MCHB/EPI Atlanta GA Conference

December 10 - 14, 2005

Sample Size Issues and Determination in Maternal and Child Health (Part I)

BRIAN MORROW: Okay. My portion of the talk is very non-technical, there's not going to be any equations, and there's hardly going to be any numbers at all until maybe the very end. But looking at the rest of the agenda, this probably, you guys, is the last chance for that. So I was just going to talk about different kinds of sampling, advantages, sampling errors and non-sampling errors, and also factors that affect the sampling error that you can expect.

Juan actually did a great job of describing advantages of sampling. I mean, what we want is particular information about a particular population. Populations, of course, can be really huge whether--where are we at, 300 million people in this country—if we want to do a national survey. So what we do instead, we sample. We take a smaller amount in order to save time, cost, and resources. One of the advantages of a sample done correctly is, once you have the data from the sample, you've got a pretty idea of how reliable that data from the sample is, how accurately you can expect it to represent the population that we care about.

Some non-sampling errors, this course is of course focused on sampling, but some things that even if the sample is done perfectly, some things that you're still going to have to worry about are, one, non-response bias. If you sample a certain

amount of people for most applications, you can't force people to answer your questions or to fill out your questionnaire. I mean, most surveys will use analysis weights to try and adjust for the fact that not everybody is responding. Not everybody who is sampled is responding. But we are pretty sure, in most cases, people who choose not to answer a questionnaire are not exactly the same as the people who do choose. So, the best way to mitigate a non-response bias is to just use a response techniques that will make the non-response percentage as small as possible. I mean, the more people that respond, the less you have to worry about non-response bias.

Another type of non-sampling error is non-coverage bias. And probably the most prominent kind of non-coverage bias is something that just--there was a big article just last week about how, with the advent of--advent and prevalence of cell phones, surveys that depend on telephone mode only, now we've got households that only have cell phones are not being adequately covered. There are a lot of times they're not included in the sampling frame. And so, if you're trying to make inferences about a total population and households or people that have--only have cell phones or not even have a chance to be sampled, well, that can contribute to a significant error in the conclusions you've--that you're--that we're making about our population.

Another kind of non-sampling error is recall bias. That's where if you're asking a person questions after the fact about some kind of previous behavior, they may not remember what you're asking about in sufficient accuracy.

Another type of non-sampling error is mode bias. And by mode, I mean how you're getting the information from a person. Different modes can be a face-to-face interview versus a paper questionnaire that the person can fill out on his or her own time versus a phone interview. It's been proven that's--especially for certain sensitive questions, you will tend to get different answers to these sensitive questions depending on the mode of the collection of information. So all these things are--even if you do the sample completely, perfectly, these are still things that can affect how accurate your final information is.

Factors affecting sampling error size. Well, the last one there, the third one, sample size, that's what this course is about, so I won't get into that anymore during this talk. But there's other things that also affect the sampling error size. By sampling error size, most of you probably are used to measuring that by, like, the confidence interval around an estimate. You know, how reliable, how accurate can we expect our information from the sample to be in describing the population. For a fixed sample size and for a fixed population variability, the sample design will affect the size of the sampling error. I'll be getting into different kinds of sampling design, but these are things like random sampling versus

stratified sampling versus cluster sampling, and all the different combinations thereof.

For a fixed sample size and a fixed sample design, the population variability will also affect the sampling error. For stratified or cluster sampling, the higher the population variability, meaning, if you're trying to measure a certain behavior, if that behavior you're trying to measure is very different in different parts of the population, your sampling error will be larger than if the behavior you’re trying to measure is very similar in all the different parts of the population.

We'll go over some of the very common probability sampling methods. You know, simple random sampling, which I think we all know what that is, systematic sampling which is just a twist on random sampling, and then more common kinds of sampling that we use for human population such as stratified sampling and cluster sampling, and then complex sample design is just one or more of these different kind of combinations.

Simple random sampling is the kind of sample we all learn about first. It’s where you've got a population, you know what that population is, you determine your sample size, and then you employ a technique that ensures that everybody in the sample has an equal chance to be--everybody in the population, rather, has an equal chance to be sampled, and any particular combination of people you pick has the exact same chance of being selected as any other populations. So,

maybe if your population has 100 people and you're sampling five people out of that 100. Persons one, two, three, four and five have the same chance of being sampled as persons 96, 97, 98, 99 and 100. Same chance as persons 20, 40, 60, 80 and 100. I mean, any set of sample has the exact same chance in a simple random sample of being sampled from the population. In order to do it, you construct a sampling frame or some kind of list of everybody in the population. Then from that sampling frame, from your list of everybody in the population, you some kind of random process to select your predetermined sample size. And by now, most of us will use some kind of computer program or a computer procedure to do that. And again, the definition is any sample is equally likely to be selected as any other. Some of the advantages of simple random sampling is--well, it's called simple random sampling. It's just very simple to understand, it's simple to do, and it's simple to analyze. It does provide a baseline for which more complicated or more complex types of samples can be compared as far as the confidence interval and standard error. Another advantage is it's just very easy to analyze. I mean, some things you can do just on your hand calculator, if you want, but you can use any normal data analysis software without having to describe to the software how the sample was taken.

Disadvantages of simple random sampling, you must have your list of your entire population. You can't take a simple random sample until you know the people that have an equal chance of being sampled. You have to have that list or you have to go out and build that list. If you're taking a sample over, say, the entire

United States, you'll have, most likely, people who were sampled will be fairly, evenly dispersed all over the country. And if you're depending on, say, face-to-face surveys for your mode of collection, you'll have to go all over the country to all these different places and that can drive up your cost, of course, very, very quickly. And another disadvantage is that simply by random chance, simply by statistical chance, certain subgroups of your population may be over or under sampled. In fact, certain parts of the population may not even be sampled at all just by statistical bad luck. One way to get around some of the disadvantages of simple random sampling is to use a systematic system of sampling. And the way we do this, we still need our list of the population, but this time we sort it by characteristics that we want to be sure are equally represented in the sample or at least proportionally represented. Once the list is sorted and we've determined our sample size, we determine a--our sampling interval. Say, that we know that we want to sample one out every 100 in the population, so our sampling interval would be one out of 100 and we pick a random start somewhere between one and 100. Say, we might pick randomly the number 73. In that list, we'll sample the 73rd person then we'll sample the 173rd person in the list, the 273rd person and et cetera. We just keep sampling every 100th person. Because we use a random start, everybody in the sample has an equal chance of being selected. Because we've sorted the population by some kind of characteristic, there is not--we have a much, much better chance--well, in fact, it's guaranteed that the population characteristics will be more evenly represented in the sample we get.

It's a guarantee where in the simple random sampling that may not necessarily happen.

Other advantages of systematic sampling include--sometimes you can do it whereas simple random sample is not possible. Say, if you're getting records--set of records--set of electronic records every month and you want to take a sample every month, well, for simple random sampling, you can't really do that. To do a true simple random sampling, you'd have to wait until the entire time period was up in order to guarantee--in order to meet the parameters of simple random sampling. But for systematic, if you know that you're going to be sampling approximately one out of every 10 or one out of every 50 people, you can do that with each month's set of records, and still everybody has the same chance of being sampled.

I talked about the final sample you get will be more likely to be uniformly spread over the population. And another nice thing about systematic sampling is even though it is a little bit different technique than simple random sampling, in almost all cases, you can use the same analysis technique, just basic data analysis software.

Disadvantage of systematic sampling. The first two disadvantages listed are exactly the same as simple random sampling. You've got to have a list of everybody that you're going to sample before you can even start. And once you

get your sample, if you're over a large geographic area, you know there's the--pretty much guarantee that the cost can go way up. A disadvantage of systematic sampling that's not present in simple random sampling, you can, by bad luck or carelessness, kind of really bias your sample if the records are sorted in a cyclical fashion, meaning, they cycle, and if your sampling interval happens to match up with the length of that cycle. I mean, an example would be if your population--if your list was, like, days of the week, Monday, Tuesday, Wednesday through Sunday, and then the next week, starting with Sunday, Monday, Tuesday. And if you’re sampling interval happened to be seven, then you’re sampling the same day of the week every time, and you get a bias sample. But that’s something that’s usually pretty easy to check for to make sure that that’s avoided.

Stratified sampling starts to get into different types of techniques. For stratified sampling, the first step is to divide the population into mutually exclusive segments, or strata. We want these segments to be exhausted. You don’t want anybody to fall through the cracks. You want the strata, when you add them all up, to equal your population. The nice thing about strata is you can really choose your strata however you want. You can do it geographically, you can do it by outcome of interest, you can do it by size, you can do it--because you’re really fairly free to choose how you want to define the strata. Once you have the strata defined, then you go ahead and take a sample from each stratum. Two of the more common techniques that you’ll see will--one is equal allocation, whereas,

no matter how many of your population are in each strata, you sample the same number from each stratum. So, for instance, you may have two strata. One has a million people in it, and the other has only a 100,000 in it, but you choose to sample 400 from each strata. The other common allocation technique in stratified sampling is proportional allocation where each strata is sampled with the same sampling fraction. You don't have to restrict yourself to, kind of--to normal or even similar allocations within each strata. You might have a sample where you have four strata, and one of the strata you sampled everybody; it's not even really a sample, you just take a census. In the second strata, you may take a simple random sample within that strata. In the third strata, you may have--you might break it up into clusters and do a single-stage cluster design. And in the fourth strata, you may have a really complicated cluster, like a three-stage cluster design. And that you can do that all within one sample design. In stratified sampling, you get the best results when your strata are related to the outcome of interest; what you're trying to study. That's not always possible, but that is a goal when you're trying to set up a stratified sample. Yes?

UNKNOWN SPEAKER: Sometimes stratified sampling is used when you want to be able to get estimates within each group, and sometimes it's used to make sure that you--it's not used to make sure you have a big enough population for a sample within each to get estimates, but it's just used to make sure you have a mix. Whereas the second kind of describing is that more like proportion or is that more like--

BRIAN MORROW: Oh, both kinds are types of stratified samples. And if your goal of your sample was either of those stratified sampling with sample--with--would satisfy each of those goals. So, you're right. It can't have more than one type of purpose.

BETH: (Inaudible).

BRIAN MORROW: Yes?

BETH: I'm a bit confused between the first two, the random and the systematic sampling.

BRIAN MORROW: Okay.

BETH: You said something about sorting by characteristics. Sorting by characteristics is not my stratified (inaudible). I mean, why do we sort by characteristics in a systematic sampling? What does it add, to me, compared to the random—the simple random sampling?

BRIAN MORROW: Okay. Say, you were kind of interested in--well, I'll just use like a north-south breakout. If you had a small sample, say, your population was only 100 people and you had 50 of those are in the north and 50 of those are in

the south, and you did a random sample of 10. Well, it's possible, in a simple random sample, that you might get all 10 people from the south, or all 10 people from the north. Systematic random sampling says you would sort your populations so that all 50 people in the north were in the top of the list, all 50 people in the south were in the bottom of the list, and then you take one every 10th person so you'd be guaranteed to get five people from the north and five people from the south.

BETH: In a stratified sample, you divide them completely.

BRIAN MORROW: Yeah. Stratified sampling is much more complicated because once you have strata you can treat the two strata completely differently. In systematic, you're sampling every person in the north with the same sampling fraction as every person in the south. Stratified sampling removes that restriction.

Advantage of stratified sampling: you can get--as we mentioned earlier, you can get increase precision, smaller, standard error, smaller confidence intervals if you set up your strata to correspond to your outcome of interest. What's your name?

BETH: Beth.

BRIAN MORROW: Okay. As Beth mentioned, you can control the sample size of each stratum. An example of that would be if you're taking a population of recent

births, and you're really interested in low birth weight infants, Well, if you stratify normal birth weight versus low birth weight, you can set the sampling fraction within the low birth weight stratums that you get a sufficient number of low birth weight babies to analyze. For logistical or administrative purposes, it may just be easier to set up a stratified sample. Maybe if you, during a national population, and you set your strata for states, and each state has a different way of measuring or collecting data, this way, with a strata, you can have each state do their own sample and not have to take all the data from the state yourself and manipulate it so that it's all in the same format.

Disadvantages of stratified sampling. Well, the first two, again, are the same as we said before. You still have to have the complete population. Sometimes that's available for--but for a lot of human studies, it's just not available. And again, since you're sampling over everybody, the units may be very far spread out geographically. Other disadvantages are if you choose equal allocation, you just give up that advantage I measured earlier--I mentioned earlier about getting reduced precision. That reduced precision is only if you use proportional allocation. Similarly, if you decide on proportional allocation, then you give up the advantage of being able to choose the--a separate sample size for each stratum.

Cluster sampling is very different from the kinds of samplings that we've just discussed. All the other samplings that we discussed up to this point--we may have sorted the data first or we may have stratified the data first, but by the time

we actually got to the actual sample, choosing people, we were choosing individual units. We are choosing the people that we wanted to--that we were choosing our final units. So, our very first stage of sampling or the primary sampling unit was the person him or herself. For cluster sampling, that's not that way. For cluster sampling, we divide the population into mutually exclusive segments, just like we did for stratified sampling. And we're going to call those clusters. But instead of going to each cluster and taking the sample that we felt was most appropriate of people, we sample the clusters themselves. So our very first sample is of clusters, so the primary sampling unit--or you may have heard--abbreviated--it's PSU, is the cluster and not the person. And then once we've sampled each cluster, then we go to those clusters--only those clusters that were sampled, and we take a sub-sample of people or smaller clusters within the original clusters. And as opposed to stratified sample where the best results are when the stratum are related to the outcome of interests, whereas the best results in cluster sampling are if somehow you can define the clusters to be as similar as possible. So, it really is quite different because we're choosing—we're sampling clusters rather than people at first. Yes?

UNKNOWN SPEAKER: As you're walking through these, do you have some examples of national surveys that might be representative?

BRIAN MORROW: Yes, I do.

UNKNOWN SPEAKER: Okay. ‘Cause that might be helpful to get a picture (inaudible) seeing it but then I think, then how—where else have you seen it in practice?

BRIAN MORROW: Okay. That’s a great question because that’s next, so. Did you have a question?

UNKNOWN SPEAKER: You mentioned that you want the clusters to be as similar as possible. You don’t mean similar to each other, you mean within the cluster, you want the—

BRIAN MORROW: Similar to each others. I meant similar to each other. Yeah.

UNKNOWN SPEAKER: Okay.

BRIAN MORROW: Yeah. A lot of times, that’s not possible but that’s how you would get the best precision, the smallest confidence intervals. Okay. Finally, we don’t need a list of the entire population. I mean, a lot of times, for your population of interest, that it just doesn’t exist or it’s just so expensive to research and compile that they would use up your whole budget just dividing the sampling frame. For instance, if your clusters are states, you just need the list of the 50 states for your first stage of sampling. You don’t need all 300 million people in the country. You only need the complete list of people for the next to last stage. For

instance, if your scheme was selecting states and then within states selecting counties and then within counties selecting townships, you would only need the detailed sampling frame or the detailed list of the population within each of the townships that were sampled. I mean, it's just so much less work and cost and expense.

UNKNOWN SPEAKER: (Inaudible) geographical clusters where you can do clusters by using other properties, and then you need the list of population?

BRIAN MORROW: Clusters, as opposed to strata, are, in my experience, are almost always geographically-based.

UNKNOWN SPEAKER: Yeah, because I can't figure out how you're actually doing your samples or how you actually administer the instrument you want to administer in terms of—I can't figure how to do a cluster analysis unless in a geographically (inaudible).

BRIAN MORROW: The ones of the examples I'm going to present are going do that and the clusters are going to be geographically-based. I think the only other kind of cluster that I've even heard about is a time-based cluster, but—

UNKNOWN SPEAKER: We used days and we also really (inaudible) but we see—we identified clusters of days and then the—and if we—it's like all of those (inaudible).

BRIAN MORROW: Okay. So, time-based—

UNKNOWN SPEAKER: But the cluster was obtained.

BRIAN MORROW: Okay. Okay.

UNKNOWN SPEAKER: But I think that by definition, cluster is anything that conglomerates the subject in any form while a strata is environment-defined level, so that will tell us the main difference. So, you can cluster people by day of the week, you can cluster them by population size, you can cluster them geographically, you can cluster them—but if you use one variable in order to divide the population then you stratify them. And I believe that that would be the main difference.

UNKNOWN SPEAKER: And so, cluster became a cluster then.

UNKNOWN SPEAKER: Yeah. Schools and (inaudible).

UNKNOWN SPEAKER: Yeah.

UNKNOWN SPEAKER: (Inaudible) geography is.

UNKNOWN SPEAKER: Sure. Sure.

UNKNOWN SPEAKER: Yeah. The (inaudible) that we in MCH usually cluster people using geographic boundaries, but you don't need to do so. For instance, we use cluster sampling in a—after a Hurricane George survey of Dominican Republic, and the cluster was a list from the American Red Cross, and they had to be 35,000 families so we just cluster them—it was not geographically fine. It was just clusters of population. Are we going to use any strata to define levels of the people but—

UNKNOWN SPEAKER: I think that's the key, I mean, what Brian was saying, what is the difference between a strata and a cluster, who is homogenous and what is heterogeneous. In this strata, the people has a characteristic that is homogenous within that entire strata. So, we are talking about this supposed ethnicity that is a characteristic that we'll classify them as homogenous within that strata, so the people we got is African American. They would have a characteristic that is homogenous across that strata, right? So, we are talking about (inaudible). So, there is a characteristic that is homogenous within them, okay? Now, when we are talking about cluster, we are talking everybody is heterogeneous. So, then suppose that I go to my neighborhood where I live and I

take the block that I live, and there is people of all ages, all races, all (inaudible) kind of status, everything is a mix. So, it's heterogeneous within that block. So, that is a cluster because it's heterogeneous.

UNKNOWN SPEAKER: But you could also have a strata that's geographically--you could--

UNKNOWN SPEAKER: Right.

UNKNOWN SPEAKER: --use the four quadrants of the country and each one's a strata and then--

BRIAN MORROW: True.

UNKNOWN SPEAKER: Yeah. I think the best thing to think about strata and cluster is that strata do have a representation of all strata in the sample. So, you see that a population and you divide your population in different sub-population idea. In cluster, you have—you do some random systems, I mean, at random, and you select cluster of, let's say, a chunk of people or elements. You select at random a set of individuals. So, some cluster represent all (inaudible). So, maybe it's the best way to say when you have a cluster and you have a strata. Strata always have representation in the sample; some cluster don't have. So, the ones that you do have represent the other that don't--was selected in the sample.

BRIAN MORROW: Yes.

UNKNOWN SPEAKER: So, if you randomly select your clusters and whatnot once within your cluster, do you still have to randomly or systematically select the individuals? Or what about the convenience of individuals within--

BRIAN MORROW: You can take a census, you can do random sampling, you can take a sample of clusters within--a sample of smaller clusters within the cluster you've just sampled. You're not limited. Yeah. Eventually, you get down to the person level, but you can do that in one stage or three or four stages.

UNKNOWN SPEAKER: Well, because there was an argument for one research project that I read about where it was people going into a particular venue. And initially, they were doing every seventh person but the response rates were really low because people didn't want to be separated from who they came. So, then they changed the methodology to you could come and your buddy can take the survey too and you guys can, like, take it in in separate locations but still kind of be together. And the question was whether or not that affected the accuracy or if that was a lot, it'd still be considered like a random, like, methodology overall.

BRIAN MORROW: That sounds, like, more like cluster, because you have--people that are chosen because they're contiguous to each other. Another

advantage of cluster sampling which is--because you're only sampling out of certain clusters, everybody is not closer together but more centrally located. You only have to send your resources if you're doing face-to-face interviews. You don't have to go all over the country; you only have to go to the certain clusters that you've selected, which can be a huge, huge savings in cost. Disadvantages. Almost always, cluster sampling will be--will have less precision than any other kind of sampling. Less precision, I mean, wider (inaudible) intervals. That's not always true but it's--normally, that's true. And another disadvantage is that you need very specialized statistical software or procedures to properly compute the variances in (inaudible). You can't whip out (inaudible) to get the right confidence intervals for a cluster sample.

Complex sample design, you hear that term all the time. A complex sample is any sample that's not using strictly simple random sampling or—and I guess you could use systematic random sampling but anything with a strata or a cluster or both or anything that's not simple random sample-based is a complex sample. And you need the special analysis software in order to properly compute standard errors. Within CDC, you can use SUDAAN, STATA. If you want to use SAS, you use--you have to use the new diversion nine complex SAS procedures like survey (inaudible) or survey freak. You can also do it in SPSS if you use the special--again, they're pretty new to SPSS, the special complex software procedures. The design effect is just the ratio of the variance that you get using the complex sample to the variance you would get if that same data were taken

with a simple random sample. And the design effect depends upon the type of sample design, the variability, the population, which we discussed earlier. I can't speak for the others but I know SUDAAN will spit out the design effect just as you ask it to.

All right, examples. Real examples of CDC surveys that use different types of sampling. I'll go over these three, because I've worked with all these three. National hospital discharge survey. This survey goes back at least into the '60s, back when hospitals kept all their discharge records on paper. So, if you wanted a national sample, you couldn't really randomly sample the hospitals because you'd have to send somebody out to every hospital that was selected and it just--and it's just--it would just be too expensive. Even now, I don't know if all hospitals are required to send their electronic records to one place. So, because of this simple random sampling doesn't make sense, they're going to do cluster sampling. What they did first was they stratified every--all the hospitals into three groups. Large hospitals are their own strata and they get sampled with certainty. Meaning, they're automatically selected. It's really not a sample at all. So, for large hospitals are selected with certainty. Once that large hospital is selected, then discharges are selected randomly within each hospital. A second strata is medium-sized clusters. They call them PSU because they're for primary sampling unit. Again, and the PSU definition is actually taken from another CDC survey. I think it's was the National Health Interview Survey but it's geographically-based. If there is a lot of hospitals in a particular PSU, that makes

up a strata of PSUs. And those PSUs were sampled with certainty. Once we have that PSU, our first stage was selecting a hospital and then the second stage was selecting the discharges from that hospital. The third strata is PSUs, geographic units with much--with many fewer hospital in them. Instead of sampling every PSU like we did in that second strata, we actually do a three-stage design where the PSUs are sampled first so not every cluster is sampled, just certain ones. Once the PSU is sampled, then we select the hospitals at the second stage and then discharges as the third stage. So, that's the most complicated survey I've ever worked with because there's the different strata--all have different sampling plans within them.

The Youth Risk Behavior Surveillance System. The goal of that is to ask high school kids just all kinds of questions about different behaviors and attitudes. And because it's a nationwide survey, again, you can't--it was just impossible--not logistically possible to go out to every--to take a random sample and then go out to all these different schools. So instead, it uses a three-stage cluster design. First of all, the strata were defined based on geography and by MSA, Metropolitan Statistical Unit. So, urban schools were stratified together and rural schools were stratified together. Also, the strata were defined by the percentage of Black and Hispanic students. The three stages were the count--the PSU, so a sample of the PSUs was taken. Then once the PSU was selected, schools were sampled within those PSUs. And then once the schools were selected, a class was randomly selected within that school. Now, once the class was selected,

there is--then they went to every student in that class. So, that's not really a four-stage, this is not a sample. Once the class was selected, everybody in that class was asked to do the survey. They ensured that they had an adequate number of Black and Hispanics by--the PSUs in the schools were--had a higher selection, had a higher probability selection for Black and--for PSUs and schools that had a higher Black and Hispanic population. And also, two classes instead of one were selected for schools that had a high Black and Hispanic population. That way, they ensure that they had adequate sample size for those populations.

Third, CDC survey is the Pregnancy Risk Assessment Monitoring System called PRAMS. Here, the population is moms who have given birth. Well, since we have electronic birth certificates and since the mode of collection was not face to face but mail and phone, there was not--clustering for this national survey is not needed. We don't have to send people out to do face-to-face interviews, and we have a readily available list of the population. Again, you don't really want to do clustering unless you have to. It's got a lot of disadvantages. So, for PRMAS, there is no clustering done but PRAMS--each state defines their own stratification categories. A lot of states stratify on birth weight, low birth weight versus normal. A lot of states stratify on different race ethnicity categories. Some states do geographic regions within their state, either in a contiguous areas or urban versus rural.

Okay. Well, I'm way over. The last couple of slides are just showing numerical

examples of how big the design effects can actually be, depending on how variable your population is and your simple random sampling versus cluster sampling. And there's really no equations to do this that I know of. I just use simulations. I simulated different populations and then took different kinds of samples over and over and over and over and looked at the standard errors and confidence limits. And I'll just--the informations in your slides but for a simple random sample, you can get a standard error. For a given sample size, you get a standard error of 2.5 percent. And then, if you use cluster sampling of this last population, which is where the clusters are very, very different from each other, and you use a cluster sampling where you're only choosing a very few clusters with a lot of people per cluster, I mean, your standard error can just explode. I mean, the confidence interval for this is you basically multiply that by two so you'd have a point estimate of 50 percent with a confidence interval of plus or minus 40 percent, which is--I mean, it's almost worthless. It tells you nothing about the population that you're trying to infer. This is just an example of how big the design effect can be depending upon how different the population clusters are and what kind of sampling you use.