizes (see Table 2).

As can be seen in the last column of Table 2, the sample size of the initial sample was allocated proportionally to both strata, whereas the sample sizes of the two refreshment samples were disproportionally allocated, with an oversampling of elements stemming from the stratum "east."

Table 2 Sample sizes and allocation of the sample sizes in the synthetic data set to the strata "east" and "west"

Sample	Stratum	No. of elements	Proportion of sample
Frame 1		**21,870**	**100%**
	East	3,716	16.99%
	West	18,154	83.01%
Frame 2		**10,692**	**100%**
	East	3,366	31.48%
	West	7,326	68.52%
Frame 3		**11,502**	**100%**
	East	3,621	31.48%
	West	7,881	68.52%
Total size		**44,054**	

Let us now assume that we want to estimate the age distribution of the population based on the survey data. We can apply three strategies:

1. Use the unweighted estimates to infer the population.
2. Apply a naive weighting approach by using the design weights of the indivdual samples without adjusting for potential overlap between the frames. Design weights would then be based on the inclusion probability $\pi_i^{s_A}$ for individuals who were sampled during the initial recruitment and on the inclusion probabilities $\pi_i^{s_B}$ and $\pi_i^{s_C}$ for individuals who were sampled in the first and second refreshments, respectively.
3. Apply design weights generated according to the multiple-frame approach described above.

The first two strategies are considered here due to their potential misuse when analysts are unaware of the multiple-frame approach or the issues discussed in the section entitled "Panels With Refreshment Samples as a Special Case of a Multi-Frame Survey." These strategies can be applied when design weights are not provided or are available only on request (e.g., the LISS panel; see https://www.lissdata.nl/faq). Misapplication may also occur if the panel provider publishes incorrect design weights, as noted by Wetzel, Schumann, and Schmiedeberg (2021) in their correction for the pairfam panel. Our objective in exploring these approaches was to highlight their adverse impacts and underscore the necessity of adopting the multi-frame approach. We therefore decided to forgo any further adjustments of these weights (e.g., for nonresponse or panel attrition).

An initial evaluation of the accuracy of a design-weighted estimator involves cross-referencing the sum of (unscaled) design weights with the actual population size. As mentioned earlier, the population of all three frames together comprises 68 million elements. With the multiple-frame approach, the sum of the design weights was 67.99 million, whereas the naive approach—which does not account for the overlap between frames or the possibility of being sampled several times—yielded a total of 186.97 million. Both estimates refer to the full set of the panel's population at the second refreshment. This stark contrast makes it evident that the naive approach would substantially overestimate the population size, a consequence of the issues discussed in the preceding section. Table 3 shows the resulting estimations for the age distribution—for example, the respective percentages of population members who belonged to the 10 age categories—applying the three different weighting approaches to the original sample and the two refreshments.

Table 3 Example: Estimation of age with the synthetic data set using the three different weighting approaches

Age category	Unweighted estimation	Naive estimation	Multiple-frame estimation	True population value
13–15 years	0.79	1.16	3.18	3.38
16–17 years	1.10	1.67	2.29	2.18
18–29 years	18.41	17.66	16.21	16.04
30–39 years	15.47	14.88	13.64	13.36
40–49 years	21.07	20.11	18.50	18.79
50–59 years	18.76	17.55	16.22	16.37
60–69 years	14.45	13.80	12.54	5.16
70–79 years	6.87	8.84	11.46	11.45
80–89 years	2.67	3.74	5.14	5.16
90+ years	0.41	0.60	0.82	0.78

Comparing the estimates obtained using the three strategies with the true population values (last column), one can easily see that the multiple-frame estimates are closest to the population values and that differences are likely attributable to sampling error alone. The estimates obtained using the unweighted and naive approaches are particularly poor for the youngest and oldest age cohorts. This shows that they cannot account for the fact that individuals in these age cohorts could be sampled by only one or both of the refreshment samples, whereas individuals in the overlapping cohorts could, in addition, be sampled in the initial recruitment. Similar but less pronounced effects were found when estimating the east/west distribution, as can be seen in Table 4.

Table 4 Example: Estimation of east/west distribution with the synthetic data set

Stratum	Unweighted estimation	Naive estimation	Multiple-frame estimation	True population value
East	24.29	16.02	15.99	15.95
West	75.71	83.98	84.01	84.05

Comparing the unweighted and naive approaches, one can see that the naive approach performed much better. This is due to the fact that the design weights for the refreshment samples accounted for the oversampling of eastern Germany. The smaller discrepancy between naive and multiple-frame estimation in the case of the east/west estimation might be explained by the similar distribution of the strata across all age classes.

To conclude, using unweighted or naively design-weighted estimations misrepresents the age distribution and might severely bias population inference. As described in the next section, to test these results on an actual panel, we applied these weighting strategies to the GESIS Panel.

Applying the Multiple-Frame Approach to the GESIS Panel

The GESIS Panel, a mixed-mode panel representing the general population in Germany (Bosnjak et al., 2018), employs self-administered surveys conducted bimonthly until 2020 and every third month from 2021 onward. The initial recruitment process involved a multi-step transition from interviewer-administered personal recruitment interviews to self-administered surveys. The survey initially targeted persons living in Germany aged between 18 and 69. Two refreshment samples were drawn in 2016 and 2018 using the German General Social Survey (ALLBUS) interview as the recruitment interview. The ALLBUS applies a two-stage sampling process, but it oversamples the region of eastern Germany. The target population differs from the initial sample and is defined as persons older than 17 years living in Germany, without an upper age limit. Table 5 displays each GESIS Panel sample, its respective design, and its target population.

Table 5 Design and target population of the GESIS Panel recruitment samples

Sample	Age range	Sampling approach
2013 initial cohort	18–70 years	Self-weighted
2016 refreshment (*R*1)	17 years and older	Oversampling of eastern Germany; self-weighted within stratum
2018 refreshment (*R*1)	17 years and older	Oversampling of eastern Germany; self-weighted within stratum

Figure 2 illustrates the target populations of the various GESIS Panel samples, emphasizing changes in eligibility criteria between the initial cohort and the refreshment samples. Notably, individuals born before 1942 and a younger age cohort are included in refreshment samples, thereby expanding the panel's coverage.

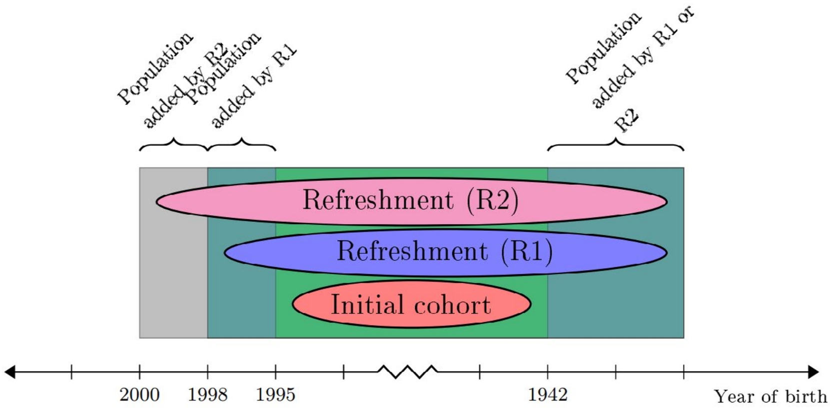

Figure 2 Target population of the initial cohort and the refreshment samples of the GESIS Panel

Deriving Design Weights of the GESIS Panel

When combining the initial cohort of the GESIS Panel with its two refreshment samples, several points must be considered. First, each sample was drawn at different times, leading to slightly different target populations. Second, the refreshment samples had different age restrictions compared with the initial cohort. Thus, individuals could potentially have been included in one, two, or all three of the GESIS Panel samples. Finally, the design weights must account for the disproportional allocation of sample size to eastern and western Germany in the first and second refreshment samples.

As already discussed, the initial cohort stems from a self-weighted sampling design. Hence, each element of the initial cohort (IC) has the same inclusion probability π_i^{IC} given by:

$$\pi_i^{IC} = \frac{n^{IC}}{N^{IC}} \tag{4}$$

In the second and third recruitment ($R1$, $R2$) the design weighting has to compensate for the disproportional allocation of sample size between eastern and western Germany in the ALLBUS sampling design. Thus, weights must be calculated separately for eastern and western Germany (GESIS, 2021). With $k \in \{East, West\}$ being an indicator for western or eastern Germany, inclusion probabilities are given by

$$\pi_{i,k}^{R1} = \frac{n_k^{R1}}{N_K^{R1}} \tag{5}$$

and

$$\pi_{i,k}^{R2} = \frac{n_k^{R2}}{N_K^{R2}}. \tag{6}$$

As described in the section entitled “Multiple-Frame Approach,” the GESIS Panel can be regarded as a three-frame design with its initial recruitment and two refreshments. A sizeable overlap of the three frames can be observed. Individuals who were born between December 1, 1942, and November 30, 1995, could—at least theoretically—have been sampled at each of the three recruitments. Hence, the inclusion probability must be adjusted in accordance with Equation (2). Moreover, due to the disproportional allocation of sample size to eastern and western Germany in Refreshments 1 and 2, an individual’s inclusion probability can be written as

$$\begin{aligned}\pi_{i,k} = \left(\pi_{i,k}^{R2} + \pi_{i,k}^{R1} + \pi_i^{IC}\right) - \left(\pi_{i,k}^{R2} * \pi_{i,k}^{R1}\right) - \left(\pi_{i,k}^{R2} * \pi_i^{IC}\right) - \left(\pi_{i,k}^{R1} * \pi_i^{IC}\right) + \\ \left(\pi_{i,k}^{R2} * \pi_{i,k}^{R1} * \pi_i^{IC}\right).\end{aligned} \tag{7}$$

Persons born before December 1942 and persons born between December 1995 and November 1998 could be recruited only in the first and second refreshments. In their case, π_i^{IC} would be zero. For persons born between December 1998 and November 2000, the equation above reduces to $\pi_{i,k}^{R2}$.

Comparison of Weighting Strategies in the GESIS Panel

In this section, we describe the application of multiple-frame weighting to the actual data from the GESIS Panel. Similar to the synthetic data example presented in the section entitled "Illustration of the Multiple-Frame Approach Based on Synthetic Data," we conducted comparisons with the unweighted and naive estimations. First, we assessed the population size estimates. Table 6 presents the design weights, gross sample sizes, and estimated population sizes for the multiple-frame weighting approach. Table 7 provides the same estimates using the naive weighting approach.

As can be seen in Table 7, the naive weighting approach yielded an overall population estimate of 192.5 million, significantly surpassing the actual population size of about 68.8 million. This overestimation stems from the naive method of iteratively calculating population sizes, resulting in inflated figures due to the repeated consideration of the intersection. When applying design weights without adjustments, the extrapolated population size equaled the sum of the three samples, whereas focusing, for example, solely on the elements and weights of the second refreshment ($R2$) yielded the correct population size of 68.85 million (56.95 million + 11.9 million). Consequently, the design weights of each sample extrapolated to the corresponding sampling frame's population size.

By contrast, the multiple-frame approach estimated an overall population size of 69.2 million (Table 6), exhibiting a slight overestimation compared with the true population size. This discrepancy may have arisen from various limitations in the real data, such as incomplete control of the population and a lack of knowledge regarding the distribution of relevant variables. Additionally, errors in element specification, reporting incorrect population information, and discrepancies in frame and gross sample sizes, coupled with the inability to retrospectively track each step of the sampling process, contributed to more pronounced discrepancies compared with the synthetic data examples. Despite the demonstrated accuracy of the multiple-frame approach under ideal conditions, these factors appear to have influenced its results in this real-data scenario. Furthermore, the observed difference might be attributable to individuals appearing in multiple groups, a possibility that cannot be ruled out due to the absence of detailed information on individual appearances across groups. We further compared the distribution of age (Table 8) and region (Table 9) in a similar way as we did for the synthetic data set.

Table 6 Distribution of weights and population size estimation with the multiple-frame approach by different age cohorts, separately for eastern and western Germany

Category	Weight	Gross sample size	Estimated population size
Born before 12/1942, west	3,710.73	2,065	7,662,658
Born before 12/1942, east	1,700.12	1,129	1,919,438
Born between 12/1942 and 11/1995, west	1,514.51	30,610	46,359,138
Born between 12/1942 and 11/1995, east	1,021.60	9,372	9,574,408
Born between 12/1995 and 11/1998, west	3,710.73	484	1,795,994
Born between 12/1995 and 11/1998, east	1,700.12	144	244,818
Born after 11/1998, west	7,226.69	202	1,459,791
Born after 11/1998, east	3,285.41	58	190,554
Overall population			**69,206,798**

Table 7 Distribution of weights and population size estimation with the naive approach by different cohorts

Category	Weight	Gross sample size	Estimated population size
Initial cohort	2,558.223	21,870	55,948,331
Refreshment (*R*1), west	7,625.969	7326	55,867,847
Refreshment (*R*1), east	3,522.321	3,366	11,856,132
Refreshment (*R*2), west	7,226.688	7,881	56,953,526
Refreshment (*R*2), east	3,285.413	3,621	11,896,481
Overall population			**192,522,317**

Table 8 Weighted distribution of age group (in %)

Age class	Unweighted estimation	Naïve estimation	Multiple-frame estimation	True population value
Born before 1943	7.25	9.96	13.85	12.71
Born between 01/1943 and 12/1995	90.74	87.05	80.82	82.02
Born between 01/1996 and 12/1998	1.43	2.13	2.95	4.02
Born after 12/1998	0.59	0.86	2.38	1.26

Table 9 Weighted distribution of region (in %)

Region	Unweighted estimation	Naïve estimation	Multiple-frame estimation	True population value
West	75.71	82.72	82.76	82.72
East	24.29	17.28	17.24	17.28

We used the 2018 German intercensal population updates ("Fortschreibung des Bevölkerungsstandes;" Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, 2021) as our source of official statistics; they also served as the sampling frame for drawing samples constituting the three cohorts. Derived from Germany's 2011 census data, this model-based estimation includes statistical uncertainty. The multiple-frame weighting approach provided the closest estimation of the true population value for most age categories, except for the youngest age group. A crucial factor contributing to this accuracy is that the multiple-frame method takes into consideration the exclusion of certain age categories in specific recruitment waves, a nuance overlooked by the naive estimation. For instance, the initial cohort does not include the oldest respondents and the second-youngest age group, thus distorting the inter-category relations when combining all cohorts. Nevertheless, the naive estimation remained closer to the population value than did the unweighted estimation.

Regarding the variable "region," both naive and multiple-frame estimations aligned closely with the true population value, which was anticipated, as this variable influences the weighting across all cohorts. By contrast, the unweighted estimation significantly deviated from the population benchmark due to the disproportional allocation of sample size to eastern and western Germany in the ALLBUS.

However, the analyses presented in this section have certain limitations. The absence of information on pertinent frame parameters, particularly concerning the population and frame distribution of the former East and West Berlin, poses challenges. We assume that during the refreshment sampling, the population of the former East Berlin was part of the “east” stratum, thereby leading to oversampling. Determining participants’ ages accurately at the time of the refreshments was also problematic, and retrospective reconstruction of this specific aspect of the frame proved unfeasible.

Discussion

In this paper, we examined how refreshment samples can be integrated correctly into panel surveys using the multiple-frame approach. The differences between multiple-frame weighting and a naive weighting approach were illustrated using a synthetic data set. We show that the estimates using multiple-frame weighting deviated only slightly and at random from the population parameters, whereas naively weighted and unweighted estimates showed large systematic discrepancies. Applying the approach to the real data of the GESIS Panel, we found the differences between the naive weighting procedure and the multiple-frame approach to be less pronounced.

The inability to fully replicate the findings from our synthetic data set when using actual panel data can be attributed to issues arising from the time gap between the calculation of weights and the sampling conducted by a third-party field agency. To achieve accurate weights, comprehensive information about the sampling process and the data used to design the survey sample is imperative. Any uncertainties or discrepancies in this information pose a potential risk to the accuracy of the weights and consequently the survey estimates. We strongly advocate for the simultaneous performance of design weighting and sampling to prevent the loss of crucial information. Furthermore, this example underscores the critical importance of transparent sampling documentation for each sample in a (panel) survey, including frame and population sizes as well as a detailed description of every sampling step. A further explanation for the inability to fully replicate the findings from our synthetic data set when using actual panel data might be that the intersections of the different sampling frames, and thus the products of inclusion probabilities of the different recruitments in particular, have only a small impact on the estimates based on panel data with refreshments.

Despite encountering challenges in generating weights for application to the GESIS Panel data, the analysis of the synthetic data set demonstrates the necessity of employing multiple-frame weighting when integrating a refreshment sample into an ongoing panel. This study employed a multiple-frame approach to recruiting respondents at three distinct points in time. Consequently, the variability observed in the design weights throughout the study did not reach a level requiring interventions such as trimming to reduce their variance. It is anticipated that a multiple-frame approach involving additional recruitments may substantially elevate the variability of the computed weights, leading to a corresponding increase in the variance of the design weights. Therefore, applying the multiple-frame approach to encompass all future waves will inevitably entail combining this procedure with a trimming approach to effectively mitigate the variability of the design weights.

The primary focus of this paper was on accurately calculating design weights in panel surveys with refreshment samples with the aim of yielding unbiased estimates in the absence of nonresponse and attrition. Consequently, we did not delve into the implementation of attrition and calibration weights. However, given that attrition is a primary driver for refreshing the panel population, it is essential to further examine the question of the optimal method for combining multiple overlapping frames and integrating attrition weights. Moreover, the multiple-frame approach discussed here aims to accurately compute inclusion probabilities used in a Horvitz-Thompson estimator, where the weights are inherently the inverse of the inclusion probabilities. Thus, the challenge lies in identifying an appropriate model specification to estimate attrition propensity rather than in the combination of the different frames.

References

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting up an online panel representative of the general population: The German Internet Panel. *Field Methods, 27*(4), 391–408. https://doi.org/10.1177/1525822X15574494

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2018). Establishing an open probability-based mixed-mode panel of the general population in Germany: The GESIS Panel. *Social Science Computer Review, 36*(1), 103–115. https://doi.org/10.1177/0894439317697949

Brick, J. M., Dipko, S., Presser, S., Tucker, C., & Yuan, Y. (2005). Estimation issues in dual frame sample of cell and landline numbers. In JSM Proceedings, Survey Research Methods Section (pp. 2791–2798). American Statistical Association. http://www.asasrms.org/Proceedings/y2005/files/JSM2005-000236.pdf

Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. (2021). *Fortschreibung des Bevölkerungsstandes 2018, On-Site-Zugang.* https://doi.org/10.21242/12411.2018.00.00.1.1.0

Gabler, S., Häder, S., Lehnhoff, I. & Mardian, E. (2012). Weighting for unequal inclusion probabilities and nonresponse in dual frame telephone surveys. In S. Häder, M. Häder, & M. Kühne (Eds.), *Telephone surveys in Europe* (pp. 147–167). Heidelberg: Springer. https://doi.org/10.1007/978-3-642-25411-6_11

GESIS – Leibniz Institute for the Social Sciences. (2021). *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften – ALLBUS sensitive Regionaldaten*. GESIS Data Archive. https://doi.org/10.4232/1.13767

Hartley, H. O. (1962). Multiple frame surveys. *Proceedings of the Social Statistics Section, American Statistical Association, 19*(6), 203–206.

Heckel, C., & Wiese, K. (2011). Sampling frames for telephone surveys in Europe. In S. Häder, M. Häder, & M. Kühne (Eds.), *Telephone surveys in Europ*e (pp. 103–119). Springer. https://doi.org/10.1007/978-3-642-25411-6_9

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association, 47*(260), 663–685. https://doi.org/10.1080/01621459.1952.10483446

Kalton, G., & Anderson, D. W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society. Series A (General), 149*(1), 65–82. https://doi.org/10.2307/2981886

Lohr, S. L. (2007). Recent developments in multiple frame surveys. In *JSM Proceedings, Survey Research Methods Section* (pp. 3257–3264). American Statistical Association. http://www.asasrms.org/Proceedings/y2007/Files/JSM2007-000580.pdf

Lohr, S. L. (2011). Alternative survey sample designs: Sampling with multiple overlapping frames. *Statistics Canada – Survey Methodology, 37*(2), 197–213. https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2011002/article/11608-eng.pdf?st=pwEXRPYa

Lohr, S., & Rao, J. N. K. (2006). Estimation in multiple-frame surveys. *Journal of the American Statistical Association, 101*(475), 1019–1030. https://doi.org/10.1198/016214506000000195

Mecatti, F. (2007). A single frame multiplicity estimator for multiple frame surveys. *Statistics Canada – Survey Methodology, 33,* 151–157. https://www150.statcan.gc.ca/n1/pub/12-001-x/2007002/article/10492-eng.pdf

Sand, M. (2018). *Gewichtungsverfahren in Dual-Frame-Telefonerhebungen bei Device-Specific Nonresponse* (Doctoral dissertation; GESIS-Schriftenreihe No. 20). GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.21241/ssoar.60293

Sand, M., & Gabler, S. (2018). Gewichtung von (Dual-Frame -) Telefonstichproben. In Häder, S., Häder, M., & Schmich, P. (Eds.), *Telefonumfragen in Deutschland. Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute* (pp. 405–424). Springer VS. https://doi.org/10.1007/978-3-658-23950-3_13

Scherpenzeel, A. (2011). Data collection in a probability-based Internet panel: How the LISS panel was built and how it can be used. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 109*(1), 56–61. https://doi.org/10.1177/0759106310387713

Singh, A. C., & Mecatti, F. (2011). Generalized multiplicity-adjusted Horvitz-Thompson estimation as a unified approach to multiple frame surveys. *Journal of Official Statistics, 27,* 633–650. https://www.proquest.com/scholarly-journals/generalized-multiplicity-adjusted-horvitz/docview/2821376567/se-2

Skinner, C. J., & Rao, J. N. K. (1996). Estimation in dual frame surveys with complex designs. *Journal of the American Statistical Association, 91*(433), 349–356. https://doi.org/10.1080/01621459.1996.10476695

Wetzel, M., Schumann, N., & Schmiedeberg, C. (2021). *New weights for the pairfam anchor data* (pairfam Technical Paper No. 17). https://doi.org/10.5282/ubm/epub.91999

Online First publication date: 01-04-2025