Alcon®
RESEARCH, Ltd.

April 4, 2006

Division of Dockets Management
HFA-305
Food and Drug Administration
5630 Fishers Lane
Room 1061
Rockville, Maryland 20852

Re: Docket No. 2006D-0044

Dear Sir or Madam:

Provided herewith are two (2) copies of Alcon's comments regarding FDA's Draft Guidance on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims..

If you have any questions regarding these comments, please contact me via e-mail at garry.heidel@alconlabs.com or via telephone at (817) 551-6813.

Sincerely,

Garry G. Heidel
Director Regulatory Compliance
Alcon Research, Ltd.

Attachment

I. General Comments:

1. Alcon is pleased to see this draft guidance which helps legitimize PRO research in the U.S. For the most part, the document describes best PRO practices and encourages quality work. Alcon offers the following constructive thoughts in the interest of strengthening the guidance.

2. The February, 2006 Mayo Clinic conference on this draft guidance seemed to provide conflicting information on two points.

   First, the guidance contains many stringent, strident requirements. However, FDA's public response to many questions was often: "it depends upon the situation, come and talk to us (the FDA) early." This response raises questions about whether this document is actually providing guidance, or if PRO work is going to be determined on a case-by-case basis.

   Second, FDA indicated that different groups in the agency will apparently be applying this guidance in different ways. So we need to ask, should this guidance be maintained in draft form until the various agency groups have established a unified application strategy?

3. **Responder Analyses are better than MIDs**

   A predefined MID employed to establish the importance of a difference in means may not be very useful. A mean is a representation of central tendency from a sample. However, there is inherent variability in the mean from a sample. For a normally distributed variable, the sample mean will lie below the population mean approximately 50% of the time irrespective of sample size. Thus, even if the population mean difference is greater than the MID, the sample mean difference from a study will be less than the MID 50% of the time by chance alone, regardless of sample size. The situation can become worse if the distribution is not normal. In a skewed distribution the observed mean could be less then the MID 70% to 90% of the time due to chance alone even if the population mean is above the MID. Conversely, in a skewed distribution the sample mean may exceed the MID more than 50% of the time due to chance even though the population mean would be less than the MID. Thus, the judgment of whether a clinical result is meaningful based on mean differences exceeding a MID is more a function of chance than of true clinical relevance.

A 'responder' analysis may be more useful in interpreting the relevance of a difference. In a responder analysis a count of how many patients exceed some threshold value is compared across treatment groups. If the threshold value is clinically meaningful then a statistically significant difference in the number of patients exceeding that threshold is inherently clinically meaningful; it means there is a non-zero difference in the number of patients who have a clinically meaningful response to therapy. If a responder analysis is used to establish relevance, then the outcome of the trial is a function of the inherent difference between treatments and the sample size rather than chance as is the case with a MID.

For a responder analysis to have value it is important that the threshold be well defined and meaningful. There are a number of ways to establish a meaningful threshold. It may be possible to develop normative tables for the PRO measure in normal subjects and establish a threshold based upon the distribution in normals. A simple threshold on an ordinal scale may simply be the zero on the scale. In this case the threshold represents patients who do not exhibit the symptom measured by the PRO. Another possible threshold could use an upper (or lower) two box analysis on a 5 point scale. For example, patients who agree or strongly agree on a Likert scale could be considered above threshold in a top two box analysis. The most important aspect of the threshold is that it be predefined in the protocol and agreed upon in advance by the sponsor and the agency.

4. Proposed Revisions to Comments on Metrics

The draft guidance states that test-retest reliability is more important than internal consistency reliability (e.g., Cronbach's alpha) which is not always accurate. Depending on the state/trait being measured, test-retest reliability may not be appropriate or expected. There may actually be a limited range of trials/indications for which test-retest reliability is essential.

The draft guidance also makes some psychometrically troubling assertions about item weighting in composites. When constructing a composite, it is important to evaluate whether the elements of that composite are actually measuring the same construct or domain which is where Cronbach's alpha, correlations between items, and/or running factor analyses would be highly relevant. The section in the guidance related to Identification of Preliminary Scoring of Items and Domains (beginning on line 409), specifically creation of weighted scores, is not justified in the psychometric literature.

Nunnally and Bernstein (1994) assert that total scores should "nearly always be obtained by weighting items equally" (p. 333). This section should be revised to accommodate a wider view of appropriate methodology for composite score weighting.

5. Consider Different Patient Populations

The draft guidance document is largely written from the standpoint of a "disease state" which encompasses only a portion of the consumers who use medical products. For example, in the case of healthy consumers, a PRO measure can provide relevant insight into the subjects' experience with the product/device that might not be available from a non-PRO measure. Careful language needs to be employed to communicate that different endpoints and standards for evaluating meaningful differences are needed for different populations and indications.

6. Wording

The guidance defines a PRO as a health status report directly from a patient, "without the interpretation of the patient's responses by a physician or anyone else" (line 32). Traditionally, parent/caregiver reports of health status for preverbal, pediatric study populations could be considered to qualify as pseudo-'patient reported outcomes'. The agency is suggesting that such types of parent/caregiver recorded assessments of children are in the scope of the document at line 692 where proxy respondents are mentioned. So it may be useful to revisit the definition of a PRO.

Table 3 of the guidance contains ambiguous language such as "...unacceptably large amount of missing data..."and "...a high percent of patients..." If the FDA has envisioned quantitative standards as benchmarks for these values, and/or minimally acceptable criteria to establish the types of reliability and validity highlighted in the guidance document, it would be helpful if these were explicitly referenced.

II. <u>Specific Comments</u>:

7. Lines 114-115 note that changes of clinical measure may not correspond with how a patient functions or feels. Alcon agrees and questions if this suggests that the MID/MCID concepts are not always relevant. Alcon proposes that the MID/MCID concept is relevant for efficacy and safety PRO assessment, but is not appropriate for HRQoL, function, or satisfaction PRO assessments. The latter three areas involve quantifying subjective patient evaluations where statistical analysis and contrasting of scores should be sufficient for label claims without adding an "important difference" hurdle on the assumption that there must be such a magic number.

8. Lines 154-156 state that a PRO used in a trial to support effectiveness should separately measure adverse consequences. Why must a PRO instrument that assesses efficacy assess adverse consequences, separately or otherwise? Shouldn't AEs be evaluated and confirmed by physicians?

9. Lines 212 to 225 present a rather hypothetical discussion of single item PROs. Lines 218-219 reference "single item questions" - a phrase that should be improved. The paragraph seems strange since it starts off talking about a reliable, validated single question PRO instrument, but then proceeds to discuss when a single item isn't enough. The latter discussion seems academic. These two concepts need separate discussion.

10. Lines 249 to 256 suggest that expected relationships among PRO items and domains need to be diagramed before the validation process. That may be difficult to do at the least, and the need and utility of the diagrams are unclear.

11. The meaning of "including excessive severity" as used in line 270 is unclear and requires explanation.

12. Lines 302 to 308 address the possibility that some questions could be irrelevant to patients. Assessments that ask the subject to respond hypothetically to questions should not be created. The responses to a given question should include an option that that implies that the category is not applicable to the subject. Usually, "not applicable" response options would handle this.

13. Lines 328 to 337 discuss appropriate recall periods for subjects. Some FDA Divisions only view a 24 hour period as producing reliable results. If the assessment is unsupervised (e.g., a diary the subject carries home with them) the measures which ensure the subject makes the entries according to study design must be determined. In other words, how can Sponsors determine that the subject completed the diary daily and not just before they came into the office visit? Also, this section suggests that Sponsors should insure that patients comply with diary completion instructions. Suggestions of how to do this would be included in the guidance. Electronic solutions may provide unchangeable date and time stamps for entries, but they may not address other issues like the quality of the entry or the identity of the person entering data. Uniformity across the agency will facilitate compliance.

14. The 100 mm VAS scale referenced in Table 2 may give a false sense of precision if only the endpoints are labeled. FDA representatives at the Mayo conference seemed to also be uncomfortable with the reliability of such a scale. This idea should be removed from the guidance.

15. Table 2 - Visual Analog Scale (VAS) -- If Sponsors use a VAS with the horizontal line and two "extreme ends" and no description of the intermediate positions, this could result in a false sense of precision. It appears that a Likert Scale or Checklist might be a better method of collection.

16. Some discussion is needed for the assertion in lines 341 to 343. It is unclear why the patient's perception of change over time is unimportant. It is helpful to know the patient's global impression of the treatment benefit. If there is a concern for validity and recall periods again, see the above comment on lines 302 to 308.

17. Lines 382 through 403: - Clarification is needed regarding the following statement "It is important that the PRO instrument format in the clinical trial be consistent with the format that is used in the instrument validation process. Format refers to the exact appearance of the instrument." The term "appearance" is unclear. Does this term mean margin settings, font, type size, line break, etc.? Does it mean if it's a Likert scale, we keep it as a Likert scale? If it's a font size 8 "yes / no" question, we keep it as a "yes / no" question? Note: The format issue arises again in lines 617 to 629. These sections should be combined for clarity.

18. Some actual examples of community preference weights used for scoring a PRO instrument are needed to illustrate the issue raised in lines 424 to 430. One would think that weightings derived from community patients representative of those in a trial would be acceptable for scoring. Weightings should not vary from trial to trial unless the trial populations are measurably different.

19. Lines 478 to 480 refer to assessing measurement properties for "future" clinical studies, (i.e. finish this assessment before using the PRO instrument). At the Mayo meeting, the thinking seemed to be that Phase II trials offer an opportunity for such assessments. However, phase II trials sometimes involve diverse objectives and/or samples too small for assessing an instrument's measurement profile. One proposal is that the Agency consider it reasonable to proceed to Phase III with an instrument that demonstrated adequate psychometric properties in a similar population prior to Phase III. Conformation could then be sought that it is also reliable and valid in the larger population of the Phase III trials; or in a parallel study conducted in the same time period as the Phase III trials. The draft guidance should allow for the results of an assessment in such a parallel study to be used to analyze the PRO results of the Phase III trials, assuming the parallel assessment conclusion was that the PRO instrument was valid, the patients in the Phase III and parallel studies were comparable, etc.

20. Validation is usually an ongoing process with each use of an instrument. Lines 579 and 664 discuss the different reasons to reconfirm the validity of a previously validated instrument. Alcon agrees with the need for reaffirming validity. There may be compelling reasons, however, to use an existing instrument with a slightly different population or a slight question modification in phase III trials. Reaffirmation of validity should be permitted to occur concurrently with the trial in a separate study outside the trial, and/or with the trial data.

21. It is unclear why an instrument developer would create a new instrument that produces results similar to an existing instrument as noted in lines 506 to 507. The Sponsor should understand why a new instrument is being developed. If a new instrument is being developed because the existing instrument has flaws, then there may be a question about using the existing instrument to establish the validity of the new instrument.

22. Line 512 refers to "meaningful discrimination". An explanation or definition of "meaningful" is needed.

23. The discussion of the establishment of a MID/MCID requires additional explanation and clarity. Lines 560 to 562 refer to using a distribution-based approach. What is the agency's opinion of using Cohen's effect size approach and substantiation by a clinical anchor point agreed to in advance?

24. Line 571 states: "There may be situations…meaningfulness of an individual's response … rather than a group's response…" These "situations" require additional discussion and explanation. Lines 572 to 577 go on to discuss the Agency's question about defining a responder. Generally speaking, PRO measures would be useful for defining a responder in health conditions where there are no hard objective clinical endpoints. In such cases, statistically significant differences in PRO score improvement, or in proportions of self-defined responders might be useful. In the latter case, the question would be something simple like: "Are you better, the same, or worse?" We would add that if the definition of a responder based on a PRO measure is agreed to, then a MID/MCID would be unnecessary. Conversely, if a MID/MCID based on a PRO measure is established, a responder definition is unnecessary. Sponsors have experienced agency opinions that MID/MCIDs differences need to make claims about differences between treatment groups. This guidance is silent on this subject. This topic should be included in the guidance. Further, achieving an accepted MID/MCID level should be sufficient for a label claim. Comparative claims and comparative MID/MCID differences should not be required.

25. The discussions of language changes and proxy respondents beginning on lines 643 and 692, respectively, are insufficient. The Agency might note that family members or friends, rather than doctors and their staff members, should generally be proxy respondents on PRO questions. Some detail on when proxy responses could be the basis for label claims would be useful (e.g. pediatric patients with a condition lacking in hard clinical endpoints).

26. The source of the statement in lines 725 to 733 suggesting patients would be biased if they saw their previous answers – but only if unblinding occurs is unclear. Lines 732 and 733 call for rigorous study of this philosophical discussion and Alcon recommends that the discussion be removed until the rigorous studies are completed.

27. Lines 807 to 809 discuss the importance of pre-specifying the method in which results will be interpreted for new instruments. Lines 927 to 929 indicate that substantial clinical experience is needed to establish interpretation rules. It is unclear if the Agency intends to work with sponsors on this issue when the instrument is new.

28. Line 940 states "...it is critical to ensure that patients enrolled in a clinical study are impaired in all domains" of a multi-domain instrument being used in the trial. Alcon agrees that if this rule is violated, a composite score using all the domains may not show a change or difference. However, Alcon does not understand why this rule needs to be followed if the a priori intent is to report selected domain scores and use them as the basis for a label claim (assuming the domains and their scores are validated). Alcon believes that that the Agency's comment here stems from the predisposition that all of the domain results must be similar before the data produced by the instrument can be used in any way. Alcon recommends that this concept be revisited.

END

Alcon®
RESEARCH, Ltd.

April 4, 2006

Division of Dockets Management
HFA-305
Food and Drug Administration
5630 Fishers Lane
Room 1061
Rockville, Maryland 20852

Re: Docket No. 2006D-0044

Dear Sir or Madam:

Provided herewith are two (2) copies of Alcon's comments regarding FDA's Draft Guidance on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims..

If you have any questions regarding these comments, please contact me via e-mail at garry.heidel@alconlabs.com or via telephone at (817) 551-6813.

Sincerely,

Garry G. Heidel
Director Regulatory Compliance
Alcon Research, Ltd.

Attachment

I. General Comments:

1. Alcon is pleased to see this draft guidance which helps legitimize PRO research in the U.S. For the most part, the document describes best PRO practices and encourages quality work. Alcon offers the following constructive thoughts in the interest of strengthening the guidance.

2. The February, 2006 Mayo Clinic conference on this draft guidance seemed to provide conflicting information on two points.

   First, the guidance contains many stringent, strident requirements. However, FDA's public response to many questions was often: "it depends upon the situation, come and talk to us (the FDA) early." This response raises questions about whether this document is actually providing guidance, or if PRO work is going to be determined on a case-by-case basis.

   Second, FDA indicated that different groups in the agency will apparently be applying this guidance in different ways. So we need to ask, should this guidance be maintained in draft form until the various agency groups have established a unified application strategy?

3. **Responder Analyses are better than MIDs**

   A predefined MID employed to establish the importance of a difference in means may not be very useful. A mean is a representation of central tendency from a sample. However, there is inherent variability in the mean from a sample. For a normally distributed variable, the sample mean will lie below the population mean approximately 50% of the time irrespective of sample size. Thus, even if the population mean difference is greater than the MID, the sample mean difference from a study will be less than the MID 50% of the time by chance alone, regardless of sample size. The situation can become worse if the distribution is not normal. In a skewed distribution the observed mean could be less then the MID 70% to 90% of the time due to chance alone even if the population mean is above the MID. Conversely, in a skewed distribution the sample mean may exceed the MID more than 50% of the time due to chance even though the population mean would be less than the MID. Thus, the judgment of whether a clinical result is meaningful based on mean differences exceeding a MID is more a function of chance than of true clinical relevance.

A 'responder' analysis may be more useful in interpreting the relevance of a difference. In a responder analysis a count of how many patients exceed some threshold value is compared across treatment groups. If the threshold value is clinically meaningful then a statistically significant difference in the number of patients exceeding that threshold is inherently clinically meaningful; it means there is a non-zero difference in the number of patients who have a clinically meaningful response to therapy. If a responder analysis is used to establish relevance, then the outcome of the trial is a function of the inherent difference between treatments and the sample size rather than chance as is the case with a MID.

For a responder analysis to have value it is important that the threshold be well defined and meaningful. There are a number of ways to establish a meaningful threshold. It may be possible to develop normative tables for the PRO measure in normal subjects and establish a threshold based upon the distribution in normals. A simple threshold on an ordinal scale may simply be the zero on the scale. In this case the threshold represents patients who do not exhibit the symptom measured by the PRO. Another possible threshold could use an upper (or lower) two box analysis on a 5 point scale. For example, patients who agree or strongly agree on a Likert scale could be considered above threshold in a top two box analysis. The most important aspect of the threshold is that it be predefined in the protocol and agreed upon in advance by the sponsor and the agency.

4. Proposed Revisions to Comments on Metrics

The draft guidance states that test-retest reliability is more important than internal consistency reliability (e.g., Cronbach's alpha) which is not always accurate. Depending on the state/trait being measured, test-retest reliability may not be appropriate or expected. There may actually be a limited range of trials/indications for which test-retest reliability is essential.

The draft guidance also makes some psychometrically troubling assertions about item weighting in composites. When constructing a composite, it is important to evaluate whether the elements of that composite are actually measuring the same construct or domain which is where Cronbach's alpha, correlations between items, and/or running factor analyses would be highly relevant. The section in the guidance related to Identification of Preliminary Scoring of Items and Domains (beginning on line 409), specifically creation of weighted scores, is not justified in the psychometric literature.

Nunnally and Bernstein (1994) assert that total scores should "nearly always be obtained by weighting items equally" (p. 333). This section should be revised to accommodate a wider view of appropriate methodology for composite score weighting.

5. Consider Different Patient Populations

The draft guidance document is largely written from the standpoint of a "disease state" which encompasses only a portion of the consumers who use medical products. For example, in the case of healthy consumers, a PRO measure can provide relevant insight into the subjects' experience with the product/device that might not be available from a non-PRO measure. Careful language needs to be employed to communicate that different endpoints and standards for evaluating meaningful differences are needed for different populations and indications.

6. Wording

The guidance defines a PRO as a health status report directly from a patient, "without the interpretation of the patient's responses by a physician or anyone else" (line 32). Traditionally, parent/caregiver reports of health status for preverbal, pediatric study populations could be considered to qualify as pseudo-'patient reported outcomes'. The agency is suggesting that such types of parent/caregiver recorded assessments of children are in the scope of the document at line 692 where proxy respondents are mentioned. So it may be useful to revisit the definition of a PRO.

Table 3 of the guidance contains ambiguous language such as "...unacceptably large amount of missing data..."and "...a high percent of patients..." If the FDA has envisioned quantitative standards as benchmarks for these values, and/or minimally acceptable criteria to establish the types of reliability and validity highlighted in the guidance document, it would be helpful if these were explicitly referenced.

II. <u>Specific Comments</u>:

7. Lines 114-115 note that changes of clinical measure may not correspond with how a patient functions or feels. Alcon agrees and questions if this suggests that the MID/MCID concepts are not always relevant. Alcon proposes that the MID/MCID concept is relevant for efficacy and safety PRO assessment, but is not appropriate for HRQoL, function, or satisfaction PRO assessments. The latter three areas involve quantifying subjective patient evaluations where statistical analysis and contrasting of scores should be sufficient for label claims without adding an "important difference" hurdle on the assumption that there must be such a magic number.

8. Lines 154-156 state that a PRO used in a trial to support effectiveness should separately measure adverse consequences. Why must a PRO instrument that assesses efficacy assess adverse consequences, separately or otherwise? Shouldn't AEs be evaluated and confirmed by physicians?

9. Lines 212 to 225 present a rather hypothetical discussion of single item PROs. Lines 218-219 reference "single item questions" - a phrase that should be improved. The paragraph seems strange since it starts off talking about a reliable, validated single question PRO instrument, but then proceeds to discuss when a single item isn't enough. The latter discussion seems academic. These two concepts need separate discussion.

10. Lines 249 to 256 suggest that expected relationships among PRO items and domains need to be diagramed before the validation process. That may be difficult to do at the least, and the need and utility of the diagrams are unclear.

11. The meaning of "including excessive severity" as used in line 270 is unclear and requires explanation.

12. Lines 302 to 308 address the possibility that some questions could be irrelevant to patients. Assessments that ask the subject to respond hypothetically to questions should not be created. The responses to a given question should include an option that that implies that the category is not applicable to the subject. Usually, "not applicable" response options would handle this.

13. Lines 328 to 337 discuss appropriate recall periods for subjects. Some FDA Divisions only view a 24 hour period as producing reliable results. If the assessment is unsupervised (e.g., a diary the subject carries home with them) the measures which ensure the subject makes the entries according to study design must be determined. In other words, how can Sponsors determine that the subject completed the diary daily and not just before they came into the office visit? Also, this section suggests that Sponsors should insure that patients comply with diary completion instructions. Suggestions of how to do this would be included in the guidance. Electronic solutions may provide unchangeable date and time stamps for entries, but they may not address other issues like the quality of the entry or the identity of the person entering data. Uniformity across the agency will facilitate compliance.

14. The 100 mm VAS scale referenced in Table 2 may give a false sense of precision if only the endpoints are labeled. FDA representatives at the Mayo conference seemed to also be uncomfortable with the reliability of such a scale. This idea should be removed from the guidance.

15. Table 2 - Visual Analog Scale (VAS) -- If Sponsors use a VAS with the horizontal line and two "extreme ends" and no description of the intermediate positions, this could result in a false sense of precision. It appears that a Likert Scale or Checklist might be a better method of collection.

16. Some discussion is needed for the assertion in lines 341 to 343. It is unclear why the patient's perception of change over time is unimportant. It is helpful to know the patient's global impression of the treatment benefit. If there is a concern for validity and recall periods again, see the above comment on lines 302 to 308.

17. Lines 382 through 403: - Clarification is needed regarding the following statement "It is important that the PRO instrument format in the clinical trial be consistent with the format that is used in the instrument validation process. Format refers to the exact appearance of the instrument." The term "appearance" is unclear. Does this term mean margin settings, font, type size, line break, etc.? Does it mean if it's a Likert scale, we keep it as a Likert scale? If it's a font size 8 "yes / no" question, we keep it as a "yes / no" question? Note: The format issue arises again in lines 617 to 629. These sections should be combined for clarity.

18. Some actual examples of community preference weights used for scoring a PRO instrument are needed to illustrate the issue raised in lines 424 to 430. One would think that weightings derived from community patients representative of those in a trial would be acceptable for scoring. Weightings should not vary from trial to trial unless the trial populations are measurably different.

19. Lines 478 to 480 refer to assessing measurement properties for "future" clinical studies, (i.e. finish this assessment before using the PRO instrument). At the Mayo meeting, the thinking seemed to be that Phase II trials offer an opportunity for such assessments. However, phase II trials sometimes involve diverse objectives and/or samples too small for assessing an instrument's measurement profile. One proposal is that the Agency consider it reasonable to proceed to Phase III with an instrument that demonstrated adequate psychometric properties in a similar population prior to Phase III. Conformation could then be sought that it is also reliable and valid in the larger population of the Phase III trials; or in a parallel study conducted in the same time period as the Phase III trials. The draft guidance should allow for the results of an assessment in such a parallel study to be used to analyze the PRO results of the Phase III trials, assuming the parallel assessment conclusion was that the PRO instrument was valid, the patients in the Phase III and parallel studies were comparable, etc.

20. Validation is usually an ongoing process with each use of an instrument. Lines 579 and 664 discuss the different reasons to reconfirm the validity of a previously validated instrument. Alcon agrees with the need for reaffirming validity. There may be compelling reasons, however, to use an existing instrument with a slightly different population or a slight question modification in phase III trials. Reaffirmation of validity should be permitted to occur concurrently with the trial in a separate study outside the trial, and/or with the trial data.

21. It is unclear why an instrument developer would create a new instrument that produces results similar to an existing instrument as noted in lines 506 to 507. The Sponsor should understand why a new instrument is being developed. If a new instrument is being developed because the existing instrument has flaws, then there may be a question about using the existing instrument to establish the validity of the new instrument.

22. Line 512 refers to "meaningful discrimination". An explanation or definition of "meaningful" is needed.

23. The discussion of the establishment of a MID/MCID requires additional explanation and clarity. Lines 560 to 562 refer to using a distribution-based approach. What is the agency's opinion of using Cohen's effect size approach and substantiation by a clinical anchor point agreed to in advance?

24. Line 571 states: "There may be situations…meaningfulness of an individual's response … rather than a group's response…" These "situations" require additional discussion and explanation. Lines 572 to 577 go on to discuss the Agency's question about defining a responder. Generally speaking, PRO measures would be useful for defining a responder in health conditions where there are no hard objective clinical endpoints. In such cases, statistically significant differences in PRO score improvement, or in proportions of self-defined responders might be useful. In the latter case, the question would be something simple like: "Are you better, the same, or worse?" We would add that if the definition of a responder based on a PRO measure is agreed to, then a MID/MCID would be unnecessary. Conversely, if a MID/MCID based on a PRO measure is established, a responder definition is unnecessary. Sponsors have experienced agency opinions that MID/MCIDs differences need to make claims about differences between treatment groups. This guidance is silent on this subject. This topic should be included in the guidance. Further, achieving an accepted MID/MCID level should be sufficient for a label claim. Comparative claims and comparative MID/MCID differences should not be required.

25. The discussions of language changes and proxy respondents beginning on lines 643 and 692, respectively, are insufficient. The Agency might note that family members or friends, rather than doctors and their staff members, should generally be proxy respondents on PRO questions. Some detail on when proxy responses could be the basis for label claims would be useful (e.g. pediatric patients with a condition lacking in hard clinical endpoints).

26. The source of the statement in lines 725 to 733 suggesting patients would be biased if they saw their previous answers – but only if unblinding occurs is unclear. Lines 732 and 733 call for rigorous study of this philosophical discussion and Alcon recommends that the discussion be removed until the rigorous studies are completed.

27. Lines 807 to 809 discuss the importance of pre-specifying the method in which results will be interpreted for new instruments. Lines 927 to 929 indicate that substantial clinical experience is needed to establish interpretation rules. It is unclear if the Agency intends to work with sponsors on this issue when the instrument is new.

28. Line 940 states "...it is critical to ensure that patients enrolled in a clinical study are impaired in all domains" of a multi-domain instrument being used in the trial. Alcon agrees that if this rule is violated, a composite score using all the domains may not show a change or difference. However, Alcon does not understand why this rule needs to be followed if the a priori intent is to report selected domain scores and use them as the basis for a label claim (assuming the domains and their scores are validated). Alcon believes that that the Agency's comment here stems from the predisposition that all of the domain results must be similar before the data produced by the instrument can be used in any way. Alcon recommends that this concept be revisited.

END