Overcoming the discrepancies between RCTs and real-world data by accounting for *Selection criteria, Operations,* and *Measurements of Outcome (SOMO)*

Luca Marzano*[a], Adam S. Darwich[a], Asaf Dan[b], Salomon Tendler[bc], Rolf Lewensohn[b], Luigi De Petris[b], Jayanth Raghothama[a], Sebastiaan Meijer[a]

**Corresponding author: Luca Marzano, Hälsovägen 11 C 141 57 Huddinge (Sweden), lmarzano@kth.se*

[a]Division of Health Informatics and Logistics, School of Engineering Sciences in Chemistry, Biotechnology and Health (CBH), KTH Royal Institute of Technology, Stockholm, Sweden

[b]Dept. of Oncology-Pathology, Karolinska Institutet and the Thoracic Oncology Center, Karolinska University hospital, Stockholm, Sweden

[c]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY

Conflict of interest statement

The authors declare no competing interests for this work.

Funding Information

The Swedish Cancer Society (grant no. CAN 2018/597 and CAN2021/1469 Pj01) to R. Lewensohn and from the Stockholm Cancer Society (grant no. #201202 to R. Lewensohn and #174063, #201103 and #231123 to L. De Petris).

Keywords: clinical trials, oncology, trial simulation, survival analysis, population analysis, personalized medicine, chemotherapy, cancers, biostatistics, bioinformatics

Abstract

The potential of real-world data to inform clinical trial design and supplement control arms has gained much interest in recent years. The most common approach relies on matching real-world patient cohorts to clinical trial baseline covariates using propensity score techniques. However, recent studies pointed out that there is a lack of replicability, generalisability, and consensus. Further, few studies consider differences in operational processes. Systematically accounting for confounders, including hidden effects related to the clinical treatment process and clinical trial study protocol, would potentially allow for improved translation between clinical trial and real-world data and enable learning across translational activities.

In this paper, we propose an approach that aims to explore and examine these confounders by investigating the impact of selection criteria and operations on the measurements of outcome. We tested the approach on small cell lung cancer patients receiving platinum-based chemotherapy regimens (n=1,224).

The results showed that the discrepancy between real-world and clinical trial data potentially depends on differences in covariate characteristics and operations (*e.g.*, censoring mechanism, the process of pre-trial patient selection related to ECOG-performance status 2 patients).

This work builds on current approaches and suggests areas of improvement for systematically accounting for differences in outcomes between study cohorts. Continued development of the method presented here could pave the way for transferring learning across studies and developing mutual translation between the real-world and clinical trials to inform future studies design.

Introduction

During the past few years there has been discussions on the limitations of randomised clinical trials (RCTs) due to their high costs and the challenge of translating clinical outcomes between RCT cohorts and real-world patient populations[1,2]. Since the 21st Century Cures Act in 2015, the potential for translating between real-world data and clinical trials to improve regulatory decisions has gained attention[3–7]. A growing body of research has focused on the extrapolation of real-world data to inform clinical trial design, often described as emulation of RCT control arms or simulated (synthetic) data[8–12], with the purpose of reproducing clinical trial outcomes[8–11,13–17].

In an attempt to adjust for confounders, analyses have often focused on recreating the inclusion criteria of clinical trials in real-world cohorts[8–10,14,16,18–26]. The most common method is patient matching based on propensity score[27]. Propensity score is formally defined by Rosenbaum and Rubin as the conditional probability of assignment to a particular treatment[28]. In other words, the method gives a probability of a real-world patient being enrolled in a clinical trial study arm given a vector of observed covariates. Then, the real-world data cohort is adjusted by including only patients with high propensity score. Propensity score approaches, including variations or adaptation of this method, still constitute the main proposed technique for adjusting for confounders in real-world cohorts based on the conditions of control trial arms[8,10,16,19,21,22,24,26,29,30].

However, propensity score approaches have been shown to be limited in this aspect since a probabilistic empirical approach is highly sensitive to undetected confounders and biases of the data[1,8,10,11,14,25]. Indeed, the simple alignment of patient characteristics does not result in similar outcomes between the populations[10,11,16]. It has repeatedly been pointed out that there is a lack of replicability and generalisability[1,9,10,13,17], with only few clinical trials being replicable based on real-world data[10,13,17]. Results have been mixed depending on the specific case study[13,17,31] and measurements of outcome[10,16] (*e.g.*, overall survival and intermediate endpoints[10,14]).
The lack of translatability has been attributed to the differences in populations, such as baseline confounders and key eligibility criteria that are not available in the data[10,11,13,14]. The main focus of improvement has been on how patient demographics affect measurements of outcome with little focus on the operations and processes behind the data[8,10,13,23,32]. The potential operational differences between clinical practice and clinical trial protocols have been mentioned as potential confounders but yet to be fully explored[1,2,8,10,11,13,32]. There is a need to investigate the bias that is introduced by differences in investigation and clinical assessment[1,10,11,23,32] (*e.g.*, lack of pre-trial monitoring in the clinical practice[1]), and potential differences in the RCTs monitoring process compared to real-world patients, with more detailed and potentially more frequent follow-up on tumour response and adverse effects, and decision-making such as withdrawal of the therapy[1,10,11,25,32].

Translation between clinical trials and real-world populations is therefore still not completely understood[1,8–10,13,14,16,17]. Further refinement of proposed methodologies are required to realise the potential of real-world data to inform clinical trial design[6,9,10,13,33]. In the past, the added value of a mechanistic systems view of translation in drug development has been beneficial for other areas of model-informed drug development, such as quantitative *in vitro-in vivo* extrapolation and physiologically-based pharmacokinetics of metabolic drug-drug interactions[34]. Developing systems approaches for real-world evidence could enable a similar learn-confirm cycle and learning across studies, where the represented systems include not only the patient, disease and treatment, but also the operational context[35].

In this paper we propose an approach that aims to explore how real-world data could inform clinical trial design by systematically accounting for the known differences in population samples (randomisation) and operation (protocols and clinical practice). We abbreviate the approach as SOMO as it is based on exploring the effects of Selection Criteria (S), Operations (O) and study protocols, on

the replication of the Measurements of Outcome (MO). We developed and tested the SOMO approach using clinical trial and real-world data on extensive-disease small cell lung cancer (ED-SCLC) patients receiving platinum etoposide chemotherapy.

Methods

The SOMO approach

Figure 1 shows the SOMO approach and the accompanying data analysis. In short, the analysis included the following components:

(S). Selection Criteria refers to all aspects related to the baseline variables that define the population, the biology of the disease, and the inclusion and exclusion criteria of the clinical datasets.

(O). Operations and study protocols refer to aspects related to operational processes (or mechanisms) occurring during treatment. These can be grouped into longitudinal factors related to the disease (*e.g.*, tumour progression and adverse effects), the study protocol operations (*e.g.*, removal and censoring of patients that do not adhere to the study protocol), or to the potential differences from the real-world routine healthcare operations (*e.g.*, adjustment of doses, change of treatment due to relapse, patients refusing to receive treatment).

(MO). Measurements of Outcome refer to the metrics used to evaluate treatment efficacy (and safety, when available) and estimation of the feasibility of translation between real-world and clinical trials using a comparative approach. These can be one or multiple outcomes depending on the study endpoints and the statistical analysis defined by the protocol (*e.g.*, overall survival, progress free survival, toxicity, exposure, or overall response, *etc*.), as well as and the available information from the real-world cohort.

Prior to commencing the analysis, available data are retrieved and pre-processed to harmonise trials and real-world data into an aggregated dataset. The analytical pipeline can then be framed in the following steps:

Explorative analysis. An exploratory analysis is performed to define the SOMO components. First, the comparison between trials and real-world outcomes is carried out to estimate differences in measurements of outcome between cohorts. Then, a comparison between the two populations is performed to detect potential mismatch between patient covariates due to selection (and randomisation). For operational aspects, potential confounders are investigated through comparison between study protocols and clinical practice with the aid of longitudinal curves of outcomes (*e.g.*, Kaplan Meier or dose response curves), and clinical expert feedback.

Estimating the impact of factors on translation between cohorts. The potential impact of selection criteria and operations is estimated by applying matching based on the relevant variables and simulating effects of operations (*e.g.*, blinding), and comparing the impact on outcome measures. The clinical expert involvement in all the steps of the analysis allowed supplemental information on missing data, and contextualisation of the analysis components, as well as feedback for validation of the results by detecting and discussing eventual biases and confounders.

Hence, one of the main outcomes was a list of potential factors related to selection criteria and operations that could contribute to the translational gap and quantitative estimation of their effects.

Case study: extensive disease small cell lung cancer patients receiving platinum-etoposide chemotherapy.

Cohort description

In this study a mixed cohort was collated, including ED-SCLC patients that had received platinum etoposide chemotherapy and been included as part of a retrospective real-world dataset from Karolinska University Hospital (Stockholm, Sweden) between 2008 and 2016[36] (RWD KI, n patients=223). The study was approved by the institutional review boards at Karolinska Institutet and at Stockholm County Council (2016/8-31).

The comparative group originated from open data shared through the Project Data Sphere Initiative[37], including participants that were randomised into the control arm from three randomised phase III clinical trials: PDS_Amgen (NCT00119613, n=232), PDS_Allianc (NCT00003299, n=270), PDS_EliLilly (NCT00363415, n=370), and 3 phase Ib-II trials PDS_PHASE2_Alliance (NCT00453154, n=46), PDS_PHASE2_EliLilly (NCT01439568, n=41), PDS_PHASE2_G1Thera (NCT02499770, n=37). Some patients of PDS_EliLilly were censored after the study was declared futile after the interim analysis. These were removed prior to the analysis.

The aggregated mixed cohort encompassed n=1,224 patients in total. The common patient variables were age, sex, brain metastasis (BM), Eastern Cooperative Oncology Group performance status (ECOG), the cohort from which the data were retrieved (STUDY), and the label referring to if patients were from the RCT or real-world cohort (STUDY_TYPE), progression-free survival (PFS), Overall Survival (OS), and eventual censoring (CENSOR). For the real-world patients the 8th version of the TNM staging was available. The real-world cohort was re-staged in a previous validation study[38]. A summary of the cohorts is reported in Table S1.

Survival Analysis

The case study analysis was divided into two parts. First, the clinical trials were aggregated into one dataset. Then, a pairwise analysis was carried out comparing the real-world cohort and the individual clinical trials. The analysis explored differences in the survival outcomes (Overall Survival and Progression-Free Survival). This was done using Kaplan Meier Curves and Cox proportional Hazard ratios. Progression-Free Survival was reported only for a small subset of clinical trials patients, mainly from PDS_EliLilly.

For the pairwise analysis a matched randomisation was simulated by sampling a surrogate cohort at a 1:1 ratio for the two populations (Figure S1). This simulation was carried out through random sampling of n patients from the analysed trial and the real-world cohort (n=250). Then, the Cox hazard ratios between real-world and clinical trials cohorts were computed. The randomisation was repeated 100 times to estimate the variability of the outcomes. Different scenarios were simulated by intervening on the sampling by accounting for the aspects detected from the SOMO approach (e.g., sub-cohorts' stratification). The main assumption behind this analysis was that in the ideal scenario there would be no significant differences in outcome between the two populations. Hence, comparing the variation of outcomes and how these get closer to this ideal scenario, allowed the estimation of the robustness of eventual interventions in the trial design.

Table 1 summarises how the SOMO approach was applied to the ED-SCLC case study. The main techniques used during the analysis were: variable or sub-cohort stratification, propensity score matching with weights computed using logistic regression[27], and oversampling to generate simulated cohorts using the standard SMOTENC algorithm based on k-nearest neighbours[39].

The selection criteria analysis identified several differences in the real-world population as compared to the clinical trials, this included: an age distribution skewed towards older age, a more balanced sex ratio, and higher number of patients with ECOG 2 (Table S1). Baseline brain metastasis records were not available for all clinical trial patients, or not explicit mentioned as an exclusion criterion (*e.g.*, PDS_Alliance).

The operations analysis highlighted differences in the censoring mechanisms between the two settings. Real-world censoring corresponded to a few patients with long survival (right censoring). For trial patients censoring occurred with high frequency during the first months after randomisation. We explored this aspect by excluding all the censored patients, simulating the censoring on the real-world patient cohort, and with propensity score matching accounting for censoring as a covariate.
Another explored operational aspect was the pre-trial assessment for enrolment of ECOG 2 clinical trials patients. Hence, we compared the ECOG 2 patients with the real-world ECOG 3 patients receiving the same treatment to evaluate potential variations in ECOG between screening phase and randomisation.

Finally, the longitudinal shapes of Kaplan Meier curves were examined by investigating the effect of sample size and censoring on the real-world cohort. This was done by simulating a real-world population with a matched number of patients using SMOTENC oversampling, and by simulating the censoring using the time distribution of the clinical trials.

Results

We observed a significant difference in OS (hazard ratio: 0.65 [0.55, 0.75]) and PFS (hazard ratio: 0.70 [0.58, 0.85]) when comparing real-world and clinical trial patient cohorts (Table 2). Figure 2.a1 shows that the shapes of the Kaplan Meier Curve also differed between the two cohorts (*i.e.*, lower survival in the first few weeks for the trials, followed by a higher survival curve compared with the real-world sub-cohort, and a similar survival drop at 125 days). All the results of the aggregated analysis are reported in Table S2.

The key selection criteria for which similar OS was achieved for both cohorts were ECOG 2 (Table 2), the real-world patients with TNM cancer staging IVA using conventional stratification and with oversampling of IVA patients (Figure 2b). For what concern PFS, similar outcomes were observed for all ECOG stratifications, TNM IVA cancer staging of the real-world cohort, and traditional propensity score matching (Table 2).
For what concerns operations, similar OS was observed using propensity score matching informed by the censoring operations (Figure 2c.2). ECOG 0-1, age, removing censoring, and longitudinal profile analysis reduced the gap between RCTs and real-world patients (Table S2). Similar PFS was observed for trial ECOG 2 and real-world ECOG 3 patients (Figure 3). Interestingly, there was a switch in survival trend between real-world and trials when censoring was leveraged (Table 2).

In addition, Figure 3 shows an interesting adherence of the overall survival curves for the first days between ECOG 2 in the trials and ECOG 3 in the real-world. This could explain that the lower survival in first weeks for the trials in Figure 2.a1 could be related to potential variations of this baseline between the screening and the randomisation.

Figures 4 and 5 show the results for the matching simulation analysis for OS of PDS_EliLilly and PDS_Amgen, and PFS for PDS_EliLilly. The results from the pairwise matching simulation analysis agreed with the results obtained from the aggregated cohort. In addition, estimating the variability of the interventions allowed to explore eventual differences between the trials (e.g., Figure 4 shows that PDS_Amgen is a trial more feasible to the translation compared to PDS_EliLilly). Figure S2 reports the overall survival matching simulation for PDS_Allianc and the aggregated cohort of Phase I-II patients.

Discussion

In this study, we developed and tested a novel approach that aims to characterise the translatability between clinical trials and real-world patients. The was done by through the development of a systematic approach that attempts to account for known differences in population samples and operational aspects.

Previous real-world evidence studies of lung cancer have mainly focused on non-small cell lung cancer, thus leaving small cell lung cancer understudied[40]. To the best of the authors' knowledge, this is the most comprehensive real-world evidence study for small cell lung cancer disease. Analyses of selection criteria and measurements of outcome showed comparable results to previous studies, where a larger difference was seen in overall survival compared to other intermediate endpoints[10,14], and population differences in age, under-representation of ECOG 2 in the real-world, and an under-representation of females in clinical trials[41–43].

The similarity between the stage IVA real-world patients and the clinical trial patients presented an interesting aspect. The 8th TNM staging (with sub-categorisation of ED-SCLC patients in IVA and IVB stages) was not yet developed when the trials were executed, and the LD/ED staging is still largely used to define treatment intention[38]. Previous studies have pointed out significant differences in survival between IVA and IVB patients[38,44–46]. This result may be due to selection bias in RCTs, favouring younger patients with comparably better overall health and lower disease burden, which is only partially captured by the TNM staging. However, the low sample size of IVA patients in the real-world cohort is a source of potential bias and further investigation would be needed to confirm this.

Censoring in clinical trials is a known potential confounder[47]. This work underlined that this was a key operational factor with a relatively high impact, thus leading to a potential overestimation of overall survival in clinical trials. Moreover, when censored patients were removed, other variables reported in Table S2 reduced the difference in OS. Thus, indicating how this confounder is biasing the estimated difference in OS between the cohorts. Figure 3 pointed out another relevant difference between processes and how these can probably influence the outcomes: ECOG 2 trial patients showed more similar behaviour to ECOG 3 real-world patients. This could potentially be explained by differences in patient condition between the pre-trial phase and randomisation that could lead to worsening in baseline ECOG.

The results showed that the challenge of replicating outcomes in clinical trials from real-world patients depends on differences in patient characteristics and operations. This can be seen in Figure 2a where the survival curves of the trials and real-world patients present different shapes due to potential operational differences that occur during the follow-up. Results in Figure 3 could explain the high mortality in the first weeks, while the censoring could be an important contribution of the steep survival slope before the similar drop around 125 days.
This is underlined by the progression-free survival analysis, where we obtained perfect translation when considering the population variables (performance status and TNM staging), but a switch of survival trends when accounting for the censoring (Table 2 and Figure 5). This result is surprising, and we suspect it could be due to more frequent monitoring of clinical trial patients that could result in a shorter reported progression-free survival when adverse events or relapse occur.

In previous works, the dominant approach has been the propensity score informed by clinical trial data[10,24–26,30,31,46]. No in-depth examination of differences in operations has been done before. Figure 2c demonstrates the relevant contribution operations to the translational gap. Indeed, in the ideal condition of having two populations with the same support of baseline covariates and sufficient population sample, accounting for the eventual operational differences, should theoretically account for any misalignment in study outcomes [9,11,31,46]. One limitation of the propensity score is that it works from only one direction by selecting real-world patients that match clinical trials, thus not accounting for any additional findings in the real-world population. This approach is limited for prospective applications since it would blindly work only in the scenario where all relevant confounders are accounted for.

Real-world data can inform any disparities between RCTs and clinical practice[1]. Previous studies indicate that this difference is consistent and independent of the studied case[9,9,10,13,17]. One of the observations in this study related to the time-dynamic differences in Kaplan Meier curves between RCTs and RWD. Similar Kaplan Meier curves as in Figure 2a.1 have previously been found in other oncology case studies[10,14], thus suggesting that the impact of operations is not isolated only to small

cell lung cancer. Further research is needed to understand the origins of this difference[35]. It has been noted that in translating between RCTs and RWD, understanding the complexity of clinical practice and treatment processes may be instrumental to explaining this[35].

The SOMO approach aims to contribute to the research direction on establishing an analytical process to estimate the challenging trade-off between internal validity and generalizability[6,11,13,33]. In this work we proposed how such a framework could be beneficial for both translational directions: clinical trials can be improved by understanding the discrepancy with the real-world, and real-world therapies can be leveraged by comparing the discrepancies of trials from the real-world (*e.g.*, the comparison of Amgen and Eli Lilly phase III trials in Figure 4).

We are not claiming to have developed a definitive approach, rather we believe that further development of the general method will allow for learning across translational activities. This may contribute to building regulatory acceptance in the real-world evidence approach over time as has been the case with other mechanistic modelling efforts (*e.g.*, metabolic drug-drug interaction predictions[34,48]). This is the first study that try to push the real-world evidence paradigm in a similar learn-confirm cycle from which we learn case by case, reinforcing the background theory behind the methodology, and pragmatically improve the usage of real-world data in future clinical trials.

The work presents some limitations to be address in the future research. The real-world data represented only a single centre and lacked longitudinal variables (e.g., tumour progression, doses, and adverse effects). To address these issues, the collection of these variables and the extensive collection of data is being performed. Expansion of the real-world patient cohort trough national cancer registries is the next step of the analysis. Moreover, SOMO was tested on small cell lung cancer, and research on other case studies would be beneficial to increase the generalizability of the work. In this study, the experts involved represented the clinical side of the real-world domain. In the future, we will explore the involvement of experts on clinical trial in oncology and regulatory agencies.

The increasing presence of real-world data in clinical trial studies constitute a natural step for regulatory decision and future study designs. Leveraging approaches such as SOMO would pave the way to understand how we can use real-world data to close the gap between internal validity and generalizability of clinical trials.

Study Highlights

What is the current knowledge on the topic?

Real-world data have the potential to inform clinical trial design, control arms, and regulatory assessment. However, real-world evidence studies have shown poor replication and generalizability and a lack of consensus on the analytical process, thus underlining that the mechanisms that would allow the translation between clinical trials and real-world populations are still not completely understood.

What question did this study address?

What are the mechanisms that would allow translation between clinical trials and real-world? How can we design a comprehensive and systematic approach to explore the grade of translation? Does the approach work in a challenging real-world case study such as small cell lung cancer chemotherapy?

What does this study add to our knowledge?

Differences in operations and protocols have a relevant impact on the gap in clinical outcomes. These must be studied in concomitance with the selection criteria of the baselines. Previous works proposed pure empirical approaches (such as propensity score), and limitations of the findings can be related to the lack of consideration of operational differences between trials and real-world practice. Our approach allowed novel insights regarding which aspects would benefit from further investigation to improve the design of small cell lung cancer studies (ECOG 2 underrepresentation and pre-trial

biases, exploring the therapies with the new TNM staging categories, operational biases of trials censoring and progress free survival)

How might this change clinical pharmacology or translational science?
Designing a comprehensive and systematic approach to investigate how selection criteria and operations are impacting on the measurements of outcome would allow us to estimate the trade-off between internal validity and generalizability of clinical trials. Thus, pushing real-world evidence towards a learn and confirm cycle from which we learn case by case and close the translational gap between clinical trials and real-world populations.

Acknowledgments
This project is a contribution to the Centre for Data-Driven Health (CDDH), KTH Royal Institute of Technology (https://www.kth.se/sv/cddh).

Authors Contribution
L.M. and A.S.D. wrote the manuscript. L.M., A.S.D., A.D., S.T., J.R., and S.M. designed the research. L.M., A.S.D., S.T., A.D., R.L., and L.D.P. performed the research. L.M., A.S.D., S.T., A.D., R.L., and L.D.P. analyzed the data

References
1. Beaulieu-Jones, B. K. *et al.* Examining the Use of Real-World Evidence in the Regulatory Process. *Clinical Pharmacology & Therapeutics* **107**, 843–852 (2020).
2. Franklin, J. M. & Schneeweiss, S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? *Clinical Pharmacology & Therapeutics* **102**, 924–933 (2017).
3. Schurman, B. The Framework for FDA's Real-World Evidence Program. *Applied Clinical Trials* **28**, 15–17 (2019).
4. Dagenais, S., Russo, L., Madsen, A., Webster, J. & Becnel, L. Use of Real-World Evidence to Drive Drug Development Strategy and Inform Clinical Trial Design. *Clinical Pharmacology & Therapeutics* **111**, 77–89 (2022).
5. Burns, L. *et al.* Real-World Evidence for Regulatory Decision-Making: Guidance From Around the World. *Clinical Therapeutics* **44**, 420–437 (2022).
6. Baumfeld Andre, E. *et al.* The Current Landscape and Emerging Applications for Real-World Data in Diagnostics and Clinical Decision Support and its Impact on

Regulatory Decision Making. *Clinical Pharmacology & Therapeutics* **112**, 1172–1182 (2022).

7. Campbell, U. B., Honig, N. & Gatto, N. M. SURF: A Screening Tool (for Sponsors) to Evaluate Whether Using Real-World Data to Support an Effectiveness Claim in an FDA Application Has Regulatory Feasibility. *Clinical Pharmacology & Therapeutics* **114**, 981–993 (2023).
8. Wang, C. Y. *et al.* Uncontrolled Extensions of Clinical Trials and the Use of External Controls—Scoping Opportunities and Methods. *Clinical Pharmacology & Therapeutics* **111**, 187–199 (2022).
9. Wang, S. V. *et al.* Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. *Nature Communications* **13**, 1–11 (2022).
10. Jemielita, T. *et al.* Replication of Oncology Randomized Trial Results using Swedish Registry Real World-Data: A Feasibility Study. *Clinical Pharmacology & Therapeutics* **110**, 1613–1621 (2021).
11. Sreeram V. Ramagopalan, Ramagopalan, S. V., Ramagopalan, S. V., Simpson, A. & Sammon, C. J. Can real-world data really replace randomised clinical trials? *BMC Medicine* **18**, 13–13 (2020).
12. Lin, J., Liao, R. & Gamalo-Siebers, M. Dynamic incorporation of real world evidence within the framework of adaptive design. *J Biopharm Stat* **32**, 986–998 (2022).
13. He, Z. *et al.* Clinical Trial Generalizability Assessment in the Big Data Era: A Review. *Clinical and Translational Science* **13**, 675–684 (2020).

14. Tan, K. *et al.* Emulating Control Arms for Cancer Clinical Trials Using External Cohorts Created From Electronic Health Record-Derived Real-World Data. *Clinical Pharmacology & Therapeutics* **111**, 168–178 (2022).

15. Franklin, J. M., Glynn, R. J., Martin, D. & Schneeweiss, S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clinical Pharmacology & Therapeutics* **105**, 867–877 (2019).

16. Franklin, J. M. *et al.* Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. *Clinical pharmacology and therapeutics* **107**, 817–826 (2020).

17. Bartlett, V., Dhruva, S., Shah, N., Ryan, P. & Ross, J. Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence. *JAMA NETWORK OPEN* **2**, (2019).

18. Abrahami, D. *et al.* Use of Real-World Data to Emulate a Clinical Trial and Support Regulatory Decision Making: Assessing the Impact of Temporality, Comparator Choice, and Method of Adjustment. *Clinical Pharmacology & Therapeutics* **109**, 452–461 (2021).

19. Ho, M. *et al.* Examples of Applying RWE Causal-Inference Roadmap to Clinical Studies. *STATISTICS IN BIOPHARMACEUTICAL RESEARCH* (2023) doi:10.1080/19466315.2023.2177333.

20. Stewart, M. *et al.* An Exploratory Analysis of Real-World End Points for Assessing Outcomes Among Immunotherapy-Treated Patients With Advanced Non–Small-Cell Lung Cancer. *JCO Clinical Cancer Informatics* 1–15 (2019) doi:10.1200/cci.18.00155.

21. Oksen, D. *et al.* Treatment effectiveness in a rare oncology indication: Lessons from an external control cohort study. *Clinical and Translational Science* **15**, 1990–1998 (2022).

22. Rivera, D. R. *et al.* The Friends of Cancer Research Real-World Data Collaboration Pilot 2.0: Methodological Recommendations from Oncology Case Studies. *Clinical Pharmacology & Therapeutics* **111**, 283–292 (2022).

23. Cramer-van der Welle, C. M. *et al.* Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. *Scientific Reports* **11**, 6306 (2021).

24. Lakdawalla, D. N. *et al.* Predicting Real-World Effectiveness of Cancer Therapies Using Overall Survival and Progression-Free Survival from Clinical Trials: Empirical Evidence for the ASCO Value Framework. *Value in Health* **20**, 866–875 (2017).

25. Lasiter, L. *et al.* Real-world Overall Survival Using Oncology Electronic Health Record Data: Friends of Cancer Research Pilot. *Clinical Pharmacology & Therapeutics* **111**, 444–454 (2022).

26. Loureiro, H. *et al.* Matching by OS prognostic score to construct external controls in lung cancer clinical trials. *Clinical Pharmacology & Therapeutics* **n/a**,.

27. Webster-Clark, M. *et al.* Using propensity scores to estimate effects of treatment initiation decisions: State of the science. *Statistics in Medicine* **40**, 1718–1735 (2021).

28. Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).

29. Loureiro, H., Becker, T., Bauer-Mehren, A., Ahmidi, N. & Weberpals, J. Artificial Intelligence for Prognostic Scores in Oncology: a Benchmarking Study. *Frontiers in artificial intelligence* **4**, (2021).
30. Liu, R. *et al.* Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature* **592**, 629–633 (2021).
31. Franklin, J. *et al.* Emulating Randomized Clinical Trials With Nonrandomized Real-World Evidence Studies First Results From the RCT DUPLICATE Initiative. *CIRCULATION* **143**, 1002–1013 (2021).
32. Feinberg, B. A. *et al.* Use of Real-World Evidence to Support FDA Approval of Oncology Drugs. *Value in Health* **23**, 1358–1365 (2020).
33. Bolislis, W. R., Fay, M. & Kühler, T. C. Use of Real-world Data for New Drug Applications and Line Extensions. *Clinical Therapeutics* **42**, 926–938 (2020).
34. Wang, Y. *et al.* Model-Informed Drug Development: Current US Regulatory Practice and Future Considerations. *Clinical Pharmacology & Therapeutics* **105**, 899–911 (2019).
35. Greenhalgh, T. & Papoutsi, C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Medicine* **16**, 95 (2018).
36. Tendler, S. *et al.* Treatment patterns and survival outcomes for small-cell lung cancer patients - a Swedish single center cohort study. *Acta Oncol* **59**, 388–394 (2020).
37. Green, A. K. *et al.* The Project Data Sphere Initiative: Accelerating Cancer Research by Sharing Data. *The Oncologist* **20**, 464-e20 (2015).

38. Tendler, S. *et al.* Validation of the 8th TNM classification for small-cell lung cancer in a retrospective material from Sweden. *Lung Cancer* **120**, 75–81 (2018).

39. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* **16**, 321–357 (2002).

40. Pietanza, M. C., Byers, L. A., Minna, J. D. & Rudin, C. M. Small Cell Lung Cancer: Will Recent Progress Lead to Improved Outcomes? *Clinical Cancer Research* **21**, 2244–2255 (2015).

41. Jaoude, J. A. *et al.* Performance Status Restriction in Phase III Cancer Clinical Trials. *Journal of the National Comprehensive Cancer Network : JNCCN* **18 10**, 1322–1326 (2020).

42. Pang, H. H. *et al.* Enrollment Trends and Disparity Among Patients With Lung Cancer in National Clinical Trials, 1990 to 2012. *JCO* **34**, 3992–3999 (2016).

43. Hutchins, L. F., Unger, J. M., Crowley, J. J., Coltman, C. A. & Albain, K. S. Underrepresentation of Patients 65 Years of Age or Older in Cancer-Treatment Trials. *New England Journal of Medicine* **341**, 2061–2067 (1999).

44. Hwang, J. K. *et al.* Validation of the Eighth Edition TNM Lung Cancer Staging System. *Journal of Thoracic Oncology* **15**, 649–654 (2020).

45. Marzano, L. *et al.* A novel analytical framework for risk stratification of real-world data using machine learning: A small cell lung cancer study. *Clinical and Translational Science* **00**, 1–11 (2022).

46. Tan, F. *et al.* External validation of the eighth edition of the TNM classification for lung cancer in small cell lung cancer. *Lung Cancer* **170**, 98–104 (2022).

47. Gilboa, S. *et al.* Informative censoring of surrogate end-point data in phase 3 oncology trials. *Eur J Cancer* **153**, 190–202 (2021).

48. Shebley, M. *et al.* Physiologically Based Pharmacokinetic Model Qualification and Reporting Procedures for Regulatory Submissions: A Consortium Perspective. *Clinical Pharmacology & Therapeutics* **104**, 88–110 (2018).

Tables

Table 1. SOMO analysis description for mixed cohort SCLC study

SOMO component	Parameter	Analysis Description
Selection Criteria.	ECOG Performance status	Sub-cohort stratification
	Sex	Sub-cohort stratification
	Age	Sub-cohort stratification of patients older than 75 years
	Brain Metastasis	Sub-cohort stratification
	Cancer Stage. 8th TNM version (IVA-IVB) and VALSG stage (ED-SCLC)	Sub-cohort stratification of RWD TNM staging IVA and IVB Oversampling of RWD TNM IVA to balance with IVB patients
	RCT selection criteria “probability of survive more than three months”	Propensity score matching with Y = “OS >90days" (y/n)
	RWD propensity score matching	Propensity score matching with Y= “probability of being included in the RCT study cohort" (y/n)
Operations and protocols.	MO Correction	Real-world survival outcomes are calculated from the start of the therapy, clinical trials from the randomisation. Correction made from the data, otherwise imputed as random effect informed by the available corrections
	ECOG 2 RCT pre-trial enrolment bias.	Sub-cohort stratification. RCT patients with ECOG 2 were compared with RWD ECOG 2-3
	Clinical trials censoring	Inclusion the censored records for futility of the PDS_Elill study to estimate the censoring bias

		Removing all censored patients from the analysis Propensity score matching also including the censoring as variable.
	Variability and robustness of the longitudinal profile of survival curves difference.	Creation a real-world simulated synthetic cohort with same sample size and censoring effect of the clinical trials aggregated cohort.
Measurements of Outcome	Overall Survival and Progress Free Survival	Common reported outcomes. Analysis with Kaplan Meier Curves, and Cox Hazard ratios of the entire cohort
	hazard ratio variability	Hazard ratios distribution obtained from the pairwise matching simulation of interventions on the matching (Figure S1)

Table 2. Summary of the most important results for the aggregated cohort analysis[1]

MO	Parameter analysis	Total Cohort (RCT:RWD)	Hazard Ratio (ref=RWD)	Kaplan Meier curves log-rank test
Overall Survival	Baseline MO difference	1224 (996, 228)	0.65(0.55-0.75)***	p <0.0001***
	(S) ECOG 0	386 (331, 55)	0.71(0.52,0.97)*	p=0.031*
	(S) ECOG 1	669 (564, 105)	0.75(0.6,0.94)*	p=0.012*
	(S) ECOG 2	**169 (101, 68)**	**0.73(0.53,1)**	**p=0.059**
	(S) TNM staging	**1224 (RCT:996, IVA:40, IVB:188)**	**IVA:(ref), IVB:1.9(1.36,2.7)*** RCT 1.1(0.78,1.5)**	**IVA-IVB(P=3.8e-06), RCT-IVB (P=9.9e-12), RCT-IVA (P=0,85)**
	(S) Oversampling TNM-IVA stage	**1372 (RCT:996, IVA: 188 IVB:188)**	**IVA:(ref), IVB:1.72(1.40,2.1)***, RCT 0.96(0.81,1.1)**	**IVA-IVB(P=0.0034), RCT-IVB (P=9.9e-12), RCT-IVA (P=04592)**
	(O) Propensity score including also the censoring as variable	**456 (228, 228)**	**1.1(0.87,1.3)**	**P=0.54**
Progress Free Survival	Baseline MO difference	689 (461, 228)	0.7(0.58-0.85)***	p<0.00033***
	(S) ECOG 0	**261 (206, 55)**	**0.78(0.54,1.1)**	**p=0.19**
	(S) ECOG 1	**330 (225, 105)**	**0.85(0.64,1.1)**	**p=0.25**
	(S) ECOG 2	**98 (30, 68)**	**1.2(0.74,2)**	**p=0.44**
	(S) TNM staging	**689 (RCT:461, IVA: 40, IVB:188)**	**IVA:(ref), IVB:1.7(1.21,2.5)**, RCT 1.1(0.76,1.6)**	**IVA-IVB(P=0.005)*, RCT-IVB (p=2.2e-05)***, RCT-IVA (p=0,475)**

[1] Bold text: detected similarity of outcomes, Underlined text: detected switch of survival outcome (better prognosis for real-world patients). (S): selection criteria, (O): operations, MO: measurement of outcomes, RCT: randomize controlled trials, RWD: real-world data. For the hazard ratios is reported the 95% confidence level interval. (*) p-value<0.05, (**) p<0.01, (***) p<0.001

	(S) Oversampling TNM-IVA stage	**837 (RCT:461, IVA: 188, IVB:188)**	**IVA:(ref), IVB:1.6(1.26, 1.9)**, RCT 1.0(0.81, 1.2)**	**IVA-IVB(P=5.4e-05), RCT-IVB (p=2.2e-05), RCT-IVA (p=0.89)**
	(S) Propensity score matching	**456 (228, 228)**	**0.79(0.62, 1)**	**p=0.058**
	(O) ECOG 2 pre-trial effect	**216 (ECOG 2 RCT: 101, ECOG 2 RWD: 169; ECOG 3 RWD: 47)**	**ECOG 2 RCT (ref), ECOG 2 RWD: 0.79 (0.48, 1.3) , ECOG 3 RWD: 1.26 (0.74, 2.1)**	**ECOG 2 RCT - ECOG 3 RWD: 0.532**
	(O) Removing all censored patients	433 (210, 223)	1.5(1.2, 1.8)***	p=0.0001***
	(O) Propensity score including also the censoring as variable	456 (228, 228)	1.5 (1.2, 1.8)***	p=0.00018***
	(O) Oversampling RWD simulating RCT censoring	922 (461, 461)	1.5(1.2, 1.8)***	p=5e-05***

Figures

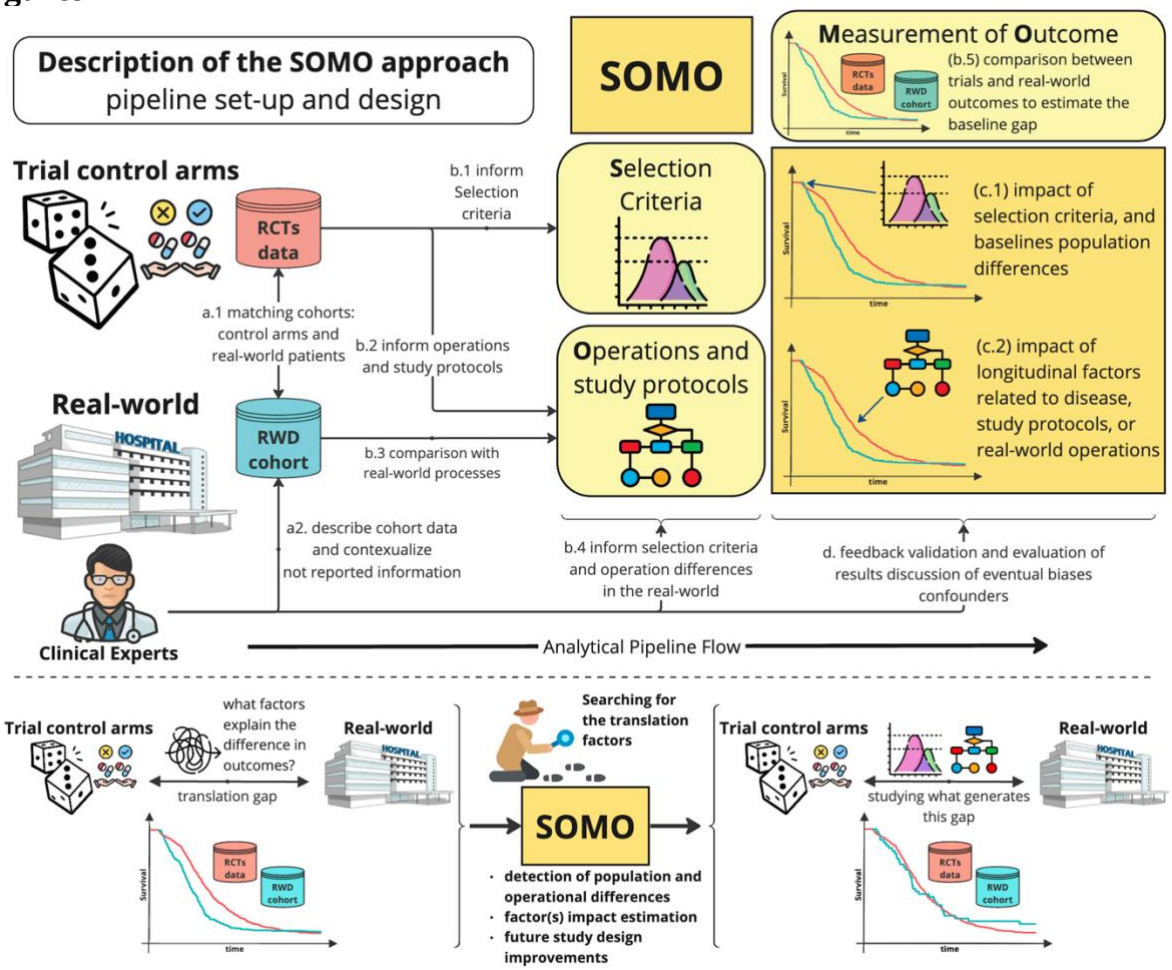

Figure 1. Summary description of the SOMO approach. Data are retrieved and pre-processes (step a), explorative analysis (step b), estimating the impact of factors on translation between cohorts (step c), validation and evaluation of the results (step d).

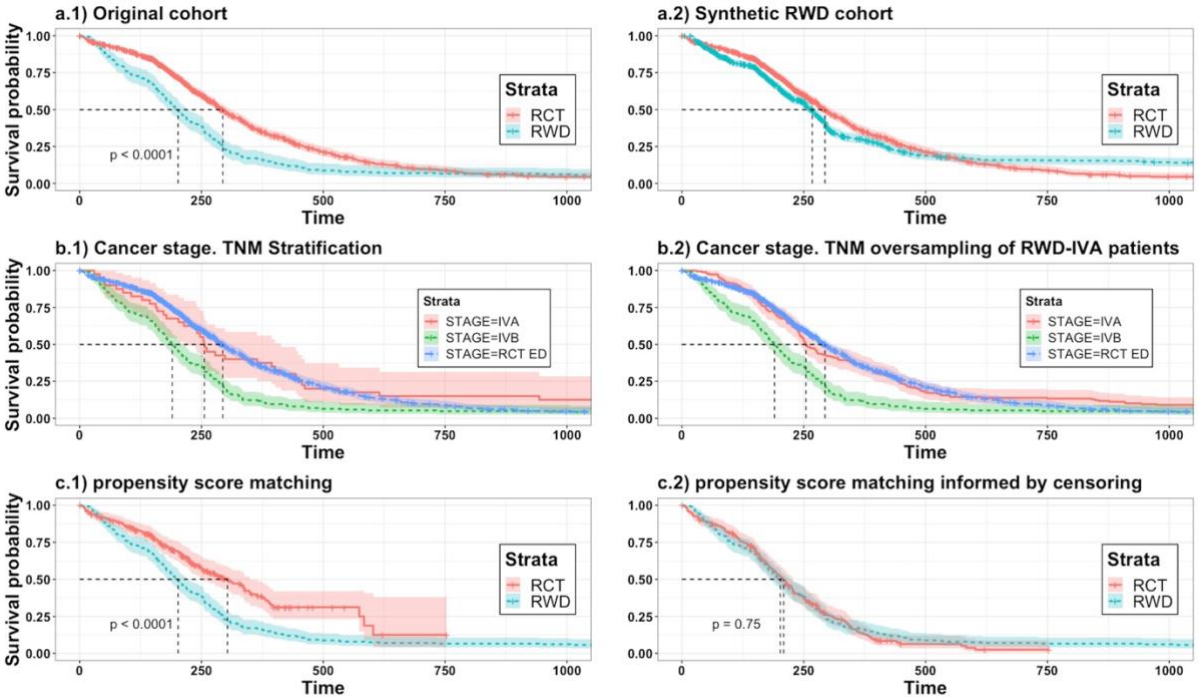

Figure 2. Overall Survival Kaplan Meier Curves of some results in Table 2. a.1) baseline survival gap (n=1224), a.2) synthetic oversampled RWD cohort with same trial censoring (n=1992), b.1) TNM stratification (n=1224), b.2) TNM oversampling (n=1372), c.1) traditional propensity score marching (n=456) c.2) propensity score accounting trials censoring (n=456). RWD: real-world data.

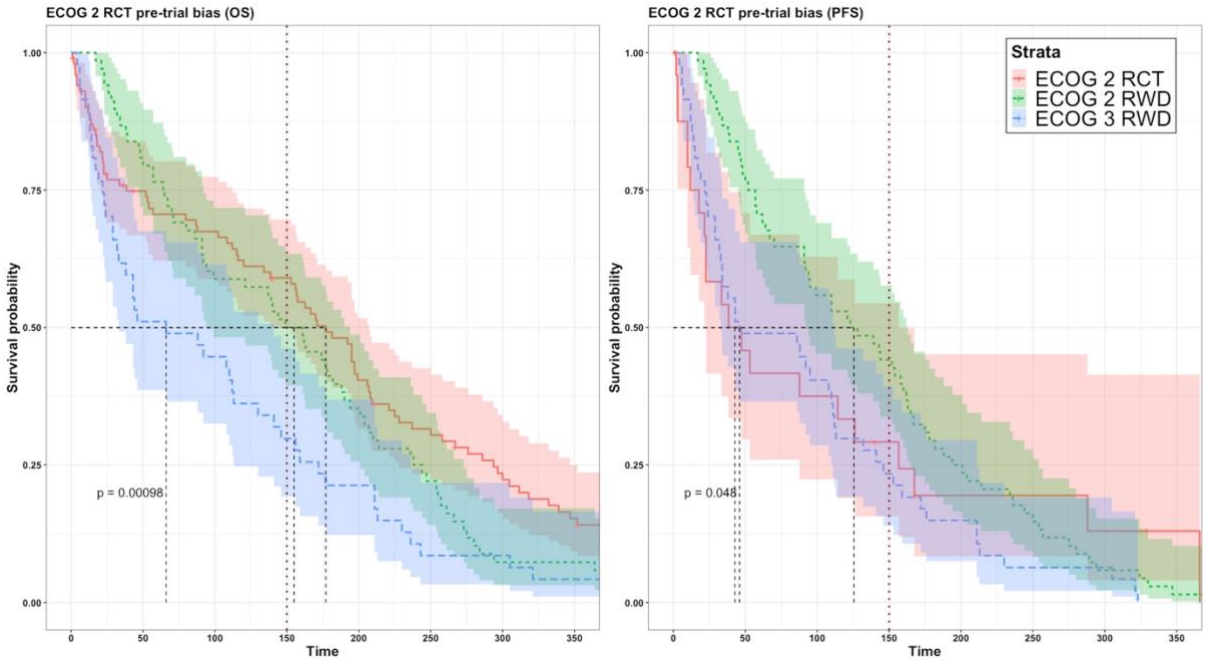

Figure 3. ECOG 2 pre-trial bias. Trial patients with ECOG 2 is compared with real-world ECOG 2-3 patients. Trial patients presents more similarities with real-world ECOG 3 patients.

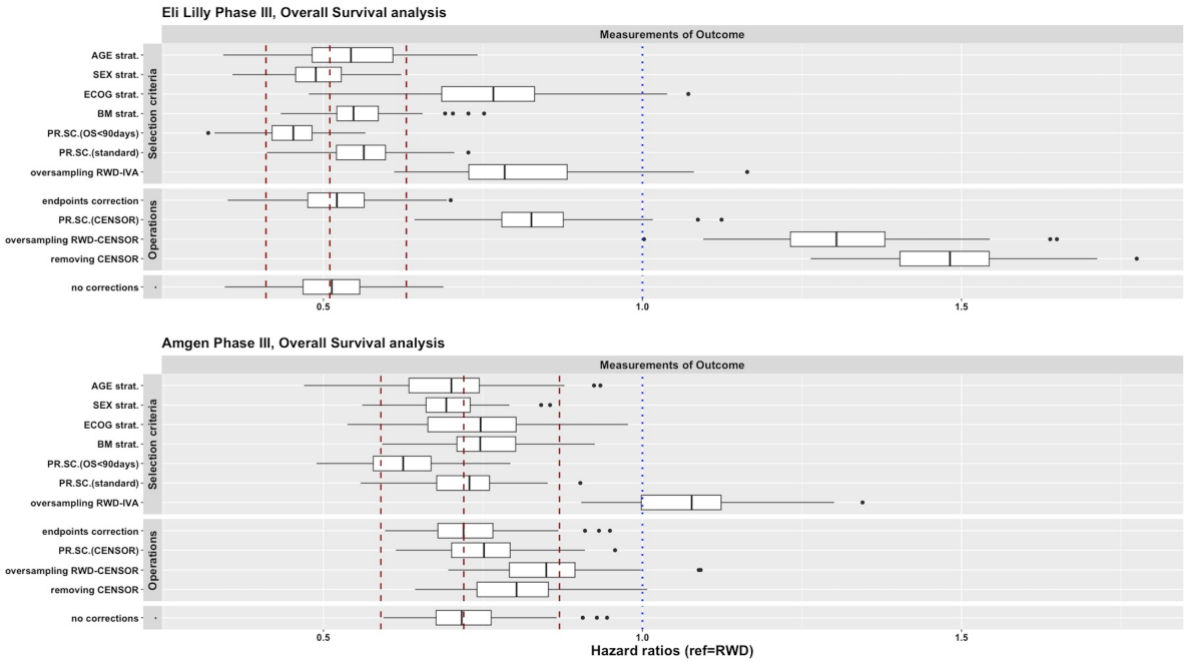

Figure 4. Matching Simulation results for overall survival hazard ratios for PDS_EliLilly and PDS_Amgen. Strat.: stratification, BM: brain metastases, PR.SC.: propensity score, RWD: real-world data. Hazard ratio and 95% confidence level interval with the whole cohort is reported in red dotted lines.

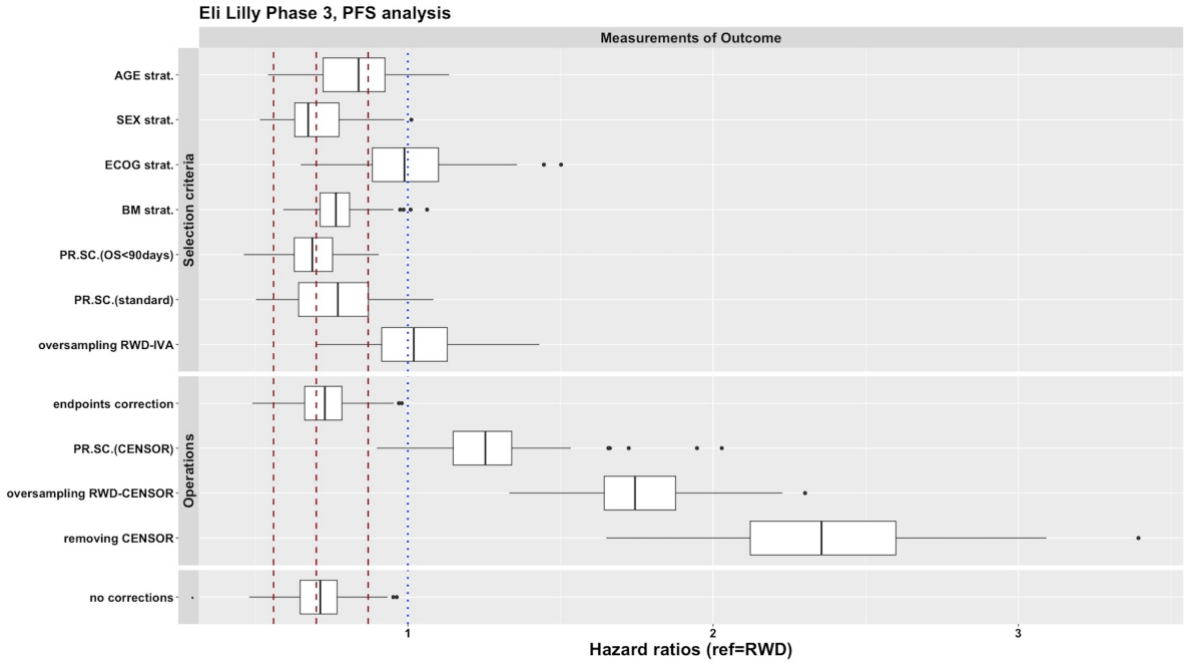

Figure 5. Matching Simulation results for progress free survival hazard ratios for PDS_EliLilly. Strat.: stratification, BM: brain metastases, PR.SC.: propensity score, RWD: real-world data. Hazard ratio and 95% confidence level interval with the whole cohort is reported in red dotted lines.

Supplementary information Titles

- Supplementary Materials for Methods section
- Supplementary Materials for Results section