Bayesian analysis of restricted mean survival time adjusted on covariates using pseudo-observations

Léa Orsini[1,2], Emmanuel Lesaffre[2], Guosheng Yin[3], Caroline Brard[4], David Dejardin[5] & Gwénaël Le Teuff[1]

[1] *CESP, INSERM U1018, Université Paris-Saclay, UVSQ, Villejuif, France, lea.orsini@gustaveroussy.fr, gwenael.leteuff@gustaveroussy.fr*
[2] *I-Biostat, KU-Leuven, Leuven, Belgium, emmanuel.lesaffre@kuleuven.be*
[3] *Department of Mathematics, Imperial College London, London, England, guosheng.yin@imperial.ac.uk*
[4] *Ipsen Innovation, Clinical Development Organisation, Les Ulis, France, caroline.brard@ipsen.com*
[5] *Product Development, Data Sciences, F. Hoffmann-La Roche AG, Basel, Switzerland, david.dejardin@roche.com*

Résumé. La différence de moyennes de temps de survie restreintes (dRMST) à un temps donné est une mesure adéquate pour quantifier l'effet traitement entre deux bras d'un essai clinique randomisé lorsque l'hypothèse des risques proportionnels n'est pas vérifiée. C'est une situation courante dans les essais d'immuno-oncologie. Plusieurs méthodes fréquentistes existent pour estimer la RMST, basées sur la modélisation et l'intégration de la fonction de survie. Une approche plus naturelle est de considérer un modèle de régression directement sur la RMST utilisant les pseudo-observations qui permettent d'éviter la modélisation de la fonction de survie. Cette approche offre aussi la possibilité d'étendre l'analyse de la RMST pour un temps donné à l'analyse jointe de la RMST à plusieurs temps. Seules deux méthodes Bayésiennes existent et modélisent la fonction de survie avec un mécanisme de priors non-paramétrique. Nous avons développé une nouvelle méthode Bayésienne, basée sur les pseudo-observations et la méthode des moments généralisées, qui offre une estimation de la RMST ajustée sur des covariables, sans avoir besoin de modéliser la fonction de survie, ce qui la rend attractive par rapport aux méthodes Bayésiennes existantes. Une étude de simulation d'essais randomisés à deux bras, avec différents effets traitement dépendant du temps et covariables, a été effectuée et montre que cette nouvelle méthode donne des résultats valides, cohérents avec les méthodes existantes et aussi une estimation plus précise après ajustement sur les covariables. A titre d'illustration, les différentes méthodes ont été appliquées aux données de l'essai Getug-AFU 15, essai randomisé de phase 3 comparant une thérapie de déprivation androgénique seule ou en complément du docetaxel chez des patients avec un cancer prostatique métastatique hormono-naif, pour analyser la survie sans progression biologique. Cette illustration montre également l'avantage de la nouvelle approche proposée pour analyser la RMST avec ajustement sur les covariables dans le cadre Bayésien.

Mots-clés. Méthode des Moments Généralisée Bayésienne, Risques non proportionnels, Pseudo-observations, Moyenne Restreinte du Temps de Survie.

Abstract. The difference in restricted mean survival time (dRMST) at a specific time point is an appropriate measure to quantify the treatment effect between two arms in randomized clinical trials when the proportional hazards assumption does not hold. This is common in the context of immuno-oncology therapies. Several frequentist methods exist to estimate RMST based on modeling and integrating the survival function. A more natural approach is to consider a regression model on the RMST directly using pseudo-observations which allows for a direct fit without modeling the survival function. This approach also has the advantage of extending the analysis for single time point of interest to the joint analysis at multiple time points. Only two Bayesian methods exist, and both model the survival function with a nonparametric prior process. We develop a new Bayesian method based on pseudo-observations and the generalized method of moments (GMM) that offers RMST estimation adjusted on covariates without the need to model the survival function, making it attractive compared to existing Bayesian methods. A simulation study of 2-arms randomized clinical trials with different time-dependent treatment effects and covariates effects was conducted, demonstrating that this new approach yields valid results, consistent with existing methods, and shows improved precision after covariates adjustment. For illustration, the methods were applied for analyzing the PSA progression-free survival of the Getug-AFU 15 trial, a randomized, open-label, phase 3 trial comparing an androgen-deprivation therapy alone or with docetaxel in non-castrate metastatic prostate cancer. This illustration also demonstrates the advantage of this new method to perform RMST analysis with covariates adjustment in the Bayesian framework.

Keywords. Bayesian generalized method of moments, Non-proportional hazards, Pseudo-observations, Restricted mean survival time, Survival Analysis.

1 Introduction

The difference of restricted mean survival time (dRMST) has been proposed as a clinically meaningful estimand of the treatment effect in randomized clinical trials, especially when the proportional hazards (PH) assumption does not hold. This situation often occurs in immuno-oncology therapies where the treatment effect is time-dependent (e.g., early or delayed effect). In this case, the interpretation of the hazard ratio as a treatment effect measure becomes difficult. Alternatively, the RMST can be used, interpretable as the expected life time experienced out of τ units of time. For example, if the dRMST between experimental and control groups at $\tau = 5$ years is 1 year, it means that, on average, the experimental treatment increases life expectancy during the next 5 years by 1 year, compared to the control treatment.

One straightforward approach to estimating RMST at a certain time τ is to numerically integrate the Kaplan-Meier curve between 0 and τ. However, this approach does not allow for covariates adjustment, which is a major limitation because omitting important covariates results in less precision, see Karrison (2018). One way to adjust the RMST estimation on covariates is to model the survival function with a parametric or semi-parametric model and integrate it, see Karrison (1987) and Zucker (1998). A more natural approach is to fit a

regression model on the RMST directly through estimating equations instead of modeling the survival function. In this case, censoring must be handled either using the inverse probability of censoring weights, see Tian et al. (2014) or pseudo-observations, see Andersen et al. (2004).

The existing Bayesian research on RMST is limited to two approaches. Recently, Zhang and Yin (2023) proposed a Bayesian nonparametric analysis of RMST for right and interval-censored data by assigning a mixture of Dirichlet processes (MDP) prior to the distribution function. However, covariates adjustment is unavailable with this approach. Chen et al. (2023) overcomes this limitation, also considering a nonparametric dependent mixture model. In this paper, we extend the method of Andersen (2004) based on the pseudo-observations to the Bayesian framework, offering multivariable dRMST estimation without the need to model the survival function, providing an attractive alternative to existing Bayesian methods. This work follows a previous manuscript by Orsini et al. (2023) that is currently under review and discusses the Bayesian analysis of pseudo-observations with the Bayesian GMM to estimate hazard ratios.

2 Methods

Suppose that $\tilde{T}_i$ the time-to-event variable for the i-th subject, Z_i a p-dimensional baseline covariate vector, A_i the treatment allocation variable, and C_i a right censoring random variable, independent of $\tilde{T}_i$, Z_i and A_i. We observe $T_i = \tilde{T}_i \wedge C_i$ and $\Delta_i = I(\tilde{T}_i \leq C_i)$ the event indicator. For a pre-specified time point of interest τ, the τ-RMST is defined as:

$$\text{RMST}(\tau) = E(\tilde{T} \wedge \tau) = \int_0^\tau S(t)dt.$$

To adjust the RMST estimation on covariates, the following regression model can be considered:

$$\mu_i = E(\tilde{T}_i \wedge \tau | A, Z) = g^{-1}(\alpha + \delta A + \beta_1 Z_1 + \cdots + \beta_p Z_p)$$

where $g(\cdot)$ is a monotone differentiable link function and $\beta = (\alpha, \delta, \beta_1, \ldots, \beta_p)^T$ the vector of unknown parameters. With an identity link function, the regression coefficient, δ, can be interpreted as the dRMST between the two arms.

We propose a Bayesian regression approach based on pseudo-observations to fit this model. Following Andersen et al. (2004), the i-th pseudo-observation is computed as

$$y_{\tau,i} = n \int_0^\tau \widehat{S}(t)dt - (n-1) \int_0^\tau \widehat{S}^{-i}(t)dt$$

with n the sample size, $\widehat{S}(t)$ the Kaplan-Meier (KM) estimator of the survival probability, and $\widehat{S}^{-i}(t)$ the KM estimator excluding the i-th subject. Because of the unbiasedness of pseudo-observations conditional on covariates proved in Overgaard et al. (2017), we can replace the non-observed (due to censoring) $\tilde{T}_i \wedge \tau$ by $y_{\tau,i}$ in the regression model.

The Bayesian generalized method of moments (GMM) is used to estimate the posterior distribution $p(\beta|y_\tau) \propto \tilde{L}(\beta|y_\tau)p(\beta)$ where the pseudo-likelihood $\tilde{L}(\beta|y_\tau)$ is defined following

Yin (2009) as:

$$\tilde{L}(\beta|y_\tau) \propto \exp\{-\frac{1}{2}U_n^T(\beta)\Sigma_n^{-1}(\beta)U_n(\beta)\},$$

where

$$\Sigma_n(\beta) = \frac{1}{n^2}\sum_{i=1}^{n} u_i(\beta)u_i^T(\beta) - \frac{1}{n}U_n(\beta)U_n^T(\beta)$$

is a $(p+2)\times(p+2)$ matrix with $u_i(\beta) = \frac{\partial \mu_i}{\partial \beta}(y_{\tau,i} - \mu_i)$ and $U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} u_i(\beta)$.

3 Simulation study

A simulation study of 2-arms randomized clinical trials with time-to-event outcomes was conducted to assess the performances of the Bayesian GMM based on pseudo-observations, to compare them with some of the RSMT estimators mentioned above (Andersen et al.(2004), Tian et al. (2014) and Zhang and Yin (2023)), and to evaluate the usefulness of covariates adjustment. Event times were simulated following a Weibull distribution, with scale and shape parameters chosen to mimic different patterns of treatment effect: PH (scenario 1), non-PH with early effect (scenarios 2 and 4), and delayed effect (scenarios 3 and 5). We also investigate the effects of covariate adjustment, drawn from uniform distribution (scenario 4) or normal and binomial distributions (scenario 5). Covariate effects were generated to be proportional. In all scenarios, a 30% censoring rate was considered, drawn from a uniform distribution with an administrative censoring at 8 years. Figure 1 displays the underlying survival curves for each scenario. The restriction time τ was set to 5 years, and 1000 replicates were generated for all scenarios. For the Bayesian approaches, noninformative priors $N(0, \sqrt{10}^2)$ were specified for all parameters of the Bayesian GMM regression model with pseudo-observations, and a mixture of Dirichlet processes prior was applied under an exponential base measure with Gamma $\Gamma(0.01, 0.01)$ mixing distribution for the method in Zhang and Yin (2023).

The main results are RMST estimation with covariates adjustment, corresponding to Scenario 4 with $n = 500$ (Table 1). The Bayesian GMM based on pseudo-observations gave similar results to the other methods. All the methods allowing for adjustment on covariates produced slightly more precise estimates. Similar results were observed in the other scenarios (data not shown).

4 Illustration on real data

For illustration, we analyzed the data from the Getug-AFU 15, a randomized, open-label, phase 3 trial comparing an androgen-deprivation therapy (ADT) alone ($n = 193$) or with docetaxel ($n = 192$) in non-castrate metastatic prostate cancer. The median follow-up time was 4.2 years. We focused on the Prostate-Specific Antigen (PSA) progression-free survival endpoint for which the PH hypothesis was rejected ($p = 0.00022$, Grambsch and Therneau). The survival curves for the two treatment groups show a diminution of the treatment effect

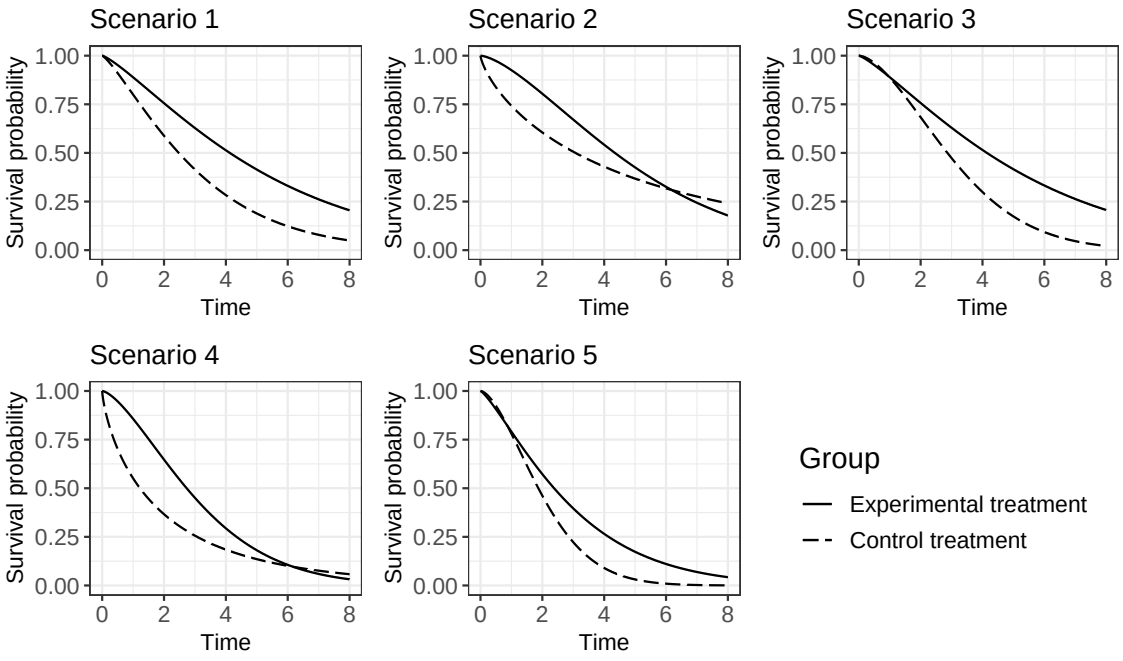

Figure 1: Theoretical survival curves for each simulated scenario.

Methods	Bias	ASE[1]	RMSE[2]	95% Coverage
Frequentist				
Kaplan-Meier estimator	0.0055	0.163	0.167	93.9
Andersen et al. (2004)*	0.0037	0.156	0.159	94.9
Tian et al. (2014)*	0.0032	0.155	0.159	94.9
Bayesian				
Zhang and Yin (2023)	0.0054	0.162	0.167	93.9
GMM*	0.0060	0.156	0.158	94.7

[1] ASE: Average Standard Error, [2] RMSE: Root Mean Square Error

* Model adjusted on the prognostic variable $Z_1 \sim U([0, 2])$

Table 1: Performance of frequentist and Bayesian methods for the estimation of the difference of 5-RMST between 2 arms in Scenario 4.

over time (Figure 2). The difference in 5-RMST was estimated around 0.58 year for all methods (Figure 3), meaning that receiving docetaxel in addition to ADT increases the life time without PSA progression during the next 5 years by 0.58 a year, compared to receiving ADT alone. Estimating unadjusted 5-RMST with the Bayesian GMM approach based on pseudo-observations gave similar results compared to the existing methods. After adjustment on four prognostic variables (gleason score, European Cooperative Oncology Group performance status, concentration of alkaline phosphatase, and presence of bone metastases), in complete-case, an increase in precision was observed.

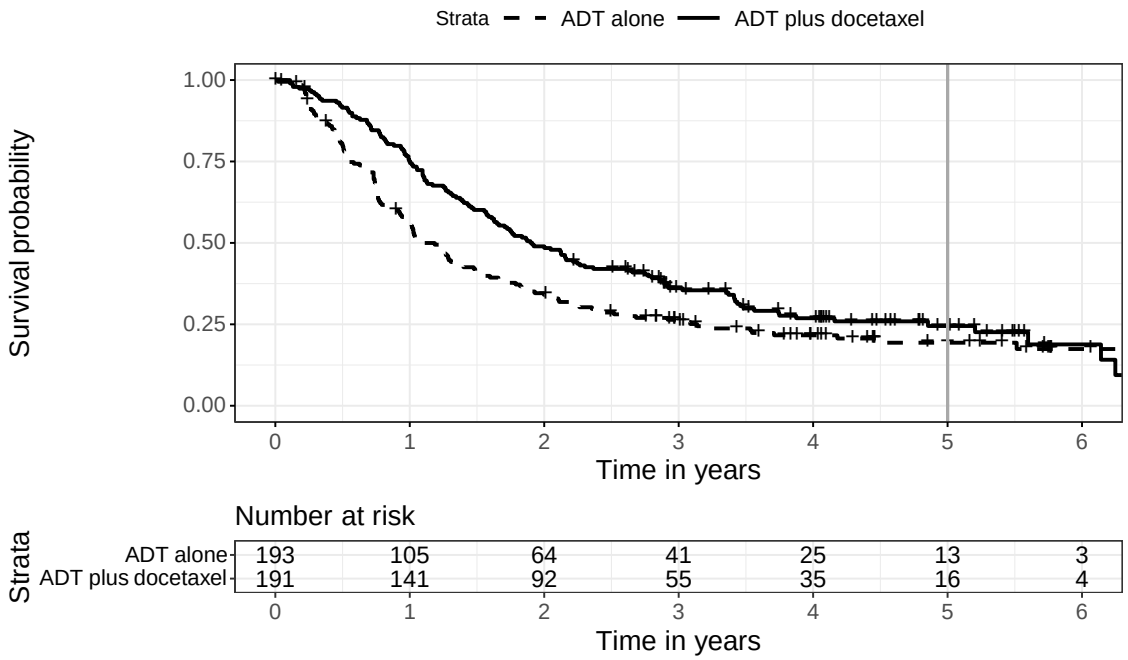

Figure 2: Kaplan-Meier curves for the PSA progression-free survival from the Getug-AFU 15 trial. The vertical grey line represents the restricted time $\tau = 5$ year.

5 Discussion

This paper introduces a novel Bayesian approach for analyzing RMST, based on pseudo-observations and the Bayesian generalized methods of moments. The first advantage is to eliminate the need to specify and integrate the survival function by fitting a regression model on the RMST directly, making this method more straightforward compared to existing Bayesian approaches. The second advantage is to provide a direct estimation of RMST adjusted on covariates in the Bayesian framework. Caution must be taken regarding the potential misspecification of the model, especially with continuous covariates, since the functional form of the relationship between the covariate and outcome must be specified. The third advantage is to provide the probability of the dRMST being above any desired cut-off, given from the posterior distribution.

Whatever the methods used for RMST analysis, the critical point is the pre-specification of the time point of interest τ. To avoid this arbitrary choice, Ambrogli et al. (2022) developed RMST curve estimations based on pseudo-observations. Further research will extend the Bayesian GMM based on pseudo-observations to jointly analyze RMST at multiple time points. This can be achieved by considering a vector of pseudo-observations.

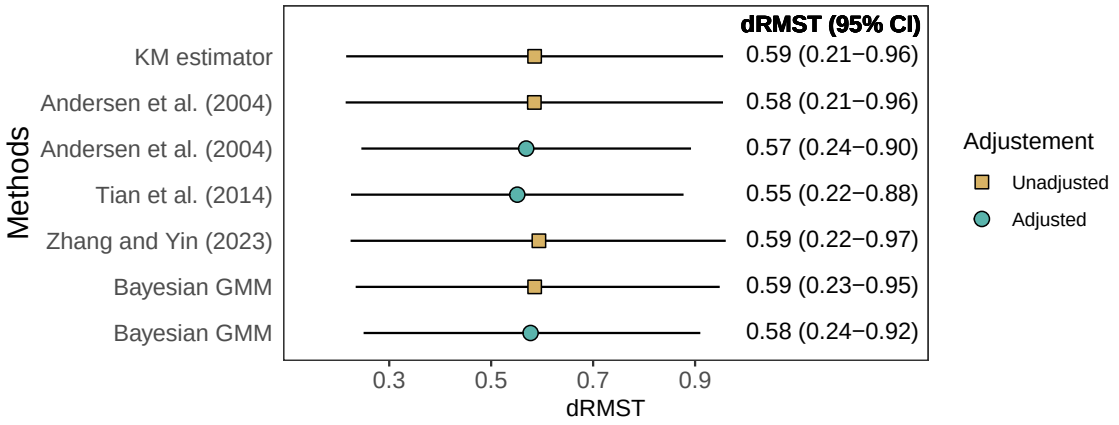

Figure 3: Estimation of the difference of 5-RMST between the ADT plus docetaxel and ADT alone groups for the PSA progression-free survival from the Getug-AFU 15 trial, the horizontal lines represent the 95% confidence or credibility intervals.

Bibliographie

Andersen, P. K., Hansen, M. G., and Klein, J. P. (2004), Regression Analysis of Restricted Mean Survival Time Based on Pseudo-Observations. *Lifetime Data Analysis*, 10(4), pp. 335-350.

Ambrogi, F., Iacobelli, S., and Andersen, P. K. (2022). Analyzing differences between restricted mean survival time curves using pseudo-values. *BMC Medical Research Methodology*, 22(1), pp. 71.

Chen, R., Basu, S., and Shi, Q. (2023). Restricted Mean Survival Time Estimation Using Bayesian Nonparametric Dependent Mixture Models. *arXiv* (arXiv:2305.14639).

Karrison, T. (1987). Restricted Mean Life With Adjustment for Covariates. *Journal of the American Statistical Association*, 82(400), pp. 1169-1176.

Karrison, T., and Kocherginsky, M. (2018), Restricted mean survival time: Does covariate adjustment improve precision in randomized clinical trials? *Clinical Trials*, 15(2), pp. 178-188.

Orsini, L., Brard, C., Lesaffre, E., Guosheng, Y., Dejardin, D., and Le Teuff, G. (2023) Bayesian generalized method of moments applied to pseudo-observations in survival analysis. *Submitted for review*

Overgaard, M., Parner, E. T., and Pedersen, J. (2017). Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations. *The Annals of Statistics*, 45(5), 1988–2015.

Tian, L., Zhao, L., and Wei, L. J. (2014), Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics*, 15(2), pp. 222-233.

Yin, G. (2009), Bayesian generalized method of moments. *Bayesian Analysis*, 4(2), pp. 191-

208.

Zhang, C., and Yin, G. (2023), Bayesian nonparametric analysis of restricted mean survival time. *Biometrics*, 79(2), pp. 1383-1396.

Zucker, D. M. (1998), Restricted Mean Life with Covariates: Modification and Extension of a Useful Survival Analysis Method.*Journal of the American Statistical Association*, 93(442), pp. 702–709.