Author's response to reviews

Title: Matching methods to create paired survival data according to an exposure occurring over time: a study based on survival data simulations with application to breast cancer

Authors:

Alexia A Savignoni (alexia.savignoni@curie.fr)
Caroline C Giard (caroline.giard@curie.fr)
Pascale P Tubert-Bitter (pascale.tubert@inserm.fr)
Yann Y De Rycke (yann.de-rycke@curie.fr)

Version: 3 Date: 14 March 2014

Author's response to reviews: see over
Original Article: "Matching methods to create paired survival data according to an exposure occurring over time: a study based on survival data simulations with application to breast cancer" (MS: 1366115683112735).

Dear Arlene Pura,

Please find enclosed a revised version of the manuscript: "Matching methods to create paired survival data according to an exposure occurring over time: a study based on survival data simulations with application to breast cancer" (MS: 1366115683112735).

Find also enclosed in this letter a point-by-point reply to the reviewers.
An English native speaker has revised the article.

Thank you for considering our manuscript for publication in BMC Medical Research Methodology.

With best regards,

Dr. Alexia Savignoni
Corresponding Author
Département de Recherche Clinique
Institut Curie
26 rue d’Ulm
75005 Paris, France
Phone: (33) 1 56 24 56 02
E-mail: alexia.savignoni@curie.fr
1 REVIEWER 1

1.1 Major compulsory revisions

1.1.1 The paper has to be shortened.

We have followed your advice and shortened the paper: some details of the simulation study are now in the appendix (see next comment), and the redundant sentences as well as the section "Conclusions" have been removed (see comment 1.1.5).

1.1.2 Restructure the paper so that only the most important details of the simulation study would be in the main part of the article while other explanatory material would be in the Appendix.

We have left the important details of the simulation study in the main part and transferred the other explanatory material to Appendix A. All the details concerning the simulation of the time of the events according to the "illness-death" model and the truncation phenomenon, the instantaneous risk functions \( \lambda_{uv}(t) \) chosen, the \( \beta_{uvk} \) triplet values and the censoring proportion are now in the Appendix Section.

1.1.3 It would be interesting to see simulation scenario where data is simulated not from the "illness-death" model but rather from a Cox proportional hazards model with a time-dependent covariate representing exposure. In this simple model, the effect of the exposure would be known directly and would salty lend itself to the comparison with obtained estimates. Please consider adding a limited simulation study of this kind.

Simulating the data through a Cox proportional hazards model with a time-dependent covariate representing exposure would have been simpler. The effect of the exposure would be known directly. However it does not make it possible to study the phenomenon of interest: the selection bias, \( i.e. \) the healthy mother effect and the interaction between our exposure (\( i.e. \) the pregnancy) and the covariates. Thus it seemed to us easier to simulate through an "illness-death" model: the healthy mother effect is modelled by the transition between the "initial state" (\( i.e. \) the cancer diagnosis) and the "exposure" (\( i.e. \) the pregnancy) and the interactions of interest were performed by creating different effects of the covariates regarding the transition \( 2 \rightarrow 3 \) ("exposure" to "final state") and \( 1 \rightarrow 3 \) (transition "initial state" to "final state").

Thus it seemed to us more relevant and especially simpler to simulate through an "illness-death" model, in order to study our particular context. However, by doing so, we could not obtain an exact algorithm to calculate the effect of the exposure adjusted on the covariates (\( HR_a(t) \)). We proposed an empirical method to evaluate this parameter, that we included in the Appendix section of the present revised version of the paper. In the following, you will find our arguments concerning this approximation.

Let’s consider \( L \) intervals \( I_l \) (\( l = 1 \) to \( L \)) inside the study interval \([0; T_{\text{max}}]\), where all the \( i \) subjects are censored at time \( T_{\text{max}} \): \( I_l = [a_{l-1}; a_l] : l = 1, \ldots, L \) with

\[
a_0 = 0 < a_1 < a_2 < \cdots < a_L = T_{\text{max}}
\]
Inside each $I_t$, the exposure variable $E_l(t)$ is as follows:

$$E_l(t) = \begin{cases} 1 & \text{if } t \geq t_{E_l} \text{ and } t \in I_t \\ 0 & \text{otherwise} \end{cases}$$

Then we define: $E(t) = \sum_{l=1}^{L} E_l(t)$.

Our model of simulation obtained from an "illness-death" model is as follows:

$$\lambda(t, Z, E) = \lambda_0(t) \exp \left( \sum_{l=1}^{L} \gamma_l E_l(t) + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + (\alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3) E(t) \right),$$

where $\lambda(t, Z, E)$ is the instantaneous risk function of state 3 occurrence according to the exposure $E$ and the covariates $Z_k$ ($k = 1$ to $3$).

From this model, we are able to calculate $HR_i(t)$, i.e. the $HR(t)$ for each profile $(Z_1, Z_2, Z_3)$:

$$HR_i(t) = \frac{\lambda(t, Z, E = 1)}{\lambda(t, Z, E = 0)}.$$ 

We approximate $HR_a(t)$ by

$$\overline{HR}_a(t) = \exp \left( \sum_{l=1}^{L} \gamma_l E_l(t) \right),$$

where the $\gamma_l$ are obtained from

$$\lambda(t, Z, E) = \lambda_0(t) \exp \left( \sum_{l=1}^{L} \gamma_l E_l(t) + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 \right).$$

We approximate $HR(t)$ by

$$\overline{HR}(t) = \exp \left( \sum_{l=1}^{L} \gamma_l E_l(t) \right),$$

where the $\gamma_l$ are obtained from

$$\lambda(t, Z, E) = \lambda_0(t) \exp \left( \sum_{l=1}^{L} \gamma_l E_l(t) \right).$$

For $\overline{HR}_a(t)$, the estimation of $\gamma_l$ is adjusted for the $\beta_k$, whereas for $\overline{HR}(t)$ it is not.

The approximation of $HR_i(t)$, $\overline{HR}_a(t)$ and $\overline{HR}(t)$ is especially relevant when $L$ and the total number of patients are large, in order to have enough exposure and events inside the transitions $2 \rightarrow 3$ and $1 \rightarrow 3$.

The approximations of $HR_i(t)$ and $\overline{HR}(t)$ could be compared to the theoretical values (See Appendix B), not of $\overline{HR}_a(t)$.

To be sure of the approximation of $\overline{HR}_a(t)$, we simulated data from a Cox proportional hazards model with a time-dependent covariate representing exposure. As we obtained directly the effect of the exposure we were able to certify that our approximation’s method was correct.
1.1.4 Introduction to the Real Data Example of section 1.3.5 should be moved to Section 2.2 leaving section 1.3 devoted to describing simulations only. Then section 2.3 would completely describe the breast cancer study. Alternatively, you could make all preliminary remarks in Section 1, simulations could be described in Section 2, the Monte Carlo simulation design as well as its results and Section 3 could include the real data example in its entirety.

We have taken this remark into account.

1.1.5 The Discussion (section 3) again reiterates some of the points that were already addressed in the previous sections and is redundant. New information from this section could be just absorbed by an appropriate earlier section. All the findings could be summarized in one section, either "Discussion" or "Conclusions". Please consider eliminating one of these sections. That will help to make the paper shorter.

The "Conclusions" section and the redundant information have been deleted or reorganized.

1.1.6 Include "matching on time dependent covariates" and "matched time-to-event data" (or "matched survival data") as key words instead of "matching paired data". You may consider replacing key word "semi-parametric models" with more specific "marginal Cox model" and "stratified Cox model".

The key words have been changed.

1.2 Minor changes and editorial revisions

We have taken into account all your remarks, from 1 to 35, which have been modified as recommended.

The last remark (number 36): "Page 26: reference#24 – capitalize Wilcoxon" was due to the LATEX compilation step but this problem has been solved.
2 REVIEWER 2

2.1 The authors proposed approach heavily relies on particular model-based numerical experiments.

2.2 We cannot find any solid perspective and usefulness in practice. Generally, the contribution is not strong enough for publication.

Any simulation studies are based on a simulation model. In the description of the results, we especially focused on the simulation scenario that was the closest to our real data. However, we also studied many other scenarios, which were realistic and relevant concerning some of the situations encountered in oncology. For all scenarios, the models of analysis (Holt and Prentice model, and Lee Wei and Amato model) are different from the simulation model (an "illness-death" model). We believe that these results can be generalized. All these simulations allowed us to assess the matching method we propose. For all the different simulated scenarios, our conclusions concerning the performance of the combination of our matching method and the model to be applied are identical. The Method 1 of matching is the common method used in the litterature when the criterion which characterizes the pair is an exposure occurring over time. In view of our simulations we demonstrated that the Method 2, our new approach, actually improves the estimation of the effect of the exposure. This work has been motivated by a real clinical pending issue. Our large simulations studies provide arguments for the validity and usefulness in practice for the proposed matching method.