SSRS Encipher Hybrid: A Case Study in Adaptation and Using Hybrid Samples to Produce Estimates for Subgroups

Often, survey researchers aim to produce estimates not only for an entire population, but also for subgroups of that population. A hybrid sample design—in which a probability sample is supplemented with a nonprobability sample—can be a good option when detailed subgroup breakouts are planned, since nonprobability sources offer a lower-cost means of increasing the sample size. However, this must be done with caution, as nonprobability samples are often subject to selection bias. In fact, experimentation by the Pew Research Center has found that selection bias in nonprobability samples may be worse within certain key subgroups than in the population at large.

SSRS's Encipher Hybrid calibration methodology is designed to minimize selection bias in the nonprobability portion of a hybrid sample, ensuring that estimates from the full hybrid sample remain representative of the target population. Our introductory white paper describes this methodology. In this case study, we demonstrate enhancements to the methodology that are useful when we need accurate estimates not only for an entire population, but also within subgroups of that population.

Figure 1 shows a commonly reported breakout: the percentage of Democrats vs. Republicans who report being "very enthusiastic" to vote in the 2022 midterms. The SSRS Opinion Panel probability sample estimates a 13 percentage-point "enthusiasm gap" in favor of Republicans, similar to other probability-based polling at the time; while the nonprobability sample, by itself, shows very little difference.

When the samples are combined, and calibrated using SSRS's Encipher Hybrid 1.0, the enthusiasm gap is in the same direction as the probability-based estimate but is still less than half as large (5.7 percentage points). In other words, while the Encipher Hybrid 1.0 methodology reduces selection bias, some bias remains in this critical subgroup breakout.

For this case study, we use a survey conducted by the Association of American Universities (AAU) in April 2022. This survey ran in parallel on (1) a nonprobability sample obtained from an opt-in Web sample vendor and (2) the SSRS Opinion Panel Omnibus, a bimonthly, multi-client probability sample. These samples were combined to create a hybrid sample. The hybrid sample was initially calibrated using SSRS's original Encipher Hybrid methodology (that is, Encipher Hybrid 1.0).

We used the AAU survey to test an Encipher Hybrid 2.0 methodology with two enhancements to better control selection bias within subgroups. First, we modified our Stepwise Calibration algorithm to incorporate subgroup estimates into the optimization procedure used to identify the best calibration model.

Figure 1: Partisan enthusiasm gap for 2022 midterms
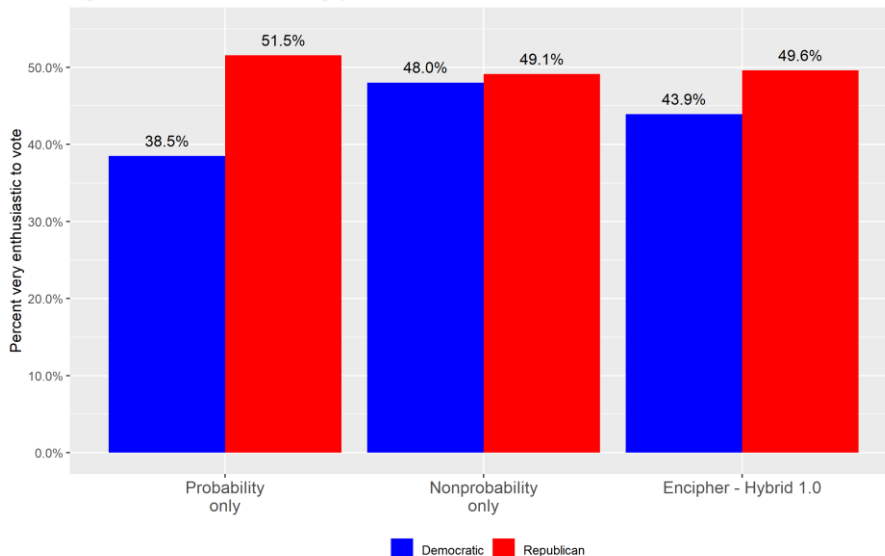
Second, we incorporated a propensity adjustment, which uses random forest models to assign a "pseudo base weight" prior to calibration. In prior research, we have found that random forest propensity models can pick up differences between probability and nonprobability samples that calibration alone may miss.

As shown in Figure 2, with our Encipher Hybrid 2.0 methodology, the hybrid sample estimates a closer enthusiasm gap (10 percentage points) to the probability-only sample. Thus, these enhancements make Encipher Hybrid 2.0 more effective at controlling selection bias in this key subgroup breakout.

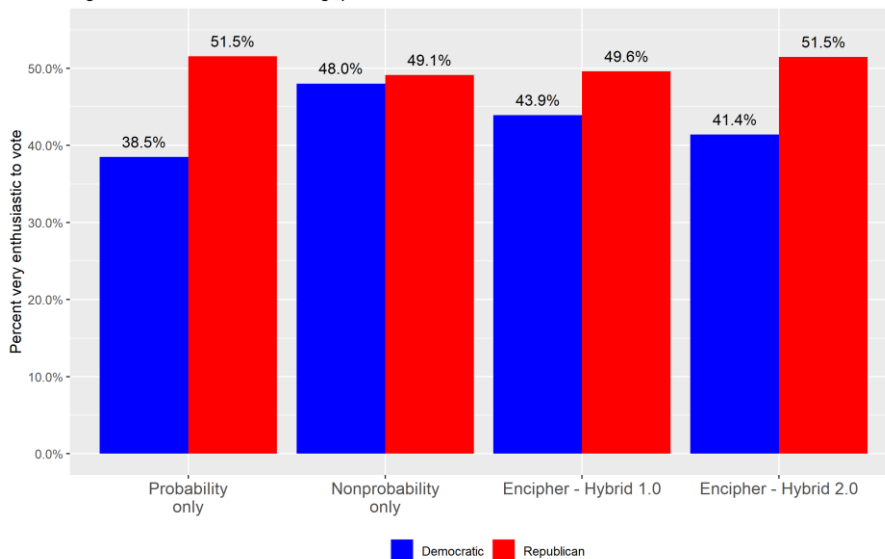Figure 2: Partisan enthusiasm gap for 2022 midterms

Figure 3 summarizes the estimated selection bias across all estimates from the AAU study, within numerous subgroups of interest. With the enhancements, Encipher Hybrid 2.0 reduces the average selection bias within most subgroups—particularly within those (such as adults ages 18 – 29 and Black adults) that showed the most bias in the uncalibrated nonprobability-only sample.

Figure 3: Selection bias across all estimates
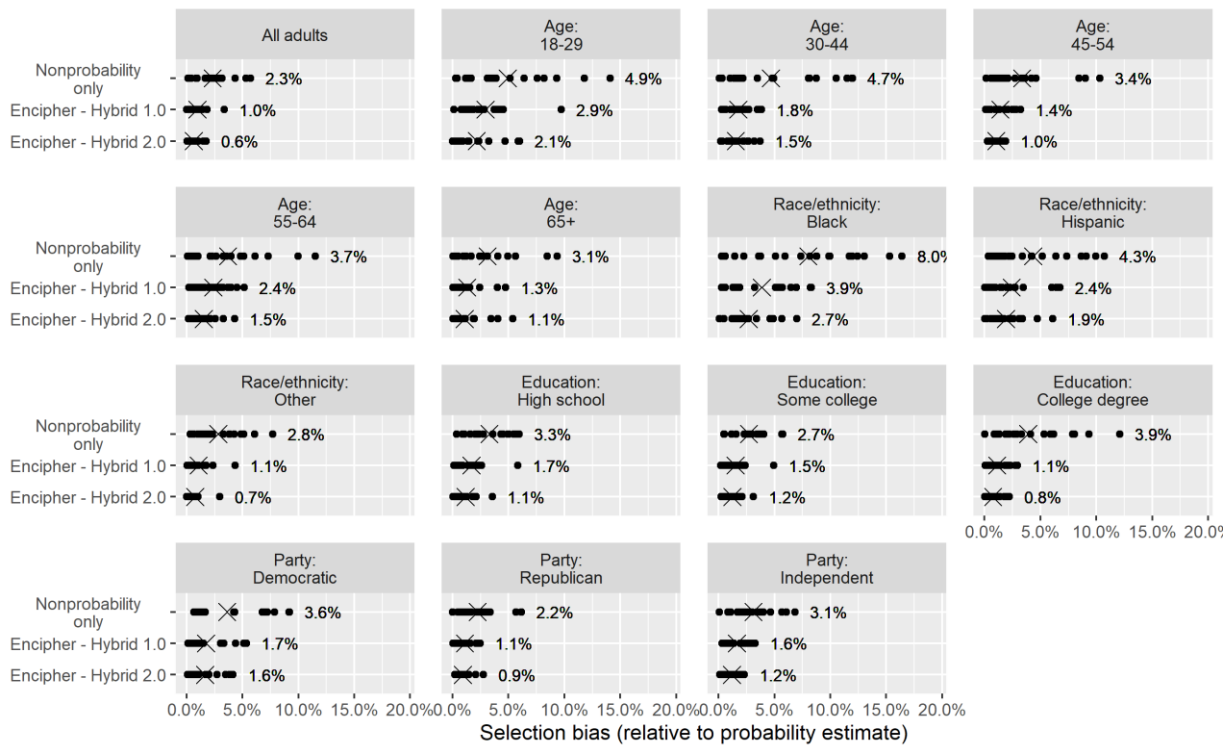
Based on these findings, our Encipher Hybrid methodology will include these 2.0 enhancements whenever sample sizes permit. This will allow hybrid samples to produce more reliable estimates both for the population as a whole and for key subgroup breakouts.

Learn more about how SSRS's Encipher Hybrid can be used for your study.

Contact our team today.