Reviewer's report

Title: Model-based estimation of measures of association for time-to-event outcomes

Version: 1 Date: 21 January 2014

Reviewer: Ralf Bender

Reviewer's report:

General regression methods based upon pseudo values are proposed to obtain estimates for the effect measures RR, OR, RD and NNT in the case of time-to-event data. The method is applied to 3 real data examples. The paper is interesting but has some limitations. In detail, the following issues deserve attention.

Major Compulsory Revisions

(1) One argument for using methods based upon pseudo values (given in the abstract, example 1 and the discussion) is that the corrected group prognosis (CGP) method applied to the Cox model yields questionable results if the proportional hazards assumption is invalid. However, as correctly described on page 4, the CGP methods can be applied also to other regression models. At first, the basic model has to be adequate which means that the Cox model should not be applied if the proportional hazards assumption is invalid. An adequate model should be chosen for the data at hand and GCP methods should be applied to this model. The interesting comparison is therefore given by CGP vs. pseudo values applied to the same model rather than CGP applied to inappropriate model vs. pseudo values applied to appropriate model. Such a comparison is performed in example 1 (figure 2). However, the presented arguments in favor of the application of pseudo values refer to the misleading comparison. This should be changed.

(2) The question whether the application of pseudo values has advantages over other existing methods to estimate absolute treatment effects for time-to-event outcomes remains unanswered by the paper in its current form. A systematic evaluation of the application of pseudo values in comparison with other methods based on the same basic model by means of a thorough simulation study is required to assess the usefulness of pseudo values. If the authors decide that this is not the scope of the paper it should be clearly stated that the paper just describes an alternative method to estimate absolute treatment effects for time-to-event outcomes but without systematic comparison with other methods.

(3) In the description of the GCP methods and counterfactuals (pages 3-4) the fact is overlooked that different averaging methods are described in the literature in dependence on the study design and the research question (see Bender & Kuss (2010) and Austin (2010b)). In my view, in the first example the averaging
over all subjects as described by Austin (2010a) should be used, whereas in the second example a separate averaging over all exposed and all unexposed persons as proposed by Bender et al. (2007) and Laubender & Bender (2010) is more appropriate. A description of the different averaging methods should be added and in each example the best method should be chosen and applied.

(4) On pages 5-7 the use of different link functions, binomial regression and GLM/GEE modeling are discussed. In this context the results of Gehrmann et al. (2010) are of importance. Gehrmann et al. (2010) showed that the GCP method applied in the context of logistic regression is the preferred method compared to approaches based upon binomial and Poisson regression. I assume that similar results could also be obtained for time-to-event outcomes. Even if a systematic comparison of pseudo values with other methods is not the scope of the paper, the results of Gehrmann et al. (2010) should be discussed. The last paragraph of the discussion also plays a role in this context.

(5) The second example deals mainly with general issues of model building rather than the use of pseudo values to estimate absolute effect measures. The main focus should be latter one. I think the 3 figures of this example can be deleted.

Minor Essential Revisions

(1) In the results section, it is described that two examples are presented. However, in fact three examples are presented in this section.

(2) Some tables summarizing the main results of the examples would be useful.

(3) I think there are too many figures. Especially, the figures dealing general model building rather than the use of pseudo values could be deleted.

Discretionary Revisions

(1) Some references are incomplete and contain question marks (???), e.g., Refs. 3, 22, 27, 28, 33, 35.

(2) There are some typing errors.

References


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

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.