Using R-Indicators to Make Case-Level Decisions for GSS 2020

Steven Pedlow[1]
[1]NORC at the University of Chicago, 55 E. Monroe, Chicago, IL 60603

Abstract
NORC uses R-indicators on projects to monitor overall representativeness and to make operational adjustments to prioritize under-represented groups through future outreach in real-time. R-indicators come from regression modeling on a response variable. They include an overall R-score (overall representativeness), unconditional partial R-scores for each predictor variable subgroup, and a conditional partial R-indicator for each variable. The overall R-indicator score measures the variation in response propensities among all cases. At the same time, the partial R-indicators split the R-indicator score into between (unconditional) and within (conditional) variation.

For the 2020 round, the General Social Survey (GSS) applied the case-level sum of the unconditional partial R-indicators to decide whether to target a case with extra effort and higher incentives, continue working a case or stop attempts to recruit a case. We collected a panel sample of 2016 and 2018 GSS respondents in summer 2020 and a new cross-sectional sample during winter 2020-2021. The panel was a unique opportunity to use R-indicators with specific information about every sample member. We used demographics and some opinion variables to inform the development of the R-indicators model. The cross-section is a more typical application where we can only use information based on where they live, though we also merged in vendor data on household characteristics.

We will show the developed models, the cases' decisions, and our R-indicators results for GSS 2020. One lesson learned is that the overall R-score is not a good measure of representativeness in isolation. Considering fewer variables and excluding good predictors will result in a higher overall R-score. Using the sum of the R-scores for the panel, we achieved a higher response rate among the cases we most needed to improve our representativeness. For cross-sectional work, the results were more ambiguous, so our research will continue.

Key Words: General Social Survey, Representativity Indicators

1. Introduction to the General Social Survey

The General Social Survey (GSS) is a nationally representative survey of adults in the United States conducted since 1972 funded by the National Science Foundation. The GSS collects data on contemporary American society to monitor and explain trends in opinions, attitudes, and behaviors. The General Social Survey (GSS) has provided politicians, policymakers, and scholars with a clear and unbiased perspective on what Americans think and feel about such issues as national spending priorities, crime and punishment, intergroup relations, and confidence in institutions. The GSS has adapted questions from earlier surveys, allowing researchers to conduct comparisons for up to 50 years.

The GSS contains a standard core of demographic, behavioral, and attitudinal questions, plus topics of special interest. Among the topics covered are civil liberties, morality, psychological well-being, social mobility, and stress and traumatic events. The GSS is the single best source for sociological and attitudinal trend data covering the United States. It allows researchers to examine the structure and functioning of society in general and the role played by relevant subgroups while comparing the United States to other nations.

The GSS aims to make high-quality data easily accessible to scholars, students, policymakers, and others, with minimal cost and waiting. The GSS has carried out an extensive range of methodological research designed to advance survey methods in general and ensure that the GSS data are of the highest possible quality. In pursuit of this goal, the GSS Methodological Reports series has published more than 130 papers.

Due to COVID-19, we replaced in-person interviewing with two separate mail push-to-web data collections. First, we fielded a panel of 2016 and 2018 cross-sectional respondents from August to September 2020. Then, we fielded a cross-sectional sample from January through April 2021. Most of this paper will present results from the panel survey, but we also include some cross-sectional results.

2. Introduction to R-Indicators

Representativity Indicators (or R-Indicators for short), introduced in Schouten et al. 2009, are outputs of a regression equation. The dependent variable is whether a case has been completed so far or not. We use the independent variables to predict which cases we have completed.

Table 1 below shows the thirty-two independent variables we used for modeling the panel completes. We have thirteen demographic variables, fifteen interview questions, and four paradata items. The interview questions are all questions on governments spending priorities. Each question asks whether the government should spend less, the same, or more for one type of spending. The paradata items include when and how the interview was done and a control variable BALLOT. We randomly assigned each household to one of three ballots and this assignment should not be significant in the models. It is important to note that since this is a panel survey, we had all of this individual data to use for modeling. Cross-Sectional Surveys do not have this luxury.

There are three levels of R-Indicators:

1. The Unconditional **CATEGORY** R-Indicator measures how under-represented or over-represented a category, for example, college graduates, is among the respondents so far. Unconditional Category R-Indicators have a value between 0.0 and ±0.5. A value of 0 indicates that the respondents and sample have the same proportion of respondents in that category. Negative Unconditional Category R-Indicators indicate under-represented categories among completed cases, while positive Unconditional Category R-Indicators indicate over-represented categories.

2. The Unconditional **VARIABLE** R-Indicator measures how related a variable is to the response status. In a sense, it measures how much a variable's categories differ in response rate. Unconditional Variable R-Indicators have a value between 0.0 and 0.5. A value of 0 indicates that each category has a similar percentage in the sample and respondents.
3. Finally, the **OVERALL** Sample R-Indicator is often described as an overall measure of the sample's representativeness. The Overall Sample R-Indicator has a Value between 0 and 1. A value of 1 indicates that the configuration of the respondents is the same as the configuration of the whole sample for all target variables (perfect correlation). However, as discussed later, our project shows that the Overall Sample R-Indicator has some undesirable properties.

Table 1: The Thirty-Two Variables Used in Our Model

Demographic Variables (13)	*Questionnaire Data (15)*	*Paradata (4)*
Gender	Alternative Energy	Original Interview Year
Age Category	Childcare Assistance	Panel Completion Mode
Hispanic Origin	Conditions of Blacks	Early/Late Respondent
Race	Drug Addiction	Questionnaire "Ballot"
Highest Education	Education	
Born in the USA	Environment	
Marital Status	Foreign Aid	
Have Children?	Health Care	
Household Size	Roads/Bridges	
Employment Status	Mass Transportation	
Income Category	Military Defense	
Own/Rent/Other	Parks/Recreation	
Political Affiliation	Scientific Research	
	Social Security	
	Space Exploration	

3. Results for the Panel Sample

After we reached two-thirds of our completion goal, about a month into our data collection, we created a model. The dependent variable was whether a panel sample member had completed the interview and the independent variables listed in Table 1 above. At the time of the intervention, we had 1,085 completes, and we eventually completed 1,771.

Figure 1 below shows the highest and lowest Unconditional Category R-Indicators. Figure 1 also shows BALLOT=2, the category closest to zero. A score close to zero indicates almost no difference between the respondents and non-respondents for this category.

Among the first 1,085 completes, College Graduates are the most over-represented group, followed by those who think the government spends too little on Roads. Retired seniors are also over-represented, as well as Midwesterners and Whites. This graph only shows the top 6 and bottom 6 with BALLOT=2 in the middle. As we hoped, BALLOT is the variable with the category R-indicator closest to 0. Hispanics are the most under-represented among our first 1,085 completes, followed by respondents without a High School Degree. Also under-represented are persons who are neither White nor Black as well as Southerners.

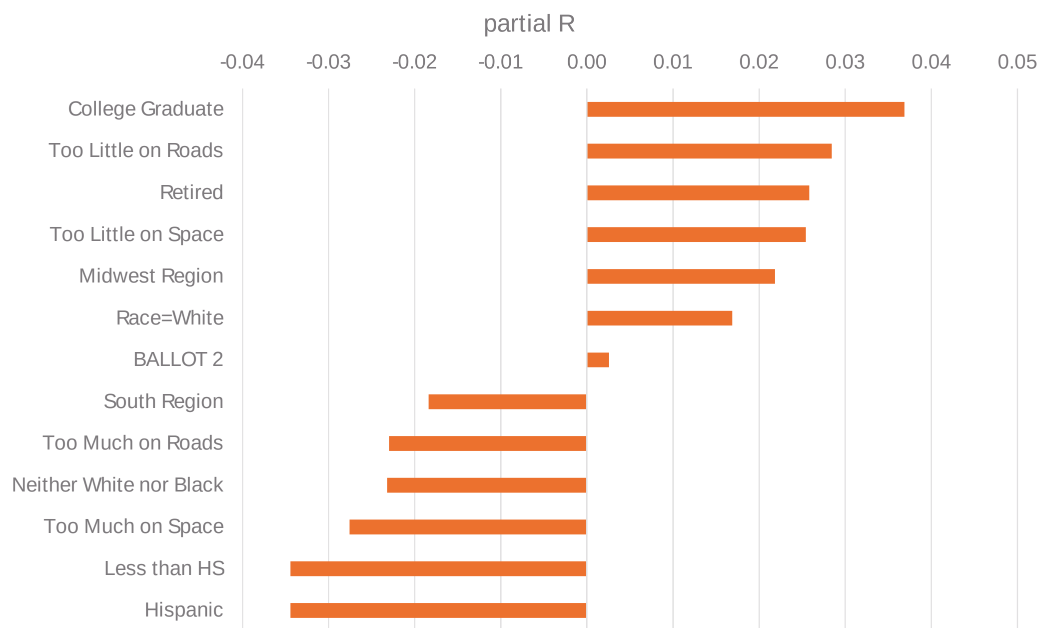

Figure 1: The Highest and Lowest Unconditional Category R-Indicators

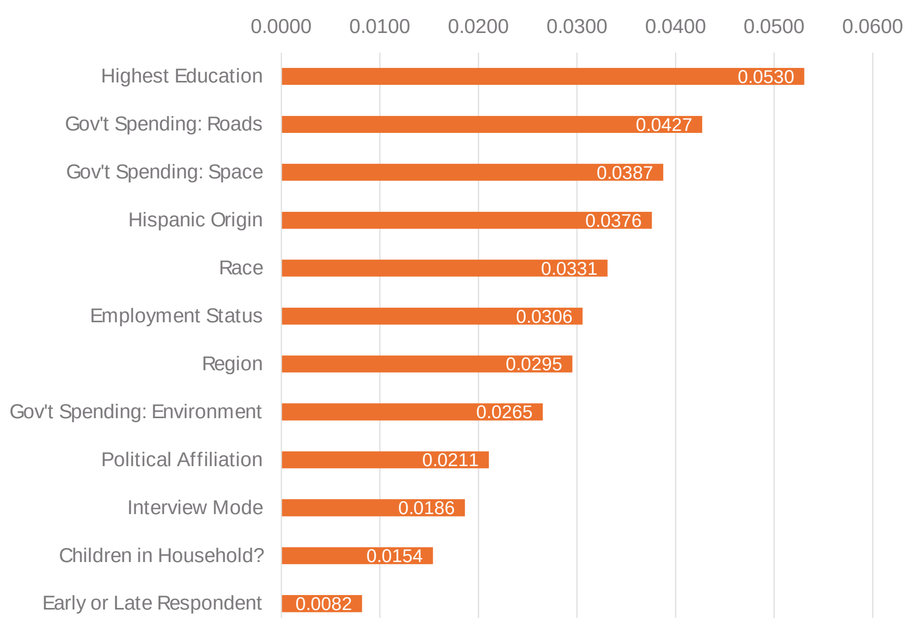

Figure 2: The Highest Unconditional Variable R-Indicators

Figure 2 shows the largest Unconditional Variable R-Indicators. Highest Education is the variable most related to response status, as shown by College Graduates and those without a High School degree in Figure 2. Two opinion variables on Government Spending on Roads and the Space Program are next, followed by Race and Employment Status.

4. The Panel Intervention

The Unconditional Variable R-Indicators indicate which variables have the most variability in response rates so far, but the Unconditional Category R-Indicators are better to plan an intervention. It is possible, for example, to choose one or more categories to add or subtract effort or incentives. For example, looking at Figure 2, we could choose to increase effort, incentives, or both for respondents in the Less than High School and Hispanic categories since these are the largest negative Unconditional Category R-Indicators. Similarly, we could subtract effort for College Graduates or Retired respondents since these are the largest positive Unconditional Category R-Indicators. We can refer to this as a one-dimensional intervention option.

For the 2020 General Social Survey, we instead chose to use a multi-dimensional option for our intervention. For each pending panel member without a panel interview yet, we use the sum of all the Unconditional Category R-Indicators as our Intervention Score. We allocated all of the pending cases into one of three action bins. The most negative one-third of sums were the most under-represented cases among the completed interviews, so we put them in Action Bin 1 to increase the incentives and effort to convert them to completes. We gave the middle-third no priority change in Action Bin 0. The most positive one-third of sums were the most over-represented among our completes, so we put them in Action Bin -1 to drop further contacting efforts but still accept any completes from prior contacts.

We made some modifications to this basic allocation system. Any cases with a previously scheduled appointment in Action Bin -1 were still. Also, as an experiment, we randomly selected two hundred Action Bin 1 cases and treated them as Action Bin 0u. The zero means no change in the efforts even though the U stands for their being under-represented. We can then compare Action Bin 0u with Action Bin 1 to measure the impact of the increased efforts. We can also compare Action Bin 0u with Action Bin 0 to compare the difficulty of completing the Action Bin 1 (under-represented) cases.

Table 2 shows the twelve most under-represented pending cases, those with the most negative Intervention Scores. Almost all fifteen cases in Table 2 do not have a High School Degree, all of them are Hispanic, and most of them are from the South.

Table 2: The Twelve Most “Under-Represented” Pending Cases

RScore	*Education*	*Roads*	*Space*	*Hispanic*	*Race*	*Employed*	*Region*
-0.2091	< HS	Just Right	Too Much	YES	OTHER	NONE	SOUTH
-0.2015	< HS	Missing	Missing	YES	OTHER	NONE	SOUTH
-0.1891	HS degree	Too Much	Too Much	YES	OTHER	NONE	SOUTH
-0.1881	< HS	Too Much	Too Much	YES	OTHER	NONE	SOUTH
-0.1866	< HS	Just Right	Just Right	YES	OTHER	WORKING	SOUTH
-0.1862	< HS	Too Much	Too Much	YES	OTHER	WORKING	SOUTH
-0.1800	< HS	Just Right	Too Much	YES	OTHER	NONE	SOUTH
-0.1782	< HS	Just Right	Too Much	YES	OTHER	NONE	WEST
-0.1748	< HS	Too Much	Just Right	YES	OTHER	WORKING	SOUTH
-0.1694	< HS	Just Right	Just Right	YES	OTHER	WORKING	SOUTH
-0.1685	< HS	Just Right	Missing	YES	OTHER	WORKING	SOUTH
-0.1642	< HS	Too Much	Too Much	YES	OTHER	WORKING	SOUTH

Table 3 shows the twelve most over-represented cases, which are those with the most negative Intervention Scores. All of them have a Graduate Degree or at least a Bachelor's

Degree. They all believe the government should spend more on roads, none of them are Hispanic, and all are White.

Table 3: The Twelve Most "Over-Represented" Pending Cases

RScore	*Education*	*Roads*	*Space*	*Hispanic*	*Race*	*Employed*	*Region*
0.2269	Masters	Too Little	Too Little	NO	WHITE	RETIRED	MIDWEST
0.1976	Masters	Too Little	Too Little	NO	WHITE	WORKING	MIDWEST
0.1894	Masters	Too Little	Just Right	NO	WHITE	RETIRED	MIDWEST
0.1847	Masters	Too Little	Too Little	NO	WHITE	RETIRED	MIDWEST
0.1810	Masters	Too Little	Too Little	NO	WHITE	WORKING	WEST
0.1810	Masters	Too Little	Too Little	NO	WHITE	WORKING	WEST
0.1783	Masters	Too Little	Just Right	NO	WHITE	WORKING	MIDWEST
0.1773	Masters	Too Little	Just Right	NO	WHITE	WORKING	MIDWEST
0.1773	Masters	Too Little	Just Right	NO	WHITE	WORKING	MIDWEST
0.1770	Masters	Too Little	Just Right	NO	WHITE	RETIRED	MIDWEST
0.1759	Bachelor	Too Little	Too Little	NO	WHITE	RETIRED	NORTHEAST
0.1721	Bachelor	Too Little	Too Little	NO	WHITE	RETIRED	MIDWEST

In summary, we had four action bins. The cases in Figure 2 are among those in Action Bin 1, in which we increased the effort and increased the incentive offered for a completed interview. The cases in Figure 3 are among those in Action Bin -1, in which we stopped contacting efforts but still accepted any interviews resulting from previous contacts. We gave the middle third of the case with scores close to zero no changes in effort or incentives. We randomly selected 200 cases from Action Bin 1 to NOT receive increased effort or incentives as an experiment. Finally, we moved the cases with any scheduled appointment from Action Bin -1 (to Action Bin 0).

Excluding the cases with appointments, there were 2,936 cases in our intervention action bins. Table 4 summarizes the results of our intervention.

Table 4: The Results of Our Intervention

Action Bin	*Pending Cases*	*Completed Interviews*	*Action Bin Conversion Rate*
1. Prioritize	520	102	19.6 percent
0u. Keep Working	186	20	10.8 percent
0. Keep Working	1,302	184	14.1 percent
-1. No More Effort	928	92	9.9 percent
TOTAL	**2,936**	**398**	**13.6 percent**

Overall, we converted 13.6 percent of these pending cases without appointments into completes. Among the most under-represented cases, we almost doubled the conversion rate from 10.8 percent for Action Bin 0u to 19.6 percent for Action Bin 1. Comparing Action Bin 0u to the 14.8 percent of Action Bin 0 shows that we would have continued to complete fewer of the under-represented cases. Even though Action Bin -1 cases were the easiest, we were able to successfully convert more in the other bins to increase the balance of our respondents.

Table 5 shows the improvement in representativeness that we were able to achieve. The left model is the original model run on September 10 used to separate the cases into Action Bins for our intervention. We report here the concordance score as well as the Overall Sample R-Indicator Score. Forcing the model to predict a case as complete or incomplete leads to the concordance score of how often the model is correct. A model with zero predictive power will have a concordance score of 50 percent because it is a random guess. Running the same model after two weeks of the intervention on September 28 decreased the concordance score from 63 percent to 59 percent, a drop of 4 percent, which means these same variables have less predictive power in determining which cases have been completed. Also, the unconditional variable R-Indicators dropped for eight of the eleven variables.

When we ran a completely fresh post-intervention model, seven of the variables from the original model dropped out, and the concordance score could only regain with other variables 1 of the 4 points dropped in the concordance score. Looking at both of the September 28 models, the partial R-scores for the variables are the same. Unlike regression parameters, the unconditional variable R-Indicators are model-independent. They would be the same no matter which other variables were in the model.

Finally, let us look at the Overall Sample R-Scores. When re-running the original model on September 28, the Overall Sample R-Score increases when the Concordance Score decreases. This combination is what we would expect, given that we had improved the representativeness of the completed cases. However, the new model on September 28 shows the lowest Overall Sample R-Score even though the lower concordance score indicates that the model cannot predict the completes as well as the Original Model on September 10. We believe the Overall Sample R-Score is only valid when comparing the same model run at different times. Once the model is changed, the concordance score is a superior measure.

Table 5: Comparison of Models Before/After Intervention on

Data Set	*Original Model 9/10*	*Original Model 9/28*	*New Model 9/28*
Concordance Score	63.4 percent	59.5 percent	60.6 percent
Overall R-Score	0.8157	0.8215	0.7959
PARTIAL R-SCORES			
Highest Education	0.0530	0.0657↑up	0.0657
Gov’t Spending: Roads	0.0427	0.0355↓	Not Significant
Gov’t Spending: Space	0.0387	0.0316↓	Not Significant
Hispanic Origin	0.0376	0.0215↓	Not Significant
Race	0.0331	0.0238↓	Not Significant
Employment Status	0.0306	0.0270↓	Not Significant
Census Division	0.0295	0.0289↓	0.0289
Gov’t Spending: Envir.	0.0265	0.0301↑up	0.0301
Political Affiliation	0.0210	0.0346↑up	0.0346
Interview Mode	0.0186	0.0071↓	Not Significant
Children in HH	0.0154	0.0059↓	Not Significant

5. The 2020 GSS Cross-Sectional Sample

We also attempted to use R-Indicators in the 2020 GSS Cross-Sectional Sample, but we did not have previous interview data or demographics to use in our models. Instead, we started with only the address. One standard technique is to use the address to merge census-tract level data from the American Community Survey or other U.S. Census Bureau data. We used Census-Tract Level Data on Education, Income, and Race/Ethnicity as three examples. However, we also merged in vendor data from Merkle, Inc. Merkle had a household match for 83.5 percent of our addresses, but they also had two or more matches for 25.4 percent. Merkle provided over 500 variables, and we used 293 as candidate variables. The most significant variable in our models was whether or not Merkle had a household match; those without matched household data had a lower completion rate.

We ran models using the American Community Survey tract-level and Merkle household variables, but we chose not to intervene. The Overall Sample R-Indicator was 0.9219 because the model had poor predictive power. Even though the Overall Sample R-Indicator was larger for the Cross-Sectional model, this does not necessarily mean we had a more representative sample than the panel. The R-square for the model was only 0.023 (correlation = 0.15) rather than 0.045 (correlation = 0.21) for the panel. The Concordance Score, however, was a higher 67 percent, which we interpreted as over-fitting. We will do more work in hopes of a Cross-Sectional intervention in GSS 2022.

Table 6 compares the Cross-Sectional Model we considered using with the Original Panel model of September 10.

Table 6: Comparison of Panel and Cross-Sectional Models

Action Bin	*Original Panel Model 9/10*	*Cross-Sectional Model 9/28*
Concordance Score	63.4 percent	67.2 percent
Overall R-Score	0.8157	0.9219
Concordance Score	0.0450	0.0229
Overall R-Score	0.2121	0.1513

6. Concluding Remarks

In conclusion, we had a successful intervention for the GSS 2020 panel. We increased incentives and effort for the one-third of cases most under-represented among the first 1,085 completes. Among the remaining 586 completed cases, we doubled the conversion rate for our high-priority under-represented cases. Meanwhile, we stopped efforts for the one-third of cases most over-represented among our first 1,085 completes and converted only 10 percent rather than 14 percent of them.

Our data became more representative after our intervention, as shown by the four percent drop in concordance score, which is a better measure than the overall sample R-Indicator. Even the best post-intervention model had three percentage points less Concordance than pre-intervention. Finally, we have more work to do for Cross-Sectional Surveys.

Reference

Schouten, B., Cobben, F., Bethlehem, J. (2009), Indicators for the representativeness of survey response, Survey Methodology, 35 (1), 101 – 113.