

ARTICLES

Reconceptualizing Survey Representativeness for Evaluating and Using Nonprobability Samples

David P Fan¹

¹ University of Minnesota

Keywords: database, survey responses, representativeness, sampling

<https://doi.org/10.29115/SP-2013-0007>

Survey Practice

Vol. 6, Issue 2, 2013

A question in recent times has been the quality of surveys using nonprobability samples. This paper approaches measurement accuracy by reconceptualizing representativeness. This approach can be used for all samples, probability or not. Current survey designs aim for representative samples with the implicit assumptions that representativeness belongs to the sample and that all measurements are equally representative. However, all real sampling methods contact or recruit members using properties of potential respondents like telephone usage. Questions correlated with sampling methods can give biased measurements. For example, a telephone survey can give biased responses for "why don't you use the telephone" while leaving unbiased responses to health related questions. An analysis recognizing different degrees of bias for different measurements de facto reconceptualizes representativeness as belonging to the measurement rather than to the sample. Accordingly, this paper proposes that every measurement has a "method dependent bias" due to the correlation of the measurement with the sampling method. A researcher can use a nonprobability sample to obtain acceptably representative results for all measurements that have method dependent biases within the researcher's tolerance. This paper describes and gives a pilot implementation of a simple method for quantifying method dependent biases.

The Goal

In the not too distant past, high quality probability surveys routinely used random digit dialed (RDD) landline samples. Since then, communication channels have proliferated and fragmented to the point that some surveys now use nonprobability samples like those obtained from Internet volunteers.

The turn to nonprobability samples has been controversial as seen in this passage from the American Association for Public Opinion Research (AAPOR) Standard Definitions of 2011 (The American Association for Public Opinion Research 2011): "Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values."

This blanket prohibition is unnecessary. Instead, individual measurements from any sample, probability or not, can be representative. The paper proposes an approach for assessing the representativeness of measurements from any chosen sampling method.

The Theory

According to (Kruskal and Mosteller 1980), Kiaer in 1895 was the first to propose that measurements could be made on a sample that is an “approximate miniature of the population” selected using a “representative method” (Kiaer 1895).

From Kiaer’s proposal through to the present, the typical goal for a survey has been the construction of a representative sample. By using representativeness to modify the word sample, the implication is that the representativeness belongs to the sample and is hence the same for all measurements. For example, the AAPOR Standard Definitions speak at length about response rates with the assumption that the representativeness of all measures increases with the response rate. Further evidence for a common representativeness for all measurements is seen in the global rejection of nonprobability samples as quoted above from the AAPOR Standard Definitions. Discussions of total survey error also generally consider representativeness to be owned by samples (Groves and Lyberg 2010).

Representativeness refers to the probability that a sample will give the same results as the total population. As noted by Groves and Lyberg (2010), representativeness includes both a variance component that gives the scatter in the measurements on repeated trials and a bias component that refers to the systematic deviation of a cluster of sample measurements from the true population value.

A useful analogy is to target shooting. Variance corresponds to the scatter in the shots while bias gives the systematic displacement of the group of shots away from the bull’s eye due to misadjustment of the gun sight.

Both variance and bias can contribute to a measurement being nonrepresentative. This paper focuses on bias rather than on the variance that is discussed at length in the survey literature.

Researchers have long recognized that a single sample can give measurements with different degrees of bias depending on the question. That is the reason for using a telephone survey for unbiased measurements about health but not for a question about why respondents do not use the telephone.

The differences in bias between health and telephone nonusage measurements can be justified from statistical theory as noted by Yates (1946): “In May 1924 the Bureau of the International Institute of Statistics appointed a commission for the purpose of studying the application of the representative method in statistics.” Yates went on to enunciate the crucial condition that “was well known to those who drew up the report to the International Institute”: “If bias is to be avoided, the selection of the samples must be determined by some process uninfluenced by the qualities of the objects sampled and free from any element of choice on the part of the observer.”

This quote states the necessary and sufficient condition for a bias-free measurement, namely that the measurement should be independent of the sampling and analytic methods. There is no requirement that all measurements from a given sample should be equally unbiased.

Overall, there are both empirical examples and statistical arguments for how a sample can give measurements with idiosyncratic biases. An analysis that includes these idiosyncrasies needs to move beyond the concept of representative samples.

In one reconceptualization, a bias due to the non-independence of a measurement from the survey methods is called a “method dependent bias.” Each measurement has its own bias thereby transferring the ownership of both bias and representativeness from the sample to the measurement.

Samples, in turn, have “isobias.” This term is analogous to isobars that give lines of constant barometric pressure on a geographical map. If all survey measurements from a sample are plotted on some surface, then an isobias corresponds to a line where all the measurement biases are the same. If biases are quantified in survey percentage points, then an isobias at zero percent would enclose all measurements having no bias and hence having complete representativeness. An isobias at one percent would encompass all measurements with no more than one percent bias. The use of isobias for designing samples and analyzing measurements will be discussed below.

A true random sample has a zero percent isobias that surrounds all potential measurements because no measurement of interest would ever be logically dependent on a set of mathematically chosen random numbers. This *a priori* independence is the statistical justification for random samples being ideal.

Clearly, the analysis becomes more complex when there is a separate bias for every pair of a sampling method and a measurement. Fortunately, modern electronic databases can hold the information for easy access. This paper proposes that such a database can be populated easily and at low cost if ongoing surveys would do nothing more than add a single question about sampling methodology.

A Calculation for Method Dependent Bias

Since 1936, candidate preferences from pre-election polls have consistently predicted actual votes in elections with accuracies in the range of sample size error (Chang and Krosnick 2009; Woolley and Peters 2012). By extension, a reasonable premise is that a survey response can give a good indication of a respondent’s likelihood of using a survey method without performing empirical measurements.

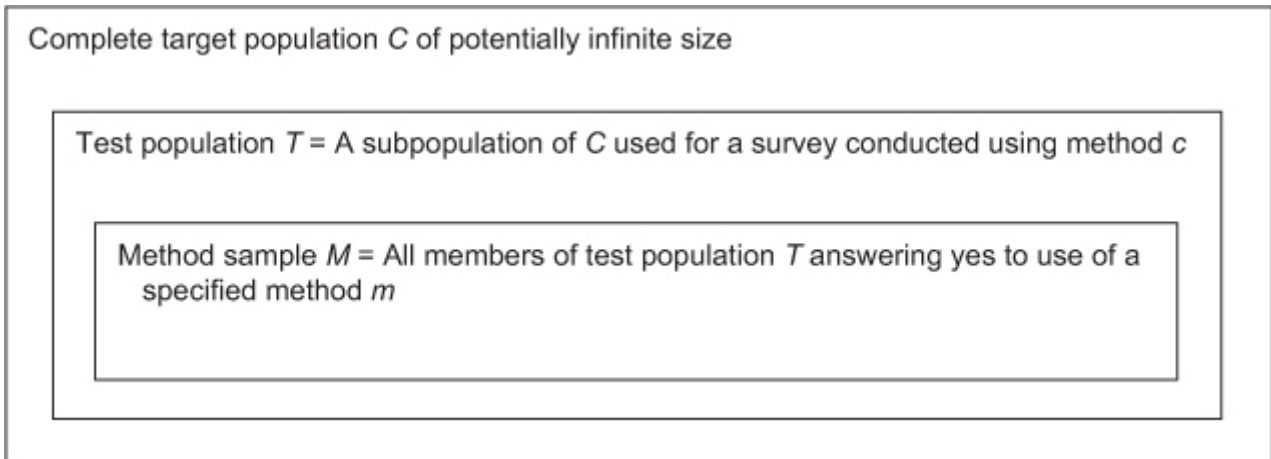


Figure 1 Relationships between populations used for computing a method dependent bias. Test population T is a subpopulation of a complete target population C . Test population T includes all respondents in a survey conducted using method c . Method sample M includes all members of test population T responding affirmatively to use of method m .

Given this premise, one approach to assessing a method dependent bias is to analyze two measurements from some survey with a telephone survey being an example. One measurement would be about a respondent's reported usage of a sampling method like the Internet. The other measurement would be about a substantive characteristic like employment.

A comparison would be made for the employment measurement among all respondents and among just those using the Internet. A significant difference would indicate that employment measurements are likely to be biased in an Internet survey.

In more detail:

Step 1: The process would begin with the researcher specifying a complete target population C (Figure 1). A typical population C would be all the adults in a country.

The researcher would also specify a substantive response s such as employment and a sampling method m like Internet usage.

Step 2: The researcher would conduct a survey of C using a method c that could be RDD telephone sampling. All respondents of the survey would be assigned to be members of a test population T so T is a subpopulation of C (Figure 1).

The survey conducted using c would include a question about Internet usage method m . Sample M would be all respondents of T saying that they use m (Figure 1). The test would be to see if the probability of measurement s is the same in both test population T and its sample M .

True independence of substantive measurement s from method m in complete population C requires that the independence should be found in every subpopulation of C . Therefore, a bias test can be performed on one such

subpopulation, test population T . Bias of s by m in T would mean that s is biased by m in C . Therefore, T can be analyzed as a separate population without concern for the rest of C .

Test population T can include respondents from any survey, even one not designed to be representative. However, method c for selecting T should not overlap too much with method m . Thus, c should not be an Internet method when m is Internet usage. The obvious problem is that all members of T would then respond affirmatively when asked about use of the Internet. The result would be complete overlap of T and M . Without a difference between T and M , there is no way to tell if m affects s .

Overall, c should be sufficiently different from m so that a sizable proportion – but not all members – of T will respond as users of m .

Step 3: A method dependent bias is computed from the completed survey.

For test population T (Figure 1):

R_T = all respondents in T .

N_T = the number of respondents R_T .

R_{Ts} = those respondents R_T giving response s .

N_{Ts} = the number of respondents R_{Ts} .

$P_{Ts} = N_{Ts}/N_T$

= the proportion of respondents R_T giving response s .

σ_{Ts}^2 = the variance of P_{Ts}

= 0 because test population T is the total population analyzed and all members of T are measured so P_{Ts} is known without error.

Proportion P_{Ts} from the total test population T is now compared with the equivalent proportion among all respondents in sample M drawn from T (Figure 1). For that comparison:

R_{TM} = all respondents in M .

N_{TM} = the number of respondents R_{TM} .

R_{TM_s} = the subset of respondents R_{TM} giving response s .

N_{TM_s} = the number of respondents R_{TM_s} .

$P_{TM_s} = N_{TM_s}/N_{TM}$

= the proportion of respondents R_{TM} giving response s .

σ_{TMs}^2 = the variance of P_{TMs}

$= [P_{TMs}(1 - P_{TMs})/N_{TM}] [(N_T - N_{TM})/N_T]$ where the first term is the usual binomial variance for an infinite population and the second term is the finite population correction from theorem 2.2 of Cochran (1977) to account for test population T being of finite size.

As noted in the outline above, a method dependent bias can be based on a significant difference between P_{TMs} and P_{Ts} . Given $\sigma_{Ts}^2 = 0$ from above, the test for significance can use the one sample Z-score

$$Z = (P_{TMs} - P_{Ts}) / \sigma_{TMs}. \quad (1)$$

routinely used for comparing fractions.

If $|Z| \leq 1.96$, then P_{Ts} and P_{TMs} are so close together that there is no significant difference at 95 percent confidence. On the other hand, if $|Z| > 1.96$, then P_{TMs} is significantly different from P_{Ts} . In that case, the bias can be eliminated by using an adjustor σ_{TMs} which will quantify how much P_{TMs} must be shifted for the difference in the two proportions to be no longer significant. In the target shooting analogy, that would mean making the minimum adjustment to the gun sight for the shot to be within the bull's eye at 95 percent confidence.

Replacing P_{TMs} by the adjusted $(P_{TMs} - \delta_{TMs})$ in (1) gives

$$Z = [(P_{TMs} - \delta_{TMs}) - P_{Ts}] / \sigma_{TMs}$$

Multiplying through by σ_{TMs} and rearranging leads to

$$\sigma_{TMs} = P_{TMs} - P_{Ts} - Z\sigma_{TMs}$$

Now, let $\delta'_{TMs} = |P_{TMs} - P_{Ts} - Z\sigma_{TMs}|$. Method dependent bias B_{TMs} can then be defined as

$$B_{TMs} = \delta'_{TMs} \text{ if } \delta'_{TMs} > 0 \text{ and } P_{TMs} > P_{Ts}, \quad B_{TMs} = -\delta'_{TMs} \text{ if } \delta'_{TMs} > 0 \text{ and } P_{TMs} < P_{Ts}$$

When $B_{TMs} = 0$, P_{TMs} is close enough to P_{Ts} that the two values are not significantly different so test population T would provide no evidence that m biases s . If there is a significant difference, then the sign of B_{TMs} indicates whether the biased estimate P_{TMs} is higher or lower than the unbiased P_{Ts} , and the magnitude $|B_{TMs}|$ gives the survey percentage points by which P_{TMs} must be shifted for P_{TMs} to be no longer significantly different from P_{Ts} .

A Sample Data Application

For a pilot computation of B_{TMs} , complete target population C was defined as all United States adults, method m was use of the Internet, and substantive response s was employment (Step 1 from above).

Test population T (Step 2) was all respondents in a combined landline and cell phone survey from the Roper Center's iPoll data archive (Social Science Research Solutions/ICR–International Communications Research 2011). For brevity, this survey will be referenced as SSRS/ICR.

Among all 1959 respondents in T , 1511 (77 percent) were in sample M corresponding to users of the Internet. Step 2 allowed T to be assembled using any arbitrary method so there was no need to consider response rates or the fact that the SSRS/ICR survey oversampled African Americans and Hispanic Americans.

Following Step 3, proportion P_{Ts} for each measurement s was computed for all members of the full test population T . In addition, proportion P_{TM_s} was calculated for just the members in M saying that they used the Internet. Then B_{TM_s} was computed to compare the two proportions using (2).

For the computation of B_{TM_s} , all null responses including the don't knows were omitted so a different NTs was used for each measurement s . Also, each response choice was treated as a separate response s with the result that a question with two choices was scored to have two separate measurements. That led the SSRS/ICR survey to give B_{TM_s} values for a total of 272 responses.

Among all the measurements s in the SSRS/ICR survey, 161 (42 percent) provided no evidence of method dependent bias because $B_{TM_s} = 0$ survey percent. Another 72 (26 percent) had $0 \text{ percent} < |B_{TM_s}| \leq 1 \text{ percent}$; 30 (11 percent) had $1 \text{ percent} < |B_{TM_s}| \leq 2 \text{ percent}$; and 53 (20 percent) had $2 \text{ percent} < |B_{TM_s}| \leq 19 \text{ percent}$. These $|B_{TM_s}|$ percentages are specific to the SSRS/SRI survey. Other studies will be needed to generalize the results.

$|B_{TM_s}|$ values were above five survey percent for measurements about currently being employed, having an individual retirement account, having a high income or being younger.

On the other hand, Internet usage gave no demonstrated bias for measurements of optimism, pessimism, job loss, residence in a particular region of the country, identification as politically Democratic or liberal, or a variety of economic views including comparisons of respondents' standards of living with those of their parents.

Discussion

This paper has advanced a method for evaluating the “fitness for use” (Groves and Lyberg 2010) of all survey methods, probability and otherwise. All samples have characteristic isobiases that delimit the measurements that are inside and outside user specified zones of tolerable method dependent bias.

The only differences among samples are their isobiases. All samples aside from true random samples will have some measurements outside acceptable isobiases. Any sampling method can be used for representative measurements

so long as the desired measurements lie within specified isobias. There is no need to exclude categorically nonprobability samples, or any other type of sample, for that matter.

The extensive mapping of method dependent biases is crucial to this isobias based approach. This paper has suggested using B_{TMs} values to quantify biases.

The SSRS/ICR study discussed above hints at possible applications of isobias for sampling methodology. First, the largest $|B_{TMs}|$ was 18.45 survey percent for email usage. That was not surprising since email and Internet access are expected to be highly correlated. Given this a priori expectation and its confirmation by a large $|B_{TMs}|$, no survey researcher is likely to use a web mode to assess email usage in the general population.

Among the economic measurements in the SSRS/ICR analysis, Internet usage also significantly biased measurements about the actual financial conditions experienced by respondents. These conditions included being employed or having a high income. This was an expected relationship because Internet usage does require a minimum amount of funds.

Given this information, a survey of a person's current financial situation could avoid an Internet method. Alternatively, the researcher might consider correcting individual measurements by their B_{TMs} because these values are calculated as adjusters for biases.

This is but one way to correct. Another possibility is to compare survey percentages after weighting by respondent characteristics as is currently done. Detailed analyses of weighting and other corrective strategies are subjects of future research.

The SSRS/ICR data gave no evidence that Internet usage biased abstract impressions such as perceptions of economic conditions in the country at large or expectations that the future would be better or worse. This finding agrees with earlier findings that mass media content is a major influence on and has been an important predictor of consumer confidence in the United States and elsewhere (Alsem et al. 2008; Fan and Cook 2003). Bias was also not observed for relative economic circumstances such as knowing someone who had been laid off or having had a better job in the past. The absence of this type of bias was consistent with there being no compelling reason why alterations in – rather than the absolute state of – a person's fortunes should be related to Internet use.

The lack of observed bias in these subject areas suggests that web surveys could be used without any correction to give representative measurements for questions about general knowledge or about changes in the financial conditions of respondents.

These examples are illustrative only because the data came from just one test population T and that population was oversampled for minorities. Similar data from other test populations would indicate consistencies in inferences across different segments of society.

Furthermore, the SSRS/ICR data only provided information about Internet usage which is but one component of any Internet sampling design. A question more directly relevant to web polls could ask whether a respondent had taken an Internet survey. There is no single best question for a survey method. Instead, a variety of similar questions could shed light on the extent to which variations in methodology might affect consensus isobias.

The methodologies explored can extend well beyond the Internet mode. Other possibilities include preferences for lottery incentives. It may even be useful to use a landline survey for T to ask about cell phone usage. Also mail surveys for T can be employed to examine method dependent biases due to landline or cell phone usage. Such bias data would comment on the extent to which RDD phone surveys can give unbiased measurements.

A researcher running a survey with a sample size error of three survey percent might decide to tolerate all method dependent biases of less than one poll percent. In that case, the researcher could prefer methods that give measurements within the one percent isobias for all desired survey responses.

Elaborating on the arguments of Step 2 above, test population T can come from any arbitrary subpopulation including snowball samples along with their recent incarnations in the form of social media friends or mall intercepts which have river samples as their descendents (Farrell and Petersen 2010). Websites hosting river samples ask users to fill out short surveys before they proceed to the websites' contents. Thus a river sample contains people who happen to visit websites much like mall intercepts capture individuals who happen to be shopping at particular stores. Furthermore, each B_{TM_s} calculation only needs responses to two questions, well within the lengths of typical river surveys.

The B_{TM_s} computed in this paper is but one plausible metric. The key is to gather data that can then be used to compute biases using any statistical approach.

One caveat about this paper's computations is that people might not always behave in accordance with their statements. Therefore, it would be useful to check biases computed from survey responses against biases calculated by comparing survey results with measurements on the entire population. Tested population data have included election results and administrative data such as passports issued by the United States (Chang and Krosnick 2009; Yeager et al. 2011). Unfortunately, most survey measurements have no equivalent reference data measured on the total population. In those cases, a survey based method like the one in this paper may be the only recourse for computing method dependent biases.

The approach in this paper mines for responses that do not show associations. That makes use of data that are typically discarded because most researchers are more concerned about associations than about non-associations. The non-report of uninteresting data has been called the “file drawer problem” (Rosenberg 2005). The ignoring of weak associations is characteristic of not only decision tree analyses such as those using the CHAID (Chi-squared Automatic Interaction Detector) algorithm (Kass 1980) but also other statistical methods including meta-analysis (e.g., Ye, Fulton, and Tourangeau 2011).

The detailed B_{TM_s} data from this paper are available from the author. Also, B_{TM_s} values from other surveys are being added to a database under development. Please contact the author for details about any interest in accessing the database as well as any willingness to contribute data by adding just one methodological question to planned surveys.

REFERENCES

- Alsem, K.J., S. Brakman, L. Hoogduin, and G. Kuper. 2008. "The Impact of Newspapers on Consumer Confidence: Does Spin Bias Exist?" *Applied Economics* 40 (5): 531–39.
- Chang, L., and J.A. Krosnick. 2009. "National Surveys via RDD Telephone Interviewing vs. the Internet: Comparing Sample Representativeness and Response Quality." *Public Opinion Quarterly* 73 (4): 641–78.
- Cochran, W.G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley and Sons.
- Fan, D.P., and R.D. Cook. 2003. "A Differential Equation Model for Predicting Public Opinions and Behaviors from Persuasive Information: Application to the Index of Consumer Sentiment." *Journal of Mathematical Sociology* 27 (1): 1–23.
- Farrell, D., and J.C. Petersen. 2010. "The Growth of Internet Research Methods and the Reluctant Sociologist." *Sociological Inquiry* 80 (1): 114–25.
- Groves, R.M., and L. Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74 (5): 849–79.
- Kass, G.V. 1980. "An Exploratory Technique for Investigating Large Quantities of Categorical Data." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29 (2): 119–27.
- Kiaer, A.N. 1895. "Observations et Experiences Concernant Des Denombrements Representatifs." *Bulletin of the International Statistical Institute* 9 (2): 176–83.
- Kruskal, W., and F. Mosteller. 1980. "Representative Sampling, IV: The History of the Concept in Statistics, 1895-1939." *International Statistical Review/Revue Internationale de Statistique* 48 (2): 169–95.
- Rosenberg, M.S. 2005. "The File-Drawer Problem Revisited: A General Weighted Method for Calculating Fail-Safe Numbers in Meta-Analysis." *Evolution* 59 (2): 464–68.
- Social Science Research Solutions/ICR–International Communications Research. 2011. "Kaiser Family Foundation/Washington Post/Harvard School of Public Health Poll: Race and Recession Survey." [Computer file USICR2011-WPH027]. Henry J. Kaiser Family Foundation/Washington Post/Harvard School of Public Health. Storrs, CT: The Roper Center, University of Connecticut.
- The American Association for Public Opinion Research. 2011. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7th ed. AAPOR.
- Woolley, J., and G. Peters. 2012. "Election Year Presidential Preferences: Gallup Poll Accuracy Record: 1936-2012." <http://www.presidency.ucsb.edu/data/preferences.php>.
- Yates, F. 1946. "A Review of Recent Statistical Developments in Sampling and Sampling Surveys." *Journal of the Royal Statistical Society* 109 (1): 12–43.
- Ye, C., J. Fulton, and R. Tourangeau. 2011. "More Positive or More Extreme? A Meta-Analysis of Mode Differences in Response Choice." *Public Opinion Quarterly* 75 (2): 349–65.
- Yeager, D.S., J.A. Krosnick, L. Chang, H.S. Javitz, M.S. Levendusky, A. Simpser, and R. Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Nonprobability Samples." *Public Opinion Quarterly* 75 (4): 709–47.