

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

Please note: The transcript for this activity is based on the actual webinar recording. Minimal editorial/formatting changes have been made to the transcript text.

SUSAN JACKSON: Good afternoon and welcome to the Strong Start Webinar. This is Susan Jackson. I'm going to be moderating this call. Thank you again, everyone for joining. My name is Susan Jackson, and I'm part of the Patient Care Model Group here at CMS Innovation Center. We're really excited to have everyone joining us today for this webinar, entitled "Potential Data Sources for Generating Baseline Data." I would like to let everyone know that all lines have been muted, and we will take questions through the online CAT function after our presenters have finished their presentation.

Just a few more housekeeping items. This call is being recorded and will be posted on the Innovation Center's website within a week. Also the audio and the slides of today's call will be posted within a week on our website. This is a call for potential Strong Start applicants. This is not for the press. If you are a member of the press, please call the CMS Media Relations Group if you have questions.

The purpose of this webinar is to provide expert guidance by entities with limited experience identifying baseline data for comparison to an intervention population where care has changed. The Strong Start application requires all applicants to state their commitment and demonstrate their ability to provide gestational age and birth weight for infants of the mothers participating in the applicant's proposed intervention and a baseline period that spans at least two years prior to the start of the Strong Start intervention.

The expectation is that the two years of baseline data will represent the applicant's own Medicaid and/or CHIP population for the two years prior to the start of the intervention. CMS views an alternative baseline data source as inaccessible only when an applicant does not have two years of their own baseline data on the population that they will serve under the Strong Start initiative, because the applicant is a startup, has been in operation for less than two years, or already provides intervention services but to a non-Medicaid and/or non-CHIP population. Under these special circumstances, an applicant may propose an alternative data source that may include aggregate information. The individual level data will be considered higher quality than aggregate data. The applicant must demonstrate that these alternative data represent a population that is sociodemographically similar to the population that will be served under the Strong Start initiative. FEMA expects applicants to propose the best quality baseline available to that applicant. Upon award, CMS will work with awardees to collect an appropriate baseline for the awardee-specific circumstances.

For this subset of potential applicants who have special circumstances and limited experience identifying baseline data for comparison to the intervention population, where care is changed, this presentation will provide an overview of approaches identifying a baseline population and potential sources for baseline information when the prior two years of baseline data from a patient's own past experience are not available or are not appropriate. In addition to data that would be incorporated in an application, the data sources and approaches that will be covered in this webinar may also be helpful to potential

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

applicants as part of an applicant's needs assessment as the applicant develops their intervention design and targeting strategy.

CMS believes that each source and approach to identifying a baseline have both strengths and weaknesses and typically stronger designs have historical data from the same population. Applicants should be mindful that baseline data from national sources may make it more challenging to identify a population similar to the population that will be served by Strong Start. This presentation is not intended to endorse any particular source or approach to identifying baseline data specific to the Strong Start funding opportunity at a particular prenatal site. We recognize that potential applicant's circumstances vary and appropriate available sources of baseline data for each applicant will also vary.

Following the presentation of this webinar of potential sources for generating baseline data, there will be time for questions and answers with our presenters on the information that has been presented. We will not be responding to Strong Start program requirement-related questions during the question and answer session. We encourage individuals and entities with Strong Start specific questions to send their questions to the Strong Start inbox: strongstart@cms.hhs.gov. We will answer these questions as quickly as possible. Again, if you have Strong Start specific questions, or if we are not able to get to your question on this webinar, you can always email us at strongstart@cms.hhs.gov. In addition, new frequently asked questions and more information about the Strong Start initiative is available at our website at innovation.cms.gov under the What We're Doing section at the top of the website.

Our presenters for this webinar, "Potential Data Sources for Generating Baseline Data," are Dr. Julie C. Jacobson Vann and Dr. Thomas L. Schlenker. Dr. Jacobson Vann has a Ph.D. in public health from the University of North Carolina, and is a licensed registered nurse in the state of North Carolina. She is currently a senior researcher at the American Institutes for Research and serves as a clinical assistant professor and healthcare systems coordinator at the University of North Carolina School of Nursing from 2007 to 2011. Dr. Schlenker is an M.D., MPH and Director of Health for the San Antonio Metropolitan Health District. Prior to his appointment with the city of San Antonio in Texas, Dr. Schlenker was a Public Health Director for Madison and Dane County in Wisconsin from 2006 to 2011. I would now like to turn this webinar over to our presenter, Dr. Jacobson Vann.

JULIE JACOBSON VANN: Thank you, Susan. Hi, everyone and thanks for joining us. Dr. Schlenker and I are pleased to talk with you today. We'll each be leading parts of the presentation, and we may interject comments in each other's presentations throughout.

Today we're going to talk about what baseline data are, the purpose of collecting baseline data, some of the characteristics of baseline that may make it more or less useful for your specific purpose or a specific purpose, and we'll also review a number of potential sources for obtaining baseline data, and we'll discuss some of the features of these data sources and how you might obtain them.

CMMI Strong Start: Potential Data Sources for Generating Baseline Data

July 25th, 2012

3:00pm-4:15pm (EST)

So what is baseline data? Well, baseline data are the information you might collect before a program, service or intervention begins. Generally we think of baseline data as being the outcome or outcomes that your program or intervention intends to change. And outcomes might be things such as cost, health status, health behaviors, knowledge, or even biomarkers. One example might be birth weight, which is measured in grams or pounds and ounces. Baseline data may also include characteristics about groups of people that are being served by an organization or program. For example, health services scientists often collect basic demographic information about a population of interest, such as age, race and ethnicity, sex, and educational level.

Baseline data may be presented in a number of ways. One really simple way to present baseline data are frequencies or counts. For example, you may want to count the number of teens in a school who have never smoked cigarettes. That would be a very simple frequency measure. A more informative way to present baseline data may be with the use of percentages, proportions, prevalences or rates. These types of measures use a denominator, which makes them relatively easy to compare between different groups. And one example is the percentage of people on this webinar who get at least 30 minutes of intense physical activity per day. So is everyone raising their hand? I am.

Another type of measure might be an average or mean value. For example, you might want to collect the average number of pounds of pregnancy weight gain for women your program serves. We often stratify these measures by certain characteristics, and by stratifying, I mean that we divide a population into levels or categories, such as different age groups. And then we can calculate and present the measures, such as percentages, for the different age groups or strata.

This table is an example of hypothetical baseline data that I made up for Beachville County. And I thought of Beachville County because I'm looking forward to going to the beach next week. In this slide, there's data presented for three time periods. So you can assume that all three of these time periods in this particular slide occur before the implementation of a new and exciting maternal and child health program. The data presented are percentage of live births that occur prior to 37 weeks gestation, so each one of the numbers or percentages of this table are births, the percentage of births occurring prior to 37 weeks gestation, and they're stratified by dividing them into four age groups. So I have one group less than 20 years of age, another 20 to 29 years of age, another 30 to 39 years of age, and the bottom row is greater than or equal to 40 years. Now, I arbitrarily put these age limits in here or age divisions, so don't think this is scientific. It's completely arbitrary. And all these data are made up. And each of the columns are for a specific year of data. So for instance when you're developing baseline data, you may want to show it for multiple years separately or combined.

So why do you want to collect baseline data? Why would anyone want to collect it? Baseline data are often used to compare what happens before a program or intervention begins with what happens after it starts, to assess the effect of that program or intervention. Baseline data are helpful for evaluating or assessing whether a change occurred after a

CMMI Strong Start: Potential Data Sources for Generating Baseline Data

July 25th, 2012

3:00pm-4:15pm (EST)

program or intervention was implemented. And they're a foundation for showing performance improvement. Another reason that we want to collect baseline data is for conducting a needs assessment. By collecting these data, we may be able to justify the need for a new program or service, and we may be able to target an intervention to people who are identified as having the most substantial need.

Here's another sample table. Again, this is totally hypothetical and it's in Organic County and I've named it that because I'd love to live in a county that's fully organic. I think of that as being a healthy environment. So this particular slide or table, is different than the previous one because this shows pre-intervention data and post-intervention data, so you can compare. The first row shows pre-term birth proportions for Center A, and the second row shows pre-term birth proportions for Center B. We don't know about the different programs in Centers A and B. Those are just two arbitrary centers.

The columns show the proportion of pre-term births for different time periods. So the column to the left shows data for several years occurred right before the new program started. The middle column shows the percentage of pre-term births that occurred during the intervention period, after the program begins.

So looking at this table, we can assume that the intervention began on January 1, 2013, so I'm a little bit ahead of time in this example. And the column to the right is one way of comparing pre-term birth proportions that occurred before and after the study intervention begins. And these are percentage point changes, which is nothing more than subtracting the post-intervention percentage from the pre-intervention percentage. And in this example, it looks like Center A may have done a little better job of reducing pre-term births than program B, but we really don't have enough information to know whether or not that's an important difference or not. So again, this is just an example of how we might use baseline data to compare to intervention data to see whether or not the program or intervention has made a difference.

Now, I've used the word intervention and intervention group quite a bit. And the intervention group is just a group that's expected to receive an expected service or a treatment, or something that's different from what another group is getting or others in the past have gotten. And many times we consider the intervention group to be the group that's assigned or allocated to receive a program or service, even if some people don't show up or they drop out or they fail to comply. And we call this concept intention to treat, or intent to treat. And it may be important to use this concept of intention to treat for measuring data in the intervention group, because we often want to know how our programs and services work in the real world, where people are not always compliant.

In some clinical trials, people only measure the effect of the people that actually got the full intervention, but again, in the real world we want to know if our programs are really reaching people and making a difference, so we often use that intention to treat concept. So what kind of baseline data do you want to collect? You probably want to collect data that ties in with the goals and objectives of your project, so that's an important starting point. So

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

you should collect the types of information that would be appropriate for measuring changes in accordance with the objectives of your program or intervention. As a starting point you'd typically want to collect some type of outcome measure. And in some types of studies we may be interested in the average gestational age measured in weeks and days, and this measure might be reported for different groups or strata of participants. We might want to know in some studies about the cost of care, such as the cost of inpatient care. We might want to know about rates and duration of breastfeeding in some type of study. We also would want to measure something like characteristics of your population or sample.

So when we're trying to assess how well a program is working, we often want to collect information that describes the sample or population, and by collecting information about characteristics of the sample or population, this should be able to help you identify a baseline group that's similar to the intervention group based on whatever characteristics you have identified as being important. These characteristics may be those that are risk factors for that specific outcome.

And some of the types of sample characteristics that we measure are race and ethnicity, age, educational attainment, sex, geographic area, maybe health belief systems or marital status, support systems, and type of health insurance coverage. And again, it depends on what you determine your needs to be. In addition, we can think of something called proxy measures. When specific measures that you really want or need are unavailable within your time and other constraints, sometimes other data may be used to provide a proxy for those measures, or indicators or groups. So a proxy is a substitute.

For example, if you have scientific evidence to suggest that maternal education is a strong measure of socioeconomic status in your specific group and this measure isn't in the dataset that you have, then you may need to use a different measure of socioeconomic status, such as income. So income would be the proxy for maternal education. I'll give you another example. If you want to measure gestational age in weeks and days, but gestational age is only available as a categorical variable, so it's in three categoricals such as 37 weeks and older, 32 weeks to less than 37 weeks, and less than 32 weeks, then this categorical data may be a proxy measure or a substitute for a more specific gestational age.

And then one final example. If you're studying birth outcomes in a Medicaid population and don't have access to a dataset that specifically identified which people have Medicaid coverage, then you may want to use some other type of demographic information or risk factor information, to help you select a group that serves as a proxy for the Medicaid group. So this proxy data would then represent the population you intend to serve.

So as you plan a study, you need to determine what outcomes are important to measure and what characteristics are needed to assess risk for that outcome, or to help you select a population that's similar to the population you plan to serve. And these characteristics might vary for each of your different programs or settings. So where should these baseline data come from? Ideally, as Susan mentioned, these data should come from an entity's own program prior to an intervention begins, if possible. And this assumes that the other

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

population characteristics are similar to what the characteristics of the group served prior to the program. And if you don't have an appropriate comparison data internally, then you may need to seek data from other sources. And when this is true, often it's helpful to obtain baseline data from one source, if you can. And one reason to limit your sources of baseline data to just one source is because it can be time-consuming and costly to obtain data from more than one source, and it can be challenging to combine multiple sets of data, especially if variables are different or are defined differently.

About an hour ago I was on a telephone call with a group of people at a Medicaid program who were doing a cost analysis of pre-term infants in a Medicaid program. And in this study we're actually matching Medicaid claim files with birth certificate records, and it's a very time-consuming processing. And after about a month of matching, we're still at 93.2 percent of the two datasets matching. So that's a concrete example of if you can avoid having to use multiple datasets, it can often be easier. Alternative baseline data may be obtained in several different ways. One approach may be to obtain publicly available data. So what does that mean? Some publicly available data are in the form of actual datasets that you can obtain by paying for them or requesting them. Others may be in the form of existing tables or charts of data that you find online, and some may be available as online query systems that allow you to request somewhat customized reports within whatever constraints they have.

A second approach might be to go directly to some organization and submit a data request, asking for information from their organization. And that organization might be a government agency or it might be a local healthcare provider. A third approach might be to partner with a local organization to facilitate access to data or sharing of data. So these are just some of the ways that you might obtain data. Now, some of the considerations for collecting baseline data are listed here. In your search, you're probably going to find several possible sources of data, and you want to know which one to choose. Well, that's a tough decision, but one important consideration is to accept whether or not the data can be obtained within the designated time frame. So if the source of data with everything you need won't be available for another two years and you need it in two months from now, that's something to consider. A second consideration is whether or not the data are relatively current. For example, if the data source of information for the years 1995 through 2007, will this meet your needs if you really need data from 2009 through 2011? So you need to decide whether or not that's really a good enough proxy for the years of data you need.

And a third consideration is the accessibility of the data, how easy is it to obtain them? For example, can you obtain the data online in prepared tables, or do you need to methodically abstract data from health records? Do you need to go through an extensive approval process, obtain a dataset and then clean and analyze the data? And this sort of accessibility may be more or less important to different organizations and may depend on your circumstances. A key consideration is whether or not the data source contains the data

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

elements that you need, and if the ideal data are not available, then is there a reasonable proxy?

And a final assessment point is to consider whether or not the source of data provides you with a strong comparison for your intervention group. One way you might find a strong comparison is to obtain a large dataset with an extensive number of variables that are important to you, and then select a baseline sample from this dataset that has similar characteristics to the population that your program or service plans to serve. And typically this dataset would come from a geographic service area, similar to what you, where you work. And this approach may allow you to select a pretty strong comparison or baseline sample, yet it might be a very time-consuming process, when compared to some of the alternatives. Alternatively, you might find a data source that provides statistical information for specific groups. And in this case you'd need to look at the geographic coverage and characteristics of the sample to assess whether the group is similar to that served by your programs.

So in summary, strive to select a data source that provides you with a strong comparison group to the study sample who you're expected to serve. And second, your data sources should provide you with the data elements that you need or reasonable proxy measures, and you should consider factors such as accessibility and timeliness.

Dr. Schlenker and I will now review a number of potential sources of baseline data. There may be advantages and disadvantages to each of these data sources. We aren't endorsing any specific source, and the usefulness of these sources may vary between programs. A few of the sources may not be strong, especially those I'm going to talk about at the very end, yet I'm discussing those specifically just very briefly because they were mentioned in a report on underutilized maternal and child data sources. So given that I found the information, you might as well have it so that you know something about them.

On this slide is a list of potential data sources that we're going to present. It's not an exhaustive list, but it provides you with many key sources. So Dr. Schlenker will now cover some of these data sources.

DR. THOMAS SCHLENKER: Hi, everyone. This is Tom Schlenker. I'm the Health Officer in San Antonio, Texas, and I want to talk a bit about some of the alternate data sources that you might consider, and they are alternate of course, because this assumes that you don't have an ongoing program where you have a population that already has a baseline. So these are alternative to that.

I think it's important to know that in the United States, all births are registered and receive a U.S. birth certificate, and this is a national certificate that's standardized all across the country. It contains about 100 different data elements. Each state administers their own birth certificates, and they are allowed to add a few elements, but essentially they are all the same nationwide, and they all look more or less like this. This one occupies two pages, so we have to split it up among several slides.

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

And as you can see, it talks about the child, the mother and the father, giving data for each. And we should all know that there are some unique qualities about births that sometimes can be confusing. It's good to know those from the start. One is that the key person involved here is the mother. She is always present at the birth. The father may or may not be there, and so even though there is many data elements addressing a father on a birth certificate, it's not used all that much, because often it's blank. So the data's just missing or perhaps a question of value.

The mother is also key, because where the child is born does not determine where the child's birth is registered. What determines where the child's birth is registered is the residence of the mother. That's important to keep in mind. If you plan on using baseline data that corresponds geographically to where you're at, which makes a lot of sense, you don't need to worry of there's a big OB referral hospital in your neighborhood where pregnant women comes from miles away to deliver because of complications of pregnancies or other problems. You don't need to worry that those births from women who live outside your jurisdiction will affect your local rates, because they won't. They'll be counted wherever the residents of the mother is.

So going down the birth certificate, we notice that there's a large section on the mother, and it gives her demographics, like her level of education, her race/ethnicity, and the type of medical insurance she has. So Medicaid insurance, if it applies, is listed on every birth certificate, and it's right there in a box near the top so it's easy to identify. If you are going to be focusing on a specific race or ethnicity, the race and ethnicity of the baby born equals the race/ethnicity of the mother. So that is an important thing to keep in mind, and of course that's listed right here.

And then what follows is a fairly long section on the mother's health information, which you may be interested in. It's not required for the baseline data for this particular proposal, but certainly it affects the baby and birth outcomes in many, many different ways, so you may be interested in that, and there's really a lot of information. But addressing the elements that must be contained in your baseline data, you would then go to the information on the newborn. And specifically birth weight, which is the very first box, and that might be listed in pounds and ounces, or it might be listed in kilograms, but the birth weight is right up there near the top. And that's followed by the estimated gestational age in weeks. So that's where the specific information that you all need is on the birth certificate. That is what's collected in every city and town in the United States, and compiled and sent to the state, and the state organizes that and then sends it on to the federal government. And regardless of where you access this data, whether it's at the local level, the state level or the federal level, it's the same data coming from the same source, may be organized in different ways, and certainly it's accessible in a variety of ways, but it comes from the same source.

And I think with practically any question you might have in public health, the best place to start is with the CDC. That's certainly true I think, when it's related to births in the United States. And that's pretty easy to do. You go to the CDC website, which is CDC.oov, and then that opens up to you what is probably the best public health resource in the entire world. I

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

think it's a very good website, and it contains an enormous amount of information on every subject you can imagine. It's well-organized, it's attractive, it's easy to navigate. And I'm sure some of you are very, very familiar with it; others maybe not so, but we should all become very familiar with it because it is such a wealth of information and so useful.

So if you go to the CDC website you'll come to their home page, and there will be lots of buttons you can click on it. One is called Data and Statistics where you might think of starting, which would be a good place to start. And at Data and Statistics, they have a category called Fast Stats, and they will give you sort of key statistics on every subject under the sun, so you click on Fast Stats, and then among the list of options they give you, births is one of them, so you click on that. And there you will have the latest information on low birth weight and pre-term infants born percentages in the United States. So it's very reliable, fairly current, but it's obviously a national figure, so it doesn't directly help you with your baseline, except what it does is help you understand the context in which you're operating and whether you might be high or low. But then below the individual Fast Stats statistics that are offered, there's a long list of more data, and in that you can dig deeper into whatever items that have to do with births you're most interested in, and explore in that way.

But really, I think the easiest way to use the CDC website is that on the home page, across the top of the page they have the alphabet, A, B, C, D, E, F, et cetera. And you choose what you're interested in, so the way I usually enter is to click on "B," and it gives every disease that you can imagine that begins with a "B," including other non-disease items like births. So I go to "B," I click on births, and doing that, you are automatically brought to the National Center for Health Statistics site. The National Center for Health Statistics is the repository of all vital statistics in the United States. They are the authority, the last word, and they are a component of the CDC. So really, the CDC doesn't off you anything on birth statistics other than this National Center for Health Statistics, where they exist.

So you will automatically be brought to that page. And again, it's sort of a hoe page type of atmosphere. There are a few things you can click on. If you would like a copy of the birth certificate that we reviewed a few slides back, then there's a standard forms item that you can click on right there and it'll give that to you. And there are many other ways that you can enter into the data, but I think what is most helpful would be to go to the National Center for Health Statistics report, called "Births: Final Data for 2009," which is again, clickable from their home page, and then the very long and very formal annual report that contains the last word on birth information for that particular year. And it talks about numbers of birth, distribution, characteristics of the mother in terms of age and marital status and many, many other things. But what you want to do then is scroll down through the report till you get to Infant Health Characteristics.

And there they present the items that you're interested in: pre-term, birth, and low birth weight births, and they have it broken down in a number of different stages of gestational age and birth weight categories. And they have it for a number of years in a row, so you can see the trends. So that's a little bit more contextual information for you to watch to the

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

national trends over the years. They also provide a very nice map of all the states according to the percentage of all births that are born pre-term. So you can look at and download this color-coded map that will help you situate yourself in the entire country to get an idea of what your situation is there locally. And then they have a number of tables that may be useful to you that track birth weight and estimated gestational age by race and ethnicity and other factors, and those are in Tables 20 to 25 within this annual report. I point out also on page 69 there's a nice table that transforms pounds and ounces into kilograms, which may be of use to you.

Now, you'll notice that 2009 is the most recent data that they have a final report for, but you can also look at the preliminary report from 2010, which looks exactly format-wise like the report from 2009. Even though it's called preliminary, it has the same wealth of data, and I think would be acceptable to the grantors and to yourselves in that it's about as good as the data's going to be, but it's not quite finalized yet.

And then lastly, at the bottom of the home page for National Center for Health Statistics they have a link to the State Health Department, so that makes it very easy for you to click on that link to the state where you live to find out what your state is offering you. Again, it's the same data, but it's just what way your state has decided to display it, and what amount of access to offer you.

So what I did was click on Texas, where I currently live, and on the Texas website they have a link called Health Data, so I went in there. And then they have another option within Health Data called Customized Queries. And within Customized Queries, you are allowed to ask questions in kind of a structured format that have to do with births, so that you can pick an indicator, for example, premature birth, you can pick a county where you live, and then you get to ask for either the frequency of the occurrence or the percent or both, and you press a button and they give you the answer, so that you then can know what the percentage of premature births are in your county. On the Texas website, I found that there were some glitches to this, and it did not work completely for the question I asked it and I had to do a little bit of calculation by myself, but it could be useful to you in your state if they offer the same kind of function.

Also on the Texas site, they offer a nice map of all the counties in Texas, and race, both frequencies and rates of premature births and low birth weight. And just a caution is that it can be confusing, and you need to keep in mind whether you're dealing with just the number, the total number of births, or whether you're looking at the rate, and in most cases what you want is the rate. Also, while I was in Wisconsin, I did a lot of work in this area, and know that they present -- the state of Wisconsin presents this data somewhat differently in that it's much more of a large, formal report rather than just the query system that Texas has, and they present lots of tables and graphs and maps on an annual basis. The most current one in Wisconsin is for 2010, and that's their final report. So in contrast to the national, which is still preliminary, some states have already published their final data.

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

And in Wisconsin they have published their final data from 2010, and that was made available in January of this year, and that's something to keep in mind. There's nowhere in the United States that you're going to find more recent final data than from 2010, so don't waste a lot of time looking for 2011, because it's not compiled and finalized yet. And so it takes about one year for the state to do that, and that's when they work efficiently, and many states are not quite there yet. So you may or may not have 2010 data available to you. But the way that it's organized is fairly standard in formal reports in that birth and fertility rates are presented, the characteristics of the mother is presented, characteristics of pregnancy and delivery, prenatal care, et cetera, and then characteristics of newborn, which is where you want to go, and that will be relatively advanced in the report, but you should look for that. And that's where you are going to find the two elements of information that you need, which is birth weight and gestational age.

And then states may vary whether they offer statewide rates or whether they break that down into local and regional. And some states divide up into regions. And I think most states will report on county-level data, but it kind of depends what they offer you as readily available through their website or paper publications.

But beyond that, you can ask your state for help. You can make special requests to the state. You could ask them for a specific report on your county or a set of zip codes or a set of census tracts or of a portion of your population, only Medicaid or only African American or only Asian. You can ask for that, and they do have all of that data. Whether you'll get it or not and how soon is another question. That is depending on many, many different things. But you can certainly ask, and you might get lucky and they may have that at their fingertips. And maybe they even have a report that they could send you, so I think it's worth asking.

Another approach would be to ask the state for the entire electronic database for all births and all the information associated with births for one or more years. That's what I have done in my research, and it's very, very useful and a very rich source of information, but you do need an infrastructure to handle that. You need an epidemiologist who can organize that electronic database and analyze it, and it actually requires quite a bit of work, so you have to have the level of expertise within your staff to do it. And also, the state may charge you for that. A few thousand dollars would not be an unreasonable charge to transfer that data to you, even though it's your own data. Even though the data for your own county, there may be a charge associated with it, but that's something you might consider.

Certainly, you need to know what time period you're interested in. Certainly, a few years are better than just one. It will give you a more accurate average and will allow you to see if there are any trends going on. Of course, if you do make a request to the state, there may be a time element involved. And they may not be able to tell you right away, I can have this ready in three weeks, or I can it ready for you in three months. They might just say, well, we'll do the best we can, and you may be waiting forever. So it's quite a bit of uncertainty in some cases and may not be conducive to getting a proposal in on time. And then like we mentioned, there may be some cost associated with it.

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

But you should also remember that you all do have local health departments. Some of the people on this conference call right now, you may be within a local health department, but for those of you who are not, you are still in a community that has one, and they can be quite a resource to you. Here in San Antonio, the health department that I head, we have a website and we have a number of reports on the website. One is entitled Health Profile 2010, which is just the broad sweep of the community. But it does contain a chapter on maternal and child indicators, and within that section, there is some information on low birth weight and premature births. So that may be enough and the right type of information for your proposal, or maybe not quite, but at least you should see what's readily available.

We have another report called Birth Changes from 2000-2010 -- that's a little typo on the slide. It's ten years of data. Again, with pertinent information, and because it's over many years, it gives you a flavor of what direction things are going in locally, which will matter to you if you intend to change that direction. But then in addition to already available material from the local health department, you can request help from them, and I think in general they are easier to reach than the state health department, and perhaps easier to find out who within your local health department could help you.

I know we respond to data requests all the time and do the best we can, and usually there's no charge. And especially with birth data, we already have that on hand, so it's not too difficult to crunch that data, so that not only could we provide county-wide data on percentages of low birth weight and the gestational age by various segments, we could also give that data for exclusively Medicaid people, or we could give it for one or another race or ethnic group, or we could even break it down by zip code and census track.

We presumably, and other health departments like us, could give baseline data that would be a fairly good match to many of your intervention populations. It's definitely worth a looking into and asking about. And another component of local public health are WIC offices, although many times WIC programs are independent and are their own separate corporations. They are often part of the local health department, but they are very standard and operate in the same way all across the country. And I think most of you are familiar with the WIC program for infants, young children and their families. They are offered to low income families exclusively, and so this may be a good proxy if you cannot get Medicaid data. If you can get data from ma WIC population, it's true that not everybody in WIC is on Medicaid, but more or less everybody in WIC would qualify for Medicaid based on their income. So you could make the case that it's a very, very similar population to Medicaid.

And WIC is very much a service operation, where they see many, many clients every day, very high volume, providing services. They are not a research or analytical group, so they don't have any internal capability usually, to do any kind of analysis, but they do collect information -- they have to -- every single baby born within any time period in WIC, and that does include birth weight and gestational age, so that data does exist. But just an example, last week I wanted to test this out on my own WIC program, and we have a very large one here locally. We have about 60,000 clients per year, and I asked my director if it would be too much trouble to give me the birth weight and the gestational age on babies born within

CMMI Strong Start: Potential Data Sources for Generating Baseline Data

July 25th, 2012

3:00pm-4:15pm (EST)

WIC. And she said no, and within a couple of days she got that information back to me. And so out of a total population of 60,000 clients, what that consists of is about 8,000 or so babies that are born within WIC every year, and she gave me two and a half years of data, which was then almost 20,000 babies, their birth weight and their gestational age. But the way she gave me the data was just one very, very, very long list of individual births, their birth weight, and their gestational age arranged by date, which was exactly the information I wanted but was really useless to me because it wasn't organized in any way that I could use it.

So to really get some utility out of the WIC data, at least for me locally, what I would have to do is import that data into some kind of software database model and manipulate it and organize it so I could draw some conclusions from it. So that's an example of how the data's there, but it may not really be accessible for the purposes that you would want.

And I think now I will turn it back over to Julie, who will talk about a few more sources, and then at the end, we will both be available for comments and questions.

JULIE JACOBSON VANN: Thanks, Tom. I think -- it's 4 o'clock. I have five or six sources to go through, so I'll just try to spend about a minute per source that we have time for questions. And the first one I'm going to talk about is the Pregnancy Risk Assessment Monitoring System, or PRAMS, which is a surveillance project of the CDC and state health departments. So we go back to the same data source. But this data file includes some of the fields from the birth certificate data and some survey data about pregnancy attitudes and health behaviors and prenatal care.

And to access that data, there is a link or a program called CPONDER -- CPONDER, C-P-O-N-D-E-R. And this link is another one of those query systems where you can answer a series of questions, like you pick a state level, you pick years between 2000 and 2008, which is the most current data of this system, and then you select one of 27 topics such as smoke exposure, infant health, infant mortality or smoke exposure. And if you select infant morbidity, you can select one option, which is low birth weight, either 2,500 grams or greater -- excuse me -- less than or equal to 2,500 grams or greater than 2,500 grams. So it's just sort of a dichotomous or categorical variable.

So on the next slide I'm going to just show you just a sample of the kind of data you might be able to get from this query system. Again, it's at the state level as opposed to smaller geographic areas. The most recent data are 2008, and really, the outcome is low birth weight, which is dichotomized into two measures of less than 2,500 grams or greater than, and there aren't really specific measures. So this is one particular source.

Another source is the pregnancy nutrition surveillance system, or PNSS. And this is a public health surveillance system, and it monitors risk factors associated with infant mortality and birth outcomes among low income women who participate in federally funded programs such as WIC and Title V. So the data from this system comes from federally funded clinics, but the submission is voluntary, so not everybody has to submit data, and they submit it to

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

their state agency who submits it to the CDC. And these are data that are indicators of maternal health and maternal health behaviors, including things like pre-pregnancy weight, parity and diabetes, and these data are presented as published tables that can be found at the link listed. Only three states actually make the data more widely available to the public, and those are North Carolina, California and West Virginia.

So again these are table format and they're state-level data and some national data, and birth weight is just presented again in categories of low, very low birth weight, low birth weight, normal and high, so the specificity of these data are pretty limited. And the years, it's a little bit more current than the previous one, up to 2010. And here is an example of a sample data table, which again is birth weights broken down into low and high, and it's in percentages and rank of the state, and pre-term, just the percent of infants that are born pre-term. And I looked online for North Carolina, and you can obtain the same types of data at the county level for North Carolina and perhaps California and West Virginia.

Now, moving on to another data source. Local hospitals. Of course, most babies, not all, are delivered at hospitals, and these hospitals have birth data and may submit data to the states or county health department or local health department, so they have birth data. They may be in different formats. Some may have electronic records, some may have hard copies records yet and they're moving in the direction of electronic records, some are in the process of going through the transition. So to obtain data from a local hospital, you have to do some groundwork. And so if you want to look at what possible data are available at your hospital, one of the things you might want to do is think about any contacts you may have at a local hospital, and talk with them and find out about what the process might be for obtaining data and what the data might be and what format they're in.

And then a second step is you probably need to contact their research office or office of medical information management, and again, then you can find out what data are available and in what format and how you might obtain them. But before you make these contacts, really sit down and plan what you want to say. Plan an outcome and plan an outline of what you'll need in terms of specific data elements, the timing, the population, and mention that it pertains to research. Make sure you ask about how soon data would be available and what the specific process is.

And for an example, I recently obtained data for a chronic kidney disease project at a local hospital. It was for Medicaid beneficiaries and we needed a huge dataset to identify people in need of a certain type of care. So we wanted every single lab value that the hospital had tested, a specific type of lab value for every Medicaid beneficiary. And to go through that process, I first had to fill out a formal data request, listing all the fields and the years of data, and then had to provide evidence that everybody that would be working on the project had completed the human subjects training and then actually needed to go before a special data board. And it's not the same as an IRB or Institutional Review Board. It was the Hospital's board that reviews data requests.

CMMI Strong Start: Potential Data Sources for Generating Baseline Data

July 25th, 2012

3:00pm-4:15pm (EST)

And then after I got approval, it took four to six weeks to get the first datasets, and they would only release it through a special secure website that I could access remotely and was not able to copy the data to another source, had to use it on that website. So it is possible to obtain data from local health organizations, but you may need to go through a series of steps to get these data.

All right, I'm going to skip this one, because this is just an example of one way that one hospital in Wisconsin reports data. State Medicaid program -- that's another potential source of data. Medicaid programs collect some information routinely such as claims and eligibility files, and sometimes they collect data through special programs. For example, I'm working on a cost analysis of pre-term infants who were taking the drug, palivizumab, which is a very costly drug to prevent RSV hospitalizations in preterm infants. And sometimes, Medicaid programs, rather than collect their own birth outcome data, they use existing data sources, for example, going to the state for birth certificate records.

One of the things that's unique about using Medicaid claims is that there isn't necessarily the specificity of gestational age or birth weight, but if you are sophisticated with using claims data, you may be able to analyze the data and find certain levels of for instance, prematurity and birth weight -- I won't go into the details. But generally to use these data, you need a fairly high level of sophistication in analyzing them.

Also, going back, Medicaid claims data, there are several ways to get them, and one might be to go directly to your state Medicaid office. And then there's another source, and it's called MAX data, which is available from the Center for Medicare and Medicaid Services. And the MAX data, which is the Medicaid Analytic Extract data, they are currently available for 34 states through 2009 and all 50 states through 2008. One of the limitations of MAX data, besides not being fully current, is there is a charge for them. And so that's one potential limitation.

And again, paying attention to time, I will move on to March of Dimes. And March of Dimes has a free online resource called PeriStat, and it provides perinatal statistics. And it's a free resource that you can obtain data in a form of pie charts or tables. The data are available as recently as 2009. One of the limitations, besides the fact that they're not fully current, is the data for preterm birth, are defined categorically as not preterm, moderately preterm, or very preterm. And birth weight are also listed in categories. These data are presented for specific geographic areas such as regional level, state level, some counties, some cities, but you may not be able to obtain the data for specific insurance categories or risk categories.

And this quickly, is one example of a chart that you might be able to obtain from PeriStats. So this just presents the percent of live births in each of these different race and ethnicity categories. I'm going to just very quickly mention, the Pediatric Surveillance System was mentioned in the article called "Underutilized NCH Data Sources." Typically these data are just for the state of New York, and the data are again in categories rather than in specific measures of birth weight. And one final source is the National Health Survey of Children's

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

Health, and I only mention that because again, it was an underutilized NCH data source and there is no birth weight or gestational age data on this source.

In summary, when seeking baseline data, you need to consider sources that align with your program objectives. Your baseline data need to be obtained from a group that's similar to your intervention group in meaningful ways. The sources we presented today are not an exhaustive list. Some have advantages and disadvantages, or each of them have some, and some sources may provide you with more comprehensive information, yet may require more effort to obtain them and analyze them. So you need to examine your options and assess the advantages and disadvantages of each within your specific setting. So I guess it's time for questions.

SUSAN JACKSON: Hi, thank you, Julie, again, for your presentation. I wanted to reiterate some of our opening comments on this and then I'll open it up to our presenters for comments. So like I said at the beginning of this webinar, the purpose of this webinar is to provide expert guidance to entities with limited experience identifying baseline data for comparison to an intervention population where carriage changed.

The Strong Start application requires all applicants to state their commitment and demonstrate their ability to provide gestational age and birth weight for intervention and baseline period that spans at least two years prior to the start of the Strong Start intervention. The expectation is that two years of baseline data will represent the applicant's own Medicaid and/or CHIP population for the two years prior to the start of the intervention. CMS views an alternative baseline data source as acceptable only when an applicant does not have two years of their own baseline data on the population that they will serve under the Strong Start initiative, because the applicant is a startup, has been in operation for less than two years, or already provides intervention services but to a non-Medicaid and/or non-CHIP population.

Under these special circumstances, an applicant may propose an alternative data source that may include aggregate information, but individual level data will be considered higher quality than aggregate data. The applicant must demonstrate that these alternative data represent a population that is sociodemographically similar to the population that will be served under Strong Start initiative.

CMS expects applicants to propose the best quality baseline available to the applicant. Upon award, CMS will work with awardees to collect an appropriate baseline for the awardee's specific circumstances. For this subset of potential applicants who have special circumstances and limited experience identifying baseline data for comparison to an intervention population where care has changed, this presentation should have provided an overview of approaches identifying a baseline population as potential sources of baseline information when the prior two years of baseline data for an intervention site's own past experience is not available or are not appropriate. In addition to data that would be incorporated in an application, these data sources and approaches were covered in this webinar.

CMMI Strong Start: Potential Data Sources for Generating Baseline Data
July 25th, 2012
3:00pm-4:15pm (EST)

CMS believes that each source and approach to identifying a baseline has both strength and weaknesses, and typically stronger designs have historical data from the same population. An applicant should be mindful that baseline data from national sources may make it more challenging to identify a population similar to the population that will be served by Strong Start. This presentation is not intended to endorse any particular source or approach to identifying baseline data specific to Strong Start funding opportunities by a particular prenatal care site.

We recognize that potential applicants' circumstances vary, and appropriate available sources for baseline data for each applicant will also vary. If you have any Strong Start-specific questions, we encourage you to submit them to our Strong Start inbox at StrongStart@cms.hhs.gov. We will answer these questions as quickly as possible. Additional information is available on our website at innovations.cms.hhs.gov. Under the What We're Doing tab, click on Strong Start.

At this time we would like to open this up for questions -- we're actually out of time, so we don't have time for a few additional questions, but we encourage you to submit them to our inbox, and hopefully these slides will be up in about a week, within the week, couple days to a week. Thank you and have a good day. Bye.

END OF FILE