Trials

METHODOLOGY

Open Access

REVERSE model: a novel approach in medical research

Luca Saba[1*], Gianluca De Rubeis[2] and Francesco Pisu[1]

Abstract

Background Randomized controlled trials are considered the gold standard but they are limited by high costs and external validity. The REVERSE model is introduced to address these challenges.

Methods The REVERSE model encompasses two sequential phases. First, in the data mining phase, compatible datasets are identified and merged by using matching or stricter inclusion/exclusion criteria, thereby reducing selection bias. Second, a randomization phase addresses the inherent biases of the selected datasets. For a dichotomous scenario, the data are organized into four sub-cohorts according to the concordance with the original and new assignments: two concordant and two non-concordant. New decision factors are tested in concordant groups. Patients in non-concordant cohorts were excluded. ROMICAT-II was used to reproduce the findings from both the ROMICAT-II and ROMICAT-I trials, with results reported as the median of 10,000 applications.

Findings The REVERSE model successfully replicated the results of ROMICAT-II and ROMICAT-I using only ROMICAT-II data. For ROMICAT-II, the median (interquartile range) of all median differences between length of hospitalization stay with cardiac computed tomography angiography (CCTA) and standard diagnostic strategy after 10,000 applications matched the trial's findings 100% of the time (18.06 h [17.76–18.32] vs. 18.1 h; $p < 0.05$). For ROMICAT-I, median of all REVERSE plaque prevalence (PP) at CCTA matched the observed PP at CCTA from ROMICAT-I (49.63% [48.2–51.2] vs. 49.7%). The REVERSE PP fell within 49.63% ± 5% in 9733 (97.33%) applications.

Conclusion The REVERSE model allows repurposing existing datasets to explore novel research questions while mitigating inherent biases through stringent inclusion criteria matching and randomization.

Introduction

Randomized controlled trials (RCT) represent the gold standard in medical research for rigorously testing hypotheses [1]. However, they suffer from various limitations, such as inadequate funding, complex regulatory requirements, excessive monitoring, privacy restrictions, and suboptimal methodology [2]. A systematic review found that the average cost per patient enrolled in an RCT ranged from 43 to 103,254 USD [3]. Furthermore, the strict inclusion criteria of RCTs can affect both the feasibility of conducting a suitable trial and external validation of results in real-world applications [1, 4].

To partially overcome these limitations, a registry-based randomized controlled trial (rRCT) methodology was proposed [5]. This approach involves embedding an RCT in a prospective registry [5, 6]. rRCTs offer several advantages over traditional RCTs, including lower costs, enhanced generalizability of findings, rapid consecutive enrollment, and the potential for comprehensive

*Correspondence:
Luca Saba
lucasabamd@gmail.com
[1] Department of Medical Imaging, Azienda Ospedaliero Universitaria (A.O.U.) of Cagliari-Polo Di Monserrato,, University of Cagliari, Cagliari, Italy
[2] Department of Diagnostic UOC of Diagnostic and Interventional Neuroradiology, San Camillo-Forlanini Hospital, Rome, Italy

follow-up [1]. However, they also present challenges related to registry data quality, ethical considerations, requirement for hard endpoints (e.g., overall survival), and methodological challenges [5, 7]. Furthermore, Lasch et al. [8] identified patient selection as another limitation. Although rRCTs have inherent more external validity than RCT owing to their pragmatic nature, selection bias is still present [2]. This occurs because RCTs are embedded in a registry with inherent selection biases (registry selection bias level) and there is systematic selection bias during the randomization process caused by inclusion/exclusion criteria of rRCT (RCT selection bias level) [2]. However, all measures for enhancing internal validity in RCT may be implemented in rRCT [9].

RCTs present a unique challenge in unlocking the full potential of collected data. The randomization process prior to the trial precludes investigators from exploring additional hypotheses, thereby leaving a wealth of information untapped. At present, there is no proven method for overcoming this challenge. Furthermore, retrospective analyses are frequently dismissed for their perceived lack of value, despite offering an abundance of data that, if properly analyzed, could unlock important insights. The opportunity for high-level evidence should not be overlooked.

We present the REVERSE model, a novel methodology that addresses the limitations of RCTs, rRCTs, and registries by allowing the repurposing of existing datasets through stringent inclusion criteria matching and randomization to explore novel research hypotheses that extend beyond the original purposes for which the data were collected.

Methods

Methodological proposal

A novel methodology called the "REVERSE model" is presented. This approach allows to repurpose existing data to explore novel research questions and optionally merge multiple compatible datasets by inclusion/exclusion criteria matching. By using this methodology, it is possible to conduct a new RCT by using existing data without the need for new collection.

The REVERSE model comprises two sequential phases: a data mining phase and a post hoc randomization phase (Fig. 1).

The data mining phase begins with identifying appropriate datasets, which may include a broad spectrum of study designs such as RCTs, retrospective observational studies, prospective observational studies, and big data. After one or more datasets have been identified, two key assumptions must be satisfied: (a) the presence of two or more distinct treatments or diagnostic strategies to evaluate and (b) the inclusion/exclusion criteria of the REVERSE should either match or be more stringent than those of the selected dataset(s).

In other words, two or more datasets can be merged and used in the REVERSE model if they can be matched by treatments or diagnostic strategies and all REVERSE inclusion/exclusion criteria can be evaluated and satisfied for all datasets. For instance, if participants aged 18 and older should be included and one dataset does not have the age feature, it cannot be included in the REVERSE model.

The randomization is applied after the data mining phase to address inherent limitations in the selected datasets. Random assignment of participants helps mitigate accidental bias in observational studies [10] and enables the reuse of existing RCT data to explore new research hypotheses.

Following randomization, the study participants are grouped into several sub-cohorts according to the original and new assignments. In case of two distinct treatments/diagnostic strategies, the following four sub-cohorts will be identified (Fig. 1):

- Sub-cohort 1: patients treated with treatment/method A and randomized to treatment/method A (*concordant*)
- Sub-cohort 2: patients treated with treatment/method B but randomized to treatment/method A (*non-concordant*)
- Sub-cohort 3: patients treated with treatment/method A but randomized to treatment/method B (*non-concordant*)
- Sub-cohort 4: patients treated with treatment/method B and randomized to treatment/method B (*concordant*)

Two distinct groups can be identified, grouped into two sub-cohorts each: one concordant with the randomization process (sub-cohorts 1 and 4) and another displaying non-concordance (sub-cohorts 2 and 3). To maintain the integrity of our analysis, we will exclude cohorts 2 and 3, as they do not meet our criteria for true randomization. This leaves us with cohorts 1 and 4, which accurately represent the results of a genuine randomization process (Fig. 1). By implementing the REVERSE approach, the population size is approximately halved by excluding the non-concordant sub-cohorts.

Moreover, the randomization process can be repeated multiple times, each iteration generating a slightly different population, which allows for the estimation of results variability and stability.

Compared with other reuse strategies—such as (i) a one-off concordant analysis, where the concordant cohort

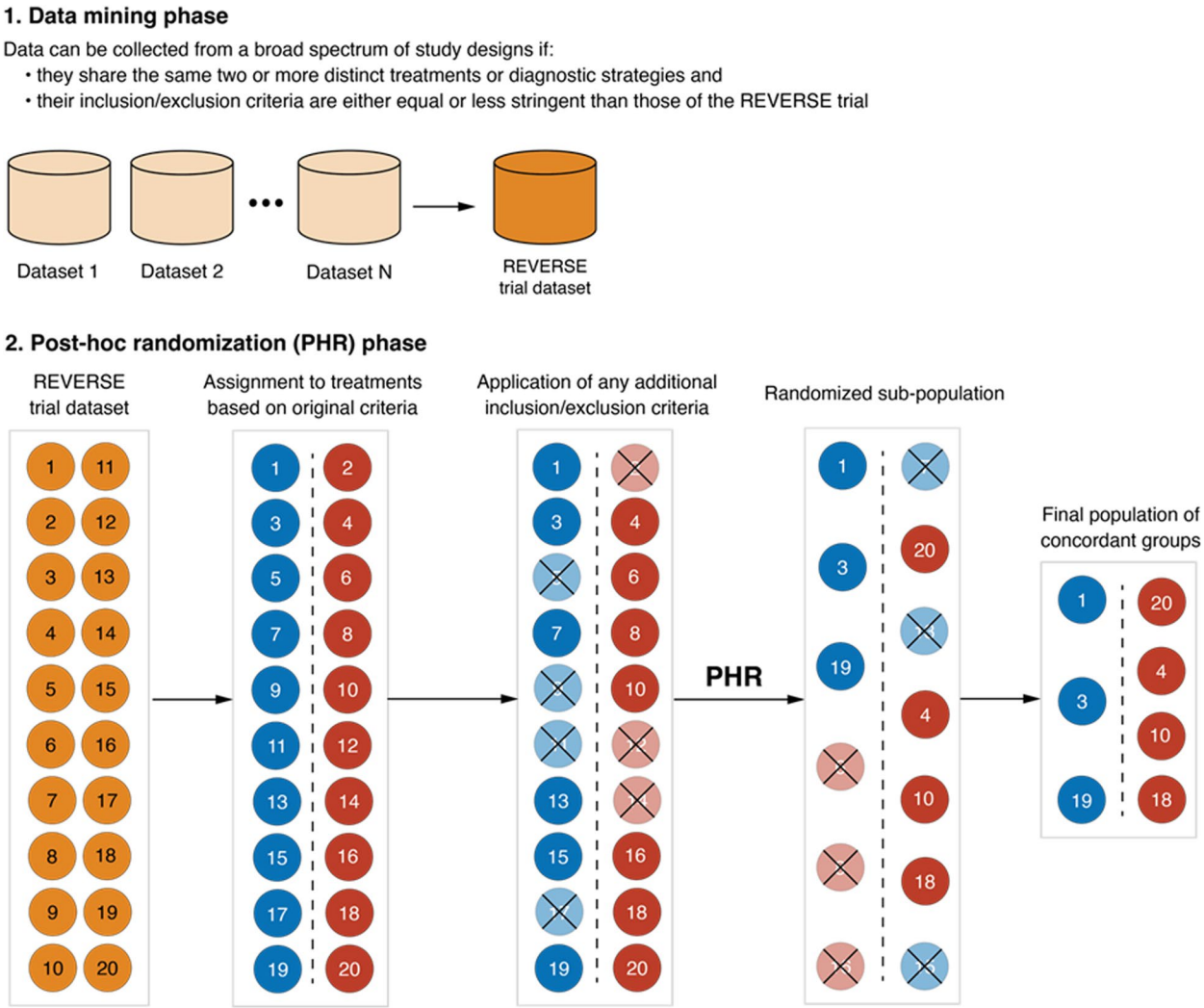

Fig. 1 The REVERSE model

is examined only once and any chance imbalance persists, and (ii) propensity-score matching, which discards unmatched patients and inflates variance in proportion to the sample-size loss—the REVERSE framework retains the randomization-based unbiasedness of an RCT even after strict eligibility filtering. By averaging independent post-filter re-randomizations (K=10,000 in this study), it reduces sampling variance to σ^2/K, thereby achieving lower bias and tighter confidence intervals (additional details in the Supplementary Material).

Case example—its application and robustness

To demonstrate the feasibility of the REVERSE model, we conducted a test using data from a previous trial published in the *New England Journal of Medicine*. We obtained access to the data through authorized channels from the National Heart, Lung, and Blood Institute (NHLBI): the ROMICAT-II trial [11] (BioLINCC RMDA V02 1d20120806).

The ROMICAT-II trial [11] was an RCT that demonstrated the benefit of including cardiac computed tomography angiography (CCTA) in the triage strategy for patients with symptoms suggestive of acute coronary syndromes. This inclusion of CCTA resulted in a reduction in the length of stay (LOS) compared to the standard diagnostic strategy (SDS).

In our study, we conducted two distinct experiments. The first experiment involved applying the REVERSE methodology to reproduce the results of the ROMICAT-II trial [11] utilizing the original data. Specifically, we calculated the number of times, out of the 10,000 randomizations, the mean and median LOS

of the concordant CCTA sub-cohort matched the mean (± 1 h tolerance) and median LOS as observed in the ROMICAT-II.

The second experiment aimed to reproduce the results of the ROMICAT-I trial [12], another RCT assessing the utility of CCTA in patients with acute chest pain. The trial's primary endpoint was different from ROMICAT-II and it was to determine the prevalence of coronary artery disease (CAD) using CCTA and its diagnostic performance in comparison to standard evaluation methods. We utilized the original data from ROMICAT-II to emulate the findings of ROMICAT-I. Our objective was to determine, across 10,000 randomizations, the frequency of plaques (± 2.5%, ± 5%, and ± 7.5% tolerances) among the concordant CCTA sub-cohort matching the observed frequency of "any plaque" at CCTA from ROMICAT-I.

Statistical analysis

To evaluate the REVERSE model methodology, we conducted 10,000 randomizations by randomly shuffling the subjects and assigning the first half to CCTA and the second half to SDS. We chose 10,000 randomizations because, at that size, the maximum Monte Carlo standard error of any permutation p value is only 0.005 (0.5%), which occurs in the worst-case scenario when $p = 0.5$. For each simulation, we identified the concordant and non-concordant groups by comparing them with the original diagnostic strategy assignments. For each concordant sub-cohort (i.e., CCTA and SDS), we computed both the mean and median LOS. To assess the statistical significance of the differences between these sub-cohorts, we employed either the independent two-sample t-test or the Wilcoxon rank sum test, as appropriate. Cohen's d was chosen to measure the effect size for the two-sample t-test. This metric quantifies the standardized difference between group means, normalized by the pooled standard deviation. It ranges from 0 to infinity, reflecting the degree of mean difference in terms of standard deviations (SD). The noncentrality parameter method was used to estimate the 95% confidence intervals for Cohen's d. For non-parametric differences (rank sum), we utilized Cliff's delta (or rank-biserial correlation for the two-sample scenario) as the effect size. This correlation measure gauges the discrepancy between the proportions of favorable and unfavorable pairs (or signed ranks). Values range from − 1 to + 1, where − 1 indicates complete dominance of the second sample (larger LOS with SDS) and + 1 indicates complete dominance of the first sample (larger LOS with CT). Confidence intervals for Cliff's delta were estimated through the normal approximation using Fisher's transformation. Furthermore, we calculated the frequency of plaques among subjects of the concordant CCTA sub-cohort only.

The code used to perform randomization simulations is included as supplementary data for this study (Section S1 in the Supplementary Material). All analyses were performed using Python software (version 3.9.6, Python Software Foundation) and R software (version 4.3.1, R Core Team). For a comprehensive list of packages utilized in the analyses, refer to Section S2 in the Supplemental Material.

Results

Results from ROMICAT-II

Following 10,000 randomizations, the preserved concordant population exhibited a range of 437 to 557 patients, with a mean (SD) of 499 ± 15.9 (Fig. 2). The mean count ± SD of concordant patients allocated to CCTA was 250.5 ± 8.0, while for SDS it was 249.5 ± 8.0. Additionally, the average number of non-concordant patients assigned to CCTA was 249.5 ± 8.0, and for SDS, it was 250.5 ± 8.0.

The median (interquartile range [IRQ]) of all the median LOS for CCTA and SDS were 8.58 h (8.5–8.77) and 26.7 h (26.42–26.85), respectively. For comparison, the median LOS observed within the ROMICAT-II trial were 8.6 h (6.4–27.6) and 26.7 h (21.4–30.6).

Across all REVERSE model simulations, the median (IQR) of all the differences between median (IQR) LOS with SDS and CT was 18.06 h (17.76–18.32). Most importantly, a consistent pattern emerged: CCTA consistently led to a notably shorter LOS when compared to SDS. Specifically, when evaluating the median LOS, the CCTA subgroup consistently demonstrated a significantly reduced duration in contrast to the SDS subgroup ($p < 0.05$) across all randomized scenarios (10,000/10,000, 100%, refer to Fig. 3). Furthermore, we observed that the REVERSE grand median (IQR) aligns closely with the difference between median LOS reported by the ROMICAT-II trial (18.06 [17.76–18.32] vs. 18.1 h) as illustrated in Fig. 4.

An intriguing aspect to consider stems from the post hoc randomization process, where roughly half of the original population is discarded. Consequently, the resultant reduction in the sample size for statistical calculations could potentially magnify the susceptibility to the influence of outliers (see Fig. S1 in the Supplementary Appendix for the observed distributions of mean and median LOS with respect to both CCTA and SDS after 10,000 randomizations). This, in turn, underscores the rationale behind our deliberate choice of employing the median as a robust measure. Further results involving the mean as the chosen statistic are detailed in Section S3 in the Supplementary Appendix.

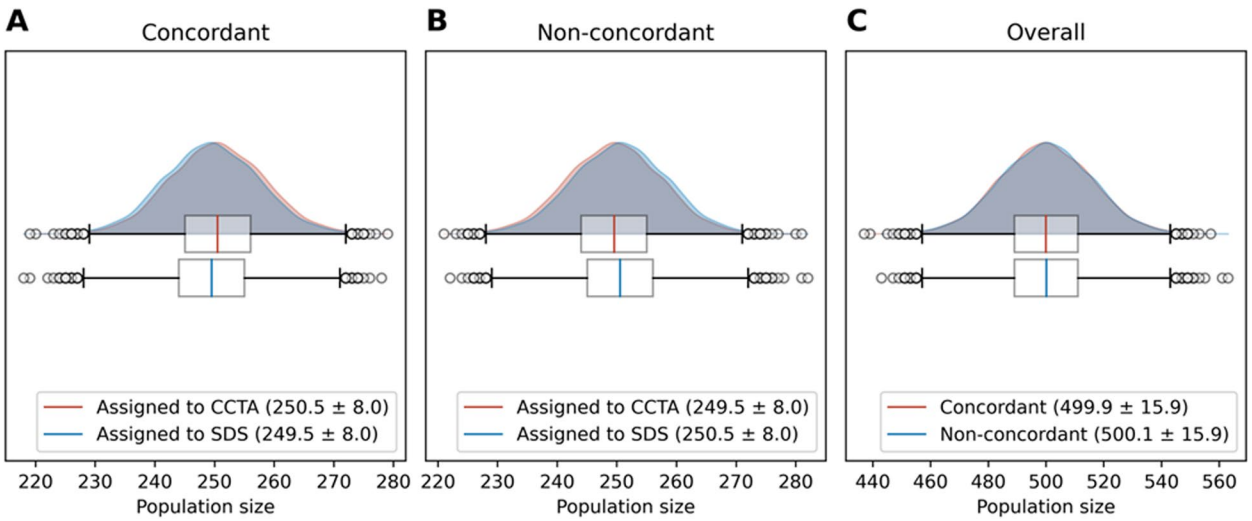

Fig. 2 Mean count ± SD

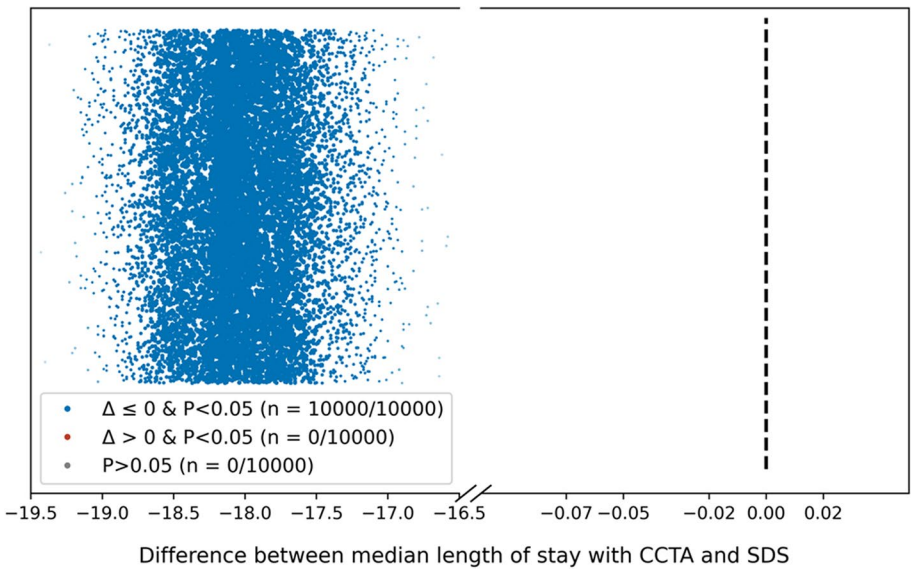

Fig. 3 Median LOS

The comprehensive results originating from the 10,000 randomizations along with statistical tests results are fully detailed in Sections S4 and S5 in the Supplementary Appendix.

Results from ROMICAT-I

Following 10,000 randomizations, the median (IQR) of all REVERSE plaque prevalence at CCTA closely corresponds to the observed PP at CCTA in the ROMICAT-I trial (49.63% [48.2–51.2] vs. 49.7%) (Fig. 5). Similar results were observed for the mean (SD) (49.69% ± 2.2) (Fig. S4 in the Supplementary Appendix). Notably, when examining the range of PP at CCTA ± 2.5% of ROMICAT-I, it is observed that 7363 REVERSE PP values out of 10,000 (73.63%) fall within this interval, with 9733 (97.33%) fitting within ± 5% and an even greater 9991 (99.91%) within ± 7.5%, thereby achieving comprehensive coverage within a tolerance of ± 10%.

Discussion

The REVERSE model offers a framework for repurposing data from multiple studies to explore new hypotheses, while mitigating inherent biases through sample size expansion, rigorous inclusion criteria matching, and

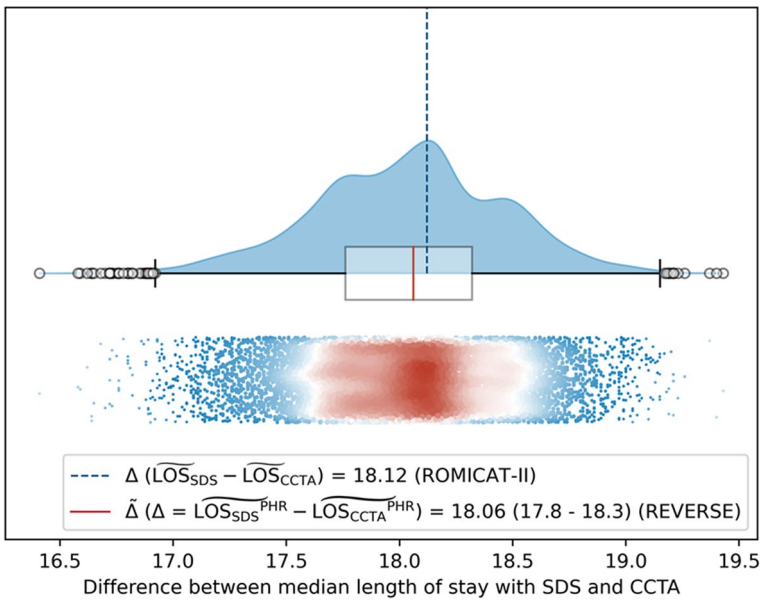

Fig. 4 Median (IQR)

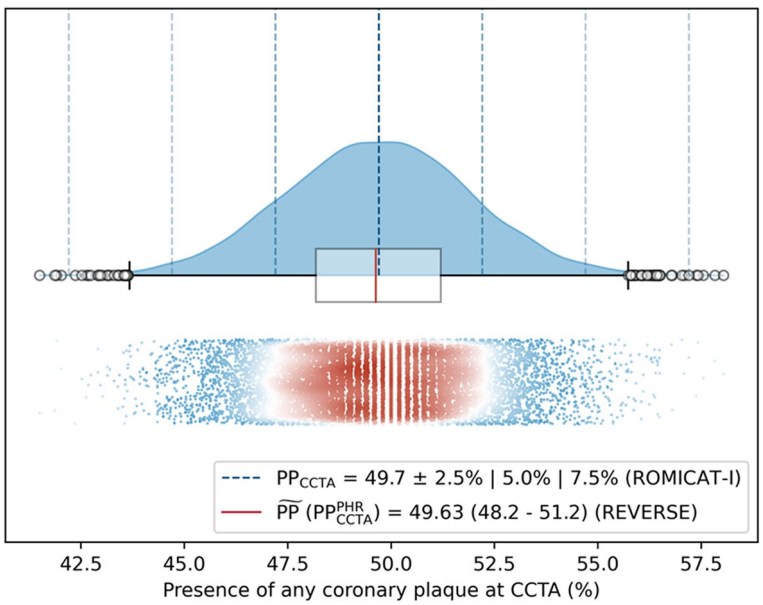

Fig. 5 Median (IQR) of all REVERSE plaque prevalence

randomization. There is a pressing need for an analytical model that can reuse data from previous studies for different purposes. This requirement is driven not only by cost considerations but also by the need to explore new research questions that were not considered in the original studies.

The effectiveness of the REVERSE model was assessed by reproducing the findings of two renown RCTs, namely the ROMICAT-I and ROMICAT-II, by using data from ROMICAT-II only. By applying the REVERSE model on the original ROMICAT-II trial data, we observed a median of all median differences in LOS with CCTA and SDS consistently matching the trial's median difference in LOS in all 10,000 iterations. We chose 10,000 randomizations because that sample size places the permutation distribution firmly in the domain where the central-limit theorem applies, so the resulting p values have minimal Monte Carlo error and provide a robust

test of the trial's findings. However, the reproduction of the results of the ROMICAT-I trial, which shares the same inclusion/exclusion criteria, was also consistently reproduced using data from ROMICAT-II, with the PP at CCTA across within ± 5% the PP at CCTA observed in the ROMICAT-I trial in 9733 REVERSE applications (97.33%). This second finding could represent a breakthrough in research methodology. In fact, we demonstrated that by using existing data from ROMICAT-II, it is possible to obtain the results of another trial with similar levels of scientific rigorousness. To validate the REVERSE framework, we selected the ROMICAT trials due to their near-identical inclusion criteria and endpoints, providing a rare opportunity to compare REVERSE-derived outcomes against a known RCT benchmark. This alignment was intentional: it enabled us to isolate the model's internal validity under controlled conditions. While broader applications across heterogeneous datasets are a natural next step, no existing trial infrastructure currently permits a comparable gold-standard validation. As such, the ROMICAT setting serves as a necessary and foundational proof-of-concept for this novel methodology. This means that it is possible to emulate multiple RCTs using a single well-curated dataset. In other words, with high-quality structured data and strict eligibility criteria, REVERSE enables the design of several RCT-like analyses without requiring new patient enrollment, each targeting distinct research questions.

Selection bias in healthcare studies is complex and enduring due to its nature [13, 14]. While randomization plays a key role in reducing selection bias, initial imbalances in the study population may prevent the findings from being widely applicable [15]. In simpler terms, if the population undergoing randomization starts off unbalanced, selection bias persists, thereby limiting the generalization of the results [15]. As highlighted by Lasch et al. [8], selection bias in rRCTs takes on a dual form: firstly, the inclusion and exclusion criteria predetermined for the registry, which are fixed and established "a priori," irrespective of subsequent RCT designs (registry selection bias level). Secondly, it stems from the inclusion and exclusion criteria governing the RCTs themselves (trial selection bias level). Because registry criteria remain unmodifiable, addressing selection bias at this level becomes a challenge. The REVERSE model offers potential solutions to mitigate selection bias by allowing the modification of inclusion/exclusion criteria in relation to the chosen dataset and by mixing more datasets, effectively mitigating the inherent "a priori" selection bias (referred to as registry selection bias level). In instances involving retrospectively generated datasets (retrospective study), the establishment of inclusion and exclusion criteria was guided by the randomization. Conversely, within the context of prospectively developed biobanks (RCTs or prospective studies), the criteria for randomization can either align with or be more stringent than those of the selected biobank. Both these approaches are crucial in effectively addressing trial selection bias.

The effectiveness of the REVERSE model lies in making the original characteristics of the data less important by mitigating inherent biases and establishing a new population for investigating novel research hypotheses. This is enabled by merging data from multiple datasets, both retrospective and prospective across diverse data types such as registries, RCTs, biobanks, and large-scale datasets. Once inclusion and exclusion criteria are defined, the subsequent step involves identifying the population meeting these criteria and proceeding with randomization. In the current medical landscape, expansive datasets and biobanks hold great potential, yet their value is often constrained due to diverse origins and hypotheses. The REVERSE model provides a tool for maximizing the utility of existing data by allowing the exploration of research questions that differ from the original scope of the selected studies, while minimizing the risks associated with small sample sizes. Nonetheless, the REVERSE model's results depend on meticulously collected high-quality data, similar to how the strength of RCTs lies in their recruitment quality and careful inclusion/exclusion criteria. Therefore, when merging datasets or dealing with rRCTs or retrospective data, precise attention to inclusion and exclusion criteria becomes crucial.

The REVERSE approach inherently leads to a reduction in cohort size. For instance, after 10,000 randomizations of the ROMICAT-II dataset, the resulting populations ranged from 467 to 576, a decrease from the original 1000 patients. This raises questions about the implications of such reductions for hypothesis testing. Power analysis can be used to determine if the reduced sample size is sufficient for the new hypotheses, which may require fewer subjects than the original study.

Furthermore, once inclusion/exclusion criteria are established and the initial REVERSE population is identified, the capability to apply randomization multiple times enables the derivation of the "REVERSE index." This index acts as an indicator of result robustness; consistent outcomes across multiple randomizations suggest a significant effect size. Thus, it is advisable to apply the randomization multiple times to rigorously evaluate the strength and repeatability of the results.

The repurposing of existing data for new research objectives—as enabled by the REVERSE model—raises important ethical and regulatory considerations. In

future applications, datasets intended for merging or secondary use should be anonymized or pseudonymized in accordance with established frameworks such as the GDPR and the Declaration of Helsinki, which permit secondary research use under appropriate safeguards, including data minimization, secure handling, and ethics oversight. Oversight by an ethics committee or institutional review board remains essential, even when data are de-identified. To support ethical reuse, we recommend that future trials include broad consent or future-use clauses in their informed consent processes.

The REVERSE model has certain limitations. First, the data quality must be rigorously assessed and validated through appropriate methods [16] to ensure its integrity, which is crucial for the model's effective implementation. Second, the study design may be constrained by the extent of information documented in patients' medical records or available through data mining. Third, the model's structure inherently leads to a roughly twofold reduction in sample size when non-concordant cases resulting from the randomization are excluded. While this may be a significant limitation for RCTs, its impact is comparatively less pronounced on registries or large-scale datasets due to their inherently larger sample sizes. At this stage, we have validated the REVERSE model on a single dataset and have not yet explored its full potential for combining multiple databases. Our immediate goal was to establish proof-of-concept, after which we intend to extend the framework to blended datasets in future work.

In conclusion, the REVERSE model effectively leverages existing data to address new research questions, demonstrating its robustness by replicating findings from major RCTs and offering a practical solution to minimize selection biases through adaptable inclusion criteria and iterative randomization, although it requires careful data quality control and management of inherent reductions in sample size.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13063-025-08974-9.

Supplementary Material 1.

Authors' contributions

LS: conceptualization, data curation, formal analysis, investigation, methodology, supervision, validation, visualization, and writing review and editing. GDR: validation, visualization, and writing—original draft. FP: supervision, validation, visualization, and writing—review and editing.

Funding

None.

Data availability

Not applicable.

Declarations

Ethics approval and consent to participate

We obtained access to the data through authorized channels from the National Heart, Lung, and Blood Institute (NHLBI), specifically from the ROMICAT-II trial (BioLINCC RMDA V02 1d20120806). The original trial received ethics approval from the relevant institutional review boards, and all participants provided written informed consent. Our secondary analysis was approved under the data use agreement established with NHLBI.

Consent for publication

Not applicable. This study is a secondary analysis of anonymized data from a publicly available dataset (ROMICAT-II) and does not include any individual-level data or images requiring consent for publication.

Competing interests

LS: none. GDR: none. FP: none.

Transparency declaration

The author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Received: 6 November 2024 Accepted: 11 July 2025
Published online: 18 July 2025

References

1. Sibbald B, Roland M. Understanding controlled trials: why are randomised controlled trials important? BMJ. 1998;316(7126):201. https://doi.org/10.1136/bmj.316.7126.201.
2. Lauer MS, D'Agostino RB Sr. The randomized registry trial–the next disruptive technology in clinical research? N Engl J Med. 2013;369(17):1579–81. https://doi.org/10.1056/NEJMp1310102.
3. Speich B, von Niederhäusern B, Schur N, Hemkens LG, Fürst T, Bhatnagar N, et al. Systematic review on costs and resource use of randomized clinical trials shows a lack of transparent and comprehensive data. J Clin Epidemiol. 2018;96:1–11. https://doi.org/10.1016/j.jclinepi.2017.12.018.
4. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015;16(1):495. https://doi.org/10.1186/s13063-015-1023-4.
5. Li G, Sajobi TT, Menon BK, Korngut L, Lowerison M, James M, et al. Registry-based randomized controlled trials - what are the advantages, challenges, and areas for future research? J Clin Epidemiol. 2016;80:16–24. https://doi.org/10.1016/j.jclinepi.2016.08.003.
6. Mikita JS, Mitchel J, Gatto NM, Laschinger J, Tcheng JE, Zeitler EP, et al. Determining the suitability of registries for embedding clinical trials in the United States: a project of the Clinical Trials Transformation Initiative. Therapeutic innovation & regulatory science. 2021;55(1):6–18. https://doi.org/10.1007/s43441-020-00185-5.
7. Foroughi S, Wong H-I, Gately L, Lee M, Simons K, Tie J et al. Re-inventing the randomized controlled trial in medical oncology: the registry-based trial. Asia-Pacific Journal of Clinical Oncology. 2018;14(6):365–73. https://doi.org/10.1111/ajco.12992.
8. Lasch F, Weber K, Koch A. Commentary: on the levels of patient selection in registry-based randomized controlled trials. Trials. 2019;20(1):100. https://doi.org/10.1186/s13063-019-3214-x.
9. Mathes T, Buehn S, Prengel P, Pieper D. Registry-based randomized controlled trials merged the strength of randomized controlled trails and observational studies and give rise to more pragmatic trials. J Clin Epidemiol. 2018;93:120–7. https://doi.org/10.1016/j.jclinepi.2017.09.017.
10. Altman DG. Randomisation. BMJ. BMJ. 1991;302(6791):1481.
11. Hoffmann U, Truong QA, Schoenfeld DA, Chou ET, Woodard PK, Nagurney JT, et al. Coronary CT angiography versus standard evaluation in acute chest pain. N Engl J Med. 2012;367(4):299–308. https://doi.org/10.1056/NEJMoa1201161.

12. Hoffmann U, Bamberg F, Chae CU, Nichols JH, Rogers IS, Seneviratne SK, et al. Coronary computed tomography angiography for early triage of patients with acute chest pain: the ROMICAT (Rule Out Myocardial Infarction using Computer Assisted Tomography) trial. J Am Coll Cardiol. 2009;53(18):1642–50. https://doi.org/10.1016/j.jacc.2009.01.052.
13. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. The Cochrane database of systematic reviews. 2007(2):MR000012. https://doi.org/10.1002/14651858.MR000012.pub2.
14. Hallberg K, Cook TD, Steiner PM, Clark MH. Pretest measures of the study outcome and the elimination of selection bias: evidence from three within study comparisons. Prev Sci. 2018;19(3):274–83. https://doi.org/10.1007/s11121-016-0732-6.
15. Fives A, Rusell DW, Kearns N, Lyons R, Eaton P, Canavan J, et al. The role of random allocation in randomized controlled trials: distinguishing selection bias from baseline imbalance. Journal of MultiDisciplinary Evaluation. 2013;9(20):33.
16. Houston L, Probst Y, Martin A. Assessing data quality and the variability of source data verification auditing methods in clinical research settings. J Biomed Inform. 2018;83:25–32. https://doi.org/10.1016/j.jbi.2018.05.010.

Publisher's Note