Baseline Characteristics Module Rebecca J. Williams, Pharm.D., M.P.H., Assistant Director, ClinicalTrials.gov National Library of Medicine

This presentation will cover the baseline characteristics module within the results database at ClinicalTrials.gov. I will be providing an overview of the module as well as describing the required data elements and the review criteria specified by ClinicalTrials.gov as well as including a few examples of common errors that we've encountered in our experience with the system.

The scientific purpose of the baseline characteristics module is really to provide a summary of important attributes of the participants who are enrolled at the start or the baseline of the study. This module is based on the requirement in FDAAA to present a table of the demographic and baseline data collected both overall and for each arm of the clinical trial.

The baseline characteristics module itself then is composed of a table because that is what FDA required. It includes information about the overall trial population as well as information for each arm or comparison group. The baseline characteristics module can accommodate different data types, both continuous and categorical, and as the table is being specified by the sponsor, this information about the characteristics of the data itself will be inputted. You may be familiar with Table 1 in a journal article, which specifies the baseline characteristics of the population. Our information in ClinicalTrials.gov is very similar to the Table 1 in a journal article. Each baseline measure will just need to be specified using the data elements within ClinicalTrials.gov. This is the public display of a ClinicalTrials.gov baseline characteristics table. It includes the overall number of participants who are analyzed at baseline as well as the description of the key characteristics that are being described.

The data elements for entering information in the baseline characteristics module include information to construct the table. The columns of the table consist of arms or groups and a title and a description must be provided for each arm or group. It is also necessary to specify the analysis population. What were the overall number of baseline participants who are analyzed both overall and per arm? For each baseline measure, the baseline measure title specifying what it is that is being reported must be specified and if any additional description about that measure is necessary that can be provided in the baseline measure description. Additional data elements include the measure type and measure of dispersion. This relates to the characteristics of the data that will actually be presented, a unit of measure to correspond with the baseline measure as well as the data for each cell that corresponds to each row and column of data.

ClinicalTrials.gov actually provides some default tables as the data is being entered through the interactive mode. These default tables are listed here and, of the tables that are provided, age and gender are required for every study that is entered into the database. How that data is described can be modified by using either continuous or categorical data for age and for gender it is typical male or female, but in rare situations it can also be customized. We make it easy, however, to provide other important demographic information such as race and ethnicity. We have tables for the NIH/OMB format or it can be customized to correspond to the way in which the data was collected for that study. It is also possible to specify the region of enrollment in one of the default data tables. This slide just shows a tabular representation on this slide would need to

Baseline Characteristics Module Rebecca J. Williams, Pharm.D., M.P.H., Assistant Director, ClinicalTrials.gov National Library of Medicine

be repeated for each additional baseline measure that is entered specifying the title, the units of measure, the measure type, as well as the measure of dispersion.

Although the minimum requirements for entering data into the baseline characteristics module include only age and gender, it would also be considered a best practice to provide other relevant demographic characteristics of the patient population as well as clinical measures that were relevant to the study. Typically, this will include clinical characteristics that were involved in the outcome measure assessment in subsequent modules. It may also be important to provide information on prior and concurrent treatment characteristics. As you will recall, I showed a table 1 from a journal article. Generally, it would be the baseline characteristic information you would normally include in a journal article and that would be considered best practice for entering information in the ClinicalTrials.gov database.

Here, I've shown just a couple screenshots of how the information is actually entered. This is the overall number of baseline participants, entering in the data cells the number of participants that were analyzed for each arm. This screen is showing how you actually specify for each baseline measure, the type of information that will be entered, so here in this case, it is age continuous. It is specified as a measure type of mean with a measure of dispersion of standard deviation and the unit of measure, years, is consistent with the baseline measure title of age continuous. This slide is just showing how the actual data are inputted once the measure type and measure dispersion information has been specified for this particular outcome measure. This would be repeated for each subsequent outcome measure.

The ClinicalTrials.gov review criteria, we have specific review criteria for the baseline characteristics module, but generally there are criteria that apply across all the modules and those are explained here. The first is that abbreviations should be expanded the first time that they are used, so that someone not familiar with that abbreviation could understand the information that is being presented. The record should be evaluated for spelling errors and the arm/group information, sort of the columns of the table should be informative both the titles and the descriptions of the information presented. In addition, the information should be consistent with other sections of the record or discrepancies should be explained and in general, there should be no written results or conclusions. All the information should be available in the tabular format.

Looking specifically at consistency across the record, one of the things that will be evaluated is whether or not the overall number of baseline participants is consistent with the number started in the participant flow module. If the numbers are not consistent, explanatory comments should be provided in the baseline measure description or it may be possible to also include an informative milestone within the participant flow module. In addition, we will be evaluating whether or not the unit of measure is consistent with the baseline measure of title and the previous example I showed you a measure title of age would generally have a unit of measure of age which could be years or months or some other measure that would be appropriate for this study. Finally, the data appear to be valid based on the information provided.

In this example, this is a screenshot from a ClinicalTrials.gov reviewer's vantage point. It is looking at the overall number of participants that were analyzed for this baseline measure as well

Baseline Characteristics Module Rebecca J. Williams, Pharm.D., M.P.H., Assistant Director, ClinicalTrials.gov National Library of Medicine

as information related to the age categorical measure and the number of participants within each of the categories. As you can see here, the total number of participants in the category is 139 plus one plus one is 141, which corresponds to the total number of participants analyzed within this module. This would be acceptable.

I mentioned that one of the review criteria relate to the validity of the data. It is not always to check validity, but in this case, a unique opportunity came along. In the age categorical section, it specified that most of the participants, 8 out of 10, were greater than 65 years of age yet age continuous was also provided and the mean age was specified as 12 and a half years. Those two pieces of information cannot both be true. It would yield a comment from our ClinicalTrials.gov review staff asking for clarification on the validity of the data.

Here is another example showing a two-arm study. Overall, the measure that was being specified was age. It was being specified as a continuous variable. You would expect the data to be 61 years or months of age; however, the units of measure were entered as participants. That unit of measure is not consistent with the measure, title, and the data that were provided. Likely, the units need to be updated to include a time of years or months or whatever was relevant for this study.

Additional criteria also look at specific information related to categorical data. The categories should not overlap and all possible categories should be presented. If number is the measure type, then generally, participants will be the unit of measure and overall, as we saw in one of the previous examples, that the number of participants in each group should match the overall number of participants analyzed for the baseline characteristics module. If it is not matching, then additional explanation would need to be provided. The unit of measure should also be specified in the category title if appropriate.

For continuous measures, we will evaluate the information specified on the slide. In general, the total column must be completed and zeros cannot be entered as you may recall. One of the requirements of the law was to present information overall and per each arm of the trial. For measures that were obtained using a scale, it is important to provide contextual information to be able to interpret the data that is being reported based on the results of the scale. Providing the name of the scale as well as the range and direction of scores indicating what would be considered a good value and what would be considered a poor value on that scale and specifying the unit of measure as "units on a scale" if there were no other units for that particular measure.

This was an example of a baseline measure that was submitted to the ClinicalTrials.gov database. It was specified as GOG performance status. There were categories of zero, one, and two. Looking at this information, I actually have no idea what it means. I don't know if zero is bad. I don't know if zero is great. I have no sense of even what the measurement was trying to assess. After revision, the sponsor actually updated this baseline measure to provide actually some very informative information. I now know that this relates to performance status and it is measuring the ability of patients to perform various daily activities. You can see now very clearly that zero being fully active is actually a good score. Four, the highest level within the scale, is actually

Baseline Characteristics Module Rebecca J. Williams, Pharm.D., M.P.H., Assistant Director, ClinicalTrials.gov National Library of Medicine

considered to be the worst score on this scale. It is possible to provide informative information about this scale in a very short period of space.

For additional information about the baseline characteristics module, you can visit the Web site listed on this slide. If you have any questions as you are entering data, please feel free to email register@clinicaltrials.gov. We would be happy to provide assistance.