



115 Carriage Road  
North Barrington, IL 60060 11/22/04 19:37  
22 November 2004

Dockets Management Branch (HFA-305)  
Food and Drug Administration  
5630 Fishers Lane, Room 1061  
Rockville, MD 20852

Dear Sir or Madam:

On behalf of the CDISC Submission Data Standards team, I would like to submit the attached comments on the Draft Guidance for Industry: Collection of Race and Ethnicity Data in Clinical Trials, Docket 02D-0018. While I realize the docket has closed, we have been evaluating the representation of race and ethnicity information as part of our work on the CDISC Study Data Tabulation Model used to submit clinical trial data to the FDA, and have found some serious issues which are discussed in the attached document. We would appreciate it if these comments could be reviewed by those responsible for issuing the final guidance.

Very truly yours,

A handwritten signature in black ink that reads "W R K".

Wayne R. Kubick  
Team Leader  
CDISC Submission Data Standards Team  
[wkubick@cdisc.org](mailto:wkubick@cdisc.org)  
847 842-1846

02D-0018

C 24

## Comments on Draft Guidance for Industry: Collection of Race and Ethnicity Data in Clinical Trials

The CDISC SDS Team has particular concerns around the recommendation to allow the selection of multiple racial categories. While this practice may "facilitate comparisons across studies analyzed by FDA and with data collected by other Federal agencies" (Lines 100-101), it will not "enhance the early identification of differences in physiological response among racial and ethnic subgroups during the evaluation of safety and effectiveness of FDA-regulated products" (Lines 97-99). The practice of allowing the selection of multiple categories will, in fact, actually decrease the ability to correlate physiological differences with race or ethnicity. The reasons for our conclusion are described in the following paragraphs.

The selection of multiple categories will cause subjects who are not genetically alike to appear as if they are, and correlations that are indeed due to race will be missed (false negatives). If two subjects who are, respectively, 75% Asian, 25% Caucasian and 75% Caucasian, 25% Asian are both allowed to check the CRF boxes designated "Asian" and "Caucasian", then these two subjects will be considered racially equal for all analyses. Had the two subjects been allowed to self select one category (as is the current practice at most companies), they would have likely chosen their predominant race. In such a case, any differences between them could have been examined to see if they were correlated with race. Under the guidance, however, any differences between them would never have been attributed to race because the only information available would have indicated that their racial categories were the same.

The selection of multiple categories can also result in people that are genetically similar appearing less so, leading to physiological differences being falsely attributed to race (false positives). Two people who are 75% Caucasian, but with some other, different racial backgrounds (one may be 25% Asian, while one is 25% Pacific Islander) will appear different for all analyses when in fact they are more similar than they are different. Physiological differences between these individuals may be mistakenly attributed to race because, on paper, they "appear" to be racially different.

The selection of multiple categories could make analyses difficult due to the number of combinations of racial categories. There are more than 30 possible combinations of the five recommended racial categories, which will likely result in some categories having small numbers. In order to perform statistically meaningful analyses, some sponsors may decide to collapse some categories, and each may decide to do so differently. One sponsor may decide to group small numbers of people designating Caucasian-Black-Asian with Caucasian-Black, while one might decide to group them with Caucasian-Asian.

It has been suggested by some that allowing subjects to designate a percentage for each racial category would alleviate the problems mentioned above. Doing this would, however, further increase the subjective nature of the data collected, and further confound the ability to make predictive correlations between study findings and race.

The current practice of selecting only one category usually results in the subject selecting the predominant racial background. This allows analyses to be performed based upon that predominant background. This practice is less subject to false positive and false negatives than that of allowing subjects to select multiple categories, and is less subjective than allowing subjects to designate percentages for multiple racial categories.

The CDISC SDS Team requests that the FDA consider recommending the collection of race data in two questions: one to ask for the race that the subject most closely identifies with, and another to capture additional racial designations, acknowledging that these serve different purposes. The first question would be used for statistical analyses of the actual clinical-trials data, while the latter would provide information on racial representations in clinical trials.