

Random Digit Dialing: Why Sampling the General Public by Telephone Has Become More Complicated

By Douglas Cox

There was a time when conducting a telephone survey of the general public was relatively straightforward. If you wanted to survey residents of a particular state, county, or metropolitan area, you could obtain a sample of randomly generated telephone numbers associated with that geographic area and begin calling.

It wasn't a perfect system, but it generally worked reasonably well.

Then cell phones came along, and something important changed. The relationship between a telephone number and the geographic location of the person using that number became much less reliable.

That might not sound like a particularly serious problem. But if you're conducting a survey and want the results to represent people living in a specific geographic area, it can be a very serious problem indeed.

How Random Digit Dialing Used to Work

For many years, random digit dialing, commonly called RDD, was a standard approach to selecting respondents for telephone surveys of the general public.

Suppose I wanted to conduct a survey of residents in an area covered by four telephone area codes. I could contact a telephone sample vendor and request, say, 5,000 randomly generated telephone numbers.

The vendor would prepare the sample, distributing the numbers across the appropriate area codes and three-digit telephone exchanges. The distribution could be proportioned to reflect the relative size of the four-digit telephone-number blocks associated with the geographic area.

The actual randomization took place within those four-digit telephone-number blocks.

There were limitations, of course. Some numbers belonged to businesses rather than households. Others were disconnected or otherwise unusable. And not everyone had a landline telephone.

Nevertheless, the relationship between telephone numbers and geography made it possible to construct a sampling frame that reasonably approximated the population we wanted to reach.

Cell phones changed that relationship.

Your Telephone Number No Longer Tells Us Where You Live

With traditional landline telephones, an area code and three-digit exchange could tell a researcher quite a bit about where a household was located.

If I lived in North Carolina and moved to Montana, my landline telephone number would ordinarily change.

But with cell phones, that's no longer necessarily the case.

I can live in North Carolina with a 336 area code, move to Montana, then to California, and finally to Florida, and keep that same telephone number for the rest of my life.

At the same time, someone who originally lived in the Bronx with a 718 area code might move to North Carolina and keep that New York number.

Both individuals have perfectly legitimate telephone numbers. The problem is that those numbers no longer reliably identify where they live.

Now imagine that I'm conducting a survey of North Carolina residents and order a random digit sample using only telephone numbers associated with North Carolina area codes and exchanges.

Two problems immediately arise.

The first is under-coverage. Some North Carolina residents will be missing from my sampling frame because they moved to the state from somewhere else and retained their previous cell phone numbers.

The second is over-coverage. My sample may include people who have North Carolina telephone numbers but no longer live in the state.

In other words, the sample includes people who shouldn't be there while excluding some of the people who should.

And that creates a problem for the reliability of the survey results.

Can't We Just Screen Out the People Who Don't Belong?

That's a reasonable question, and there is a relatively simple way to address part of the problem.

Early in the interview, we could ask respondents, "Do you currently live in North Carolina?"

If someone says no, we terminate the interview.

That takes care of the over-coverage problem among the people we contact. We're no longer including respondents who live outside our geographic area of interest.

But what about under-coverage?

Remember our resident who moved from the Bronx to North Carolina and kept the 718 area code?

That person was never included in our original North Carolina telephone sample. We can't screen someone into a survey if we never had an opportunity to contact them in the first place.

This is an important distinction.

Screening can remove people who don't belong in the survey, but it cannot recover people who were missing from the sampling frame to begin with.

And if those missing individuals differ in meaningful ways from the people we do reach, our survey findings may be biased.

What About Using Address Information to Identify Cell Phone Users?

Telephone sample vendors have developed ways to improve the geographic identification of cell phone numbers.

Rather than relying exclusively on area codes and exchanges, vendors may attempt to associate telephone numbers with additional information, such as billing ZIP codes or other address-related records.

This can certainly help.

Instead of assuming someone lives in North Carolina because they have a North Carolina area code, the vendor is attempting to determine where that person actually resides.

But there is still a problem.

That additional information isn't always available, and when it is available, it isn't automatically accurate or current.

People move. Records become outdated. Some telephone numbers cannot be reliably matched to residential addresses.

So while this approach can improve geographic coverage, it doesn't guarantee the elimination of coverage error.

We're still faced with the question of whether our sampling frame adequately represents the people we want to study.

An Alternative: Address-Based Sampling

One way to approach the geographic coverage problem is to stop using telephone numbers as the starting point.

Instead, we can use physical addresses.

This approach is known as address-based sampling, or ABS.

Rather than building the sampling frame from telephone numbers associated with a geographic area, we draw a sample from residential addresses within that area.

Suppose, once again, that we want to survey residents of North Carolina.

Using address-based sampling, we could select residential addresses throughout the state and send an invitation to each sampled address.

The invitation might be a letter or postcard explaining the purpose of the research, requesting participation, and providing a website address or QR code along with a unique access code to complete the survey online.

Notice what's different.

We're no longer concerned about whether a resident has a 336, 718, or some other telephone area code. We're selecting households based on where they are located, not on the telephone numbers they happen to use.

That can offer a substantial advantage when geographic coverage is especially important.

Of course, address-based sampling brings its own challenges. Not everyone responds to mailed invitations. Some addresses may be vacant or difficult to classify. And if the survey is completed online, we still need to consider whether sampled residents have adequate access to the internet and are willing to participate.

There are also decisions to make about selecting individuals within households and following up with nonrespondents.

So ABS isn't a magic solution that eliminates every potential source of survey error. It does, however, provide a sampling frame that can be much better suited to surveys where the geographic location of the population is especially important.

The Lesson: Begin With the Population, Not the Method

There's a broader lesson here that extends well beyond telephone surveys.

Too often, discussions about survey research begin with the method of data collection.

Should we conduct the survey by telephone? Should we send invitations by email? Should we put a questionnaire on our website?

Those may be important questions, but they shouldn't be the first ones we ask.

We should begin by defining exactly whom we need to hear from.

Once we've identified that population, we can determine what sampling frame gives us the best opportunity to reach its members. Only then should we decide how we're going to collect the information.

The order matters.

A well-written questionnaire and a carefully administered survey won't necessarily produce reliable findings if important segments of the population were never given an opportunity to participate.

And that is especially troubling when leaders are relying on those findings to make important decisions.

The goal isn't simply to collect responses. It's to collect responses from the right people so that the evidence we produce can be trusted.

That's where good survey research begins.

About the Author

Douglas Cox is President of Accurus Research and has more than 40 years of experience in survey research. Today, he helps organizations strengthen their survey research capabilities, close critical information gaps, and turn reliable evidence into better business decisions.

www.accurusresearch.com

AI Disclosure: *This article is based on Douglas Cox's professional experience, knowledge, and perspectives developed over more than 40 years in survey research. AI was used as a supplementary tool for organizational and editorial assistance. All core ideas, insights, judgments, and conclusions are his own.*