

RESEARCH

Open Access



KMSubtraction: reconstruction of unreported subgroup survival data utilizing published Kaplan-Meier survival curves

Joseph J. Zhao¹, Nicholas L. Syn¹, Benjamin Kye Jyn Tan¹, Dominic Wei Ting Yap¹, Chong Boon Teo¹, Yiong Huak Chan^{2*} and Raghav Sundar^{1,3,4,5,6*}

Abstract

Background: Data from certain subgroups of clinical interest may not be presented in primary manuscripts or conference abstract presentations. In an effort to enable secondary data analyses, we propose a workflow to retrieve unreported subgroup survival data from published Kaplan-Meier (KM) plots.

Methods: We developed *KMSubtraction*, an R-package that retrieves patients from unreported subgroups by matching participants on KM plots of the overall cohort to participants on KM plots of a known subgroup with follow-up time. By excluding matched patients, the opposing unreported subgroup may be retrieved. Reproducibility and limits of error of the *KMSubtraction* workflow were assessed by comparing unmatched patients against the original survival data of subgroups from published datasets and simulations. Monte Carlo simulations were utilized to evaluate the limits of error of *KMSubtraction*.

Results: The validation exercise found no material systematic error and demonstrates the robustness of *KMSubtraction* in deriving unreported subgroup survival data. Limits of error were small and negligible on marginal Cox proportional hazard models comparing reconstructed and original survival data of unreported subgroups. Extensive Monte Carlo simulations demonstrate that datasets with high reported subgroup proportion ($r = 0.467$, $p < 0.001$), small dataset size ($r = -0.374$, $p < 0.001$) and high proportion of missing data in the unreported subgroup ($r = 0.553$, $p < 0.001$) were associated with uncertainty are likely to yield high limits of error with *KMSubtraction*.

Conclusion: *KMSubtraction* demonstrates robustness in deriving survival data from unreported subgroups. The limits of error of *KMSubtraction* derived from converged Monte Carlo simulations may guide the interpretation of reconstructed survival data of unreported subgroups.

Background

Advances in secondary data analysis of survival data has been made by Guyot et al. [1]. This enables retrieval of individual patient data (IPD) from reported Kaplan-Meier (KM) plots [2, 3]. However, this is only amenable for KM plots presented in the original publication. Often, subgroups of interest in clinically negative randomized controlled trials (RCTs) may not be presented in KM plots. Where subgroups of interest are not reported (or more commonly presented as a summary statistic

*Correspondence: medcyh@nus.edu.sg; mdcragh@nus.edu.sg

² Biostatistics Unit, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

³ Department of Haematology-Oncology, National University Cancer Institute, National University Hospital, 1E Kent Ridge Road, Singapore 119228, Singapore

Full list of author information is available at the end of the article



in forest plots), IPD retrieval utilizing such algorithms becomes impossible.

In this era of precision medicine, where molecular markers or gene expression levels are increasingly available to clinicians, treatment should ideally be rendered to the patient when their clinicopathological profile is predictive of a good treatment response [4–6]. This is critical in deciding approvals for unique subgroups, or conversely for discouraging use in the opposing subgroup – where negative biomarkers have demonstrated the lack of benefit [7]. In addition, cost-effectiveness analysis (CEA), a major consideration in deciding against therapeutic strategies at a public health level, requires IPD data for the appropriate analysis. While new biomarkers for therapeutic regimens have been sought after and studied in trials, they remain exploratory and are often underpowered for interpretation. To overcome this, meta-analytic techniques may be exploited to pool data together to increase statistical power.

In an effort to enable secondary data analyses in such circumstances, we propose a workflow to retrieve unreported subgroup survival data from published KM plots. This method was first utilized to retrieve unreported overall and progression free survival of patients treated with first-line immune checkpoint inhibitors with advanced gastroesophageal adenocarcinoma with a tumor combined positive score (CPS) of 1–4 and 1–9 from CheckMate-649 and KEYNOTE-062 respectively [8–10]. Here, we demonstrate the robustness of this novel approach and discuss parameters that may influence its limits of error.

Methods

KMSubtraction is an R-package (<https://github.com/josephjzhao/KMSubtraction/>) designed to retrieve survival data of patients from unreported subgroups. A Shiny app is also made available and accessible through the GitHub user guide. An overview of the *KMSubtraction* workflow is illustrated in Fig. 1. The workflow of *KMSubtraction* is detailed as such:

Step 1 Reconstruction of time-to-event outcomes

A graphical reconstructive algorithm is exploited to estimate time-to-event outcomes from KM plots by an iterative algorithm based on KM estimation described by Guyot et al. [1, 2, 11]. The implementation of this step is described by Liu et al. in the package *IPDfromKM* [11]. KM plots describing the overall cohort and a known subgroup is required.

Step 2 Matching of patients – *KMSubtractionMatch()*

KMSubtractionMatch() is a wrapper function that utilizes raw reconstructed time-to-event outcomes to match

patients from subgroups among the overall cohort. Minimal cost bipartite matching with the Hungarian algorithm [12] was adopted as the primary matching algorithm. The minimal cost bipartite matching algorithm aims to match patients from the overall and subgroup cohort by minimizing follow-up time differences between matched patients [12]. The Hungarian algorithm (or the Huhn-Munkres algorithm) is a combinatorial optimization algorithm that solves the problem of assigning optimal pairs in polynomial time, and is implemented in the *RcppHungarian* package in R. Matching through Mahalanobis distance matching (Mahalanobis) and nearest neighbor matching by logistic regression (Logistic) with the *MatchIt* [13] package are provided as well. In Mahalanobis distance matching, pairs are assigned based on a scale-free Euclidean distance to minimize covariate differences within each pair [14]. In nearest neighbor matching, pairs are assigned based on the propensity score difference estimated via logistic regression, without reference to how other units will be or have been paired, and therefore this method does not aim to optimize any criterion [14, 15]. The three methods of matching were comparatively evaluated as described below. Patients with events and censorships were matched separately and are referred to as “patients with events” and “patients with censorships” in this study. By excluding matched patients, the opposing unreported subgroup may be retrieved.

Evaluation of matching – *KMSubtractionEvaluateMatch()*

The quality of matching was evaluated by comparing matched patients. This was conducted by inspecting the following parameters and quality of match test statistics (1) Bland-Altman [16] plots to explore discrepancies between matched pairs with the *blandr* [17] package in R (2) Empirical cumulative distribution functions and the Kolmogorov–Smirnov Test of follow-up times between matched pairs (3) KM plots of matched patients from the overall cohorts versus directly reconstructed subgroup cohorts, along with hazard ratios and log-rank tests from a marginal Cox proportional hazard model.

Potential limits of error of each task – *KMSubtractionError()*

In light of the variation in each implementation, the limits of error surrounding each task would be different. It would therefore be of paramount importance to ascertain whether the implementation of *KMSubtraction* would be appropriate in each context. This may be especially relevant in the interpretation of derived data in situations where there is a sizable proportion of missing data in the opposing subgroup. *KMSubtractionError()* conducts Monte Carlo simulations to evaluate the limits of error of *KMSubtractionMatch()* given parameters surrounding

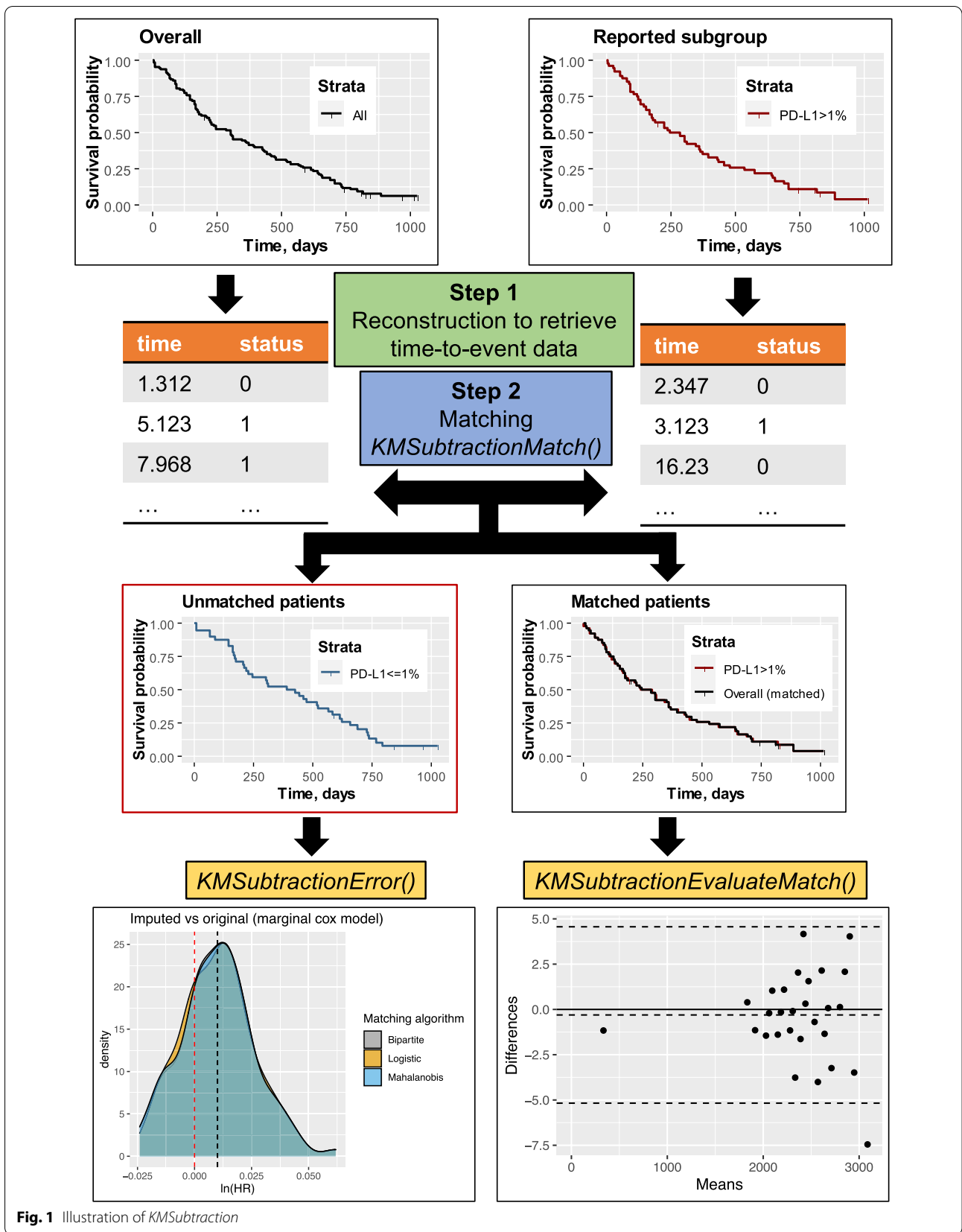


Fig. 1 Illustration of KMSubtraction

the reconstruction task required. Follow-up time was modelled by a random weibull distribution of common shape parameter of 1.000 and scale parameter of 5.000. Reconstructed and original survival data were compared by means of marginal Cox-proportional hazard models and restricted mean survival time difference (RMST-D) [18]. The similarity may be summarized by the natural logarithmic of the hazard ratio, $\ln(\text{HR})$ and RMST-D. Deviations from the true value is reflected by the absolute of $\ln(\text{HR})$ and RMST-D, where $|\ln(\text{HR})|$ and $|\text{RMST-D}|=0$ is ideal.

Simulation exercise

Next, we conducted Monte Carlo simulations to evaluate the effect of the sample size, reported subgroup proportion, proportion of missing data, censorship proportion in the overall and subgroup cohort, and number of number-at-risk table intervals on the quality of reconstruction. Five thousand iterations were conducted for each scenario. Follow-up time was modelled by a random Weibull distribution of common shape parameter of 1.000 and scale parameter of 5.000. To facilitate the high magnitude of simulations performed, curve coordinates retrieval from KM plot images were automated by identifying RGB coordinates of each rasterized curve with a prespecified color and rescaling them with prespecified maximum and minimum xy coordinates. This was conducted using the *magick* [19] package in R. The extracted xy coordinates of curves were thereafter incorporated into *IPDfromKM* [11] for reconstruction. Utilizing *KMSubtraction*, patients from the opposing subgroup are identified.

Thereafter, reconstructed KM plots were compared against original plots using a marginal Cox proportional hazards model. The absolute value of natural logarithmic transformed hazard ratios ($|\ln(\text{HR})|$) was inspected as the main summary statistic. Running $|\ln(\text{HR})|$ mean plots were inspected to evaluate for simulation convergence. For each scenario, means and Wald's 95% confidence intervals were derived per outcome. The Pearson's product-moment correlation coefficient r was inspected to evaluate the proportion of the variation in $|\ln(\text{HR})|$ that is predictable from the upstream varying independent variables. Cutoffs were determined by the intersection between $|\ln(\text{HR})|=0.03, 0.04$ & 0.05 and smoothing splines of 15 degrees of freedom generated with the primary matching algorithm. Comparisons between the different matching algorithms were undertaken using two-way analysis of variance (ANOVA) with Tukey multiple pairwise-comparisons between each algorithm. Besides minimal cost bipartite matching, Mahalanobis distance matching, and nearest neighbor matching by logistic regression were explored as well. Finally,

we sought to evaluate associations between the quality of match and the limits of error. Correlograms reflecting associations between $|\ln(\text{HR})|$ and quality of match test statistics of *KMSubtractionEvaluateMatch()* were inspected to help interpret and identify tests that may assist in prognosticating reconstruction accuracy.

The R code for the simulations is provided in [Supplementary Code](#), and the parameters utilized for each simulation scenario are provided in [Supplementary Table 1](#). Running $|\ln(\text{HR})|$ mean plots to demonstrate simulation convergence is shown in [Supplementary Figure 1](#). All analyses were conducted in R-4.1.0.

Implementation

Scenario 1

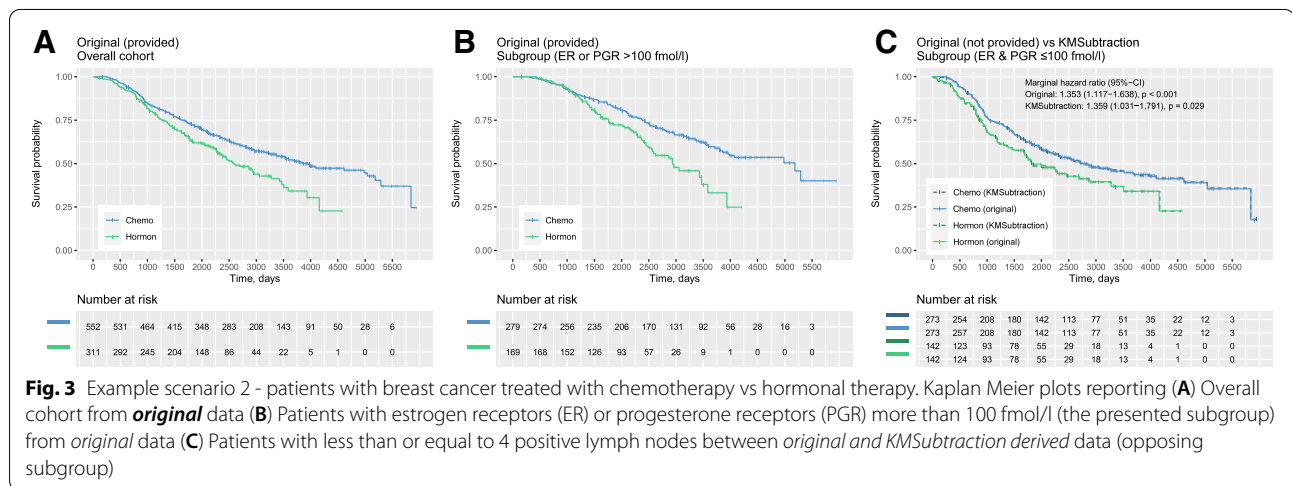
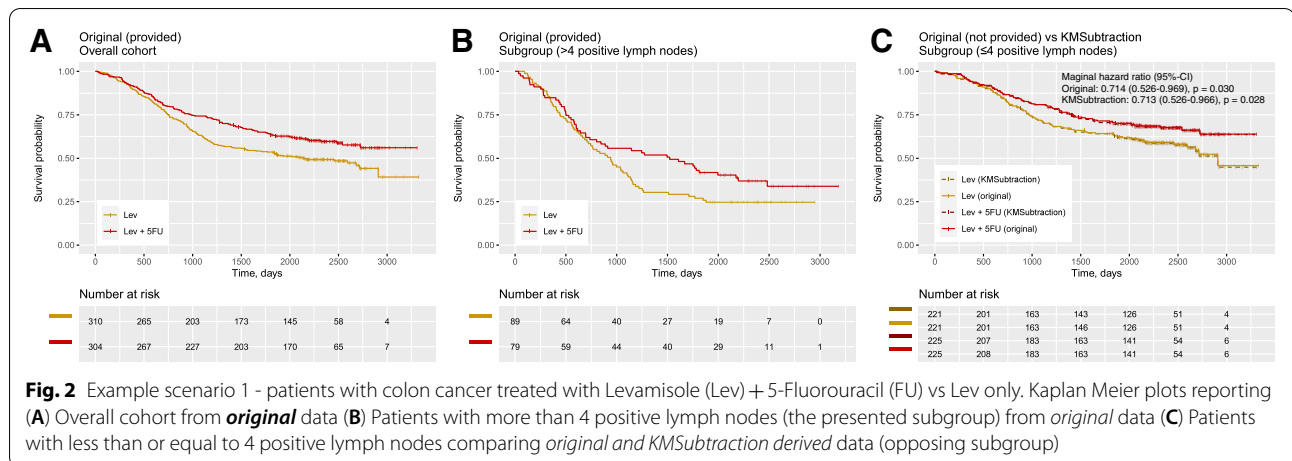
The first scenario utilizes time-to-death data of patients with stage III colon carcinoma treated with fluorouracil (5FU) plus levamisole (Lev) versus levamisole only after resection in a randomized controlled trial by Moertel et al. [20]. The dataset was retrieved from the *survival* package in R. The hypothetical scenario is designed as such: KM plots describing survival outcomes for patients treated with Lev vs Lev+5FU is presented for the overall cohort ($n = 614$) (Fig. 2A) and for patients with more than 4 positive lymph nodes ($n = 168$) (Fig. 2B). Survival outcomes for patients with less than or equal to 4 positive lymph nodes is however unreported ($n = 446$) (Fig. 2C, solid line).

Upon reconstruction of Fig. 2A and Fig. 2B, and processing with *KMSubtractionMatch()* with minimal cost bipartite matching, the retrieved subgroup yielded plots (Fig. 2C, dashed line) similar to the original data. Hazard ratios on the marginal Cox model were similar as well (original-HR=0.714, 95%-CI: 0.526–0.969, $p = 0.030$ vs *KMSubtraction*-HR=0.713, 95%-CI: 0.526–0.966, $p = 0.028$) (Fig. 2C).

Following, 1000 Monte Carlo iterations with *KMSubtractionError()* were conducted per arm, yielding mean (standard deviation) $|\ln(\text{HR})|$ of 0.00960 (0.00780) and 0.01120 (0.00866) for Lev and Lev+5FU respectively. This suggests that the limits of error of *KMSubtraction* in the above scenarios are likely small and negligible.

Scenario 2

The second scenario utilizes time-to-death data of patients with primary breast cancer from the Rotterdam tumor bank by Royston and Altman et al. [21]. The dataset was retrieved from the *survival* package in R [22]. The hypothetical scenario is designed as such: KM plots describing survival outcomes for patients treated with chemotherapy-only (Chemo) vs hormonal therapy-only (Hormon) is presented for the overall cohort ($n = 863$)



(Fig. 3A) and for patients with estrogen receptors (ER) or progesterone receptors (PGR) more than 100 fmol/l ($n = 448$) (Fig. 3B). Survival outcomes for patients with ER and PGR less than or equal to 100 fmol/l is however unreported ($n = 415$) (Fig. 3C, solid line).

Upon reconstruction of Fig. 3A and Fig. 3B, and processing with *KMSubtraction* with minimal cost bipartite matching, the retrieved subgroup yielded plots (Fig. 3C, dashed line) similar to the original data. Hazard ratios on the marginal Cox model comparing Chemo vs Hormon among patients with ER and PGR less than or equal to 100 fmol/l were similar as well (original-HR = 1.353, 95%-CI: 1.117–1.638, $p < 0.001$ vs *KMSubtraction*-HR = 1.359, 95%-CI: 1.031–1.791, $p = 0.029$) (Fig. 3C).

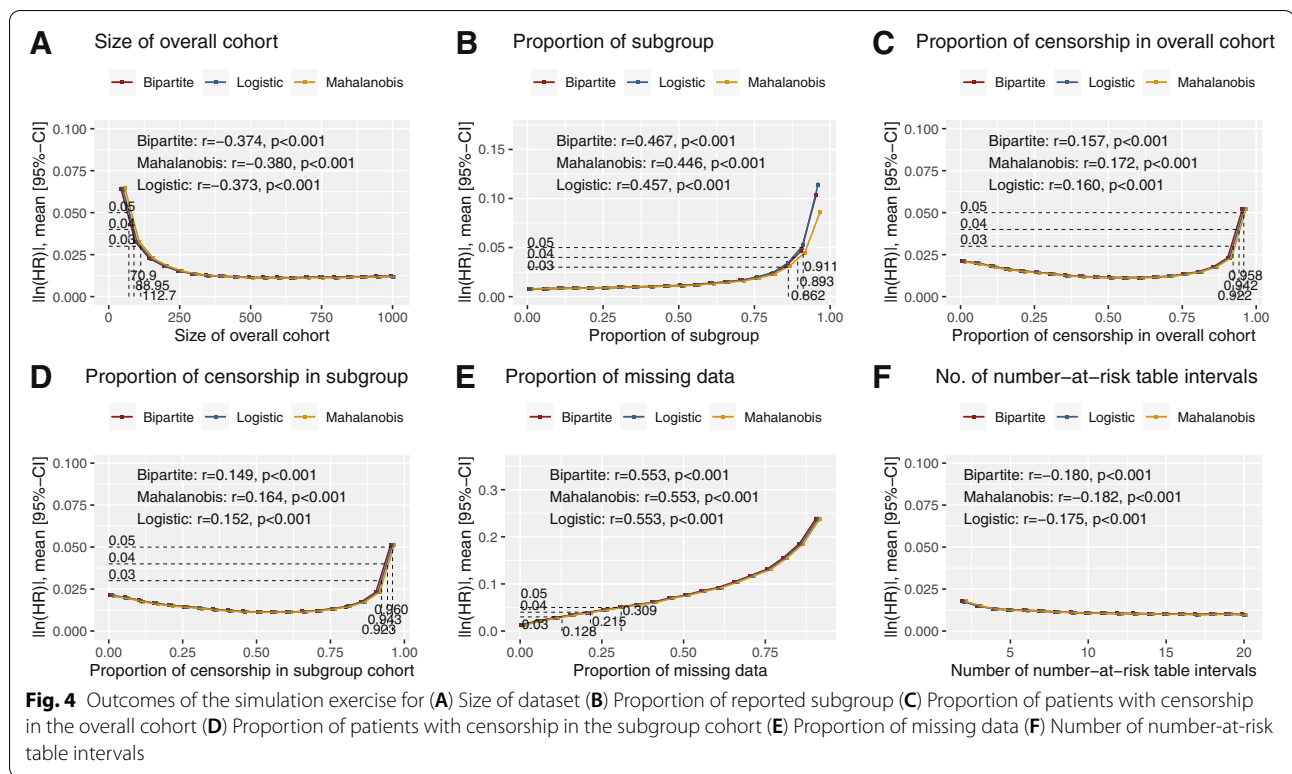
Following, 1000 Monte Carlo iterations with *KMSubtractionError()* were conducted per arm, yielding mean (standard deviation) $|\ln(\text{HR})|$ of 0.01211 (0.00818) and 0.01271 (0.01125) for Chemo and Hormon respectively.

This suggests that the limits of error of *KMSubtraction* in the above scenarios are likely small and negligible.

Simulation results

Size of dataset

Twenty scenarios of intervals $n = 50$ were created from sample sizes ranging 50 to 1000. $|\ln(\text{HR})|$ was negatively associated with sample size (Bipartite: $r = -0.374$, $p < 0.001$; Mahalanobis: $r = -0.380$, $p < 0.001$; Logistic: $r = -0.373$, $p < 0.001$), suggesting that small data sets are likely to yield high limits of error (Fig. 4A). No significant differences between the matching algorithms were demonstrated on the Tukey pairwise comparisons (all $p > 0.05$). Datasets smaller than $n = 112, 89, 71$ are likely to yield a mean $|\ln(\text{HR})|$ of 0.03, 0.04 and 0.05 respectively with minimum cost bipartite matching. The mean of absolute differences in follow-up time among patients with events appears to be weakly associated with the limits of error (Fig. 5A).



Proportion of reported subgroup

Twenty scenarios of intervals 5% were created from proportions ranging 1 to 96%. |ln(HR)| was positively associated with the proportion of reported subgroup (Bipartite: $r = 0.467, p < 0.001$; Mahalanobis: $r = 0.446, p < 0.001$; Logistic: $r = 0.457, p < 0.01$), suggesting that large proportion of reported subgroups are likely to yield high limits of error (Fig. 4B). The Mahalanobis matching algorithm yielded significantly lower limits of error when the proportion of reported subgroup is beyond 86% (Supplementary Table 2). Proportions of reported subgroups larger than 86.2, 89.3 and 91.1% are likely to yield a mean |ln(HR)| of 0.03, 0.04 and 0.05 respectively with minimum cost bipartite matching. In scenarios with high proportions of subgroups, the |ln(HR)| between matched patients of the overall and subgroup cohort, log-rank test and mean of absolute differences in follow-up time among patients with censorships appears to be associated with the limits of error (Fig. 5B).

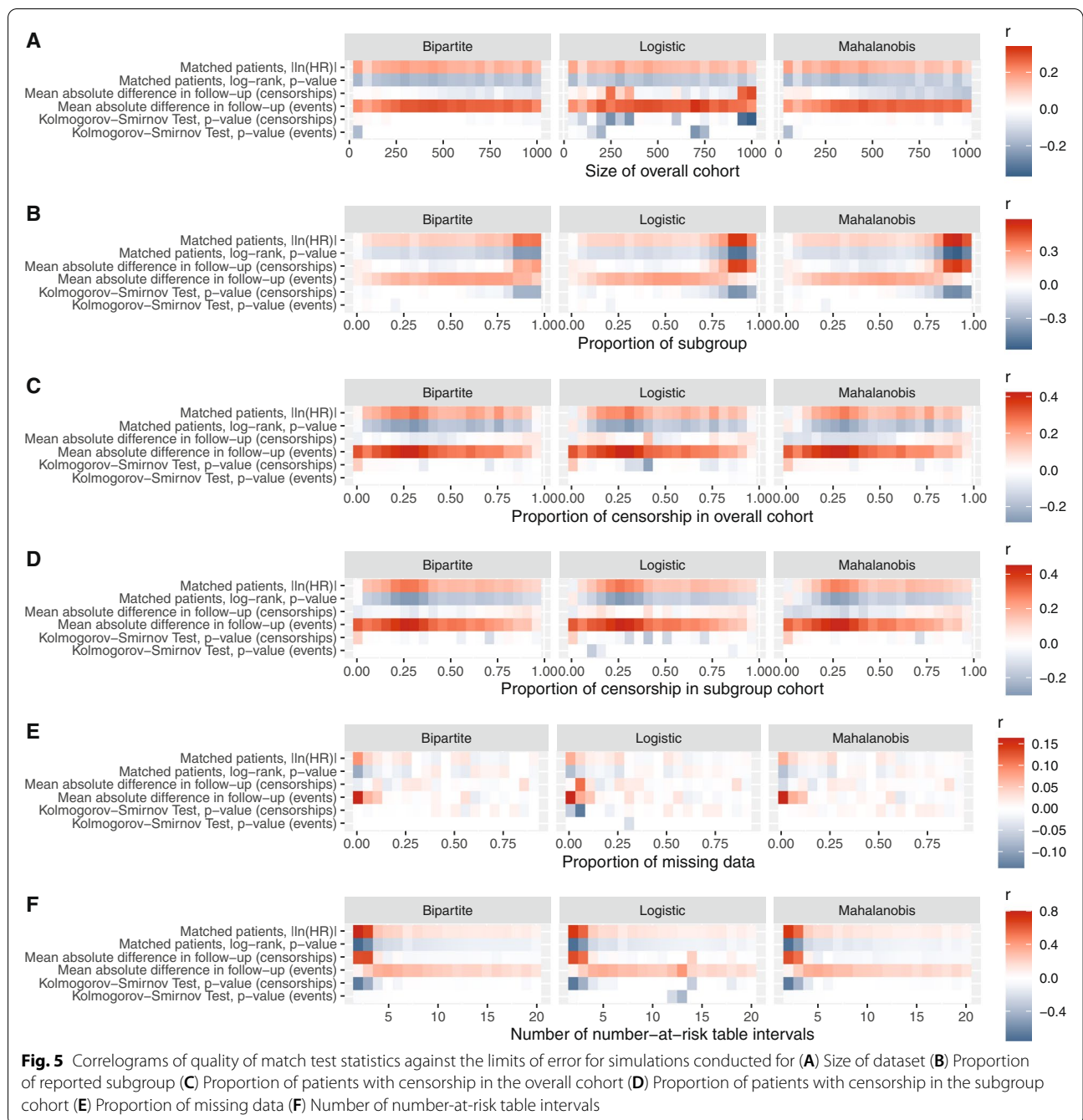
Proportion of censorship in overall cohort

Twenty scenarios of intervals 5% were created from proportions ranging 1 to 96%. |ln(HR)| was weakly associated with the proportion of censorship (Bipartite: $r = 0.157, p < 0.001$; Mahalanobis: $r = 0.172, p < 0.001$; Logistic: $r = 0.160, p < 0.001$) (Fig. 4C). No significant differences between the matching algorithms were

demonstrated on the Tukey pairwise comparisons (all $p > 0.05$). Proportions of censorships in the overall cohort larger than 92.2, 94.2 and 95.8% are likely to yield a mean |ln(HR)| of 0.03, 0.04 and 0.05 respectively with minimum cost bipartite matching. In scenarios with smaller proportions of censorship in the overall cohort, the mean of absolute differences in follow-up time among patients with events appears to be weakly associated with the limits of error (Fig. 5C).

Proportion of censorship in subgroup cohort

Twenty scenarios of intervals 5% were created from proportions ranging 1 to 96%. |ln(HR)| was weakly associated with the proportion of censorship (Bipartite: $r = 0.149, p < 0.001$; Mahalanobis: $r = 0.164, p < 0.001$; Logistic: $r = 0.152, p < 0.001$) (Fig. 4D). No significant differences between the matching algorithms were demonstrated on the Tukey pairwise comparisons (all $p > 0.05$). Proportions of censorships in the overall cohort larger than 92.3, 94.3 and 96.0% are likely to yield a mean |ln(HR)| of 0.03, 0.04 and 0.05 respectively with minimum cost bipartite matching. In scenarios with smaller proportions of censorship in the subgroup cohort, the mean of absolute differences in follow-up time among patients with events appears to be weakly associated with the limits of error (Fig. 5D).



Proportion of missing data

Nineteen scenarios of intervals 5% were created from proportions ranging 1 to 91%. $|\ln(HR)|$ was positively associated with the proportion of reported subgroup (Bipartite: $r = 0.553, p < 0.001$; Mahalanobis: $r = 0.553, p < 0.001$; Logistic: $r = 0.553, p < 0.001$), suggesting that large proportion of missing data in unreported subgroups are likely to yield high limits of error (Fig. 4E). No

significant differences between the matching algorithms were demonstrated on the Tukey pairwise comparisons (all $p > 0.05$). Proportions of reported missing data larger than 12.8, 21.5 and 30.9% are likely to yield a mean $|\ln(HR)|$ of 0.03, 0.04 and 0.05 respectively with minimum cost bipartite matching. Across the range of simulations, all quality of match test statistics were poorly associated the limits of error (Fig. 5E).

Number of number-at-risk table intervals

Nineteen scenarios were created from intervals ranging from 1 to 20. $|\ln(\text{HR})|$ was weakly associated with the number of number-at-risk table intervals (Bipartite: $r = -0.180$, $p < 0.001$; Mahalanobis: $r = -0.182$, $p < 0.001$; Logistic: $r = -0.175$, $p < 0.001$) (Fig. 4F). No significant differences between the matching algorithms were demonstrated on the Tukey pairwise comparisons (all $p > 0.05$). Across all scenarios, mean $|\ln(\text{HR})|$ was under 0.03 with all matching algorithms. Where the number of number-at-risk table intervals were small, the $|\ln(\text{HR})|$ between matched patients of the overall and subgroup cohort, log-rank test, mean of absolute differences in follow-up time among patients with censorships and the Kolmogorov-Smirnov Test of follow-up time between patients with censorships appears to be strongly associated with the limits of error (Fig. 5F).

Discussion

Secondary analysis through meta-analysis has become increasingly relevant in this era of precision medicine, where definitive conclusions on biomarkers of disease or treatment regimens are required to facilitate clinical decision making. The upstream analyses demonstrates that *KMSubtraction* is robust and reliable. Other real-world implementations of *KMSubtraction* were demonstrated to yield similar if not identical treatment effects in KEY-NOTE-590 [23] among patients with advanced gastroesophageal adenocarcinoma dichotomized by CPS and in CheckMate-227 [24] among patients with advanced non-small cell lung cancer dichotomized by PD-L1 expression levels [10].

The simulation exercise established that the limits of error of *KMSubtraction* were small and likely negligible in most circumstances. Apart from the proportion of missing data from the unreported subgroup, the reproducibility of reconstructed data was largely not affected by the other parameters studied. This is perhaps because when the reported subgroup is larger relative to the original cohort, the likelihood of identifying the correct patients in the opposing subgroup decreases. Interestingly, there was some evidence that Mahalanobis distance matching proffered a significantly smaller error when the proportion of the reported subgroup is beyond 86%. Nonetheless, given that *KMSubtraction* would be discouraged in scenarios where the proportion of reported subgroup is high, there is unlikely a scenario for this advantage to be exploited. There were otherwise no other appreciable differences on the Tukey HSD pairwise comparisons between the matching strategies investigated.

Extensive Monte Carlo simulations demonstrate that datasets with high reported subgroup proportion ($r = 0.467$, $p < 0.001$), small dataset size ($r = -0.374$,

$p < 0.001$) and high proportion of missing data in the unreported subgroup ($r = 0.553$, $p < 0.001$) were associated with uncertainty and are likely to yield high limits of error with *KMSubtraction*. Users are advised against implementation of *KMSubtraction* in such situations, or if need be, transparently acknowledge the ensuing limitations and likely compromise in quality of reconstructed data. The limits of error of *KMSubtraction*, as reflected by the mean $|\ln(\text{HR})|$ from converged Monte Carlo simulations may be arbitrarily interpreted as small (< 0.03), moderate (0.03–0.05) and large (> 0.05), and may guide the appropriateness of implementing *KMSubtraction* in each context.

In implementing *KMSubtraction*, the quality of match may be the only raw parameter derived from the manually retrieved survival information. From our simulations, we found a wide variation of how the quality of match influences the limits of error of *KMSubtraction*. As such, in circumstances where the simulated limit of error is shown to be high, users are advised to inspect the association of each quality of match test statistic against the simulated limits of error.

This approach is not without its limitations. The described method is also unable to retrieve patient-level covariates for adjustment, or deal with high-dimensional survival data. Hence, biomarker investigation through this analysis should be interpreted with caution as no further analysis of causal inference may be conducted. Further, the described method is only able to handle dichotomous variables, rather than continuous variables. This would unfortunately limit meta-analytic pooling of unreported subgroup data after *KMSubtraction* when different cutoffs are used instead.

Given the demonstrated effect of the above parameters on the quality of *KMSubtraction* derived data, it would be instructive for those utilizing *KMSubtraction* to report (1) Proportion of reported subgroup among the overall cohort (2) Size of dataset (3) Proportion of censorship in overall and subgroup cohorts (4) Proportion of missing data (5) Number of number-at-risk table intervals (6) Mean and standard deviation of $|\ln(\text{HR})|$ from *KMSubtractionError()* simulations.

Conclusion

While *KMSubtraction* demonstrates robustness in deriving survival data from unreported subgroups, the implementation of *KMSubtraction* should take into consideration the aforementioned limitations. The limits of error of *KMSubtraction*, as reflected by the mean $|\ln(\text{HR})|$ from converged Monte Carlo simulations may guide the interpretation of reconstructed survival data of unreported subgroups. This approach hopes to

enhance the quality of secondary analysis of subgroups and biomarkers in the era of precision medicine.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-022-01567-z>.

Additional file 1.

Acknowledgements

The authors would like to thank Associate Professor Hyungwon Choi, Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, for guiding us through the simulations and development of the package.

Authors' contributions

RS, JJZ, YHC, NLS were responsible for the concept and design of the study. JJZ, RS, YHC were responsible for the development of the novel KM Subtraction methodology to retrieve unreported patient subgroups from reconstructed time-to-event outcomes. JJZ, NLS and YHC conducted the statistical analyses. JJZ, NLS and YHC interpreted the data. RS, JJZ, NLS drafted the manuscript. All authors critically revised the manuscript for important intellectual content. RS supervised the study. All authors had full access to all the data in the study, and RS had the final responsibility for the decision to submit for publication. The author(s) read and approved the final manuscript.

Funding

RS is supported by the National Medical Research Council (NMRC) (NMRC/Fellowship/0059/2018 and NMRC/TA/0014/2020). As part of the Singapore Gastric Cancer Consortium, the study is funded by the NMRC Open Fund-Large Collaborative Grant (OFLCG18May-0023). All other authors have no funding to declare.

Availability of data and materials

The code utilized to create simulation data is provided in the supplementary materials.

This study does not involve patients or participants.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

RS has received honoraria from Bristol-Myers Squibb, Lilly, Roche, Taiho, AstraZeneca, DKSH and MSD; has advisory activity with Bristol-Myers Squibb, Merck, Eisai, Bayer, Taiho, Novartis, MSD and AstraZeneca; received research funding from MSD and Paxman Coolers; and has received travel grants from AstraZeneca, Eisai, Roche and Taiho Pharmaceutical. NLS is an editorial board member on *BMC Medical Research Methodology*. The other authors declare that they have no competing interests.

Author details

¹Yong Loo Lin School of Medicine, National University of Singapore, 21 Lower Kent Ridge Rd, Singapore 119077, Singapore. ²Biostatistics Unit, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ³Department of Haematology-Oncology, National University Cancer Institute, National University Hospital, 1E Kent Ridge Road, Singapore 119228, Singapore. ⁴Cancer and Stem Cell Biology Program, Duke-NUS Medical School, Singapore, Singapore. ⁵The N.1 Institute for Health, National University of Singapore, Singapore, Singapore. ⁶Singapore Gastric Cancer Consortium, Singapore, Singapore.

Received: 4 December 2021 Accepted: 25 February 2022
Published online: 03 April 2022

References

- Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012;12(1):9.
- Syn NL, Cummings DE, Wang LZ, Lin DJ, Zhao JJ, Loh M, Koh ZJ, Chew CA, Loo YE, Tai BC, Kim G, So JB, Kaplan LM, Dixon JB, Shabbir A. Association of metabolic-bariatric surgery with long-term survival in adults with and without diabetes: a one-stage meta-analysis of matched cohort and prospective controlled studies with 174 772 participants. *Lancet*. 2021;397(10287):1830-41. [https://doi.org/10.1016/S0140-6736\(21\)00591-2](https://doi.org/10.1016/S0140-6736(21)00591-2). Epub 2021 May 6.
- Zhao JJ, Tan E, Sultana R, Syn NL, Da Zhuang K, Leong S, Tai DWM, Too CW. Intra-Arterial Therapy for Unresectable Colorectal Liver Metastases: A Meta-Analysis. *J Vasc Interv Radiol*. 2021;32(11):1536-1545.e38. <https://doi.org/10.1016/j.jvir.2021.05.032>. Epub 2021 Jun 22.
- Amatya AK, Fiero MH, Bloomquist EW, Sinha AK, Lemery SJ, Singh H, et al. Subgroup Analyses in Oncology Trials: Regulatory Considerations and Case Examples. *Clin Cancer Res*. 2021;27(21):5753-6.
- Tsimberidou AM, Fountzilas E, Nikanjam M, Kurzrock R. Review of precision cancer medicine: Evolution of the treatment paradigm. *Cancer Treat Rev*. 2020;86:102019.
- Ashley EA. Towards precision medicine. *Nat Rev Genet*. 2016;17(9):507-22.
- Sundar R, Tan IBH, Chee CE. Negative predictive biomarkers in colorectal cancer: PRESSING ahead. *J Clin Oncol*. 2019;37(33):3066-8.
- Janjigian YY, Shitara K, Moehler M, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *Lancet*. 2021;398(10294):27-40.
- Shitara K, Van Cutsem E, Bang Y-J, et al. Efficacy and safety of pembrolizumab or pembrolizumab plus chemotherapy vs chemotherapy alone for patients with first-line, advanced gastric cancer: the KEYNOTE-062 phase 3 randomized clinical trial. *JAMA Oncol*. 2020;6(10):1571-80.
- Zhao JJ, Yap DWT, Chan YH, Tan BKJ, Teo CB, Syn NL, et al. Low Programmed Death-Ligand 1-Expressing Subgroup Outcomes of First-Line Immune Checkpoint Inhibitors in Gastric or Esophageal Adenocarcinoma. *J Clin Oncol*. 2022;40(4):392-402.
- Liu N, Zhou Y, Lee JJ. IPDfromKM: reconstruct individual patient data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2021;21(1):111.
- Kuhn HW. The Hungarian method for the assignment problem. *Naval Res Logistics Q*. 1955;2(1-2):83-97.
- Ho D, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. *J Stat Softw*. 2011;42(8):1-28.
- Elizabeth AS. Matching methods for causal inference: a review and a look forward. *Stat Sci*. 2010;25(1):1-21.
- Thoemmes FJ, Kim ES. A systematic review of propensity score methods in the social sciences. *Multivar Behav Res*. 2011;46(1):90-118.
- Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307-10.
- Datta, D. blandr: a Bland-Altman Method Comparison package for R. 2017. <https://doi.org/10.5281/zenodo.824514>.
- Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013;13(1):152.
- Ooms J. Magik: Advanced Graphics and Image-Processing in R. 2021. <https://cran.rproject.org/web/packages/magick/magick.pdf>. Accessed 1 Dec 2021.
- Moertel CG, Fleming TR, Macdonald JS, et al. Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: a final report. *Ann Intern Med*. 1995;122(5):321-6.
- Royston P, Altman DG. External validation of a cox prognostic model: principles and methods. *BMC Med Res Methodol*. 2013;13(1):33.

22. Therneau T, Grambsch P. Modeling survival data: Extending the Cox model. New York: Springer; 2000. ISBN 0-387-98784-3.
23. Sun J-M, Shen L, Shah MA, et al. Pembrolizumab plus chemotherapy versus chemotherapy alone for first-line treatment of advanced oesophageal cancer (KEYNOTE-590): a randomised, placebo-controlled, phase 3 study. *Lancet*. 2021;398(10302):759–71.
24. Hellmann MD, Paz-Ares L, Bernabe Caro R, et al. Nivolumab plus Ipilimumab in advanced non-small-cell lung cancer. *N Engl J Med*. 2019;381(21):2020–31.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

