

What's in a Match?

Ashley Amaya, Karen Wooten, Ben Skalland

Tags: survey practice

DOI: [10.29115/SP-2010-0027](https://doi.org/10.29115/SP-2010-0027)

Survey Practice

Vol. 3, Issue 6, 2010

What's in a Match?

As researchers try to compensate for the growing undercoverage of landline RDD frames, alternative designs are increasingly being considered. One such alternative is address-based sampling (ABS). As such, much research is underway from a variety of organizations (Dekker and Murphy 2009; Link et al. 2008) to develop best practices for the use of ABS as a replacement or supplement to landline RDD frames. However, this research has lacked a fundamental component - matching telephone numbers to addresses. Because an accurate match is necessary to conduct interviews by telephone, the success (or lack thereof) in matching a telephone number to an address can dictate the data collection mode and have cost, response rate, and mode effect implications. We must therefore answer the question: What is a match? While this question may seem easy to answer – if you have an address and you have a telephone number, you have a match – it is not so simple. In this paper, we discuss different possible definitions of a “match.”

METHODOLOGY AND ANALYSIS

All analyses were conducted using data collected for the National Immunization Survey – Address-Based Sample Experiment conducted in 2009 (NIS-ABS). Sponsored by the Centers for Disease Control and Prevention's National Center for Immunization and Respiratory Diseases (CDC/NCIRD) and conducted by NORC at the University of Chicago, the NIS is designed to provide continuous, high-quality, timely data on up-to-date vaccination rates among children age 19 through 35 months and teens age 13 to 17 years in the United States.

A national sample of addresses in the 50 States plus the District of Columbia was drawn from the USPS Delivery Sequence File as provided by a commercial address frame vendor (Valassis). An oversample was selected for other analytic purposes in Bexar County, TX. The tables in this paper display weighted (by the reciprocal of the probability of selection) percentages to account for this oversampling. Addresses flagged as sole businesses, drop points, simplified addresses, and P.O. boxes that were not flagged as the sole source of mail receipt were excluded from selection; all other addresses had a non-zero probability of selection. In total, 69,123 addresses were selected – 12,879 in Bexar County

and 56,244 from the rest of the U.S.

To contact an ABS sample via telephone, each address must be matched to a telephone number. Matching was attempted on all NIS-ABS sample lines to identify one or more telephone numbers for each address. NORC first sent all sampled addresses to Marketing Systems Group (MSG), a sample vendor, for matching to telephone numbers. MSG returned a maximum of one telephone number per address and provided a dichotomous variable to evaluate the quality of the match – either “exact” (matched to a street address for single unit buildings or street address + unit for multi-unit buildings) or “inexact” (matched only to a street address for multi-unit buildings). Any address that was not returned with an “exact match” was sent for an additional attempt to locate a telephone number to a respondent locating service, Accurint, which provided up to three telephone numbers per address. If multiple phone numbers were returned, Accurint ranked them in order of quality. Addresses with an “exact match” from MSG were not attempted in Accurint since NORC believes the accurate telephone number was already matched.

In some instances, the same telephone number was matched to multiple addresses by MSG and/or Accurint. This was primarily due to multiple units within the same building (i.e., apartments) being linked to the same telephone number. This is caused when one or more of the matches was inexact. If the same telephone number was linked to multiple addresses, it was retained for the address for which it had the best quality rank and dropped for the other addresses; if it had the same quality rank for multiple addresses, it was dropped for all such addresses.

An algorithm was created to rank multiple phone numbers for the same sampled unit from the two sources. It used the ranking order provided by Accurint, the exact/inexact quality indicator from MSG, and whether or not the phone number was reported by both vendors versus just one. The best telephone number available for each case was loaded into the CATI system for dialing. Prior to dialing, addresses with at least one available telephone number were mailed an advance letter. Telephone data collection occurred May through July, 2009.

Given different vendors and different quality indicators, an initial “match” could be defined in different ways. Two possibilities are:

- ***Any Match.*** Define “match” as having at least one phone number returned from either vendor.
- ***Exact and Inexact Match.*** Define “match” based on the vendor's reported quality of the matches – for example, only MSG “exact matches” could be considered “matched” for a given study.

Table 1 shows the proportion of addresses matched to any telephone number by type of address. While approximately 74 percent of addresses could be

matched to a telephone number, the “any match” rate differed greatly by type of address. Not surprisingly, P.O. boxes and rural route addresses were difficult to match to a phone number, with match rates of 9.6 percent and 29.9 percent, respectively.

Table 1 “Any Match” Rate by Address Type.

Address Type	N	Match Rate
All Address Types	69,123	73.6%
Single-Unit Building	51,616	69.0%
In Multi-Unit Building	16,565	92.8%
P.O. Box	723	9.6%
Rural Route	219	29.9%

The “any match” rate on apartments was exceptionally high – 92.8 percent. Table 2 sheds light on this phenomenon by parsing out the quality of the match as defined by the vendor. Apartments received a high proportion of inexact matches. As mentioned above, these are multi-unit buildings with a telephone number matched on street address but not on unit number. As there are most likely several telephone numbers associated with the building as a whole, it becomes more likely that one of them will be matched (correctly or not) to the sampled unit, thus increasing the “any match” rate among multi-unit buildings.

Table 2 Addresses by Type of Match Given Address Type: All Matched Cases.

Address Type	MSG Exact*	MSG Inexact Only	MSG Inexact + Accurint	Accurint Only
All Address Types	77.8%	15.2%	6.7%	0.3%
Single-unit building	97.6%	2.0%	0.2%	0.2%
In multi-unit building	29.9%	47.2%	22.4%	0.5%
P.O. box	100.0%	0.0%	0.0%	0.0%
Rural route	100.0%	0.0%	0.0%	0.0%

*Cases with a MSG exact match were not attempted in Accurint. Therefore, a MSG exact match plus Accurint match is not possible.

Because the quality of the address (as reported by the vendors) differs so greatly by address type, it is important to gauge the real accuracy of the matches. In order to evaluate the accuracy of the vendor-supplied telephone number, respondents were asked “Just to check, we sent the letter to [street address (and unit number if applicable)]. Is that still your address?” If a contacted household reported that their address was not the sampled address (including cases with the correct street address but wrong unit number), NORC terminated the telephone interview and moved the sample line to mail production; however, when available, attempts were made to contact the sample address via a different telephone number. This methodology ensured that interview data were collected for sampled addresses only.

Table 3 shows the observed match rate, working number rate, and accuracy rate for MSG exact matches. Table 4 shows the same information using the “any matched” definition. Note that both the working number rate and the accuracy rate are higher (1.8 percentage points and 3.9 percentage points, respectively), especially for apartments (15.1 percentage point increase in the accuracy rate), for MSG exact matches than for the set of any matches. (The resolution rate of the telephone number has been excluded from these tables as no difference between “exact match” and “any match” was found. Both types of matches achieved approximately a 79 percent telephone resolution rate; that is, the telephone number was resolved as a working residential number or non-working/business number.)

Table 3 Addresses by Status of First Dialed Phone Number Given Address Type: MSG Exact Matched Cases.

Address Type	MSG Exact Matched as Percent of All Released Cases	Working as a Percent of Resolved	Accurate as a Percent of Working	Matched, Working, and Accurate (in Percent)*
All address types	57.3%	81.2%	95.7%	44.5%
Single-unit building	67.3%	82.3%	96.2%	53.3%
In multi-unit building	27.7%	72.6%	90.3%	18.2%
P.O. box	9.6%	91.9%	92.0%	8.1%
Rural route	29.9%	89.3%	100.0%	26.7%

*This is an estimate of the proportion of sampled addresses for which a matched, working, and accurate telephone number was available, and is equal to the product of the match rate, the working number rate (given a matched telephone number), and the confirmed accurate rate.

Table 4 Addresses by Status of First Dialed Phone Number Given Address Type: Any Matched Cases.

Address Type	Any Matched as Percent of All Released Cases	Working as a Percent of Resolved	Accurate as a Percent of Working	Matched, Working, and Accurate (in Percent)*
All address types	73.6%	79.4%	91.8%	53.6%
Single-unit building	69.0%	82.1%	95.8%	54.3%
In multi-unit building	92.8%	72.3%	75.2%	50.5%
P.O. box	9.6%	91.9%	92.0%	8.1%
Rural route	29.9%	89.3%	100.0%	26.7%

*This is an estimate of the proportion of sampled addresses for which a matched, working, and accurate telephone number was available, and is equal to the product of the match rate, the working number rate (given a matched telephone number), and the confirmed accurate rate.

Also shown in these tables is the product of the match rate, the working number rate, and the accuracy rate. This rate can be interpreted as the estimated proportion of sampled addresses for which a matched, working, and accurate telephone number was available in the sample. This should not be interpreted as the observed rate, as additional factors such as unresolved telephone numbers and non-response must also be accounted for. Overall, we observe a large drop in this rate when only looking as MSG exact matches (8.1 percentage points). This is driven exclusively by the drop in the overall match rate which, in turn, is driven almost exclusively by the reduction in the match rate among apartments. These data suggest that the decision of whether or not to confirm the accuracy of the telephone match must be made in conjunction with the decision about what quality of matches (as reported by the vendors) to accept.

DISCUSSION

In address-based sampling studies that include a telephone data collection mode, decisions must be made as to what quality of “matched” telephone number to accept and whether or not to confirm the accuracy of the match upon contacting the household. The initial match rate can vary greatly by the level of match quality accepted and the address type; these match rates are related to the true accuracy of the match – there is thus a trade-off between achieved initial match rate and the accuracy of the matches. Additional analyses will be needed to evaluate this trade-off in relation to bias, cost, and transparency. For example, as “exact matches” have demonstrated high accuracy, it may be acceptable to skip confirming the sampled address with little risk to selection bias. However, this approach would also have cost implications (i.e., fewer telephone numbers, hence fewer interviewers but more mailings and data entry staff). Models that consider each of these areas should be constructed to allow for an informed methodological choice either by survey and/or as an industry standard.

REFERENCES

Dekker, K., and W.E. Murphy. 2009. "Address-Based Sampling and Address Matching: Experience from REACH US." Presented at the American Association of Public Opinion Research Annual Conference. Hollywood FL.

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: 6-27.