So, Scott is at the Pew Research Center, and, obviously, as you can imagine, one of the key themes that we care about these days is the fact that many people believe survey response rates have been dropping over time. The question is how serious of a problem is that for survey accuracy, and we are very fortunate to have Scott with us to talk about this. Thank you.

Scott Keeter:

Thank you, Jon. Let’s see. This one is working. I’d like to put my notes down, but it’s going to make that come on, so . . . Well, good morning, and thank you very much for inviting me, and thanks to NSF for sponsoring this conference and putting it together. This is a wonderful opportunity for those of us who live and die by surveys to try to understand what are the issues facing us and how can we do better.

I do have some trepidation about talking to you today, because, especially about this particular topic, because there are probably a dozen people in the room who have done more research on this topic than I have. It brings back the classic scene in Annie Hall when the bore in the line is talking about Marshall McLuhan’s work, and suddenly Marshall McLuhan steps out of the line and says, “I’m Marshall McLuhan, and you know nothing of my work,” so just hoping that Stanley Presser or Roger or someone won’t do that to me today.

However, since I am no longer an academic – I’m sort of a recovering academic, and I’m a practicing pollster – I’m going to talk to you largely from the perspective of someone who is doing political polling, particularly in the kind of environment that we live in these days.

It has not escaped anyone’s attention that over the past couple of weeks there’s been a lot of criticism of the polls in the presidential race, that we are oversampling Democrats. We’re getting too many liberals in our polls, and thus the leads that most of the polls have shown for Barack Obama in the election is either malpractice on the part of pollsters, non-response bias, or downright conspiracy to pump up the Obama campaign and depress Republicans.

So, this is the environment in which I’m doing my work, and the question of non-response bias is something that’s in front of us constantly and that we worry about. So, while I will talk a bit about the literature in this field and particularly some studies that I think have made wonderful contributions to our understanding of non-response bias, I’m also going to talk a good bit about a study that we at Pew Research did earlier this year to try to get a better
handle on the issue as it affects people who do the kind of social and political surveys that we do, surveys that tend to have very low response rates, particularly much lower than they used to be but much lower than the government surveys that have been the basis for a lot of our understanding.

I was reminded in preparing today that the National Science Foundation had a conference about 40 years ago, which was described as addressing the question of whether non-response have reached a level or are growing at a rate that poses a threat to the continued use of surveys as a basic tool of survey research. So this is not a new question, of course, and it’s one that concerns us, has concerned the field for decades.

Of course, anybody in the room who is on the upper half of the 50/49 age break knows that when we went through our training and survey research the idea that you could make reasonable inferences from surveys with response rates below 60 percent or 70 percent or certainly 50 percent was just kind of laughable. Today, many of the important data collections that are funded by the federal government have fallen below that level, and, of course, many of the surveys of the sort that I’m involved in are now down in the single digits with response rates.

However, there is some good news, and I appreciate Jon’s warm-up here with talking about the accuracy of respondent reports. We do know, especially in the political world, that our surveys have continued to perform well in terms of predicting how voters are actually going to vote.

This is, of course, a perennial anxiety for us in the field. I’m five weeks away now from my final exam when we have to make a final pre-election prediction in the presidential election, and it will be a very visible success or failure if we’re right or we’re wrong, but, fortunately, polls have in forecasting elections been right a lot more often than they’ve been wrong.

You think about the number of different elections that are being polled. We’re really talking about a very sizeable body of evidence about this. You have state level races. You have the electoral college vote at each state level, and the polling aggregators and the individuals that I think many of you, if you’re interested in politics, are familiar with, Nate Silver, Mark Blumenthal, and others, the real clear politics, do keep score cards.
After elections you hear them talking about for the last three or four president elections that really good records have been kept of this sort. Polls have done extremely well in forecasting the outcome and not just in terms of saying who’s going to win or who’s going to lose but getting within a point or two when you take the collective average of the polls of the margins.

So, there is a good record. There is evidence that non-response, as serious as it is, is not creating a bias, at least for that particular survey statistic, but, of course, that’s part of the key and part of the big finding from the literature, and that is not that there is no non-response bias – there is non-response bias all over – but that it is very survey- and very statistic-specific. It tends to occur on some kinds of measures and not on others.

I produced a little handout with just a few graphs, probably all of which you’re familiar with, but let’s just review the evidence about the rise of non-response just for a moment. Again, this is not anything new to anybody.

I just grabbed two or three graphs that I thought was indicative. The one on the front page is the non-response in the National Household Education Survey. This is a telephone survey. You can see that the level of non-response has been growing from around 40 percent back in 1996 to near 70 percent.

On the basis of this trend and the problems that it was creating in terms of the data collection, the decision was made to experiment with an alternative method for data collection, different mode using address-based sampling. But this is one – this is a very typical graphic.

On the next page is the same kind of graph, but this is from the National Immunization Survey. The slope of the line is not as great there, but the trend is unmistakable. One reason, though, that this slope is not so great is the fact that this particular survey largely consists of very, very short screener instrument that poses very little burden. This is also a survey on a health topic, and health topics tend to have lower levels of non-response, and so this particular data collection, because of these two features, at least, has not been affected as seriously as some others.

When you shift over to the world of face-to-face interviewing, there is evidence that more effort is involved and that there is potentially greater non-response in many of these data collections, but when you look at the NSF-funded General Social Survey you
actually see a kind of plateau. There was a drop in response rates a
decade or more ago, and then for the last several data collections—I
don’t think I have the 2012 numbers. I’m not sure if those are
available, but it’s been flat at about 70 percent for roughly ten
years.

When you look, though, at the kind of surveys that we do at the
Pew Research Center, you look at something like random digital
telephone surveys that have to be conducted over a very short field
period, the nature of the business that we’re in. The next graphic
at the bottom sort of tells a pretty serious tale of woe for us.

So what I’ve done here is I’ve taken some representative surveys
from our data collection going back to 1997, and I’ve just broken
out the contact rate, the cooperation rate, and then the overall
response rate, the APOR Response Rate 3. Two things are
immediately evident. This is true, I think, for lots of surveys but
maybe not to the degree that it is here.

We go from a 90 percent contact rate in 1997 for a five-day survey
down to 62 percent his year in 2012. The cooperation rate has
fallen from 43 percent down to 14 percent, and that translates into
response rates going from 36 percent down to 9 percent.

The nine percent number when we released this in May made quite
a splash. A number of people who are in the survey business,
some of our competitors, said, “Oh, our response rates are not that
low.”

They said, “Our response rates are in the mid to high teens,” and I
had to laugh that, “Okay, you’ve got a bigger response rate than we
do,” but still by the standards that we used to hold this is a pretty
scary number.

You know, we could go on through lots of different – the Survey
of Consumer Attitudes, University of Michigan, has had the same
kind of trend line. It’s been documented.

So, what is – what do we know about the consequences of
decreasing non-response bias? What is the potential impact here? This literature is happily very vast, and
it is, for people in this field, anyway, it’s compelling reading,
because there’s lots of clever work that has been done using a
variety of different techniques, some of which Jon talked about in
his introduction.
Probably the best way to summarize it is the graphic that’s on the next page, which comes from a 2008 POQ article by Bob Groves and Emelia Peytcheva. This simply shows the level of bias for individual measures across a large collection of surveys, meta-analysis that they conducted, distributed across different levels of survey non-response.

A couple of things are immediately evident here. One is that there is no obvious trend in this in terms of seeing a bigger cluster at one end of the graph than the other, which suggests that taken as a whole the level of non-response is not associated with the level of bias.

The second observation is that at any given level of non-response the level of bias varies considerably. Many of these little stacks of points are actually stacks within the same survey so that within the same survey at a given level of non-response you may have lots of measures that are spot on and validated and lots of measures that have significant amounts of bias.

The devil is trying to figure out what kind of measures are apt to have the greatest likelihood of being biased by non-response, but, of course, the problem is even more complicated than that, because the type of bias that can occur can be a result of any number of different factors, including topic salience, how the survey is introduced, what your population is that you’re drawing from relative to the nature of the questions, the measures that are being employed. So it’s very difficult to come up with any clear generalizations about when you can expect non-response bias to occur and how severe it’s likely to be.

Now, given this difficulty of being able to predict, I mean, we do have some expectations, and I’ll talk about those in a moment, but given the nature of this, the Pew Research Center has been interested in this question for a long time. We begin trying to get at the question of non-response bias back in a study that we did in 1997 that we published in 2000 in Public Opinion Quarterly, which looked at the impact of greater effort in a survey, essentially trying to do a survey by gold standard methods, as opposed to the five-day data collection that is our standard practice.

In that study, we came up with the conclusion, which was a bit surprising at the time, certainly to us, that the level of effort did not seem to make any difference in most of the survey estimates that we came up with. We did have a couple of estimates that seemed
different in the high effort/higher response rate study and the low effort/lower response rate study.

One set of – one set of measures had to do with racial bias or animosity, that is, holding of attitudes that were negative towards racial minorities. The second one was the level of reported volunteer activity, particularly reports of high levels of volunteer activity. Those appeared to have a bias. The bias wasn’t large. It was only about five points, comparing the lower and the higher response rate methods, but it was clearly there.

Upon further examination, it turned out that the bias that appeared to be present in terms of racial attitudes, that is, that the survey that got the higher response rate turning up more racial animosity, was compromised somewhat by some design features in the study. The interviewers who conducted most of the interviews in the extended field period portion who did the refusal conversions and so forth were overwhelmingly male and white.

The ransacking of the data that was done largely by Stanley Presser and Bob Groves as partners on that study discovered that, in fact, what had happened was that there was a race of interviewer effect that accounted for much but not all of the supposed non-response bias in racial attitudes in that study, but on the whole the study was reassuring and somewhat surprising. Other studies at the same time, including some that Stanley was – one that Stanley was involved in, showed that varying levels of effort in the Consumer Confidence Survey also did not produce differences in the estimates.

So, there was some reassurance in that that at least within the ranges of survey response rates that we were talking about that we were not seeing significant differences in the accuracy of survey measures. That is not at all the same thing as saying that there was no non-response bias, but at least within the boundaries defined in the study it was a good finding.

We repeated the study in 2003 with basically the same results, and then we had a lengthy hiatus as we turned our attention to the issue of non-coverage in cell phones, but we returned to the non-response bias issue this year to try to look at it again. Given what had happened in the literature over this period of time, it was clear to us that a comparison of a low-effort and a high-effort survey was not really going to be definitive in answering the kinds of questions that we had, in part because I think the model underlying
the notion of comparing the higher and the lower effort surveys was wrong.

The idea of continuum of resistance model just didn’t seem to fit the data in our study or many of the other studies that have come in the meantime. We also knew that even the high-effort survey was not going to get a very respectable response rate.

So, instead, we decided we would do a multi-prong investigation in which we did not only a different – we did greater effort in a rigorous survey, but we would also do a very careful effort to benchmark many of the findings in the study to high response rate U.S. government surveys. We all do this when we look at our demographic characteristics, but we went beyond it to a number of other measures, which I’ll talk about momentarily.

Then, finally, taking advantage of the growth in the availability of voter and consumer databases that we have now that we really didn’t have at the time that we did our previous studies, we purchased a couple of those databases, and we made an effort to match our sample, at least from the land line sample, to the records in these databases so that we would have a basis for judging whether respondents were different from non-respondents.

So, if you turn the page, I’ll take you through a few of the salient findings here. We won’t dwell too much on it, because I knew we want to keep moving here, but, first of all, what do we get out of the extra effort that was involved in doing the high-effort survey? The answer is in some cases you get a fair amount in terms of additional contacting, but you don’t get all that much improvement in cooperation, and response rates still were pretty miserable.

So, if you look at this box on the left on page four, you see each section has a line for the Standard Survey and the high-effort survey. In terms of the contact rate, we found that if you give yourself enough time, you can actually make contact in almost as many households now as we could in our previous two surveys. We ended up being able to contact 86 percent in the land line frame and 84 percent in the cell frame with multiple calls stretched over a long period of time.

We were able to get the cooperation rate up but not to particularly high levels and with the consequence that the response rates, while higher – we ended up with a 22 percent response rate in the high-effort survey this time compared with the 9 percent in the Standard
Survey – that’s still just one 1 out of 5 households successfully interviewed.

You’ll see one other thing here, and that is the differences between response rates in the land line and the cell frames. The cell phones are posing considerable difficulties to us in terms of getting cooperation, both in terms of getting people on the phone and getting information from them.

So, what did we find in the substantive analysis? First of all, let’s look at the comparisons between the Standard Survey’s estimates and the government benchmarks. The picture is quite mixed here. For a lot of the estimates our numbers are within the margin of sampling error of what the government statistics are.

One place where we fell short is on Social Security payments. We got 32 percent in the Standard Survey, just 27 percent taken from the government survey.

In terms of food stamps, we got more people reporting getting food stamps or nutrition assistance than in the government survey. We did our best to try to make sure that measurement error would not confound these. I’m not sure that we succeeded on this particular measure because of the way in which those questions are asked.

We actually ended up doing okay in terms of voter registration, but I think that we’re all very skeptical about the voter registration numbers that are in the Current Population Survey’s post-election supplement, and so depending on how you calculated it, it’s either 67 percent, or it’s 75 percent. So while this looks like a victory, it may not be.

The final three measures, though, are the ones that are most problematic. Contacting a public figure in the last year, we get an estimate that’s three times what the Current Population Survey gets. In terms of volunteering, we get 55 percent compared with 27 percent in the Volunteering Supplement. In talking with neighbors in the past week, we get a 17-point inflation in the reported.

So these three measures of social connectedness exhibit very large bias. I’ll come back to that here in a moment, because this is something that we did expect to see from work that Katharine Abraham and Sara Helms and Stanley Presser had done. On the next page, we take a look at the databases that we were able to
incorporate into the design and compare the responders and the non-responders.

Just as an aside, we did do a fair amount of work to try to validate the estimates. That is, when we got respondents and we had the data from the database and the data from the respondents, we found very high levels of accuracy, very much as Jon was reporting in his introductory remarks. That was true on party affiliation. It was true on many of these other kinds of measures to the extent that you could line them up.

Here the picture is really quite positive. On a number of different measures of financial circumstance, including the extremes, the high wealth and high income and the low wealth and low income, the measures don’t appear to have any particular systematic bias.

Also very gratifying to us was the fact that in terms of the split in party affiliation between the respondents that we were seeing – we see virtually no bias there at all. That’s the result we would expect if we’re able to correctly predict the outcome of elections, but it was gratifying to see that, in fact, it showed up.

This particular set of comparisons comes from a large commercial credit database. The ones on the right come from a voter database, and here we do see some other issues. The reported voting in the survey, 54 percent, the people who responded – the people who responded voted in the database at a 54 percent rate, those who didn’t at a 44 percent rate.

That, I think, is evidence of exactly what Jon was talking about that there is a non-response bias here, not necessarily an over-reporting. In terms of party registration or party affiliation, the splits in the people who responded to the survey and those who didn’t were virtually identical.

The last box on the page breaks this down according to the types of non-response that we had, but in the interest of time I’ll skip that.

Let’s turn to the back page, because here is the – here is the topic that I want to conclude on. This is a table from the 2009 AJS article that Katharine Abraham and Sara Helms and Stanley Presser wrote on volunteering rates. This is based upon an analysis of the Current Population Survey, a Volunteering Supplement, and then a subsequent analysis of a subset of respondents who were recruited to participate in the American Time Use Survey.
The second – the top line here shows that 29 percent of respondents reported that they volunteered in the Volunteering Supplement to the Current Population Survey. The respondents who actually took part in the Time Use Survey, the third line, they report volunteering at a 36 percent rate, and the non-respondents, the people that were not able to be successfully interviewed, report volunteering at a 20 percent rate.

So, for context here, the Volunteering Supplement respondents basically are talking about an 80 percent or an 81 percent response rate. For the Time Use Supplement respondents, you’re talking about a 53 percent response rate. In the article, they make the strong case that the difference in the reported volunteering rates from people who did and didn’t respond to the Time Use Survey is a function of non-response bias.

This is, I think, the largest known kind of bias that we see in the sort of surveys that we in the field that I’m in do, and it’s a very worrisome one, because I’ve been doing volunteer and civic engagement for many years, and I think a lot of what I’ve discovered is probably not right. Certainly any kind of trend estimates over time in an era of growing non-response are going to be suspect if this conclusion is correct.

There is a silver lining here, which I think is important to note, and that is that the correlates of volunteering were seemingly unaffected by this bias in the overall reported level. That is good news, because it suggests that you can use a survey that has a lower response rate.

You may have a massive inflation in the level of reported civic engagement, as we apparently do in our surveys, 55 percent, just to remind you of what our estimate was, but it may be the case that any analysis that we do on what predicts it and what causes it and what its correlates are is not going to be biased to the extent that what Stanley and his colleagues found is true.

Let me conclude with just a couple of remarks about what we can do about this and maybe one other observation about the consequences here. Much of the focus of the research in the area of non-response bias has been on what is the bias? How sizeable is it? I think it’s also important to note that something else is happening here.

With these trends in non-response, all of us in the business are having to devote considerable effort to fighting against those
trends. It is much harder to contact people today on the telephone than it used to be, even with the greater effort, the bigger calling regimens and the use of incentives and other things that we may or may not be doing. We’re still sort of losing the battle in terms of the slope of the line, but we’d be losing it a lot worse if we weren’t doing that.

But what this also means is that we’re putting a lot of effort into trying to prop up our response rates and keep our non-response down. Because survey research is fundamentally a craft of trade-offs, those resources are coming out of other things that we could be doing that might reduce total survey error.

So, we have to make a calculation of where our money is most effectively spent in the pursuit of good survey estimates. I mean, no doubt about it, nine percent response rates for public polls creates an appearance problem, a credibility problem, but 30-something percent response rates in the National Household Education Survey also does that.

The question that we have to confront as researchers and that I’m hoping that we will be able to do more work on is what could we be doing with the money that we’re using to prop up our response rates that might be more valuable to us, more pre-testing, bigger sample sizes, and the like? So, it’s really critical that we have a better understanding of the circumstances under which non-response bias occurs so that we can most effectively use our resources.

I think I’ll just say I think the kind of work that has contributed to our understanding of non-response bias needs to continue. In particular, we need to do more with the available databases matched in with our samples than we’ve done in the past. This is a resource that really wasn’t available to us five years ago or ten years ago to nearly the degree that it is now.

The quality of these databases, the amount of information contained within is growing, and it does offer us the opportunity to do more with the assessment of bias than we’ve ever been able to do in the past, but not a lot has been done with it. I mean, Jon reported some studies. I looked at some studies. We did this study, but I think that’s an area of further investigation that’s well worth supporting.

I think some smaller bore work is still very useful. The idea of seeding your samples with households with known characteristics
to observe their response propensities remains a very smart technique and one that oftentimes we don’t think about in the course of trying to get the work done.

I’m hopeful that the National Science Foundation will be a partner in the continuation of this. I think for all of us in the practicing survey business we just need to think more about how our studies that we’re doing anyway can contribute to this body of knowledge.

Let me stop there so that there is a little bit of time for questions. I’m happy to respond, but I hope that those of you who have been doing a lot of this work would share your thoughts, as well, if questions come up that I can’t answer.

Jon Krosnick: Thank you, Scott. So I’d like to ask that people who comment, again, turn on the microphone. Speak into the – by pressing the button so the light’s on, and if you could, just identify yourself before you start making your comments.

Richard Freeman: Richard Freeman, economics.

Scott Keeter: Hi.

Richard Freeman: Hi. This is a question about the correlates being – holding up, because one way you could imagine the correlates holding up would be that the response rates are falling similarly or suspiciously low among all groups, because if that’s not the case, it then makes for a complicated interaction that I’m somewhat dubious we would see.

If we get very few young men responding, you would think you would get much worse situation there, and that correlate would not give you the right results. It would only be the young men, let’s say, interested in politics or something.

Scott Keeter: I think that’s true, and it’s one of the reasons why I was gratified to see what the Abraham paper was able to do to demonstrate that the difference in reported rates of volunteering across a whole variety of categories, particularly demographic categories, was consistent, but then they went much further with multivariant analysis.

Whether that applies in other areas, for example, the clear bias that we have that people who are interested in politics, not necessarily the same mechanism that’s creating the bias in the volunteering rates, but the question of whether the interest in politics polls, as
well, when you look internally at those correlations. I have a feeling that it doesn’t.

I mean, we in the polling business are happy with the bias, because it means we get the respondents who are going to vote, and we don’t get the respondents who are not going to vote, but if you’re trying to assess, I think, as Jon said at the very end of his talk, if you’re trying to assess what it is that makes a voter versus a non-voter, you may not be able to do that in an unbiased way.

I don’t know if, Stanley, you want to speak to that question from your work or not.

*Stanley Presser:* Well, your point is a good one. I mean, at least in the context of the volunteering question we didn’t find. We tried real hard, and we did all sorts of sophisticated modeling, and essentially the inferences are the same. But I think the point that Scott made is a good one that this is kind of an N of 1, in a way, because we studied this one issue, and there are lots of others that need to be examined.

*Scott Keeter:* Let me make one further point on that, because Stanley and his colleagues made a very interesting suggestion at the end of that article, which was that to the extent that this social connectedness or willingness to help or whatever you want to call it that’s creating this bias is measureable by the reported volunteering or any other measure that you might have, you could incorporate that into your weighting if you trust the government statistic.

So we did that in our survey. We weighted the entire survey by the usual suspects plus reported volunteering, and then we reran all the estimates and compared them. What you find is that no estimate, with the exception of those that logically would be very closely connected, those social connectedness measures that I showed you earlier, no measure changed by more than about two or three percentage points in the reweighted version.

Now, the things like talking to your neighbors or contacting a public official, those were more biased, but if you could take all three of those and put them in combined as part of your post-stratification weighting, you might be able to correct to some degree for the bias that you have there, although at a consequence, given how far off we are from the parameter, at a consequence to your design effects. But a lot of things are not correlated with willingness to help or volunteering, and so that helps us a bit there. Okay?
Paul Biemer: You know, it occurs – I am Paul Biemer. Thank you, Roger. This is Roger. It occurs to me that with RDD surveys you sort of have two real problems. One is that you’re dealing with people on the phone, and so you don’t get good cooperation that way as you would, say, in face-to-face interview. We know that cooperation is higher there, but also you don’t have a lot of data on the non-respondents to make adjustments.

So, one of the things you might try to explore is ways in which you might bring in more data in a telephone survey to try to make adjustments. Andy Peytchev and I have been doing some work with looking at Census, you know, bringing in Census blocked and tracked information, attaching that to the units.

Unfortunately, that hasn’t worked too well. It only works when you have highly clustered characteristics in the population, because then the tracked or blocked information is more predictive of the – characteristic of the individual, and that’s usually for things like income and race and hispanicity and so forth.

So the other thing has to happen is that you have to – those characteristics have to be somewhat related to the outcomes that you want to adjust for. If the outcomes aren’t related to race, hispanicity, and income, that adjustment is not going to work very well, but at least for those where it is it could help reduce the bias.

The other thing is I’ve been dealing with some ways in using callback information to model the response propensities, and I think that’s showing some promise for telephone surveys, as well. So I think one direction we can go is to just look at ways in which we may be able to do better adjustments for telephone service.

Scott Keeter: Thank you. I think that’s a very good suggestion. One of the – one of the problems – I mean, you pointed out the problem, and that is that it may be that the things that you can know about a household from Census information or the like are not things that are really biasing characteristics.

We already can take care of some of this through the observed demographics and the post-stratification weighting, but one idea that intrigues me that I haven’t figured out how to make it work in my business is if you went to addressed-based sampling for your frame, you would have even better information about households that you could incorporate in.
But I think ultimately that takes you a stop down the road to having a sort of ongoing panel that you can harvest on the telephone or maybe in other modes to be able to do the kind of quick turnaround surveys that we have to do. I think that may be a direction that the field is going to have to go, because the line going down past nine percent is still pretty linear. It’s not kind of flattening out for a soft landing. Thanks.

Peter Miller: Peter Miller, Census Bureau. I’m not speaking for the Census Bureau. Jon mentioned and Scott has reinforced the point that it’s common to use external records or some sort of external criterion to validate survey measures. This was a common practice back in the seventies and before when federal surveys were designed to use a record-check study to look at how well the surveys stood up, so in the crime statistics area this was a frequently done thing.

At the same time that these studies were being done, in Chicago there was a scandal with respect to the police records in which the cops were systematically going through the records and declaring certain reports of crime to be unfounded, meaning that no crime had occurred. This was a convenient way to lower the crime rate.

There was a scandal, and the Police Department had to do something in response to the public outcry about this, and so they did a telephone survey to call up the alleged victims of crime to see whether or not there was, in fact, a crime that had occurred, and they should reinstate the record.

So, I guess the point or an issue that we may want to consider in all of these comparisons is just how good are the records to which one is comparing the survey evidence? How would we validate records without using a survey to do it?

Maybe you’ve suggested, Scott, that we should explore these records more now and that they’ve gotten better. I’d be interested to know how you know that, because it’s a continuing issue that we have to face. Thanks.

Scott Keeter: I probably can’t give you evidence that they’re better. I can give you evidence that they’re bigger, that there is a lot more information in them, but I share the concern that Bob Groves has expressed and his predecessor at the Census expressed that administrative records and other databases have not been subjected to the same level of scrutiny.
They don’t have the same standards of transparency that surveys have, and I do think that that is a problem. That’s another area where some help from funders could be really wonderful if we could figure out a program for doing that, because I do think that these databases are important adjuncts to our work, but I share your concern about the accuracy of the information in them.

Yes?

**Randy Olsen:** Randy Olsen, Ohio State. This is an area where NSF could potentially play a helpful role. One of the problems in doing record check validations is that of late we seem to have become very nearly paranoid on the whole issue of confidentiality and privacy and so forth, making it very difficult to access administrative records to do the sort of cross-checking, unless you happen to be in one of these hallowed environs, which raises the whole issue of replicability, which is usually thought to be a good thing in science.

So, if we’re going to make progress on this and quantify what’s going on and maybe even fix it, it might help if we lightened up just a little bit on the ability to access these records. We’ve done some work comparing wage reports in the NLS with UI earnings match data, and the results look pretty good. It kind of confirmed what Jon said.

Collecting factual information works pretty well, and, yeah, by the way, the administrative data have their problems, but that entire effort has become really, really compromised because of the way we now treat Social Security numbers, which is the only feasible way to do a good match of that sort.

**Scott Keeter:** I mean, there are two aspects to that. One of them is there is a cultural shift that would be helpful in terms of the public’s willingness to be less paranoid or concerned, and I think people are probably justified to some degree.

There is a bit of a generational shift underway. I don’t study this all that carefully, but it certainly seems to me that there may be a different attitude about privacy and the like among younger people than there is among older people, whether that will persist over time.

But the other aspect is something that Peter said in his talk last week at the POQ conference here in Washington, and that is that we need a federal statistical law that compels agencies to share
data under strict rules of confidentiality. Peter, I don’t know if you have a soapbox or if you want to get up on it for a minute, but that is something that we in the profession could potentially be a force for pushing Congress down the road to do.

Peter Miller: This is Peter again. So, we face this issue in spades just within the building in Suitland, of course. It is extremely difficult to get data, MOUs, between various parts of the Census Bureau, let alone getting them between the Census Bureau and the Treasury Department or NCHS or what have you.

So I share this concern about being able to do this in some way that facilitates our being able to understand things without waiting a couple of years to get through the thicket of – again, I’m not speaking as a government official here. I’m speaking as somebody who is just trying to work.

Scott Keeter: Well, I’m not a government official, so I can say that it may be that AAPOR and the Market Research Association see more – Kasrow and others are going to have their hands full over the next few years just fighting a battle to preserve the resources that the Federal Statistical Service currently has, but we might also consider as these professional associations an effort to try to push this issue and to bring the U.S. practice more into line with a lot of the practices elsewhere in the world, not to get all political on us here.

Eleanor?

Question: Just to comment, I think we can assume, and I think there is evidence for it – you mentioned the scandal in Chicago. I think we can assume that when there is motivation the administrative records will be biased just in the same way that the survey reports would also be biased if there is a motivated reason for the bias.

So, there is this paper by Joe Saxhow, and if I’m misreporting maybe Frauka can correct me, but this was a study comparing survey reports of diabetes or, sorry, yes, diabetes with blood tests and Medicare reports of diagnosis. If I – as I recall it, and my recall may be biased, because I’m motivated to report in a certain way, the survey reports were closer to the blood test, which is the gold standard in that case, than they were to the Medicare diagnosis codes, because there is motivation to report a different code, you know, for billing purposes, right.
So I think these are reasonable – I mean, if we use our heads a little bit, we can assume that there are going to be certain errors, biases, not errors. I mean not random errors but biases in all of these data sources.

Now, how you deal with that is a different question, but I don’t know how many more studies we have to do. That is, we have to identify the sources of error in administrative records, and nobody has ever bothered, I think, to ask what those might be. There would be random errors, and there would be biases, right, but what you do about it, how you calibrate these things to try to get at “truth” is still another question.

Scott Keeter: Thank you. Good point.

Frauka Krytev: Since Eleanor already mentioned me, I’m going to jump in here for a second. I would second what you said, Eleanor, that in working with the administrative data in Europe I’ve certainly seen biases, uncertain variables, just like in the survey, too. It’s not – I wouldn’t say it’s the entire admin data source, but it’s specific variables that can suffer in a similar way.

The other element I think that we have to keep in mind is that there is always a timeliness issue, so if you plan on using them for the weighting adjustment in case they were good, they will always lag behind, and that is some problem with the Medicare data, too. You don’t – the blood test and the survey they’re sort of taking at the same point in time, but the admin data will always lag behind, because there is a big processing effort inside the agencies to make them as good as possible, and that delays things.

So it will always work for us in retrospect to evaluate data, but it’s much harder to think of it to use it in ongoing production to have a timely estimate that is – I mean, there is certainly research to be needed how that could be done. How much processing do we need for our survey purposes? Maybe that can be a different set of administrative data than the one that is later on published by the agency.

Scott Keeter: That’s an excellent point. We confronted this in the work that we did with the credit database that we obtained. The company that produces it reports on the frequency with which the information is updated, but they’re not very transparent, and it’s really a black box. They have an indication of income or home value. They say that they update them every year or two years, but we don’t
necessarily know that that’s the case. It’s very difficult to know the quality.

I mean, as I said, we were able to verify by asking people some of the same kinds of questions in the survey that the records were pretty accurate, but there’s a lot of slippage there. I think the fear is that if some of the records are tied to characteristics that are susceptible to bias that’s motivated in some way, as Eleanor says, we would not be able to detect that in the kind of validation effort that we did.

Question

The lesson, don’t unplug your plug. So I would actually like to ask Matt DeBell, ask the last question. This is building on some work that he has done that links, I think, with, Scott, what you talked a little bit about that when I looked at your table – I guess it doesn’t have a number, but page four on the right-hand side – that it looked like the government survey was corresponding well with your survey for things like U.S. citizen and homeowner and that the really striking differences at the bottom of the table involved contacting a public official in the past year, volunteering, and talking with neighbors.

You alluded to something that I think Matt could elaborate on a little bit, which is concerns about the CPS supplements that maybe in this case it’s not the case that your survey is inaccurate, that maybe the benchmark is inaccurate, which, of course, picks up on a theme everybody is talking about. Matt has done work in this area that is actually not on the program of this conference, and now seems like a good time to describe it and maybe kind of ask that question.

Scott Keeter: I’m all ears.

Matt DeBell: This is Matt DeBell. Well, Jon, I’m not sure I’m reading your mind quite accurately, but I do have a question related to this very thing, so maybe this is what you have in mind. So, Scott, your survey, I presume, is all self-reports.

Scott Keeter: Yes.

Matt DeBell: And the CPS makes heavy use of proxy reporting.

Scott Keeter: Yes.
Matt DeBell: And I wonder if you have any thoughts on the role that proxy reporting might play in that, in the differences that appear in this table. Is that even at all related?

Scott Keeter: Did he ask the right question? Yes. To be honest with you, I don’t think we have a hypothesis about that. You know, I’m not an expert on proxy reporting. I do think that it has lots of potential for error, but I don’t know that I can specify it for any of these. Do you have some theories? Because I’d be really interested if you do.

Matt DeBell: There is a second related issue to this, which you may have alluded to when you were referring to the voter registration proportions. You said there are two different numbers that you might report. When the Census Bureau reports voter turnout rates, they have used two different methods.

One is the method they used for a long time was to treat non-response on the voter turnout item as non-voting, and if you do that, the numbers look quite accurate. But if you leave out the non-responding cases, then the error becomes much bigger. Is that what you’re referring to?

Scott Keeter: Yes.

Matt DeBell: Okay. Is that issue – I’m not sure if that issue applies to the other items on contacting a public official or volunteering. I haven’t looked at those items specifically.

Scott Keeter: I don’t think that it does.

Matt DeBell: Okay. I don’t know what the item non-response rates are, but it could be something else to look at, I suppose.

Scott Keeter: Yeah.

Question: Let me supplement Matt’s question. So Matt has discovered that in the – so I used to think the CPS was a terrific benchmark, and I’m now not speaking as a federal government employee. I no longer have the confidence in it that I used to have because of what he showed me.

So one of the issues of the supplements is – of the study in general is that it’s not a survey of a randomly selected household member. Whoever in the household is willing to answer the questions at any given time is providing reports, and the majority of reports within
multi-person households then end up being proxy reports, as opposed to self-reports.

There are a lot more “Don’t knows” in the proxy reports, even probably for these items, and you can easily imagine that if somebody is asked, “Did your husband volunteer for an organization?” that people would answer that by saying, “Can I think of an instance in which he volunteered for an organization?” and there would be instances they wouldn’t be aware of, and this would exacerbate the nos.

In addition, there are – there is this sort of striking aspect of the supplements, which is that the interviewer gets to decide whether to administer the supplement or not. The interview is given the opportunity to decline to administer the supplement if he or she thinks that doing so might tax the respondent in a way that would reduce the likelihood that he or she would participate in the next wave of the CPS.

So, now we have “Don’t knows” happening for the proxies, because I don’t know whether or not my husband or my wife did things. Secondly, we’ve got lots of instances in which the thing is not administered at all, because the interviewer chose not to administer it, and you could easily imagine that they are choosing not to administer it based on inclination to be nice or to be generous with time or whatever.

So, I guess, I just wanted to supplement your pointing to the proxies with this other issue. Is that – is there one more source of error that I’m overlooking in what you looked at?

Matt DeBell: This is Matt again. Yes, I think those are the key issues, and I would say that I would expect that proxy reports of these activities like contacting a public official would be lower than self-reports, because if you don’t know, you’re not going to report something, and you’re not going to be as familiar with those activities by somebody else, so I think you might be more accurate. Your results might be more accurate than these numbers imply they are.

Scott Keeter: That’s a nice thing to think, although I don’t know on what basis I would be able to make that claim. I mean, I guess, just simply doing a pair of estimates based on different assumptions about the no-answer categories would be the way to do it.

Question: Can I? It’s fundamentally – it’s speculative, right? I mean, we don’t know, but I think what works in your favor is that there is
reason to doubt the benchmark for these particular items, so we shouldn’t highlight this as a big error. There are other errors to worry about, so I would focus on them, rather than the one where the benchmark is a little more suspect. That’s all.

Scott Keeter: Stanley, you confronted this, didn’t you? It’s been a little while since I read your paper.

Stanley Presser: Let me chime in. We actually examined this issue, and so we redid all our analyses in single-person households to be able to completely remove the effect of proxy reporting, which, as you say, could really be important in these.

For our analyses, there were no differences, so that suggests that at least for this one measurement about volunteering over the course of the last year that probably there is a bias, because it persists in all our analyses when we just use the single adult households. That’s not to say that these issues that you are raising more generally, because there is no question that proxy reports can be important some of the time.

This issue of the fact that interviewers may have an incentive not even to – even in the single-person households may have an incentive not to administer it, because they’re worried about the impact of this longer interview, there are a lot of important issues that I think really do need to be addressed, but at least with this one finding on volunteering I think there probably is something going on.

Scott Keeter: I just want to make sure I understood what you said. You said you did it on single-person households but that you also did it using the full – the other cases that involved proxy reporting and that you would have come to the same conclusion either way about the bias?

Stanley Presser: That’s correct.

Scott Keeter: Okay. Thanks.

Jon Krosnick: So, amazingly, we are right on schedule here, and it is time for a coffee break, and then we will come back at five of ten for Curtis Cobb, who will talk about proxy reporting. How perfect.