Contemporary Clinical Trials Communications 33 (2023) 101142

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Testing the "RCT augmentation" methodology: A trial simulation study to guide the broadening of trials eligibility criteria and inform on effectiveness

Clementine Nordon [a,d,*], Benoit Sanchez [a], Mei Zhang [b], Xiaowei Wang [c,1], Phillip Hunt [d], Mark Belger [e], Helene Karcher [f], on behalf of The GetReal Initiative

[a] Formerly LASER Research, Paris, France
[b] Sanofi R&D, Bridgewater, NJ, United States of America
[c] Formerly GSK R&D Biostatistics, Collegeville, PA, United States of America
[d] AstraZeneca, Gaithersburg, MD, United States of America
[e] Eli Lilly, Bracknell, United Kingdom
[f] Novartis, Basel, Switzerland

ARTICLE INFO

Keywords:
RCTs
Exclusion criteria
Effectiveness
Real-world
Machine learning

ABSTRACT

Background: Exclusion criteria that are treatment effect modifiers (TEM) decrease RCTs results generalisability and the potentials of effectiveness estimation. In "augmented RCTs", a small proportion of otherwise-excluded patients are included to allow for effectiveness estimation. In Hodgkin Lymphoma (HL) RCTs, older age and comorbidity are common exclusion criteria, while also TEM. We simulated HL RCTs augmented with age or comorbidity, and explored in each scenario the impact of augmentation on effectiveness estimation accuracy.

Methods: Simulated data with a population of HL individuals initiating drug A or B was generated. There were drug-age and drug-comorbidity interactions in the simulated data, with a greater magnitude of the former compared to the latter. Multiple augmented RCTs were simulated by randomly selecting patients with increasing proportions of older, or comorbid patients. Treatment effect size was expressed using the between-group Restricted Mean Survival Time (RMST) difference at 3 years. For each augmentation proportion, a model estimating the "real-world" treatment effect (effectiveness) was fitted and the estimation error measured (Root Mean Square Error, RMSE).

Results: In simulated RCTs including none (0%), or the real-world proportion (30%) of older patients, the interquartile range of RMST difference was 0.4–0.5 years and 0.2–0.3 years, respectively, and RMSE were 0.198 years (highest possible error) and 0.056 years (lowest), respectively. Augmenting RCTs with 5% older patients decreased estimation error substantially (RMSE = 0.076 years). Augmentation with comorbid patients proved less useful for effectiveness estimation.

Conclusion: In augmented RCTs aiming to inform the effectiveness of drugs, augmentation should concern in priority those exclusion criteria of suspected important TEM magnitude, so as to minimie the proportion of augmentation necessary for good effectiveness estimations.

1. Introduction

Pre-authorization randomized controlled trials (RCTs) remain a gold standard to assess drug efficacy for clinical research and for regulatory approval of a new therapy. However, they often include a highly selected population poorly representative of patients who will receive the intervention in routine clinical practice [1–4] hence the risk for an efficacy-to-effectiveness gap [5]. This issue is increasingly recognized by regulators [6] and health technology assessment bodies [7] that call for broadening eligibility criteria [8]. On the other hand, exclusion criteria are often chosen to protect the safety of patients participating in clinical trials and minimize the risk of treatment effect dilution. Broadening a trial population is not always feasible and therefore, the use of predictive modeling techniques is an appealing option to predict effectiveness using data generated by the pre-authorization RCT and extrapolate its results to the real world [9,10]. However, when eligibility criteria lead

[*] Corresponding author. Formerly LASER Research, Paris, France.
E-mail address: nordon.clementine@gmail.com (C. Nordon).
[1] Currently Biostatistics, Moderna Inc., Pennsylvania, United States of America.

https://doi.org/10.1016/j.conctc.2023.101142
Received 26 January 2023; Received in revised form 4 April 2023; Accepted 12 April 2023
Available online 14 April 2023

to totally excluding specific patient phenotypes and to modifying the distribution of key treatment effect modifiers (TEM) [11], predictive modeling techniques may be of limited use.

The "RCT augmentation" is a trial simulation technique that provides a compromise between totally excluding patients based on exclusion criteria that are also well-known or suspected TEM on the one hand, and totally including them on the other hand. Through trial simulation, various proportions of patients with the exclusion criteria of interest in the RCT are "re-included". Then, the impact of broadening these criteria is evaluated in regards to treatment effect size – and statistical power – and the ability of RCT data to inform on the real-world effect of the investigational therapy. The principle of an augmented RCT is to include a small proportion of patients with the exclusion criteria of interest in the RCT. This proportion should be small to minimize the risk incurred by completely allowing patients into the trial, but yet sufficient for effectiveness prediction purposes.

Before implementing this methodology in a genuine RCT, we set-out to test its potentials and characteristics through different case studies. A first case study used observational data from a cohort of schizophrenia patients. Multiple RCTs comparing two antipsychotic drugs were simulated using this dataset, and augmented with increasing proportions of patients having specific characteristics often used as exclusion criteria (e.g., history of suicide attempt, alcohol use disorder). RCTs augmented with 10–20% patients having these characteristics allowed for good effectiveness prediction, highlighting the relevance of this methodology to better predict treatments effectiveness. The results of this case study also suggested the possibility to augment a trial population while also maintaining statistical power [12].

In the present study we conducted a second case study on the "RCT augmentation" method, to explore in more details how the magnitude of treatment effect modification impacts the proportion of patients to re-include and the estimation accuracy of the model. We thus used fully-simulated data to set different magnitudes of treatment effect modification to the exclusion criteria used.

2. Material and methods

2.1. General methodology

For simplicity, we will call the estimated comparative effect from data similar to a traditional RCT "efficacy" while the observed comparative effect in real world will be called "effectiveness".

The RCT augmentation methodology is a step-by-step process starting at the trial design stage when one or several exclusion criteria need to be applied (e.g., for safety, or ethical reasons) although they are known or suspected TEM. The general methodology is explained below and the specific example used in the present case study is depicted in Fig. 1. In a first step, a preferably large real-world dataset (e.g., observational cohort, electronic healthcare records) is used to estimate the effectiveness of the investigational treatment considered as the target measure for estimation. Second, from this real-world dataset patients who fulfil the RCT eligibility criteria are randomly selected. From this, multiple augmented RCTs are simulated by re-including patients who would otherwise be excluded at different proportions of augmentation. In these simulated RCTs, patients who would otherwise be totally excluded owing to specific characteristics are re-included in the RCT at different proportions of augmentation. Third, a predictive model is trained on each simulated augmented RCT and tested on real-world data. Estimation accuracy is then measured, thus providing an estimation of the smallest necessary proportion of patients that would be necessary to augment the RCT so as to allow for good estimation of effectiveness. Ultimately, following the conduct of the genuine augmented RCT, the actual RCT data can be used, not only to measure the efficacy of the investigational treatment, but also to estimate its effectiveness.

2.2. Scope of the case study

Oncology was chosen for the present case study because oncology trials often use restrictive eligibility criteria [13–19] and initiatives have been taken recently to tackle this issue [20]. Namely, older age, frailty and the presence of severe comorbidity are commonly used as exclusion criteria while also being TEM.

We used the example of adult patients suffering from relapsed/refractory Hodgkin's Lymphoma (rrHL) at a stage III or IV in the Ann Arbor Staging Classification [21]. They initiated either standard therapy (Adriamycin, Bleomycin, Vinblastine and Dacarbazine), thereafter called drug A, or the investigational combination of a monoclonal antibody with chemotherapy, thereafter called drug B.

2.3. Study population

A simulated dataset of $N = 10^6$ patients suffering from rrHL was generated, using the Python programming language (Python Software Foundation, https://www.python.org/) [22] (details in the Supplemental material). Patients' characteristics were generated at a patient-level using information on their distribution as found in the literature (Table S1): age [23–26] (categorized into 5 groups), gender [27], disease stage (III vs. IV) [28], severe comorbidity (vs. none) [23]. The correlation between age and comorbidity was taken into account [23]. We purposely used a simple framework (directed acyclic graph, Fig. S1) in which patients were randomly allocated to either of the two therapeutic options (drug A or B) with a 1:1 ratio. Patients' characteristics were not associated with treatment choice, that is, there was no

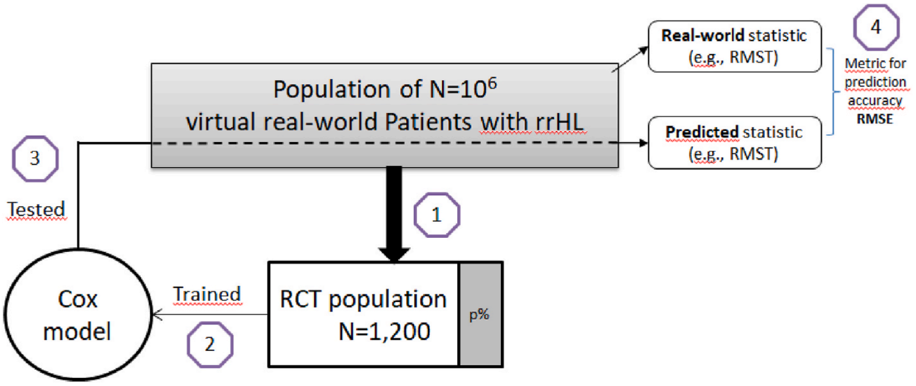


Fig. 1. Predictive modeling using augmented RCT data and providing an estimation of the statistic in the real world, and error made (RMSE) when compared with the "true" statistic
rrHL, relapsed/refractory Hodgkin's Lymphoma; RMST, Restricted Mean Survival Time; RMSE, Root Mean Square Error
1: Augmented RCTs are simulated by extracting 1000 random samples of 1200 patients from the simulated real-world population; for each set of simulations, RCTs are including increasing % older patients (or patients with severe comorbidity)
2: An exponential Cox model is trained using the RCT dataset, providing model parameters
3: The model is tested on the real-world population
4: The estimated statistics are compared with the "real" statistics; estimation accuracy is measured using the RMSE

* A practical use of the methodology requires a large real-world dataset.

confounding by indication.

The endpoint of interest was Progression-Free Survival (PFS). Relevant confounders and TEM were selected based on the literature. Effect modifiers used as exclusion criteria of interest (and criteria for augmentation) for the case study were older age (age≥60 vs. <60 years old) and severe comorbidity (any vs. none). Finally, each simulated patient's PFS was generated using the Cox survival model with time independent hazard function defined by (Equation 1):

$$\lambda = \lambda_0 \left(\lambda_{drug}\right)^{drug} \left(\lambda_{age}\right)^{age} \left(\lambda_{stage}\right)^{stage} \left(\lambda_{com}\right)^{com} \left(\lambda_{drug\_age}\right)^{drug\times age} \left(\lambda_{drug\_com}\right)^{drug\times com}$$

λ_0 being the hazard when all covariates are 0 and λ_k being the association related to covariate k, i.e., the multiplier of the hazard when the covariate increases by one unit. In this model, a covariate has no effect when $\lambda_k = 1$. Equation 1 is equivalent to the more commonly used: $\lambda = \lambda_0 e^{\beta_{age} \times age}$ with $\lambda_{age} = e^{\beta_{age}}$.

Constraints were used, based on information from the literature to find realistic parameters because these are not directly available in the literature. In turn, these lambda parameters were used to generate individual PFS values [25,28–32]. The values of lambda parameters in Equation 1 are provided in Supplemental material. The drug-age interaction parameter was $\lambda_{\text{drug age}} = 1.43$ and the drug-comorbidity interaction parameter was $\lambda_{\text{drug com}} = 1.25$. However, because age had 5 possible categories and comorbidity only 2, the magnitude of interaction with age was in fact substantially more important than the one with comorbidity. The characteristics and median PFS in the simulated patients populations initiating drugs A or B are summarized in Table 1.

2.4. Simulation of augmented RCTs

From the simulated dataset of real-world patients, a sample of n = 600 patients in each therapeutic arm was randomized in a 1:1 ratio (typical sample size for oncology Phase-3 RCTs [29]). Follow-up lasted 3 years. No loss to follow-up was considered because dropping out from oncology trials is infrequent [29]. No patient died prior to cancer progression or end of follow-up. The primary endpoint was PFS. The main statistic of interest was the between-arm difference in Restricted Mean Survival Time (RMST) at 3 years, expressed in years and thereafter called the RMST difference. RMST remains valid for non-proportional hazards [33] and is now often recommended as a primary endpoint in clinical trials [34–36]. It represents the average survival from Time 0 to a specified point in time (here, 3 years) and may be estimated as the area under the survival curve up to that point; the RMST difference is RMST under treatment B minus RMST under treatment A, at 3 years. The hazard ratio (HR) was used as an alternative statistic of interest.

In order to explore the impact of different magnitudes of TEM on estimation accuracy we used the two exclusion criteria separately: when age≥60 was the exclusion criteria used in the simulation, the presence of severe comorbidity was not (i.e., patients with severe comorbidity were not excluded) and vice versa. The augmentation consisted in increasing the heterogeneity of the RCT population by re-including increasing proportions *p* of patients with either of the two exclusion criteria of interest thereby creating a series of augmented RCTs. The augmentation of an RCT with these specific patients was made while preserving a constant trial size: when older patients were added, younger ones were removed randomly. The same applied for severe comorbidity. The RCT population was augmented from $\text{p}_{\text{min}} = 0\%$ (i.e., no augmentation) to p_{max} (proportion of older/severely comorbid patients in the simulated real-world population, i.e., exclusion criteria not applied). The proportions of older patients used to augment the RCT were successively: 0%, 1%, 2%, 3%, 4%, 5%, 6%, 8%, 10%, 15%, 20% and 30%. The proportions of severely comorbid patients used to augment the RCT were successively: 0%, 1%, 2%, 3%, 4%, 5%, 6%, 8%, 10%, 15%, 20% and 29.7%.

All simulations were performed with R software [37]. The R script and libraries used are provided in Supplemental material.

Table 1
Characteristics of the simulated real-world population suffering from rrHL and effect of drugs A and B in this population.

	Simulated real-world population
	N = 1,000,000
	%[a]
Age (years)	
15–29	29%
30-44	22%
45-59	19%
60-74	19%
75-90	11%
Male sex	58.5%
Disease stage	
Stage III	36%
Stage IV	64%
Number of comorbidities	
0	70%
≥1	30%
Initiates drug B (vs. A)[b]	50%
Median PFS[c], in years	
In patients initiating drug A[d]	1.37
In patients initiating drug B	1.94

rrHL, relapsed/refractory Hodgkin's Lymphoma; PFS, Progression-Free Survival:

[a] Numbers are not provided due to the population size;

[b] Drug B is the investigational drug compared to drug A;

[c] The difference in median PFS between patients initiating drug A or B is solely due to the drug effect since as there is no difference in population characteristics;

[d] In the simulated population, 50% of patients initiated drug A (n = 500,000), same for drug B.

2.5. Analyses and modeling

The simulation process is described in Fig. 1 and detailed in Supplemental material. For each simulation of an augmented RCT, the process was repeated 1000 times to minimize sampling error. In each simulated RCT, an exponential Cox model was trained and ridge regularization was used [38]. The model was then tested in the simulated dataset thus providing estimated RMST difference and HR.

The error made using the model was measured using the Root Mean Square Error (RMSE) [39], i.e., the mean distance between the real value of the statistic and the estimated statistics,

$$\text{RMSE} = \sqrt{\frac{1}{\text{K}} \sum_{j=1}^{K} \left(\widehat{s}_j - s\right)^2}$$

where K is the number of estimated statistics (here, K = 1000), the real value of the statistic and $\widehat{s}_j$ the estimated statistics. RMSE was computed for each percentage *p* of augmentation. It decreases as estimation accuracy improves, i.e., as the estimated statistics approach the actual statistic in the simulated dataset. No statistical test was performed to compare RMSE.

2.6. Added value of predictive modeling

We called the "standard RCT reading" the simple reading of the augmented RCT results, that is, without any predictive modeling (Fig. 2). When using the standard RCT reading, results obtained in an augmented RCT differ from those in a traditional, non-augmented RCT only because the populations are different. The standard RCT reading was explored to highlight the added value of predictive modeling. We anticipated that (i) standard RCT reading would be as good as predictive modeling to estimate effectiveness when $\text{p} = \text{p}_{\text{max}}$, and (ii) predictive modeling would provide better estimation of effectiveness for $\text{p} < \text{p}_{\text{max}}$.

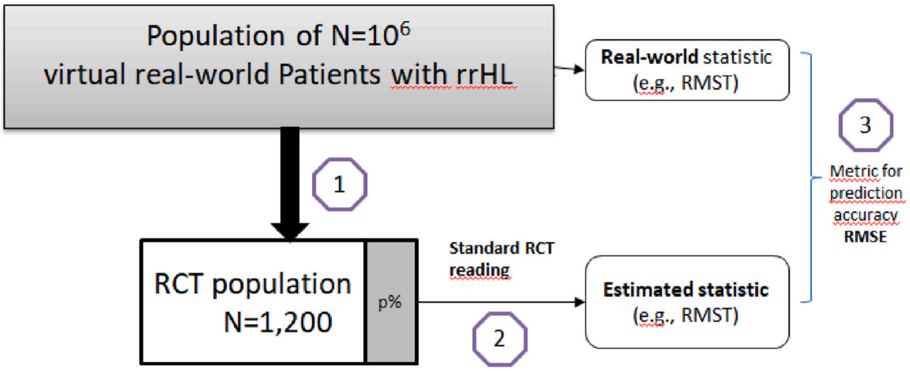

Population of N=10^6
virtual real-world Patients with rrHL
Real-world statistic (e.g., RMST)
1
3
Metric for prediction accuracy RMSE
RCT population N=1,200
p%
Standard RCT reading
2
Estimated statistic (e.g., RMST)

Fig. 2. Standard RCT reading providing an estimation of the statistic, and error made (RMSE) when compared with the "true" statistic
rrHL, relapsed/refractory Hodgkin's Lymphoma; RMST, Restricted Mean Survival Time; RMSE, Root Mean Square Error
1: Augmented RCTs are simulated by extracting 1000 random samples of 1200 patients from the simulated real-world population; for each set of simulations, RCTs are including increasing % older patients (or patients with severe comorbidity)
2: The results of the RCTs are calculated in a standard manner to estimate the statistics of interest
3: The estimated statistics are compared with the "real" statistics; the error made is measured using the RMSE.

3. Results

By design, 57.6% of patients from the simulated population were eligible to a traditional, non-augmented RCT; 30.0% of patients were ineligible due to age≥60 years old and 29.7% due to having severe comorbidities. The RMST at 3 years in the entire patient population, and then stratified by age and severe comorbidity are provided in Table 2. Drug B was more effective than drug A overall (RMST difference = 0.26 years), more effective than drug A in patients <60 years old (RMST difference = 0.46 years) but less effective in patients aged ≥60 years (RMST difference = −0.21 years), illustrating the impact of a strict application of exclusion criteria in the presence of TEM.

3.1. RCT augmentation with older patients

In terms of effect size, the standard RCT reading of 1000 simulated RCTs augmented with p = 0%, p = 5% and p = 30% of older patients provides an estimation of the (distribution of) RMST difference that would be obtained directly from RCT data (drug B efficacy, Fig. 3a): simulated RCTs totally excluding older patients yielded RMST differences around 0.4–0.5 years (interquartile range, IQR), compared with IQR of 0.2–0.3 years when including 30% of older patients. RCTs augmented with 5% of older patients (e.g., 30 older patients replace 30 younger ones in each treatment arm) yielded RMST difference very close to the RMST difference obtained from RCTs totally excluding older patients. In turn, when the same augmented RCTs are used to estimate drug B effectiveness (Fig. 3b), the results (IQR of RMST difference) obtained with an RCT augmented with only 5% of older patients are very close to those obtained in an RCT not excluding older patients (p = 30%). In other words, the data obtained from an RCT augmented with 5% of older patients would provide a good estimation of effectiveness. This latter point is also illustrated by Table 3 and Fig. 4a showing the error (RMSE) made in estimating the effectiveness using data from RCTs augmented with p = 0%, p = 1%, p = 10% and p = 30% of older patients. In RCTs totally excluding older patients, the error made in estimating the RMST difference in the real world was at maximum for both standard RCT reading (RMSE = 0.209 years) and predictive modeling (RMSE = 0.198 years). In RCTs allowing 30% of older people, the error made in estimating the RMST difference in the real world was at minimum for both standard RCT reading (RMSE = 0.060 years) and predictive modeling (RMSE = 0.056 years). Predictive modeling did not add information to standard reading for this proportion of augmentation and in both cases, the remaining error (i.e., RMSE>0) was due to sampling variance. Between those two extremes (p = 0% and p = 30%), the accuracy of predictive modeling improved rapidly as p increased (Fig. 4a). For instance with 5% of older patients in the RCT, RMSE = 0.076 years which is close to the RMSE = 0.056 years with 30% of augmentation. For proportions of augmentation >5%, the accuracy of predictive modeling did not improve in a relevant manner.

To summarise, in an RCT augmented with 5% of older patients, the RMST difference is very close to this obtained in a traditional RCT (standard reading, Fig. 3a). In addition, predictive modeling using data from this augmented RCT would allow for accurate estimation of effectiveness (Fig. 3b).

Table 2
Treatment effect of drugs A and B expressed in terms of RMST at 3 years, in patients from the entire real-world population, and then stratified by age (< vs ≥ 60 years old) and presence of severe comorbidity (none vs ≥ 1).

	Entire population (N = 1,000,000)	Real-world population stratified by age		Real-world population stratified by severe comorbidity	
		<60 years old	≥60 years old	None	≥1
% of the entire population	**100%**	**70.0%**	**30.0%**	**70.3%**	**29.7%**
RMST[1] at 3 years (in years)					
Patients initiating drug A[1]	1.55	1.74	1.10	1.67	1.27
Patients initiating drug B	1.81	2.20	0.90	2.06	1.22
Difference in RMST[2] (drug B vs. A), in years	0.26	0.46	−0.21	0.39	−0.05

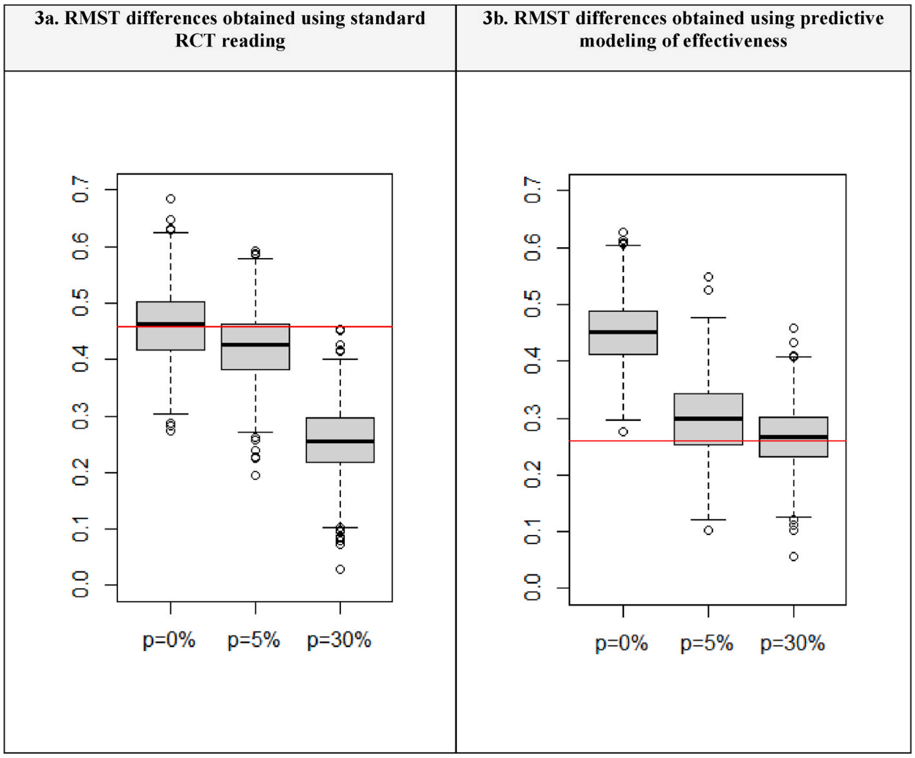
RMST, Restrictive Mean Survival Time; [1]Drug B is the investigational drug compared to drug A; in the simulated population, 50% of patients initiated drug A (n = 500,000); [2]The difference in RMST between patients initiating drug A or B is solely due to the drug effect since as there is no difference in population characteristics.

3.2. RCT augmentation with severely comorbid patients

RCTs augmented with 0%, 5% and 30% of comorbid patients would provide treatment effect size (RMST difference) that are shown in Fig. 5a. RCTs augmented with 5% vs. 0% would generate data allowing for predictive modeling of similar accuracy (Fig. 5b).

Errors (RMSE) of standard RCT reading and of predictive modeling in estimating the effectiveness, with augmented RCTs are displayed in Table 3 and Fig. 4b. When the RCTs allowed patients of any age but totally excluded patients with severe comorbidity (e.g., no augmentation with severe comorbidity), predictive modeling provided a better estimation of the RMST differences compared to standard RCT reading, with RMSE = 0.093 years and RMSE = 0.148 years, respectively. This result suggests that the model can learn on the comparative effectiveness of the two therapies even in the total absence of comorbid patients. Further augmenting the RCT yielded moderate accuracy improvement.

The results on hazard ratio (HR), as an alternative metric to RMST for survival, and the complete results of estimation errors (RMSE) are provided in Supplemental material.



3a. RMST differences obtained using standard RCT reading

3b. RMST differences obtained using predictive modeling of effectiveness

Fig. 3. RMST differences (in years) obtained from 1000 simulated RCTs augmented with p = 0%, p = 5% and p = 30% older patients, when using the standard RCT reading (Fig. 3a) or predictive modeling of effectiveness (Fig. 3b)
Boxplot providing the median (central black line), interquartile range (square), minimal and maximal values of 1000 Restricted Mean Survival Time (RMST) differences; the red horizontal line is the true RMST difference on a population with no older patients (3a) and the full population (3b); the X-axis is the proportion of augmentation with older patients; the Y-axis is the RMST difference, in years.

Table 3
Error, expressed as RMSE (in years) in estimating real-world between-arm difference in RMST (primary trial endpoint), as obtained from 1000 simulated RCTs that include increasing proportions of older, or comorbid patients.

Proportion of augmentation (broadening of eligibility criteria)	Augmentation with older patients		Augmentation with comorbid patients	
	Standard RCT reading	Predictive modeling	Standard RCT reading	Predictive modeling
	RMSE in years	RMSE in years	RMSE in years	RMSE in years
0%	0.2091	0.1981	0.1482	0.0931
1%	0.2007	0.1209	0.1432	0.0818
5%	0.1737	0.0763	0.1248	0.0690
10%	0.1456	0.0653	0.1064	0.0619
30%	0.0597	0.0563	0.0626	0.0549

RMSE, Root Mean Square Error, expressed in years.

3.3. Interpretation

RCTs augmented with older participants (while severe comorbidity is not an exclusion criterion) revealed that [1] the absence of older patient in the RCT does not allow for satisfactory effectiveness estimation using RCT data, and [2] the inclusion of just 1–5% older patients is sufficient to improve estimation accuracy. Although the real-world proportion of older patients (30%) is similar to that of comorbid ones (29.7%), results are different regarding comorbidity: [1] in the absence of comorbid patient in the RCT, predictive modeling already enables a good effectiveness estimation, and [2] the augmentation of RCTs with comorbid patients improves estimation accuracy in a less meaningful manner than with older ones. In other words, allowing a few older patients in the RCT is more important for effectiveness estimation than allowing a few comorbid patients. There are two underlying reasons for this: the correlation between age and comorbidity, and the larger magnitude of the magnitude of the drug-age interaction than this of the drug-comorbidity interaction. In the absence of comorbid patients, the proportion of older patients is reduced due to correlation: out of 30% older patients, 18% remain. This has a strong impact on standard RCT reading, which therefore provides poor effectiveness estimation. In addition, because of the magnitude of the drug-age interaction, the estimation of effectiveness through predictive modeling depends mainly on age, and good estimation accuracy can be achieved. Since the drug-comorbidity interaction is of smaller magnitude, the absence of comorbid patients in the RCT has a smaller impact on estimation accuracy. Likewise, adding more patients with comorbidity results in smaller gains.

3.4. Impact on statistical power

Post-hoc analyses were conducted to explore the impact on the statistical power of augmenting an RCT with 5% older patients, using the RMST difference. The criterion "older age" was chosen because of the high magnitude of the drug-age interaction, and potential impact on statistical power. We chose 5% of augmentation because this percentage was found sufficient for prediction purposes and thus, there would be no reason to include a higher percentage of older people. We computed the 1000 p-values of the tests on the RMST difference in the 1000 simulated RCTs (standard RCT reading). The number of p-values<0.05 ($\alpha = 0.05$) is an estimation of the statistical power. We found that for the 1000 simulated RCTs with 5% of older patients, the highest p-value found was <0.0002, meaning that the statistical power was close to 100%.

4. Discussion

In the present case study we simulated RCTs augmented with patients meeting relevant and carefully chosen exclusion criteria, so as to find the smallest possible proportion of patients with a particular criterion to include for effectiveness prediction purposes.

4.1. Key findings

The main result of our study is that the magnitude of the expected

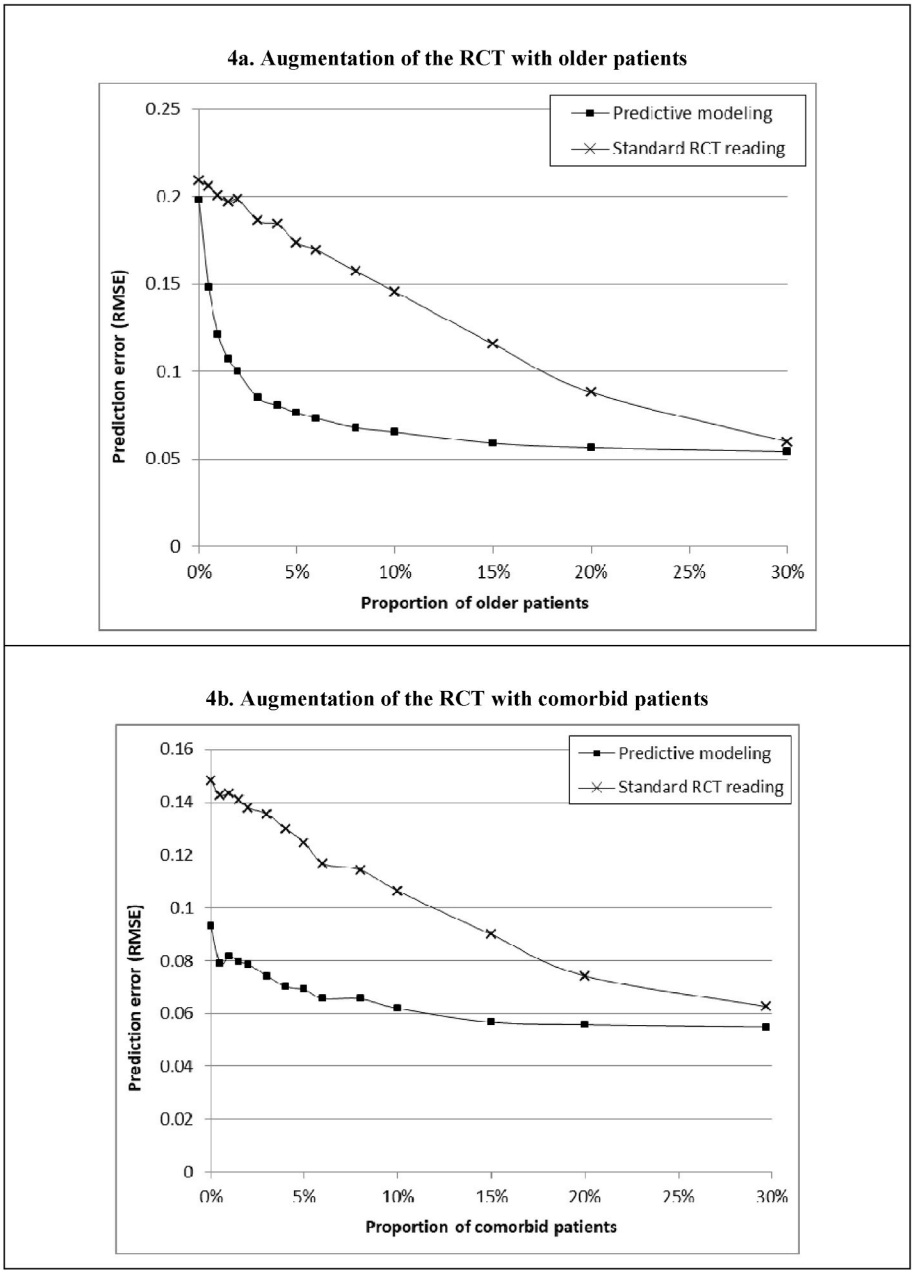
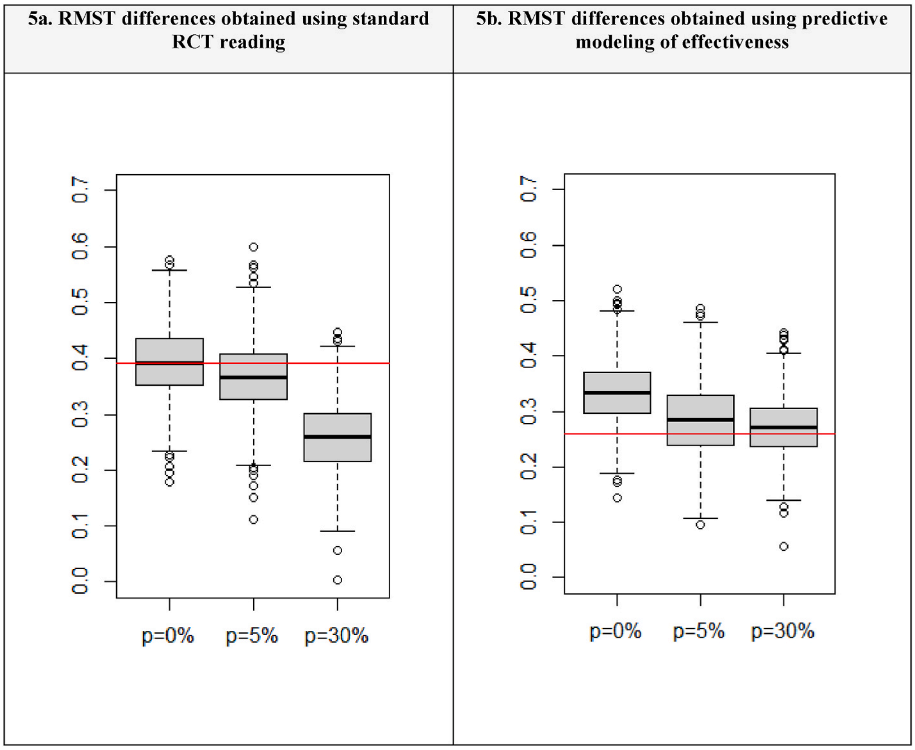


Fig. 4. Error (RMSE) in the estimation of the real-world between-arm difference in RMST using data from 1000 simulated RCTs that include increasing proportions of older patients (4a) or comorbid patients (4b)
RMSE, Root Mean Square Error; RMST, Restricted Mean Survival Time.

effect modification between the investigational treatment and the exclusion criterion is a key element to take into account when considering the use of the RCT augmentation methodology and the choice of eligibility criteria to be relaxed. A higher magnitude of effect modification not only makes the exclusion criterion more relevant to relax, but also provides the possibility to augment the population by a very small proportion and remarkably increase learnings about the investigational treatment's future effectiveness. For the example with older age, we demonstrated that the replacement of 6–30 younger patients per arm in a two-arm trial of 1200 patients (corresponding to 1%–5% proportion of patients) by the same number of older ones in the RCT was sufficient to generate data for sound effectiveness predictions. This result is in line with our previous simulation study [12] suggesting the benefit of augmenting the RCT being mostly gained from re-including the first few real-world patients. The inclusion of this small number of patients is likely to minimize the risk of safety issues and of treatment effect dilution. Moreover, these patients replace others, keeping the trial sample size constant thus adding minimal operational costs.

To our knowledge this methodology is the first to propose the augmentation of a pre-authorization RCT for the purpose of conducting effectiveness predictive modeling using RCT data. Recently, Liu et al. [40] have presented an innovative methodology using artificial intelligence and aiming at optimizing the choice of inclusion criteria in RCTs. Contrary to our methodology the choice of which criterion to relax is automatized whereas we minimize the number of modifications to be made to a standard RCT in recommending that exclusion criteria are carefully chosen *a priori* using scientifically-based hypotheses. Moreover, unlike Liu et al.'s our methodology does not aim at increasing or decreasing the investigational treatment effect size (although it could be a consequence). Rather, it provides the possibility to generate evidence


5a. RMST differences obtained using standard RCT reading

5b. RMST differences obtained using predictive modeling of effectiveness

Fig. 5. RMST differences (in years) obtained from 1000 simulated RCTs augmented with p = 0%, p = 5% and p = 30% comorbid patients, when using the standard RCT reading (Fig. 5a) or predictive modeling of effectiveness (Fig. 5b)
5a. RMST differences obtained using standard RCT reading 5b. RMST differences obtained using predictive modeling of effectiveness.

on its effectiveness in a robust manner. Of note, in Karcher et al. [12] the re-inclusion of a few recently-diagnosed schizophrenia patients, classically excluded from pre-authorization trials, enabled the prediction of an improved treatment effect once this would be prescribed in routine clinical practice. Finally, augmenting a pre-authorization RCT and being able to provide estimations of treatment effectiveness using RCT data may benefit both regulatory approval and access to broader population for market authorization.

The present simulation study assessing an RCT with an augmentation of 5% of older patients shows that the RCT statistical power is close to 100%. This is due to the large effect size of drug B vs A and a rather large number of patients in the RCT. Measuring decrease in statistical power would be more relevant in a context with smaller treatment effect size.

4.2. Limitations of the method

Several limitations need to be pointed out. First, we used the same exponential Cox model to generate the PFS in the generated dataset as in the predictive modeling phase. The model was therefore already partially adapted to the data, which may have yielded overly optimistic results. Second, because our primary focus was to gauge how the magnitude of treatment effect modification impacted predictive model accuracy, we used the two exclusion criteria one at the time to compare two different cases. We did not explore the case in which older age and severe comorbidity are simultaneously used as exclusion criteria. This type of simulation introduces additional statistical questions and requires further exploration. Finally, further studies are needed to explore the question of statistical power.

4.3. Perspective for the implementation of the methodology

The possible implementation of the RCT augmentation methodology in a pre-authorization setting raises several practical questions, in addition to technical ones.

4.3.1. Do I need to design an augmented RCT? Is the trial feasible?

Conducting an augmented RCT may present advantages over a conventional RCT in very specific situations, namely when some exclusion criteria cannot be removed entirely (e.g., for safety reasons) although their strict application would lead to totally excluding specific patient phenotypes and to modifying the distribution of key TEM (hence a risk for biased treatment effect estimation). The relevance of conducting an augmented RCT has to be appraised in the light of the risk of selection bias and of stakeholders expressing concerns about an efficacy-to-effectiveness gap. The first step is thus to explore whether specific exclusion criteria are also possible TEM. The feasibility of this methodology is related to the availability of adequate real-world data source (s). In the present study fully-generated data were used for illustrating and further exploring specific aspects of the methodology. In a real situation, one should use a large real-world dataset, e.g., electronic medical records, containing information on the interaction between treatment and key eligibility criteria. Because the treatment of interest might not been launched yet (e.g., first pre-approval Phase 3 RCT), information on treatment effect modification may be estimated using real-world data on the treatment as prescribed in another indication or on a similar treatment or a treatment of the same class, if this is sensible to believe that treatment effect modification will be similar to that of the investigational treatment. In the absence of adequate data, it is still reasonable to consider *a priori* that augmenting the RCT with 1–5% otherwise-excluded patients will provide sufficient information for effectiveness prediction purposes. This is important to note also that the smallest necessary proportion of patients allowing good effectiveness prediction depends not only on the magnitude of the drug-variable interaction, but also on the RCT sample size and the complexity of the predictive model used.

4.3.2. Will my augmented phase 3 RCT be accepted by regulators, payers and patients?

The population to be included in a pre-authorization RCT is typically discussed with regulatory authorities, sometimes payers and patients, while designing the trial. Regulatory bodies are now explicitly encouraging the broadening of trial populations and the use of adaptive trial designs or trial enrichment [8]. The RCT augmentation methodology could be an option to meet stakeholders' expectations although the acceptability of the methodology by stakeholders is still to be explored.

5. Conclusion

Augmenting RCTs with a few patients of relevant characteristics provides the possibility to generate accurate predictive models of effectiveness with the data they generate. When planning to augment an RCT, the choice of the most relevant exclusion criteria to relax should take into account the suspected importance of treatment effect modification.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2023.101142.

References

[1] M. Hordijk-Trion, M. Lenzen, W. Wijns, P. de Jaegere, M.L. Simoons, W.J. Scholte op Reimer, et al., Patients enrolled in coronary intervention trials are not representative of patients in clinical practice: results from the Euro Heart Survey on Coronary Revascularization, Eur. Heart J. 27 (6) (2006) 671–678.

[2] H.G. Van Spall, A. Toren, A. Kiss, R.A. Fowler, Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review, JAMA, J. Am. Med. Assoc. 297 (11) (2007) 1233–1240.

[3] T. Kennedy-Martin, S. Curtis, D. Faries, S. Robinson, J. Johnston, A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results, Trials 16 (2015) 495.

[4] C. Nordon, T. Bovagnet, M. Belger, J. Jimenez, R. Olivares, H. Chevrou-Severac, et al., Trial exclusion criteria and their impact on the estimation of antipsychotic drugs effect: a case study using the SOHO database, Schizophr. Res. 193 (2018) 146–153.

[5] C. Nordon, H. Karcher, R.H. Groenwold, M.Z. Ankarfeldt, F. Pichler, H. Chevrou-Severac, et al., The "Efficacy-Effectiveness gap": historical background and current conceptualization, Value Health 19 (1) (2016) 75–81.

[6] H.G. Eichler, E. Abadie, A. Breckenridge, B. Flamion, L.L. Gustafsson, H. Leufkens, et al., Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response, Nat. Rev. Drug Discov. 10 (7) (2011) 495–506.

[7] A. Makady, A. van Veelen, P. Jonsson, O. Moseley, A. D'Andon, A. de Boer, et al., Using real-world data in health technology assessment (HTA) practice: a comparative study of five HTA agencies, Pharmacoeconomics 36 (3) (2018) 359–368.

[8] The Food and Drug Administration, Enhancing the Diversity of Clinical Trial Populations, Eligibility Criteria, Enrollment Practices, and Trial Designs, 2020.

[9] K. Panayidou, S. Gsteiger, M. Egger, G. Kilcher, M. Carreras, O. Efthimiou, et al., GetReal in mathematical modelling: a review of studies predicting drug effectiveness in the real world, Res. Synth. Methods 7 (3) (2016) 264–277.

[10] M. Happich, A. Brnabic, D. Faries, K. Abrams, K.B. Winfree, A. Girvan, et al., Reweighting randomized controlled trial evidence to better reflect real life - a case study of the innovative medicines initiative, Clin. Pharmacol. Ther. 108 (4) (2020) 817–825.

[11] N.T. Longford, Selection bias and treatment heterogeneity in clinical trials, Stat. Med. 18 (12) (1999) 1467–1474.

[12] H. Karcher, S. Fu, J. Meng, M.Z. Ankarfeldt, O. Efthimiou, M. Belger, et al., The "RCT augmentation": a novel simulation method to add patient heterogeneity into phase III trials, BMC Med. Res. Methodol. 18 (1) (2018) 75.

[13] J.H. Lewis, M.L. Kilgore, D.P. Goldman, E.L. Trimble, R. Kaplan, M.J. Montello, et al., Participation of patients 65 years of age or older in cancer clinical trials, J. Clin. Oncol. 21 (7) (2003) 1383–1389.

[14] G.L. Stark, K.M. Wood, F. Jack, B. Angus, S.J. Proctor, P.R. Taylor, et al., Hodgkin's disease in the elderly: a population-based study, Br. J. Haematol. 119 (2) (2002) 432–440.

[15] A. Thyss, E. Saada, L. Gastaud, F. Peyrade, D. Re, Hodgkin's lymphoma in older patients: an orphan disease? Mediterr. J. Hematol. Infect. Dis. 6 (1) (2014), e2014050.

[16] N. Duma, S.M. Kothadia, T.U. Azam, S. Yadav, J. Paludo, J. Vera Aguilera, et al., Characterization of comorbidities limiting the recruitment of patients in early phase clinical trials, Oncol. 24 (1) (2019) 96–102.

[17] C. Terschuren, S. Gierer, C. Brillant, U. Paulus, M. Loffler, W. Hoffmann, Are patients with Hodgkin lymphoma and high-grade non-Hodgkin lymphoma in clinical therapy optimization protocols representative of these groups of patients in Germany? Ann. Oncol. 21 (10) (2010) 2045–2051.

[18] S. Jin, R. Pazdur, R. Sridhara, Re-evaluating eligibility criteria for oncology clinical trials: analysis of investigational new drug applications in 2015, J. Clin. Oncol. 35 (33) (2017) 3745–3752.

[19] M.S. Sedrak, R.A. Freedman, H.J. Cohen, H.B. Muss, A. Jatoi, H.D. Klepin, et al., Older adult participation in cancer clinical trials: a systematic review of barriers and interventions, CA A Cancer J. Clin. 71 (1) (2021) 78–92.

[20] S.M. Lichtman, R.D. Harvey, M.A. Damiette Smit, A. Rahman, M.A. Thompson, N. Roach, et al., Modernizing clinical trial eligibility criteria: recommendations of the American society of clinical oncology-friends of cancer research organ dysfunction, prior or concurrent malignancy, and comorbidities working group, J. Clin. Oncol. 35 (33) (2017) 3753–3759.

[21] C.C. ESBBDRC, Ann arbor staging classification for Hodgkin lymphoma, in: AJCC Cancer Staging Manual, seventh ed., Springer, New York, NY, 2010.

[22] A. van Rossum, Python Tutorial, Technical report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), 1995.

[23] H. Fowler, A. Belot, L. Ellis, C. Maringe, M.A. Luque-Fernandez, E.N. Njagi, et al., Comorbidity prevalence among cancer patients: a population-based cohort study of four cancers, BMC Cancer 20 (1) (2020) 2.

[24] T.A. Eyre, E.H. Phillips, K.M. Linton, A. Arumainathan, S. Kassam, A. Gibb, et al., Results of a multicentre UK-wide retrospective study evaluating the efficacy of brentuximab vedotin in relapsed, refractory classical Hodgkin lymphoma in the transplant naive setting, Br. J. Haematol. 179 (3) (2017) 471–479.

[25] C. Pellegrini, A. Broccoli, A. Pulsoni, L. Rigacci, C. Patti, G. Gini, et al., Italian real life experience with brentuximab vedotin: results of a large observational study on 234 relapsed/refractory Hodgkin's lymphoma, Oncotarget 8 (53) (2017) 91703–91710.

[26] J. Walewski, A. Hellmann, N. Siritanaratkul, G.H. Ozsan, M. Ozcan, S. Chuncharunee, et al., Prospective study of brentuximab vedotin in relapsed/refractory Hodgkin lymphoma patients who are not suitable for stem cell transplant or multi-agent chemotherapy, Br. J. Haematol. 183 (3) (2018) 400–410.

[27] S.M. Szabo, I. Hirji, K.M. Johnston, A. Juarez-Garcia, J.M. Connors, Treatment patterns and costs of care for patients with relapsed and refractory Hodgkin lymphoma treated with brentuximab vedotin in the United States: a retrospective cohort study, PLoS One 12 (10) (2017), e0180261.

[28] Cancer Research UK, Hodgkin lymphoma survival statistics: one-year net survival by stage, Available from, https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/hodgkin-lymphoma/survival#heading-Three.

[29] J.M. Connors, W. Jurczak, D.J. Straus, S.M. Ansell, W.S. Kim, A. Gallamini, et al., Brentuximab vedotin with chemotherapy for stage III or IV Hodgkin's lymphoma, N. Engl. J. Med. 378 (4) (2018) 331–344.

[30] E.A. Zagadailov, S. Corman, V. Chirikov, C. Johnson, C. Macahilig, B. Seal, et al., Real-world effectiveness of brentuximab vedotin versus physicians' choice chemotherapy in patients with relapsed/refractory Hodgkin lymphoma following autologous stem cell transplantation in the United Kingdom and Germany, Leuk. Lymphoma 59 (6) (2018) 1413–1419.

[31] P. Kaloyannidis, M. Hertzberg, K. Webb, A. Zomas, R. Schrover, M. Hurst, et al., Brentuximab vedotin for the treatment of patients with relapsed or refractory Hodgkin lymphoma after autologous stem cell transplantation, Br. J. Haematol. 188 (4) (2020) 540–549.

[32] M. Yildirim, V. Kaya, O. Demirpence, S. Paydas, The role of gender in patients with diffuse large B cell lymphoma treated with rituximab-containing regimens: a meta-analysis, Arch. Med. Sci. 11 (4) (2015) 708–714.

[33] R.P. A'Hern, Restricted mean survival time: an obligatory end point for time-to-event analysis in cancer trials? J. Clin. Oncol. 34 (28) (2016) 3474–3476.

[34] D.H. Kim, H. Uno, L.J. Wei, Restricted mean survival time as a measure to interpret clinical trial results, JAMA Cardiol. 2 (11) (2017) 1179–1180.

[35] K. Pak, H. Uno, D.H. Kim, L. Tian, R.C. Kane, M. Takeuchi, et al., Interpretability of cancer clinical trial results using restricted mean survival time as an alternative to the hazard ratio, JAMA Oncol. 3 (12) (2017) 1692–1696.

[36] P. Royston, M.K. Parmar, 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. 13 (2013) 152.

[37] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2008.

[38] A.C. Faul, Regularization, in: C. Press (Ed.), A Concise Introduction to Machine Learning, Taylor & Francis Group, 2020.

[39] J. Li, Assessing the accuracy of predictive models for numerical data: not r nor r2, why not? Then what? PLoS One 12 (8) (2017), e0183250.

[40] R. Liu, Evaluating eligibility criteria of oncology trials using real-world data and AI, Nature. 592 (7855) (2021) 629–633.