Reviewer's report

Title: Using Relative Survival Measures for Cross-Sectional and Longitudinal Benchmarks of Countries, States, and Districts - The BenchRelSurv- and BenchRelSurvPlot-Macros

Version: 2 Date: 13 November 2012

Reviewer: J. David Beatty

Reviewer's report:

The revised manuscript has been reviewed and there are several corrections and several recommendations for clarification or improvement as follows:

1. Introduction, page 3, paragraph 2, line 10 - the second "SWI" should be "STI" with the sentence reading as follows:

"They also described a standardized work-up index (SWI) to address the issue of 'upstaging' of cases and a standardized treatment index (STI) to evaluate outcome using institutional or regional mortality (stage and overall) compared to national mortalities."

2. Data, page 4 - The authors state that the methodology requires high data quality. ("Clinical registers, cohort studies, or health care network databases are also suited if a high data quality and epidemiologic relevance is ensured [9–11]. In the application example, the data of the Surveillance Epidemiology and End Results (SEER) [12] and the Norwegian cancer register [13] fulfill these requirements.") While the reviewer agrees with the importance of high data quality, it is noted that concerns regarding the accuracy and completeness of institution cancer registry and SEER data have been documented. (Reference - Beatty JD, Adachi M, Bonham C, Atwood M, Potts MS, Hafterson JL, Aye RW. Utilization of cancer registry data for monitoring quality of care. Am J Surg. 2011 May;201(5):640-4.) In the Discussion, the authors are recommended to note the documented concern regarding data accuracy and completeness in the SEER breast cancer example used and the resultant limitation this places upon interpretations of results using this methodology for health care decision-making.

3. Standardized Case-Mix Index (SCI), page 5, sentence 2 appears to have errors as written and does not make sense. It is suggested that sentences 2 and 3 be change to:

"The products are then summed up and divided by the total number of ill patients. The index is standardized by comparing benchmark to reference objects and is defined as:"

4. Conceptional pitfalls, page 8, paragraph 2 - Stage migration and the Will Rogers phenomenon are well described in this paragraph, but the authors fail to note that this methodology provides a practical means to examine this issue. It is
recommended that the following sentences be added to the end of this paragraph:

"Comparison of SWI and STI in the method of Beatty et al (6, 7) and in the suggested method facilitates a greater understanding of the contribution of stage migration to the overall outcome. For example, if the SWI is approximately 1.0 but the STI is substantially greater than 1.0, the improved overall outcome corrected for stage mix appears to be primarily a treatment effect. On the other hand, if the SWI is substantially greater than 1.0 but the STI is approximately 1.0, there is a stage migration occurring that does not appear to have a major impact on administered treatments."

5. Weakness, page 9 - Again the authors fail to note that the methodology does provide a means to explore the stage migration and Will Rogers phenomenon. It is recommended that the following be inserted at the end of the first sentence of this section:

"The former (stage migration and Will Rogers phenomenon) is explored in the current method by comparison of SWI and STI, but lead-time bias would distort that assessment and cannot be distinguished from an apparent increase in the incidence of the disease."

6. Weakness, page 9 - The authors mention in the introduction the weakness of the SCI, that it appears in both the numerator and denominator of the OPE. However, the strength of the SCI is that the survival of each stage or group is used to weight the value of the incidence of each category employed in the calculation of the SCI. Clinical relevance is derived by combining both incidence and survival in the SCI. The eSSI is simple to calculate because it is uses the stage or group number and the incidence of each category employed. As a result, the eSSI devalues the earliest disease identified, especially stage 0, and weights the index towards the more advanced disease with the higher stage numbers (stages 2, 3 and 4) and group 5 which represents stage 'not known' (see Table 1). The use of stage and group numbers rather than survival is simple but arbitrary. This lessens the clinical value of both the eSSI and the ROPI. The authors are recommended to include this weakness in the Weakness section.

7. Perspectives, page 9, sentence 1 - This sentence has a typographical error and does not make sense. It is recommended that it be changed to the following:

"Benchmark-algorithms that compare countries, states, and districts are highly complex and require great attention to research details [40,54]."

Note: Items 1, 3 and 7 are minor essential revisions. Items 2, 4, 5 and 6 are major compulsory revisions.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable
**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

'I declare that I have no competing interests'