Outcome Measures and Statistical Analyses Module Deborah A. Zarin, M.D., Director, ClinicalTrials.gov National Library of Medicine

In this presentation, I will be reviewing the outcome measures module of the ClinicalTrials.gov results database. The purpose of this module is to display the results and associated statistical analyses for each pre-specified primary and secondary outcome measure. Other outcome measures may also be included in this module.

The law requires this module and in fact says that you need a table of values for each of the primary and secondary outcome measures for each arm of the clinical trial. In addition, the module should include the results of scientifically appropriate tests of statistical significance for each outcome measure. All of you are used to seeing outcome measures as reporting in journal articles. Here is one from a journal article that shows the three arms of the clinical trial and the rows are different outcome measures. The ClinicalTrials.gov format is quite similar. In this view, you can see the setup, which is showing that we are going to be talking about a primary outcome measure and the measure title, a little bit of description, the time frame, and then the population description. This next slide shows the actual data. It shows the three arms of the clinical trial, the number of participants analyzed for each arm, and the actual results for different categories of this particular outcome measure.

The data elements that are required in this module are as follows. First, there are a set of data elements required to set-up the outcome measure table. These include descriptions of the arms and groups as you have seen in other modules for the results database. These form the columns of the table—the number of participants analyzed for this particular outcome measure and if necessary, an analysis population description. You need a set of data elements to actually describe the specific outcome measure. These include the outcome measure title, description if necessary, the unit of measure, the time frame, the measure type, and measure of dispersion or precision. Finally, you need the actual data for that particular outcome measure. There are additional data elements that can be used in this module, including the outcome measure type. This is where you would indicate whether your measure was a primary outcome measure, a secondary outcome measure, or one of the other two types of outcome measures that we accommodate.

Here is a more intuitive way of thinking about the outcome measure data elements. This is now available as a PDF and could be used to help gather the data that you need to fill out the outcome measure module. In the future, we intend to have this as a data entry mode as sort of a wizard mode to enable you to enter the outcome measure data in a more intuitive fashion.

The general review criteria for this module are the same for other modules and basically include items meant to ensure that the information is understandable. There are also specific review criteria. Basically, the measure title, description, time frame, and unit of measure need to be logical. The measure title needs to clearly and accurately indicate what was measured, not why it was measured. We say that because many people list outcome measure titles as things such as "to demonstrate the superiority of this drug." That is the goal of the outcome measure, perhaps, but it is not what the measure is and does not help one to understand the data in the table. It is important to use precise language when describing the outcome measure. For example, some people attempt to report numbers of participants with a certain attribute such as number of participants with a myocardial infarction using words such as incidence of MI or frequency of

Outcome Measures and Statistical Analyses Module Deborah A. Zarin, M.D., Director, ClinicalTrials.gov National Library of Medicine

MI or rate of MI. This is not correct since all of those words imply a unit compared to another unit, as is rate, such as number of participants per unit time.

This is a framework for thinking about specificity of the outcome measure described. Along the left, you can see that for any level of specificity a specific time frame is necessary. Then going from top to bottom, you can see increasing levels of specificity. The highest level is the domain, and in this case anxiety, depression, schizophrenia, etc. For this example, we are going to report a measure about anxiety. That is not very specific though. The next measure is level two and is designed to say what specifically was used to measure anxiety and in this case the Hamilton Anxiety Rating Scale was used. The third is the specific metric, which is what was measured in the individual participants in each arm of the trial. It could have been the end value of the Hamilton Anxiety Rating Scale at the end of the study. It could have been the change from baseline. It could have been time to reach a certain target level of the Hamilton Anxiety Rating Scale. For this example, we are going to look at change from baseline. Finally, the question is how are the data across participants in each arm aggregated? For example, if you have 100 people in each arm, how are all the changes from baseline aggregated? They could be aggregated as continuous measure or categorical measure. In this case, categorical measure, and there are many ways of doing it, but one set of categories is the one chosen, which is noting the proportion of people in that arm with a decrease of greater than or less than 50 percent. You can see as you move down this diagram, you are getting more and more specific. It is extremely important that when registering the outcome measures and reporting the results that the information be entered with as much specificity as possible.

Other specific review criteria include the need for a specific time frame and the fact that the unit of measure must be specific and must be consistent with the measure title and description. Here is an example where the time frame was listed as "during scheduled treatment period." This is not specific and would need to be corrected to give a specific time frame. Other review criteria refer to the measure type, which includes number which is typically the number of participants with a certain event, but sometimes might be the number of eyes affected by a certain treatment for an ophthalmologic disease or something of that sort. Then, there are measures of central tendency. It could be mean, median, or other measures of central tendency. It must be noted which of these measure types is being reported.

Here is an example where the unit of measure does not make sense, so the column says "participants with multiple myeloma" and the outcome measure is transplantations that achieved a certain endpoint. The unit says "number." If you try to read the nine, which is the data with number, the answer is there were nine numbers. That doesn't make sense. In fact, what they meant to say was that there were nine transplantations as shown here. The unit of measure needs to be something that you would actually say after the number in the cell to help you to understand what the data are.

If the outcome measure is using a scale, it is important that some information about the scale be presented, so that the reader who is not intimately familiar with that scale could understand the information. The name is needed. At a minimum, the range and direction of scores, which can be usually provided by giving the worst score and the best score provided. If there are no other units

Outcome Measures and Statistical Analyses Module Deborah A. Zarin, M.D., Director, ClinicalTrials.gov National Library of Medicine

for the scale, the units of measure can be reported as units on a scale. The arms and groups need to be specified in detail and clearly as is been explained in other modules.

The analysis population description also needs to be specific. Sometimes people enter data using an abbreviation like "ITT" as an abbreviation for intention-to-treat. This is not specific enough. In fact, even saying intention-to-treat is not specific enough since that has multiple meanings and it is important to actually say whether all the participants were analyzed or how many participants were analyzed and who was excluded.

The data in the end should make sense and be meaningful. No data should be entered if zero participants were analyzed. For example, sometimes in a study there could be a secondary outcome measure that was never analyzed because of some problem with the study. The way to indicate that would be to have zero participants analyzed for that outcome measure with a description in the analysis population free text box. It is important not to enter data since nobody was analyzed for that. In addition, the data should not be clearly invalid. Some validity checks which we can do on our end include things like the fact that the median has to fall within the interguartile range or within the full range. There are some things that occur that are clearly invalid such as the fact that there cannot be 823 mean hours per day of sleep. Scores on a scale have to fall within the range of the scale as provided in the description. There should not be a fraction of participants if it is a number of participants as the outcome measure type. For example, there could be a mean number of participants that's a fraction, but there cannot be a number of participants that's a fraction. The data have to be consistent with other data in the record. There shouldn't be any placeholder numbers; although, sometimes for convenience, people will use things like a "zero" or a "999" as an indicator that the data are not available or some other indicator. This is not appropriate in our system where the data in each cell have to have meaning.

Here is an example of what we call data mismatch. The measure name is "time to disease progression" and one would expect a measure of time in the outcome measure. Here you have down at the box in Drug A in the first column, you have 40 participants. If the measure name is time to disease progression and the answer is 40 participants that actually doesn't make sense. It is important to rectify that.

For the statistical analysis part, there are a number of different data elements. These include data elements that give an overview of the statistical analysis that is being reported, which include the comparison groups selected. If there are two arms, it is frequently both arms, but for example, in a three-arm study there might be a series of pair-wise comparisons and it must be indicated which two arms are being compared. We also ask whether it is a noninferiority or equivalence analysis and that is a yes/no question. If the answer is yes, we do ask for some more details about that analysis. If a P-value is being reported for a statistical test, we require you report the method used. If a confidence interval is being reported, we require various details about the confidence interval along with the dispersion of the confidence interval.

Outcome Measures and Statistical Analyses Module Deborah A. Zarin, M.D., Director, ClinicalTrials.gov National Library of Medicine

Here is an example of a data entry screen for a confidence interval where it is defaulted at the 95 percent confidence interval, but another level can be chosen and then various dropdown menus as well as data entry boxes are provided to give all of the information that is necessary.

The statistical analysis data need to be consistent with the data reported for the associated outcome measure. As mentioned, if it is a non-inferiority or equivalence analysis, some details need to be provided. If there is a P-value, the method needs to be provided. If there is a confidence interval, the estimation parameter information also needs to be provided. Here is an example of an error where the mean difference, which is listed as nine, does not correspond with the actual data. For example, given the data provided, the mean difference is actually 1.5.

Additional information about this module as well as other modules can be found at these sites on our public Web site.