Author's response to reviews

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

Authors:

Federico Ambrogi (federico.ambrogi@unimi.it)
Elia Biganzoli (elia.biganzoli@unimi.it)
Patrizia Boracchi (patrizia.boracchi@unimi.it)

Version: 3 Date: 9 July 2014

Author's response to reviews: see over
Dear Editor

Thank you for deciding to reconsider the manuscript after the first review. We thank the Reviewers for showing some problems still into the presentation and we hope the new version of the manuscript sufficiently addresses all the requests of the Reviewers.

The revised manuscript reports in blue the main revisions (additions or modifications of former text). We hope the revision has improved the paper to a level of satisfaction.

Please find enclosed:
- Revised manuscript
- Rebuttal letter with response to the reviewers comments

Yours sincerely,
Federico Ambrogi
federico.ambrogi@unimi.it

**Reviewer's report**

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

**Version:** 2

**Date:** 21 May 2014

**Reviewer:** Ralf Bender

**Reviewer's report:**

The paper has been improved. However, most of my points are not sufficiently answered. In particular, the following problems remain.

Major Compulsory Revisions

(1) Misleading comparisons and descriptions

Although it was added that a correct model should be the basis of the CGP method, it is still stated in the Abstract that the use of pseudo values is useful when the proportional hazards assumption is not valid. In the Results section it is stated that the performance of the GCP method is problematic when the Cox model is misspecified. Similar arguments can be found in the Results. 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.
We think that there is a possible misunderstanding about the words we use, i.e. “misspecification” and “absence of proportional hazard”. In fact we are using these terms to refer to the presence of time dependent effects. We hope that the new text is clearer than the previous version about the problem of time dependence. The sentence in the abstract “This is especially useful when the proportional hazard assumption is not valid.” was modified to better stress this aspect: ”This may be especially useful when, in presence of time dependent covariate effects, it is not straightforward to specify the correct, if any, time dependent functional form.”

The question of the adequate model and GCP method moreover is more theoretical than practical. In fact it is not always easy to specify an adequate model, in particular when dealing with time dependent covariate effects. In presence of time dependence it is not at all obvious how to model the time dependence itself and the simulation inserted in the manuscript demonstrates that, if a flexible method is used, the results can be contrasted with those obtained with the PV model. In the methods section, when dealing with CGPM we say:” The CGPM can be applied in principle to whatever regression model and an adequate model must be chosen. Considering, for example, the Cox regression model, in presence of time dependent covariate effects, an interaction of the covariates with a pre-specified function of time should be specified, in order to estimate HR(t) varying during follow-up time. It is important to remark that it is not always easy to specify an adequate model in presence of time dependent covariate effects. In fact it is not always obvious how to model the time dependence itself. In general simple functions of time (linear or logarithm) or more flexible alternatives are used [ref].”

In the prostate cancer example the PV approach with time dependent treatment effect is compared with the Cox model with time dependent treatment effect. We think this is the interesting comparison, i.e., the comparison of CGP with Cox and PV both with time dependent covariate effects. In particular the presentation of the results is:

1) there is no evidence of TD effects in Cox according to Schoenfeld residuals.
2) A cox model without time dependence is used in the CGP method and presented in figure.
3) Further considerations using the Cox model suggest the possible presence of TD. Therefore, a Cox model with TD is used in the CGP method and presented in the figure. In particular a flexible way to specify the time-dependence is used (with splines).
4) There is no evidence of TD effects using PV.
5) A constant RD estimate is provided and described in the text.
6) However, as it is interesting to compare the RD(t) estimated by the CGPM, which is by definition time dependent, whatever the regression model used, with the one estimated through the transformation model letting the treatment effect varying in time, the corresponding estimate with PV is provided in the figure.

At the end all the relevant comparison are there. Note that detecting TD effects may not be so straightforward, as the example clearly shows, so speaking about the adequate model may be always questionable in practice.

At the end of the example it was said “…CGPM gives an incorrect picture of the ARR increasing during follow-up until reaching a plateau. The time-dependent estimate obtained from the pseudo-value model is effective in describing the treatment effect during follow-up
time: harmful at the beginning, when cardio-vascular deaths are more frequent, beneficial later on when cancer deaths are more frequent.” In this last sentence, it is not clearly stated which is the Cox model commented on: the one with treatment effect constant through time (proportional hazard) or the one with time dependent treatment effect. In fact, both Cox models, the one without TD and the one with TD, are not capturing the initial harmful effect, so the reference to the model with or without TD is not in principle necessary. However, to be clear, the sentence was changed in “…CGPM applied to the Cox model without a time dependent treatment effect gives an estimate of the RD increasing during follow-up until reaching a plateau. At the same time, the constant estimate obtained using PV provides a distorted estimate constant through follow-up. CGPM applied to the Cox model with a time dependent treatment effect provides an RD estimate not yet capturing the initial harmful treatment effect. The time-dependent estimate obtained from the pseudo-value model is instead effective in describing the treatment effect during follow-up time: harmful at the beginning, when cardiovascular deaths are more frequent, beneficial later on when cancer deaths are more frequent.”

Moreover as the term misspecified can be misleading, also the discussion was changed: “However the Cox model may easily not be the best regression procedure to be applied, simply because of the assumption of proportional hazards. In fact, in presence of time dependent effects Cox regression may be less appealing. This is demonstrated here through a simple simulation. When time-dependent effects in the Cox model are specified using well-known flexible methods, [ref], without committing to a specific functional form, the estimates are not optimal, especially in terms of efficiency. In these circumstances the results of the simple simulation presented here, suggest that the use of the pseudo-value model may represent a valid alternative to the CGPM. The simulation is not exhaustive and more work is needed to fully understand the properties and the relationships among the different estimation methods.”. The term misspecified is no more used into the text.

(2) Insufficient evaluation of PV methods

The question whether the application of pseudo values has advantages over other existing methods to estimate absolute treatment effects for time-to-event outcomes still remains unanswered. Although it was stated that a systematic evaluation of different methods is not the goal of the paper, a simulation small study was performed. However, the results are not sufficiently described with similar misleading arguments as in (1). The main conclusions of this simulation study are given by

(i) The CGP method applied to a correctly specified Cox model leads to comparable or better values in terms of bias, MSE, CI width, and coverage probability than the PV method.

(ii) The PV method with identity link and without time dependence estimating a constant RD leads to a strong undercoverage demonstrating that the estimation of a constant RD is misleading.

We agree on the fact that more work is needed to fully understand the properties of the different methods, and in fact we state that “A small simulation study is provided to show a
preliminary evaluation of the properties of the different estimation procedures.”

However, compared to the first version of the paper, the revised version demonstrates some preliminary answers to the question whether the proposed method has advantages over other methods.

Moreover, time-to-event analyses are really complex, there are many possible ways to simulate the data, it could be really the case that a definitive answer will never be there! The proposed simulation simply suggests that, in certain situations, there could be an advantage with pseudo-values. We agree that more work is needed to have more complete answers, although it will be difficult that simulations could be able to definitely answer all the questions.

In the simple simulation proposed, data were generated according to the Cox proportional hazard model or to the Cox model with a time dependent covariate effect. This is a very particular way of generating survival times, many others are in fact possible.

Under this simulation strategy, it is expected that CGP method applied to the generating model will produce the best results and such method was in fact used as benchmark, as it is clearly stated in the text. The results using PV are therefore commented accordingly. Here the interest is not on the correct Cox model and CGP, which are the benchmarks considered the simulation framework, but on the comparison of the benchmark with the PV and with the Cox model having, for example, a time dependence specification different from the true one. It is in fact not always obvious how to model TD effects and it could happen that there is no a “correct” simple functional form (unless in to the simulation!). Therefore for the first simulation, the relevant result is that the PV model is actually comparable with the best one (CGP with correctly specified Cox) with some problems in efficiency.

Concerning the suggested conclusions, the suggested point (i) is in fact into the text having a different subject, namely PV and not the CGP method applied to a correctly specified Cox model for the motivation just explained.” The method based on pseudo-values with Z time dependent appears effective especially in terms of bias. Confidence interval coverage is good, although the width of the confidence intervals with pseudo-values is fairly large.” When we say “effective” and “good” or “large” we are referring to the benchmark: “RD(t) estimated through the CGPM using the Cox proportional hazard regression model was used as the benchmark estimation method.”

Point (ii) was added, further commenting the constant RD estimate.

In conclusion, for the first part of the simulation we now say “RD estimated through the CGPM using the Cox proportional hazard regression model was used as the benchmark estimation method. The method based on pseudo-values with Z time dependent appears effective especially in terms of bias. Confidence interval coverage is good, although the width of the confidence intervals with pseudo-values is fairly large. It is interesting to observe the results of pseudo-values with the covariate Z not time dependent. In this case the RD estimate is constant through follow-up, a situation which can result from lack of power to detect the time-dependence of the RD. The simulation results appear very interesting for late follow-up times. At time 600, results are very similar to that of the Cox proportional model. However, the PV method with identity link and without time dependence estimating a constant RD leads to a strong undercoverage demonstrating that the estimation of a constant RD is misleading.”

In the second part of the simulation data are generated according to a Cox model with a time dependent covariate effect generated according to the logarithm of time. Here we
specified the TD effect in the Cox model used in CGPM in 2 ways: (1) an interaction with log of time, which is the benchmark in this simulation, (2) an interaction with a flexible spline function of time.

The interest is about the PV model and about the Cox model in which a flexible modelling of the TD effect is used. Actually it is not always obvious how to model a TD effect. The description of this simulation is “In this second simulation the Cox model was specified in two different ways. Specifically, the covariate Z was inserted with an interaction with log(t), the correct one, and with a restricted cubic spline with 3 knots of time. Namely the use of cubic splines for modelling time dependent effects was proposed by Hess, [46], allowing the study of possible covariate-time interactions without having to specify a specific functional form, using a limited number of parameters. The results are reported in table (2). In this simulation the Cox model with the correct specification of the time dependent effect of Z, that is log(t), is used as benchmark estimation. When the time dependence of Z is modelled using the restricted cubic spline, the performance of the CGPM is less appealing, compared to the benchmark, regarding all the parameters considered into the simulation. The pseudo-value model is really a competitor in this situation. It is in particular interesting to observe the 95% confidence interval width. The transformation model using pseudo-values with identity link is a valuable alternative to the CGPM when the time dependent effect in the Cox model is unknown and modelled using a flexible method. Model checking is therefore very important and pseudo-values can be of help also in this case, see the work of Anderson and Perme, [43].”

(3) Insufficient use of the different averaging methods

Although the different averaging methods are now added and applied in one example, the explanation and use of these methods is still insufficient. It would be adequate to shortly mention that these methods exist (with reference to Bender et al. (2007), Bender & Kuss (2010) and Austin (2010b)) but that in this paper averaging is only performed over the whole population because the focus of the paper is the presentation of the PV methods and not the comparison of different averaging methods. Moreover, Bender & Blettner (2002) is the wrong reference for NNE and EIN. The correct reference is given by Bender et al. (2007).

We apologize for having inserted the wrong reference for NNE and EIN. Now all the citations are inserted with the additional one of Laubender and Bender 2010, which actually deals with time to event data. However we are surprised by the comment about the insufficient explanation and use of the methods. The different averaging methods are mentioned in the introduction, described in the methods (end of paragraph “Corrected Group Prognosis Method and developments”), and in the results when dealing with the observational study, as was previously requested. We further added in the methods section that it is not the goal of the paper the comparison of the different averaging techniques: “A comparison of the model based estimated RD(t) with that obtained through different averaging techniques, namely NNE(t) and EIN(t) [ref], is provided in the second example. However, the focus of the paper is not the comparison of different averaging techniques which are provided only for illustrative purposes. In particular, only the estimates obtained through the averaging performed over the whole population are
compared with those based on transformation models methods.”.

(4) Insufficient discussion of different link functions in the framework of logistic regression

Although the paper by Gehrmann et al. (2010) is now shortly mentioned in the Discussion, the corresponding main results are not sufficiently linked to the arguments and results presented in the paper. The main conclusion of Gehrmann et al. (2010) is that the CGP method applied to logistic regression is the preferred method to estimate RDs and NNTs adjusted for covariates compared to binomial, Poisson and linear regression methods that directly estimates the RD (similar to PV with identity link) even if the fitted response function differs from the true response function. Whether a similar results holds for the Cox model has to be explored in a thorough simulation study (which may be beyond the scope of the paper).

The work of Gehrmann is in a different (and simpler) context. Care should be used in linking the present work to that one. We inserted the sentence suggested by the Reviewer with the specification that the setting is fairly different: “Considering a similar problem in the context of logistic regression, Gehrmann and colleagues, [49], concluded that the CGPM applied to logistic regression is the preferred method to estimate RD and NNT adjusted for covariates compared to binomial, Poisson and linear regression methods that directly estimates the RD (similar to pseudo-values with identity link) even if the fitted response function differs from the true response function. The context of time-to-event outcomes is more complex than that of logistic regression especially for the problem of time dependent effects. Whether similar results hold for the Cox model has therefore to be further explored thorough a series of simulation studies”.

Minor Essential Revisions

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

The reply that 2 tables are added showing the results of the simulation study is not a reply to my point.

Actually we thought that tables would not be necessary to the explanation of the examples, which are fully described in 5 figures, some of which composed of multiple panels. We decided instead to add two tables to explain the simulation study. This is the motivation of the previous answer. We apologize for the lack of explanation. We recognize that a table to describe the different estimates could be useful especially in the prostate cancer example. We therefore inserted a table summarizing the results of the prostate cancer example. We think the second example is already fully described by the figures inserted.

Discretionary Revisions

(2) There are some typing errors. No reply provided.

We thank the Reviewer for highlighting this problem. We hope there are no more typos in the manuscript.
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.
The revision by the authors has substantially improved the paper. The focus is now more clear with more emphasis on estimation methods and less emphasis on the association measures themselves. This is fair enough and makes sense.

However, I feel the problem of the definition of the measures has not been solved completely in an adequate manner. For example, the definitions given on page 2 are just not definitions, as they have a time dependent quantity on the right side, and a time constant quantity on the left side. I think it would contribute to the clarity of the paper, if all quantities are introduced as time dependent parameters, and that it is explicitly stated, where in the different procedures the assumption of a constant RD, HR or RR is made, or not made.

We thank the reviewer to show the erroneous definition of the different measures, which, in the most general definition, are clearly time dependent. Specifically we added the time dependency into the definitions, and the entire manuscript, including figures, tables and corresponding legends, was revised clearly specifying when the measures are time dependent or not. Moreover, after the definition of the different measures in the introduction it was added: "The different measures of effect are in general time-varying. In certain situations, however, they are estimated as constant through follow-up, as it happens for example with the Cox proportional hazard model for the hazard ratio. When the measure is assumed to be constant during follow-up the time dependence is omitted (i.e., HR(t) is written as HR)". We hope the new version makes now clear where time dependence is assumed.

Level of interest: An article of importance in its field 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.