_Jon Krosnick:_ So, we’re shifting our schedule a little bit, ‘cause we botched up who can be here when. So, fortunately, we can squeeze Roger Tourangeau in at the end of this afternoon for a high – and ending on a high point. He will talk about confidentiality and anonymity.

_Roger Tourangeau:_ Thank you for accommodating me. Why don’t you tell me how long you’d like me to talk.

[Inaudible comment]

_Roger Tourangeau:_ Okay, all right, I’ll go for 25-30 minutes. So, here’s what I’m going to talk about. I’m going to talking about the concepts of privacy, confidentiality, and anonymity. I’ll talk about sensitive questions. I’ll talk about the evidence regarding self-administration and it’s utility for collecting information about sensitive topics.

Then I’ll talk a little bit about trying to get beyond self-administration and improve reporting further if self-administration improves reporting. Then I have a couple of conclusions.

Eleanor Singer has done interesting work on the concepts of confidentiality in privacy in connection with the decennial Census, and so these are some items she’s administered to people about the decennial Census. So, she classifies items as assessing concerns about privacy versus concerns about confidentiality. And it’s a kind of a subtle distinction.

Privacy refers to your unwillingness to reveal information at all. It’s none – you know, none of your business, is what the concept of privacy is about. Whereas confidentiality is maybe a willingness to reveal the information to a researcher, but not wanting it to fall into the wrong hands. And there’s two different kinds of wrong hands. It could be that somebody in your household would overhear what you said during an interview, and they would learn something that you’d rather they didn’t know, or it could be that some third party will get a hold of the data, maybe another federal agency.

One of the themes of Eleanor’s work is that a substantial portion of the American population thinks it’s real easy to link up these datasets, and that if you give the information to the Census, the IRS will have it in seconds. Right? And they’re not sensitive to the information we heard this morning about how impossible that is.
Okay, so, for example, one of her confidentiality items is do the police and FBI use the Census to keep track of troublemakers? I like that item a lot. And then the Census is an invasion of privacy, is sort of a classic privacy item.

Okay, it turns out that a lot of interviews aren’t done in private. There’s been – there was some early work in the American National Election studies by Barbara Anderson and Brian Silver, suggesting that, you know, on the order of 40–50 percent of the interviews are done with some other household member present.

And this is from Zeina Mneimneh’s dissertation. She’s a Michigan Ph.D. student who just finished. And she looked at data from the World Mental Health interviews, which are conducted in many countries around the world. And the only thing I want you to look at is this second column of data, and you can see that in Japan, about 12.6 percent of the interviews are done with somebody else present, whereas in – I think the worst country in the world – is it India? – 70 percent are done with another person present.

These surveys are about mental health symptoms. Like, “Have you ever tried to commit suicide?” So, this is kind of sensitive stuff, and you wouldn’t necessarily want to have somebody else listening in. And according to Zeina’s results, there are – you can think of at least a couple of factors, pretty obvious factors, that determine whether or not an interview gets conducted in private.

One factor is, are there other people around? So, if you live in a household of size 1, you’re much more likely to have a private interview than if you live in a household with size 27. But beyond that, even controlling for differences in household size, you still see big differences across countries, and partly that reflects differences in cultural norms regarding privacy.

So, in some countries, respecting people’s privacy is a value; in other countries, sharing with other people – you know, collectivist cultures, privacy is not such a value. And you see that across these countries.

The third factor that she points to is there are huge interviewer variance components to this. So, it’s like some interviewers get it, that the interview is supposed to be conducted in private, but other interviewers don’t worry about that so much.

And I think this picks up on a theme we heard earlier today, in that interviewers have a lot of jobs to do, and there’s definitely a job
one, which is to administer the questions. And then the secondary
tasks, well, they may exhibit a lot of variance in carrying them out.

What difference does it make if an interview is conducted in
private? The best model I’m aware of about the consequences of a
lack of privacy are, as this model by Bill Aquilino, a drug
researcher at the University of Wisconsin.

And Aquilino says that there are two main factors that determine
whether there’s damage to candor from having a third party
present. One is, does the person already know the information? If
you smoke dope with your wife, and she’s there, and they ask you
on that NSDA, “Do you smoke dope,” you might actually feel
more obligated to tell the truth because you’d be embarrassed to lie
in front of your wife, who knows the truth.

But if the other person doesn’t know the information, if they’re
you’re parents rather than your spouse, and you hardly ever get
stoned with your parents, then you may not want them to find out.
And I actually did a study once where you asked teenagers about
whether they smoked or not, and if parents were around, you get a
much lower percentage of people reporting that they smoke. This
is the kind of thing, it turns out, that teenagers keep secret from
their parents.

And in fact, Ting Yan and I did a meta-analysis, where we looked
at a lot of studies that looked at the impact of third-party presence,
and there’s a huge effect of having parents around when young
people are interviewed.

And in general, Zeina found that there are signs of increased social
desirability bias across cultures when third parties are present
during this interview.

Okay, so I’ve talked a little bit about two of the concepts I was sort
of assigned: confidentiality and privacy. A step that’s sometimes
taken in order to give – I think to convey a greater sense of privacy
and confidentiality is not identifying the respondents. So,
collecting data anonymously.

I think it’s very hard for face-to-face surveys to do this
convincingly, and to say that the data are anonymous. If
somebody comes to your house, and they have your address,
maybe they get a signature on a consent form. I mean I think
anonymity’s out the window regardless of what you do after that.
But a few studies try to do this. Monitoring the Future is a study of high school seniors about drug use, and they mail out – or they send out questionnaires to schools, which get distributed in classrooms. And there’s no identifying information on the questionnaires. And there’s a worry about this.

Jon has been involved in a research project, and he sent me this paper and asked me to read it. And so, I read it; I did what I was told by – is it Lokus? Is that what it says? [inaudible]. They did three laboratory studies, and they find some signs that people don’t take a questionnaire very seriously. The idea is that if you’re not being held accountable for the quality of the data, then you’ll goof off and do signs of cognitive shortcuts and other forms to satisfy something like that.

And there’s evidence in those studies that you get that. I’m not convinced that I’m worried about Monitoring the Future. Monitoring the Future lines up pretty well against other drug study benchmarks, but it could be that the data quality is lower in that study than in some other studies. And there’s lots of evidence that reporting of drug use in surveys is not great.

Okay, so, those three things – privacy, confidentiality, anonymity – are often talked about in conjunction with another variable: collecting sensitive information. So, it’s presumably more important that the data be seen as confidential, that data collection be anonymous, that the data collection be carried out in private when there’s something about the questions that’s potentially bothersome to the respondents.

And in my review of the literature on this topic, there are three somewhat distinct meaning. Sometimes questions are seen as inherently offensive. Actually, Gordon once made the distinction between a sensitive question versus a sensitive answer. And some questions – like if the Census Bureau asked you what your religion is, people would worry about that. They did actually contemplated putting a religion item on the decennial Census in the '50s, but decided against it.

So, there are inherently offensive questions. There are questions that raise concerns about disclosure to third parties, such as other family members or other government agencies. People worry a lot about – the survey companies all worry about disclosure risks, and that somebody’ll hack into the dataset and figure out that so-and-so said this.
To the best of my knowledge, that’s never happened, and lots of datasets are destroyed and rendered worthless by the removal of all geographic information in an effort to make sure that that doesn’t happen in the future. I’m not sure whether it’s worth it or not, but it would only take one or two cases of somebody hacking in and blackmailing somebody, I think, to destroy the industry, so I guess I approve of this disclosure avoidance stuff.

The last meeting, the ones that people usually talk about, is questions have social desirability. That is to say, there’s a socially approved answer, and a socially disapproved answer. There are two conceptions of this social desirability bias.

The older, psychological conception, dating back to Marlowe and Crowne and Paul Paulus, is that there’s somehow a trait involved. That there are some people who are very worried about how they’re perceived, and so that they consistently present themselves in a very positive light.

I think most survey researchers take a more social/psychological take on this, and they say, “Well, an item is sensitive or raises social desirability concerns depending on whether there’s somebody there that you might be embarrassed in front of, and depending on the particular topic, and also depending on whether you’re in the undesirable category or not. So, if you asked somebody if they’ve ever used heroin, and they’ve never used heroin, this is not a particularly sensitive thing.

Here are the Crowne-Marlowe items. They’re, I think, designed especially to measure acquiescence. Oh, sorry, they’re supposed to measure social desirability. I really think these are worthless, but you still see them in studies. And basically, they consist of statements that hardly anybody would give the socially desirable answer to and be true.

So, before voting, I thoroughly investigate the qualification of all the candidates. There are two people on this planet for whom this is truly a correct statement or something. But they do have this problem, in that most of the correct answers are on the – are true, or most of the answers that are in the measure of social desirability are true. It’s a crummy scale.

So, respondents may not want to disclose the information to anybody, to an interviewer, or to outside parties. Here’s – too bad Norman has gone. This is from a study he did with Eleanor.
Singer, and the topics that are embarrassing are basically sex and drug use.

Drinking – people don’t want to tell you, but they don’t particularly see it as potentially embarrassing. I think there’s other reasons that they don’t want to tell you. And in my work, I’ve looked at three different –

I’ve completely forgotten what you told me about the number of minutes I have.

[Inaudible comment]

Roger Tourangeau: Okay. People worry about unit non-response. That if the topic of a survey is sensitive, people won’t do the survey at all. Other people worry about missing data, that people will skip offensive or embarrassing questions. And then what I’ve worried about in my work mostly is that they’ll do the survey, and they’ll answer the questions, but they won’t tell you the truth. So, that’s reporting errors.

You know, since the middle '70s, at least three techniques, maybe four if you count open items, but you can find this in the work of Norman Bradburn and his colleagues, that self-administration helps. Open items are better than closed items.

There’s some early studies suggesting the randomized response technique is useful, and then there are some papers on the bogus pipeline, and I’ll talk a little bit about randomized response technique and bogus pipeline later.

Here’s an example of some studies. These are all studies – I’ll just talk about the top line. Bill Aquilino compared a paper self-administered questionnaire to a paper-and-pencil interview. The response rates were similar. This is a mode experiment. And interview brought a paper questionnaire to the respondent, and you see some increase in the reported marijuana and cocaine in that study.

If you go down to, let’s see, Turner, Lessler, and Devore, this is a study done in conjunction with – I think it was then the National Survey of Drug Abuse, the NHSDA. You see, there’s twice as much cocaine use. These are ratios. They’re not odds ratios; they’re ratios of percentages. More than twice as many people admitted in a paper self-administered questionnaire that they’d
used cocaine in the past month than in a paper-and-pencil interview.

And there’s – you know, there are other more recent studies. So, a lot more people say they use drugs when the questions are self-administered. But do they really use more drugs? And there’s about four or five studies that have record-check data, and do these mode comparisons.

Frauke, Stanley, and I were involved in one of these studies. We compared three modes of data collection. This was a survey of the University of Maryland alumni, and in a weak moment, the registrar agreed to give us access to the records of these people. I don’t know how – it just seems like a miracle that this happened.

And we compared things like false – the false negative rate in reporting a bad GPA, for example, in the three modes. And you can see the CATI data are the least accurate on most of these variables for the negative things. So, if you ever got a D or an F. A third of the people in the caddie condition denied it, whereas only 20 percent in the Web condition denied it. And these are all contingent on what their transcript actually said.

So, in this particular study at least, the self-administered modes seem to yield not just more reporting of bad stuff, but more accurate reporting.

And then here’s a study we did. We used the Aristotle – the much-condemned Aristotle database. We started the Aristotle database, and we got people who were reported on the Aristotle database to have missed the two previous general elections. And then we phoned them up, or mailed them a questionnaire, and we asked them about their voting in the last three elections. And you can see we find an effect for mail administration relative to telephone administration. And this is Tourangeau, Groves, and Redline.

These differences in reporting are significant. And then we also looked at the percentage of people who are non-voters according to their record, and how much did they over report in the survey, and you can see they over reported some more in the telephone survey than in the mail survey. It’s not like they were perfect in the mail survey, though.

I mean half of the people classified as voters by Aristotle. We found the records pretty good. We had more than a 90 percent
match rate on birthdays, so we think that the other data might have been good, too.

But remember, these are people we found. Right? So, if people had moved out of the area, or has some of the other problems, this is a reverse record check in that we started with the records and then surveyed the people. So – okay.

“People have been doing,” this is a quote from Warner, so this – I think it’s from 1964. So, this has been around, randomized response, for 40 years. I’m not going to spend a lot of details on how to do this, but in the Werner technique, people get one of two statements.

For example, statement A might be, “I am for legalized abortion on demand,” and statement B is, “I’m against legalized abortion on demand. And they get these items with fixed probability. So, they may spin a spinner or something. The most common thing used in the field is a coin flip. Many people have coins right in their homes, so you can do this. They don’t generally have spinners. And then the respondent just says whether they agree or disagree, without revealing which statement they got.

Here’s another version. This is from Norman’s book, “Have you received a ticket for parking in the last month.” Some people get the red item. Some people get the blue item, “Is your birthday in month of June?” And there are various estimators. The third technique, which I didn’t talk about, is you’re told to say yes a certain fraction of the time. And again, as a result of some kind of randomizing device.

Oops – the data are pretty good. The randomized response technique works. You get a higher percentage of people admitting to various sensitive characteristics. There’s a pretty good meta-analysis. Jon recent did a study with Allyson Holbrook, where they got totally implausible estimates. No survey uses this; it’s a pain. You know, I don’t know. And I reported this to a number of you, but I was at the ESRA Conference in Switzerland last year, and there must have been a dozen papers on the randomized response technique.

So, just for once, I’m going to say I recommend against any further research on the randomized response technique. I’m going on record here. Don’t spend any more money on this. Statisticians love it, and there are a number of statistical journals that will publish papers on new estimators, but I don’t ever expect to see it
in a real production survey. I don’t know, you may disagree, Jon, but your experience was pretty negative.

A related technique is the item count technique. In the item count technique, you get an item that has a list of – four or five is usually the things, and you say to the respondent, “Don’t tell me which of these you’ve done, just how many.” And then one list contains the sensitive item – you know, “Murdered my wife last night,” and then the other list doesn’t contain that sensitive item. But they’re the same, and then you can do the math.

And then Allyson Holbrook and Jon Krosnick did four experiments with the voting question, where they compared. And my take-away on this was the addition of item count to a self-administered survey is not that – oops – it doesn’t get you anything more than self-administration does.

Now, the item count technique is a little bit like the randomized response technique in that you don’t know, for an individual respondent, whether they did the thing or not. And so, this complicates the analysis to no end.

And then Ting and I did – Ting Yan and I did a meta-analysis of all the studies we could find on the item count technique, and it sometimes works really well, and a lot of the time it doesn’t work at all. And then sometimes it just gives, again, totally implausible results, like percentages that are negative or greater than one. These are widely recognized, in the mathematical field, as problematic values. So, I’m not keen on that either.

I just feel like we need – one of my conclusions will be is we need to invent some new methods that are actually workable, that go beyond self-administration. But I don’t think randomized response or item count is it.

Another method that’s been around for a while is what’s called the bogus pipeline. It’s some device that the respondent thinks can detect a lie. So, I did a bogus pipeline study once. I committed a bogus pipeline once. And we had the respondent hooked up to a microphone, where if the respondent talked a bunch of lights went off, and then they were hooked up to a polygraph, and none of this worked.

But we had a trick where we knew some things about the respondent, and we said, “Lie to us about your age. I’ll tell you when you’re telling me the truth.” And so, we were – the
laboratory guy’s fiddling with the controls, and the respondent gives a false age, and the person says, “No, that’s not true.” Then that person gives a second false guess, “No, that’s not true.” Then the person tells the truth, and we say, “Yes, that’s your age.”

And then everybody’s persuaded that this hunk of junk – it was all equipment left over, laying around the University of Chicago, and we persuaded them that it worked. And there were some better studies than the one I did.

This study by Bowman and Dent interviewed a bunch of school kids about whether they smoked or not, and they took breath samples. And from a breath sample, by looking at the carbon monoxide levels, you can tell whether somebody smoked recently, supposedly. No, I think this is true.

And then half the subjects were told, “We’re going to take a breath sample. We’ll be able to figure out if you smoked recently, just wanted you to know.” And what they found you get a much higher percent agreement between the survey report and the monoxide result when you forewarn people. And the results were impressive in that all the errors were in the expected direction. So, people who do smoke deny it, but people who don’t smoke don’t say that they did in the Bowman and Dent study. Okay?

And Eric Wish and one of his collaborators, [inaudible], have looked at this in the context of urinalysis and drug results. So, you warn people, “We’re going to take a urine specimen. We’ll be able to tell if you used cocaine,” and all of a sudden you get a lot more agreement between the urine test results and the reports. So, that probably works.

Okay, we had some discussion today about the relationship between having voted, or at least having voted according to Aristotle, and whether you participate in a survey. And so I actually have some results on this. This is, again, from Tourangeau, Groves, and Redline.

So, in our entire sample that we got from Aristotle, 47.6 percent were – according to Aristotle – had voted in 2004. Of the respondents, 11 percent more had voted, so this non-response bias is the impact of differential response. So, it’s about ten percent more voters in the sample of respondents than in the original sample.
And then among the ones who responded, 80 percent said they voted, and that implies a measurement bias of about 21 percent. So, the reason why I have this chart up here is I think there need to be a whole lot more studies like this one that at least have some plausibility in saying, “Gee, the worst source of error is X,” whether it’s non-response, or sampling, or measurement error.

I mean we spend a lot of money on surveys, driving down sampling error. In fact, I’d say it’s the biggest driver of costs is we have big samples so we can drive down measurement error. But at least at this high level of aggregation, looking at the whole sample – hey, you’re off by 22 points, at least in comparison with the Aristotle data because of measurement error.

Okay, so, I think this is my last slide. People worry about privacy and confidentiality, but many people are done in the presence of other people. So, this seems like a huge problem. Measurement error can be very large in surveys on sensitive topics. So, the two examples I showed you, where I felt we had good records data and could evaluate the quality of the survey reports, they both involve sensitive topics.

So, it could be that my conclusion about, “Well, the measurement error’s a big deal, we ought to worry about it more,” is restricted to surveys where you expect there to be a lot of measurement error, because you’re asking sensitive questions.

But I think we need more studies like that, where we can say, “Gee, the problem that we ought to be working is X,” whether it’s measurement error, or coverage bias, or whatever. It just seems like we’re shooting in the dark, and we don’t really know where we should put resources in terms of reducing error.

Okay, so my third conclusion is that self-administration seems to reduce reporting error, but it hardly eliminates misreporting. So, let me remind you of my slide before. So, even when the questions were self-administered, half the non-voters said they voted. Okay? So, there’s – how should I put it? – a little room for improvement there yet.

There are lots of clever methods, like randomized response and item count, but it’s not clear whether they actually add much to self-administration, and they’re rarely used in practice because they don’t give you an individual level of that variable. And you can do clever analyses, and new analyses are being invented all the
time. There’s a little cottage industry of these studies, but I just don’t see it. I don’t see a lot of surveys.

If you had one estimate you wanted to make, it might make sense to do RRT. But if you want to correlate the answer with anything else, it becomes problematic. Not impossible, you can certainly do subgroup analyses, and look at the relationships between reporting and sex, or reporting and race, but it just really complicates things.

So, we need to devise new methods for collecting accurate information on sensitive topics. And in thinking a little bit more about what I would really recommend to the National Science Foundation, would it fund, it seemed to me they feel under three headings. I don’t have a slide for this, but the three headings are causes, consequences, and fixes. Okay?

And so, under the causes heading – now, you talked about lying; I don’t see it as lying. Frauke and I have used the phrase “motivated misreporting,” but I just don’t think we really understand this process very well. And what is it that people are thinking, and what are they doing? I think it’s, at best, semiconscious. I think people are so adept at dodging embarrassing questions in everyday life, they do it unthinkingly. And so, lying seems too strong a word.

Now, in the social psych literature on lying, most lying is of this type. The sort of “white lie” kind of lie – right? So, you’re wife asks you about the roast beef that she just burned to a crisp, and she says, “How was dinner?” And most of us, who have been married as long as me anyway, know how to answer that question without giving offense. And I think some of it is that kind of conversational skill carried over into the interview setting.

But I just don’t think we know – understand very well about the process. Everybody thinks that once you say “social desirability,” you’re done, and that that’s an explanation. But I don’t see that at all.

Okay, so that’s – a second area on the causal side, that I don’t think we have a good grasp on, is you see these lists, like I showed, of what topics people regard as embarrassing, and the presumption is that’s the – those are the ones they’d lie about.

In the study that Frauke, and Stanley, and I did, we asked people, at the end of the survey, what questions they thought they’d be embarrassed by. I just don’t think we have a good handle on what
topics are really sensitive topics. Most people just rely on their intuitions about this, and they may be fairly accurate; I don’t know.

But I think this is a topic that’s worth systematic investigation because these are topics where you get 30 percent wrong answers, or 40 percent wrong answers. And I just think this is a big thing, and we ought to spend more time on it. So, those are the causes. We need to understand the process by which people modify their answers, and we need to understand the determinants of sensitivity.

On the consequences, I’d like to see more studies like the ones I was showing you, where we have a sense of what the relative magnitude of the different sources of error are. And, you know, it’s tough to do this. But as I said, we just seem to be shooting in the dark. I don’t know whether interviewer observations are solving an important problem, and I don’t think, Brady, you do either. Well, you know that they’re not solving an important problem because they’re pretty crumby. But is that something we ought to be working on or not?

And it would be nice to have a body of a hundred studies, not two or three, where we could say, “Well, these hundred studies have looked at the relative magnitudes of these different sources of error, and they suggest that we really ought to be worrying about X, at least when the topic is Y.”

Okay, I think we need to be more clever about record check studies and other forms of validating information. So, there was some questions raised about the quality of these voting data. I still think that the best situation you could be in is when you have some source of information you regard as trustworthy, and that, you know, when you use complicated models, like latent class models, which I’m guilty of on occasions, you know, it’s just not as satisfying or convincing, even to the research, I think. You know, the results just seem suggestive.

And I love the idea of planting bad items. Tim Johnson used that. I know Fred has done that. Johnny, I think you’re guilty of planting bad items. I’ve done that. I mean it’s just nice to have something that you really believe in, in order to evaluate whether the data got better or worse. So, I’m an advocate of that

And then as regards the fixes, I don’t know what the right strategies are. I do know that we have spent a lot of time over the years on things like Marlowe-Crowne and randomized response
techniques. They’re not part of the solution. I’ll say it here again, you know, we need to go in some new directions.

Okay, so that’s it for me, thank you. Gary?

*Gary Langer:* So, a quick comment and question. The comment I think may be interesting is that I’ve had some experience in which the sensitivity around questions we ask is situational and not always anticipated. And a good example, just ‘cause I’ve just survived another election, increasingly has been well discussed. We have early voting elections.

But we found out in our surveys, starting in 2008, that while people are perfectly happy to tell you who they will vote for, they’re much more reluctant to tell you who they did vote for.

*Roger Tourangeau:* That’s interesting, yeah, I believe that.

*Gary Langer:* And so what we found we were getting, we presented on this within a larger paper at AAPOR a couple of years ago, is a high level of refusal on the vote question among already voted. So, we put in place some adaptive techniques to try to encourage them, and we were very successful. We cut refusals from more than 20 down to about 5 percent on that item. So, that’s the background, which is just kind of interest about sensitivity and the difficulty sometimes in anticipating it, but the success sometimes in dealing with it.

But the question is, more broadly, rather than – there seems to be a notion that sensitivity is discovered through wrong answers, through what we predict or expect to be socially desirable, but is it also predicted or indicated simply by an item non-response?

*Roger Tourangeau:* Yeah, we’ve tried to – let’s see, I’m not sure that we’ve ever done this, but I know – I think Becca Medway, in her dissertation, did this. She looked at three or four different ways to try and get at sensitivity. And I think she ended up using item non-response as the indicator that she liked the best. So, she got some ratings data and some other data, but she ended up using item non-response.

I don’t think it’s terrifically good – here’s why. There are some items, like the one you described, that are, I think, seen as intrusive, and people say, “It’s none of your business; I refuse to answer.” Income’s like that.
But then there are other items you say, “Well, have you used cocaine during the last month?”

They’ll say, “Well, I refuse to answer that.” This is like tantamount to saying, “Sure. In fact, I’m using cocaine right now.” It’s like the Brady subjects who were having sex while the interviewer was there – right? – and making the interviewer’s observational task a lot easier.

You know, so I think any of these indicators are flawed. But I think the flaw with item non-response is there’s some items where people – they’re very sensitive, but people don’t want to say that they refused because they’re afraid they’d be revealing something about themselves by refusing.

Let’s see, Frauke?

_Frauke Kreuter:_ I just wanted to add to the first comment that you made and what Roger mentioned that we did ask for the perceived sensitivity for these questions. I mean you’re right; it’s not just situational dependent, but also how you ask. Because we did see, just like we saw the differences in over- or underreporting across the modes, there were strong differences in what was perceived and how sensitive it was perceived across the mode. So, that, too, is not stable.

So, you know, across the surveys, or in particular in [inaudible] mixed mode surveys, you will – it’s not – it’s not – I mean my impression there was it’s not the same level of sensitivity, and then you decide to reveal it, but it just doesn’t feel that sensitive in that setting.

_Roger Tourangeau:_ Yeah. If you’re doing a Web survey, for example, it’s just a lot less embarrassing to admit that you had an F to an impersonal computer – right? – I think.

_Gordon Willis:_ Roger, do you think that there’s the possibility, with so-called sensitive questions sometimes, ones that are grossly underreported, that the key problem is not one of sensitivity and the sense of embarrassment, whatever, but it involves the nature of the language we sometimes use in sensitive questions that can just plain be misunderstood? And as an example of that, I did some cognitive testing on a simple question, “How many sex partners have you had,” and you find out there’s a disconnect.
You know, the person says like, “One,” and then you go through how many people they’ve had sex with, and it’s like six or eight, whatever. And then you ask them to resolve it, and he says, “Oh, I have one partner, the person I live with. Now I’m telling you how many people I had sex with.”

So, the imposition that we use of the term “partner” is misleading. We should just be asking perhaps, “Well, how many people,” if that’s what we mean, “have you had sex with?” You know?

Roger Tourangeau: Yeah, in my own work on this, I’ve always asked men how many women they’ve had sex with, and men how many women they’ve had sex with. So, I never use the word “partner.”

What’s amazing about these items to me, some of these sensitive items, and I can point to four or five studies like this, is the direction – the errors are all in the same direction. And as I said, nobody underreports voting. You know? At least in our study. We don’t find any evidence that anybody forgot that they voted. You know? Zero. You know, the voters all say that they voted.

It’s the non-voters that are problematic. And likewise in the Bowman and Dent study, the kids who didn’t smoke, according to the carbon monoxide test, they don’t have any trouble answering the question, and there, again, 100 percent right. Right? And I could point to three or four other examples where, you know, at least there’s plausible records data, and all the errors are always in the same direction.

I mean, this is why I don’t much put much stock in the Belli work on people forget whether they voted. Why is it that only the voters – the non-voters forget that they didn’t vote, whereas the voters don’t forget that they did vote? And it just isn’t plausible. I don’t know of any memory phenomenon that’s like that.

And so, likewise, with the misunderstandings, why would the men misunderstand the definition of a sex partner in one direction, but women misunderstand the meaning in the opposite direction? Because what happens is men report a lot more partners than women, that’s the problem in that particular domain.

You know, if it’s comprehension, why wouldn’t you see the same problem in both genders? I mean these are items where there’s a lot of measurement error. It doesn’t mean that they’re understood perfectly, or that there isn’t any forgetting or anything else, but you
see massive errors in one direction and no errors in the other direction, this says something to me.

\textit{Gordon Willis:} The finding about trying to determine early voters, that’s unfortunate from getting, you know, good estimates of that. But it’s exactly what we should have expected. When you’re asked about your vote intent, you’re asking about an opinion, and there is a presumption that you can express opinions.

But voting is a secret act. You vote in private. No one is supposed to know who you voted – and so, it’s a fundamentally different thing that you’re asking about, and you really have to work to overcome that cultural situation there.

In terms of income, yeah, you’re right about some questions – that cocaine question not giving it. But income, you get more no answers of income than all other demographics added together, which clearly shows there’s a very special reluctance to give income. So, sometimes it is a good indicator of finding something sensitive that you don’t want to share with others.

And then the last thing I’ll add is I’ll just add a little bit to your description of the Chicago bogus pipeline study, which was about sex partners, which fits in with your last comment. That incredibly impressive bunch of electronic junk we told the people were a verifasiter – okay, we invented a new word – it was a verifasiter which could tell them – and the people were overwhelmingly convinced that it, of course, could tell – it could detect whether they were telling the truth or now.

\textit{Tom Smith:} Just a couple of quick comments. Actually, to be fair, the surveys don’t actually ask people to predict how they’re going to vote. It’s a purely hypothetical, “If the election were held today, who would you vote for?” So, it’s even less real and confidential in that sense. So, that seems like a small thing.

Second thing is that I’m actually increasingly convinced by work we’ve done recently that I think a lot of the reason that people decline to answer their income is that they don’t know the answer, and that the question is asked badly, and that if we were to ask it better – so, I would caution us on assuming that all of the non-response, even if it’s coded as a refusal, is due to sensitivity.

\textit{Gordon Willis:} Just on income non-response, the current fashion for the last 15 years has been to follow a “don’t know,” or “refuse” with bracket questions. And that gives you, I think, some indication of, you
know, if they’re not entirely willing to tell you that they’re poor or rich and may sort of point to the idea that, at least in many cases, the original non-response was a “don’t know.”

*Gary Langer:* Gary Langer, if I could just add something that this may be useful just on a practical scale, what our concept was in dealing with sensitive questions is that people don’t mind, perhaps, telling you the answer, but they just don’t want to speak the word.

So, some time ago, we did a survey in which we wanted to measure sexual orientation. And then rather than ask people to tell us if they were heterosexual, homosexual, or bisexual, we asked them to tell us the number. “Are you – just say the number. Are you Number 1, heterosexual; Number 2, bisexual; Number 3 –”

That’s the exact same adaptive technique we used in attempting to obtain having voted results – not, “Tell me Obama, tell me Romney.” “Just tell me the number, “Number 1, Obama; Number 2, Romney.” And again, item of non-response in that question went from 20 – refusal went from 20 percent to 5 percent –

*Roger Tourangeau:* That’s neat.

*Gary Langer:* – simply by doing that. So, there’s way to work it, I guess, is the –

*Roger Tourangeau:* You know, one time I watched Census Bureau interviewers on a pretest for the National Health Interview survey, and they had the show card. So, the respondent didn’t have to say their income. And, you know, it went from A to ZZ or something.

And then afterwards, we debriefed with some interviews, and we were in the bar, and this one interviewer walked in and said, “I was in the nicest neighborhood. All the houses were – all the households were type ZZ households.” I mean it’s not like you’re not revealing the information; that’s amazing. But I think it is effective. I think it probably helps to reduce item non-response in that case.

All right.

*Jon Krosnick:* Well, fabulous. Another terrific day. Thanks to all the presenters, and congratulations to all whose work was cited today, Tim especially.