

Computing "e" in Self-Administered Address-Based Sampling Studies

Jill M. DeMatteis*

Keywords: response rate, standard definitions, unknown eligibility

DOI: [10.29115/SP-2019-0002](https://doi.org/10.29115/SP-2019-0002)

Survey Practice

Vol. 12, Issue 1, 2019

In recent years, address-based sampling (ABS) has become a widely used approach for obtaining samples for household surveys in the United States. In computing response rates for self-administered ABS studies, one challenge is in estimating the eligibility rate of sampled cases with unknown eligibility, denoted e in American Association for Public Opinion Research response rate calculations RR3 and RR4. The assumptions underlying standard approaches for estimating e are unlikely to hold in self-administered ABS studies. This note describes an approach for estimating e in such situations.

In *Standard Definitions* (American Association for Public Opinion Research 2016), two of the formulae for computing response rates (RR3 and RR4) include a factor, e , that accounts for the proportion of cases with unknown eligibility that are assumed to be eligible. This note addresses challenges with computing e in self-administered household surveys in which address-based sampling (ABS) was used to select the sample of addresses.

In recent years, ABS has become a widely used approach to obtaining samples for household surveys in the United States. ABS begins with a sampling frame that generally uses as its basis mailing address files maintained by the U.S. Postal Service (USPS). Because the USPS uses the database on which these files are based to manage its mail delivery services, efforts are made to keep these files up-to-date. However, the USPS database depends on input from the mail carriers. In addition, survey research organizations are only able to access these files through third-party vendors who license them. The ABS frames generally have nearly complete coverage of household addresses (Link et al. 2008); however, inaccuracies still exist due to lags in timing or failures to update statuses of addresses.

In general terms, each sampled address can be classified into exactly one of three possible address eligibility categories: (1) addresses known to be eligible addresses (where "eligible" here is defined in terms of the address being associated with a household, regardless of whether the household is eligible for the particular study); (2) addresses known to be ineligible; and (3) addresses with unknown eligibility. Eligible addresses include those who respond to the

survey (complete or partial respondents) as well as those who fail to respond but provide information to indicate that a household resides at the address (e.g. those who call in to refuse or return the survey with a comment indicating they are refusing). Ineligible addresses include those whose survey mailings (at the initial contact stage) are returned by the USPS as nondeliverable as well as those who call in (or contact the survey organization by mail, email, etc.) to indicate that the address is associated with only a business. All remaining addresses have unknown eligibility and are referred to as *unresolved*. Addresses known to be eligible or ineligible are collectively referred to as *resolved*.

Various approaches have been used to compute e (Brick, Montaquila, and Scheuren 2002; Ezzati-Rice et al. 2000; Frankel and Wiseman 1983; Kennedy, Keeter, and Dimock 2008; Shapiro et al. 1995). These approaches may be reasonable for surveys in which a large proportion of sampled cases are resolved. However, in ABS studies, it is often the case that only a minority of addresses are resolved, a situation in which the assumptions underlying the existing methods may not hold.

The calculation of e we propose involves “backing out” the estimated eligibility rate using estimated totals of eligible and unresolved addresses from the survey ($\widehat{T}_{\text{elig}}$ and \widehat{T}_{unr} , respectively) and a reliable external estimate (\widehat{T}_{ext}) such as an estimate of the number of households from the American Community Survey (ACS). It is necessary for this external estimate to be an estimated number of *households* (as opposed to housing units), for it to correspond to the geographic scope and the timing of the survey, and for it to come from a reliable source. In addition to the ACS, other sources that could be considered (depending on the particular survey) include the Current Population Survey and the decennial census. Here, $\widehat{T}_{\text{elig}}$ and \widehat{T}_{unr} are sums of weights (generally base weights) of eligible and unresolved addresses, respectively. We note that the quantity $\widehat{T}_{\text{elig}} + e\widehat{T}_{\text{unr}}$ provides an estimate of the number of eligible addresses (i.e. households), as does \widehat{T}_{ext} . Setting these two estimators equal to each other and solving for e yields the “backing out” estimate

$$e_B = \frac{1}{\widehat{T}_{\text{unr}}} \left(\widehat{T}_{\text{ext}} - \widehat{T}_{\text{elig}} \right)$$

It is mathematically possible for e_B to be negative (if $\widehat{T}_{\text{elig}}$ exceeds \widehat{T}_{ext}) or to exceed 1 (if $\widehat{T}_{\text{ext}} > \widehat{T}_{\text{elig}} + \widehat{T}_{\text{unr}}$). In such cases, we recommend setting e_B equal to 0 or 1, respectively.

AUTHOR’S CONTACT INFORMATION

Jill M. DeMatteis, Westat, 1600 Research Blvd., Rockville, MD 20850.
jilldematteis@westat.com; (301) 517-4046.

REFERENCES

- American Association for Public Opinion Research. 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 9th edition. AAPOR.
- Brick, J.M., J. Montaquila, and F. Scheuren. 2002. “Estimating Residency Rates for Undetermined Telephone Numbers.” *Public Opinion Quarterly* 66 (1): 18–39.
- Ezzati-Rice, T.M., M.R. Frankel, D.C. Hoaglin, J.D. Loft, V.G. Coronado, and R.A. Wright. 2000. “An Alternative Measure of Response Rate in Random-Digit-Dialing Surveys That Screen for Eligible Subpopulations.” *Journal of Economic and Social Measurement* 26 (2): 99–109.
- Frankel, L.R., and F. Wiseman. 1983. *The Report of the CASRO Task Force on Response Rates: Improving Data Quality on Sample Surveys*. Cambridge, MA: Marketing Science Institute.
- Kennedy, C., S. Keeter, and M. Dimock. 2008. “A ‘Brute Force’ Estimation of the Residency Rate for Undetermined Telephone Numbers in an RDD Survey.” *Public Opinion Quarterly* 72 (1): 28–39.
- Link, M.W., M.P. Battaglia, M.R. Frankel, L. Osborn, and A.H. Mokdad. 2008. “A Comparison of Address-Based Sampling (ABS) versus Random-Digit Dialing (RDD) for General Population Surveys.” *Public Opinion Quarterly* 72 (1): 6–27.
- Shapiro, G., M. Battaglia, D. Camburn, J. Massey, and L. Tompkins. 1995. “Calling Local Telephone Company Business Offices to Determine the Residential Status of a Wide Class of Unresolved Telephone Numbers in a Random-Digit-Dialing Sample.” In *1995 Proceedings of the Survey Research Methods Section of the American Statistical Association*, 975–80. Washington, D.C.: American Statistical Association.