Author's response to reviews

Title: Psychometric properties of the Osteoporosis Patient Assessment Questionnaire (OPAQ) 2.0: results from the Multiple Outcomes of Raloxifene Evaluation (MORE) study

Authors:

Wei Shen (shen_wei_x1@lilly.com)
Russel Burge (rburge@lilly.com)
April N Naegeli (naegelian@lilly.com)
Jeremy Shih (jeremy.shih@inventivhealth.com)
Jahangir Alam (alam_jahangir@lilly.com)
Deborah T Gold (deborah.gold@duke.edu)
Stuart Silverman (stuarts@bhillsra.com)

Version: 5
Date: 29 September 2014

Author's response to reviews: see over
Reviewer's report

Reviewer #1:
Major Compulsory Revisions
• Line 133/206: Were the participants informed of the results of their vertebral fracture assessments following each evaluation? This is important to clarify since knowledge of this (or not) could influence their responses to the HRQoL questions.

RESPONSE: We agree with the reviewer that this information could in fact affect HRQoL responses. Participants in the MORE trial were informed of the results from their vertebral fracture assessments, and a statement has been added to the manuscript.

• Line 192/213: In the ANCOVA analyses done for the various endpoints, what co-variables were accounted for? Several variables could influence the responses to HRQoL independent of having osteoporosis fractures. It will be important to account for those to be more likely to accurately predict changes in HRQoL based on fractures using this tool.

RESPONSE: We thank the reviewer for this insightful comment. In our original analyses, only country of origin was controlled for in the models. Accordingly, we have updated our models to include the following covariates: country of origin, age, body mass index (BMI), years since menopause, smoking status (yes vs. no), alcohol consumption (yes vs. no), and number of preexisting conditions. Tables 4 and 5 have been updated based on these new model specifications.

The statistical significance of the OPAQ domain results in Table 4 were not affected except for dressing/reaching, which is no longer was statistically significant when comparing 0 to 1+ vertebral fractures or 0 to 2+ vertebral fractures; fatigue, which is no longer significant when comparing 0 to 1+ vertebral fracture; and social activity, which is now significant for linear trend and when comparing 0-1 to 2+ vertebral fractures.

In Table 5 for nonvertebral fracture comparison of 0 versus 1 or more nonvertebral fractures, there were 5 domains that became statistically insignificant (walking/bending; standing/sitting; usual work; independence; and back pain). In the comparison of 0 or 1 versus >1 nonvertebral fractures, only 3 domains lost statistical significance (usual work; body image, and independence). There were no changes in significance for any domains when using the updated model for comparison of osteoporosis (femoral neck BMD T-score < -2.5) to osteopenia to normal (femoral neck BMD T-score ≥ -2.5). The text in the results section of the manuscript has been updated to reflect these changes, as appropriate. The results by age in Table 6 showed the greatest impact on the p-values, as 6 domains lost statistical significance at the 10% level and 8 did so at the 5% level. The results section has been updated accordingly.
• Line 256: Patients with past history of fractures (prevalent) will likely have poorer responses to HRQoL compared to those that have never had a fracture independent of their T-scores. While looking at T-score ranges and the effects of T-scores on HRQoL (table 5) were “prevalent fracture” and other variables like age, BMI and years since menopause accounted for in calculating the p-values? If yes, this needs to be clearly specified.

**RESPONSE:** We have updated our models, and results in Tables 4 and 5, to include these other variables, as suggested by the reviewer (see response to previous comment).

Minor Essential Revisions

• Line 112: The description of study population needs to be more specific with a clearer explanation for the selection criteria for the subgroup out of the main MORE study population. Why were only 1477 participants selected to complete the questionnaire?

**RESPONSE:** The text has been modified slightly to state that all 1477 patients who completed the OPAQ instrument were included for this study.

• Line 112: Description of the MORE population does not adequately describe the subgroup (1477) studied in this particular analysis. Although table 1 gave averages, it is unclear for example what the T-score ranges for the subgroup were. I suggest using the “study population” section to describe the 1477 subjects rather than the overall MORE population which was not studied in this analysis.

**RESPONSE:** We have edited this section to more clearly indicate that the sample size for this study was based on the 1477 OPAQ responders within the full MORE study population. However, we believe the summary of the full MORE study population is necessary to better understand the population for our analyses on the OPAQ instrument. Also, in Table 1 we have added minimum and maximum values for the T-scores.

• Line 117: Be specific about what “low BMD” meant using T-scores criteria to define low BMD.

**RESPONSE:** We have added text to indicate that low BMD refers to a BMD T-score less than or equal to -2.5 standard deviations below the young adult peak mean BMD.

• Line 296: It is a strong statement to claim that this questionnaire can detect disease without clarifying if other variables which could also predict disease were considered. At best, we can say it suggests an association rather than imply that it can detect disease.

**RESPONSE:** We have edited the text by replacing “detect” to “were associated with.”
Discretionary Revisions
• Line 78 can be modified to read “….healthcare professionals which shows adequate…..

RESPONSE: The text has been edited as suggested.

• Line 160: Since some responses on the OPAQ2 were reverse-scored, please briefly clarify how the transformation (line 170) was done.

RESPONSE: We thank the reviewer for identifying this lack of clarity. We have added new text which describes the reason for this recoding. In particular, 17 items were reverse coded to be consistent with other items where higher scores represent better health status. The reverse scoring was done before the calculation of scores to avoid any systematic response biases.

• Overall, I found this paper interesting and the well written by the authors. However, there were some unclear sections as outlined above. There were also some limitations that need to acknowledge including the fact that it was a post hoc, retrospective analysis hence limiting the conclusions that can be made from its findings. Also, the design involved several adjustments including dropping some questions that did not fit the models. These questions dropped were reported as not correlating with other questions (question 35) but may be more relevant in assessing the reliability of the OPAQ2.

RESPONSE: We thank the reviewer for this comment as it uncovered an oversight on our part. We reviewed our analyses and found that we actually did not drop question 35 in the social activity domain, as it did not correlate poorly with the other 2 questions in the social activity domain (questions 33 and 34). All questions were part of all further analyses. Also, we have acknowledged a limitation concerning the post hoc nature of this analysis.

Reviewer #2:

Major concerns.
For discriminant validity, 4 categorical groups were assumed by authors. However, they should describe their rationale or references. For Responsiveness, they compared OPAQ score change from baseline to study endpoint between patients with and without incident vertebral fractures. They had better compare the score between baseline and the time of vertebral fracture in the each individual.
**RESPONSE:** In response to the reviewer’s request, we have added references to support the selection of each of the four known groups used for the discriminant validity analyses.

The MORE study collected the PRO instruments and vertebral fracture assessments at baseline, month 12, month 24 and month 36. Thus, exact dates of fractures were not available for the analysis. In addition, it is important to note that in the primary publication of the MORE data (Ettinger B, et al. *JAMA* 1999;282:637-645), the primary endpoint analysis was on the proportion of patients with an incident vertebral fracture, and was not a time-to-event analysis. In the limitations section we have added as statement regarding this shortcoming in terms of lack of data to conduct the time to event analysis. “We were unable to conduct a time to event analysis on HRQoL scores between baseline and the time of vertebral fracture due to the timing of vertebral fracture assessments (baseline, and months 12, 24 and 36).”