SoftwareX 31 (2025) 102214

Contents lists available at ScienceDirect

SoftwareX

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

Original Software Publication

# Educational platform trials simulator (EPTS): Software for planning and simulating cluster-randomised, multisite and simple randomised platform trials

Mohammad Sayari [a,*], Akansha Singh [b], Germaine Uwimpuhwe [a], Nasima Akhter [c], Tahani Coolen-Maturi [a], Rashmika Gupta [d], Jochen Einbeck [a]

[a] Department of Mathematical Sciences, Durham University, Durham, UK
[b] Department of Anthropology, Durham University, Durham, UK
[c] School of Health and Life Sciences, Teesside University, Middlesbrough, UK
[d] Business School, Durham University, UK

ARTICLE INFO

*Keywords:*
Platform trials
Simulation
Cluster randomised trial
Multisite trial

ABSTRACT

Platform trials offer a robust framework for evaluating multiple interventions simultaneously against a common control group within a single master protocol. Acknowledging their increased importance in the medical and social sciences, including education, a specialised software 'Educational Platform Trials Simulator (EPTS) is designed. The software's user-friendly interface allows users to plan and execute simulations for randomised controlled trial designs, analyse data (Bayesian/ Frequentist methods, multilevel models, futility and superiority analyses), and visualise results without requiring extensive programming expertise. This practical, economical and academic tool facilitates efficient and effective research involving planning and simulating platform trials in education and other disciplines.

## 1. Motivation and significance

Platform trials are an innovative trial design where multiple interventions can be evaluated simultaneously against a common control group within a single master protocol [1]. They are primarily used in medical sciences; however, in recent years, they have gained popularity in social sciences as well. Platform trials can offer unique advantages over traditional trial designs (e.g., two-arm trials) by investigating multiple interventions simultaneously against a common control group using specialised statistical tools for allocating participants and analysing results comparing intervention effects on the specific outcome [2]. A distinguishing feature of platform trials is that they focus on a specific outcome (or disease in a clinical context), rather than on an intervention (or treatment). This enables the comparison of multiple interventions by allowing new arms to be added and ineffective ones to be discontinued, as well as the possibility to update the control arm to reflect a new standard of care, at the same time ensuring continuity of the overall study [3]. Outcomes can vary significantly depending on the context, such as disease outcomes in clinical settings or educational outcomes in educational settings. Statistical criteria used to advance or drop intervention arms to the next phase are known as decision rules [4]. Selecting appropriate decision rules is crucial in platform trials to reduce the risk of bias and inefficiency in decision-making during interim evaluations. These rules may rely on frequentist or Bayesian statistical metrics [5,6]. Frequentist metrics often include test statistics associated with P-values and conditional power, while Bayesian metrics commonly involve posterior probabilities of futility and superiority [7,8]. Platform trials are not limited to a single statistical framework; they can employ both Bayesian and frequentist approaches [1]. Platform trials can discover beneficial interventions with fewer participants, fewer failures, less time, and with greater probability of success than a traditional two-arm trial [9].

In the education context, platform trials can help comparing multiple educational interventions against the same control, which is not possible in any traditional two-arm trial designs. Since all the intervention arms are compared against the same control group, platform trials have the potential to examine which intervention is most effective in improving a specific outcome. However, it is also necessary to take control measurements at the start and the end of each intervention (as pre- and post-intervention outcomes). Platform trials allow the flexibility to consider

* Corresponding author.
*E-mail address:* md.sayari13@gmail.com (M. Sayari).

https://doi.org/10.1016/j.softx.2025.102214
Received 28 February 2025; Received in revised form 15 May 2025; Accepted 21 May 2025
Available online 30 May 2025
2352-7110/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

one single control group for two or more interventions conducted in the same period that targeted different outcomes. For example, comparisons can be made for one intervention that targeted maths while the other targeted literacy, given that all other baseline characteristics are similar and the results of these outcomes can be made available (e.g., via National Pupil Database scores). A platform trial design can also provide flexibility to add new literacy or maths interventions later in the trial and evaluate multiple literacy or maths interventions at the same time.

There is a notable lack of literature addressing platform trials in the field of education. The high level of flexibility provided by platform trials can make it difficult to obtain meaningful results in educational studies. This is primarily because education trials are typically conducted over a single academic year, where excessive adaptivity can have implications on student learning and may prove counterproductive. A major challenge is the complexity of educational interventions and their impact on student learning outcomes [10]. Although these trials have proven effective in certain domains, their use in educational research is constrained by the unique complexities and ethical considerations intrinsic to this field [11].

In recent years, there has been increased interest in software development for multi-arm and platform randomised controlled trials in clinical settings [12]. Several tools have been developed to support the design and simulation of such trials, including SIMPLE, Octopus, NCC, HECT, MAMS and the commercial software FACTS. SIMPLE is a modular simulation framework designed to improve code shareability and reusability in platform trial projects. It uses independent modules to handle aspects like recruitment, analysis strategies, and intervention management, enabling flexible and complex trial designs. The main simulation wrapper is general and makes minimal, clearly defined assumptions, allowing for varied strategies across interventions. SIMPLE supports both novice users, through accessible design, and advanced users, who can deeply customise simulations [13]. OCTOPUS is an R package designed to help drug developers simulate platform trial designs with customisable options for design, endpoints, and operational scenarios. While offering flexibility and deep customisation, effective use of OCTOPUS requires a solid understanding of its structure and codebase [14]. The NCC R package enables simulation of platform trials using non-concurrent control data, with support for continuous or binary endpoints and varying numbers of treatment arms entering at different times. It accounts for differing treatment effects and time trends, offering both frequentist (e.g., fixed/random time effects, regression splines) and Bayesian methods (e.g., time machine, meta-analytic predictive priors). The package allows flexible and complex trial design simulations with multiple analytic options [15]. The HECT is an R Shiny app for simulating platform adaptive trials with up to 10 treatment arms, supporting both binary and continuous endpoints. It allows for comparisons between treatments or against a reference, with arms being dropped or graduated based on Bayesian posterior probabilities. Response-adaptive randomisation is also supported after a burn-in period. The app provides detailed output on single trial simulations, along with type 1 error and power estimates, and includes comparisons to traditional RCTs [16]. The R package MAMS supports the design of multi-arm multi-stage (MAMS) trials with various endpoint types within a group-sequential framework. Users can customise key design elements such as number of treatments, stages, treatment effects, error rates, and interim decision rules. It also includes functions to compute operating characteristics like rejection probabilities and expected sample size for specific hypotheses [17]. FACTS is a clinical trial simulation software to support trial design and statistical analysis. It estimates operating characteristics where closed-form solutions aren't feasible and accounts for practical aspects like accrual and dropout rates. The software supports various trial types (e.g., dose escalation, treatment comparisons) and endpoints (continuous, dichotomous, time-to-event), including complex designs like basket and umbrella trials with interim analyses and longitudinal modelling [18].

However, to the best of our knowledge, there is still a lack of software designed explicitly for cluster-randomised trials (CRTs) and multisite trials (MSTs) in relation to educational research and other social science fields that often involve hierarchical data structures. To address this gap, we have developed the Educational Platform Trials Simulator (EPTS), an R-Shiny web application tailored for planning and simulating CRT, MST and simple randomised trials (SRT). This software includes functionalities for conducting multilevel analyses using both Bayesian and Frequentist methods, as well as futility and superiority analyses using Bayesian methods. By providing a dedicated interactive tool accommodating functionalities for CRT, MST and SRT designs, EPTS enables researchers to simulate complex trial designs that are more aligned with the practical requirements of educational and similar multisite studies.

## 2. Software description

### 2.1. Software architecture

The EPTS is a web application developed using R-Shiny, a package within the R and RStudio statistical software environment. R-Shiny is a web-based user interface built on the R language, enabling users to execute R functions with their data without needing to install R or RStudio. The EPTS is compatible with all web browsers and can be accessed through the following link: https://epts-app.shinyapps.io/Educational-Platform-Trials-Simulator/. All computations are performed remotely on an R-Shiny server. The software's rules are based on calculating Bayesian posterior probabilities of superiority and futility. To run the simulator, you need to input some data into the input bar manually. The software enables you to save and load simulation outputs. Fig. 1 displays the initial window of the software. For all tabs, the input bar is located on the left side of the browser window, and the outputs are displayed on the right side. All functionalities available in the web application are also included in the *epts* R package, which is available on the Comprehensive R Archive Network (CRAN) [19]. The package can be installed and loaded using the following commands:

```
R> install.packages ("epts")
R> library (epts)
```

An offline version of the application is included in the package and can be launched locally using:

```
R> runEPTS()
```

Once installed, the offline version replicates the full functionality of the web application but runs entirely on the user's local machine, making it accessible without an internet connection.

### 2.2. Software functionalities

EPTS consists of seven tabs: data simulation, multi-arm analysis, futility analysis, superiority analysis, add new intervention, plot posterior probabilities, and user manual. The manual provides detailed descriptions of each input option and includes overviews of the outputs.

#### 2.2.1. Data simulation

In this study, we extended the model from Uwimpuhwe et al. [20] and Singh et al. [21] to accommodate multiple arms. Suppose $Pre_{ij}$ and $Post_{ij}$ are the pre-intervention and post-intervention scores, respectively, for pupil $i$ in school $j$. It is noted that we speak of "pupils" and "schools" owing to the educational context within which this work is presented. However, one can consider these terms simply as a proxy for the respective lower- and upper-level units, which could more generally be termed as "individuals" and "sites" (such as class, region, hospital…). The indicator variable $T_{ijk}$ takes value 1 if pupil $i$ in school $j$ is in the intervention group $k$, and 0 if pupil $i$ in school $j$ is in the control group. The model with multiple intervention arms can be formulated as:

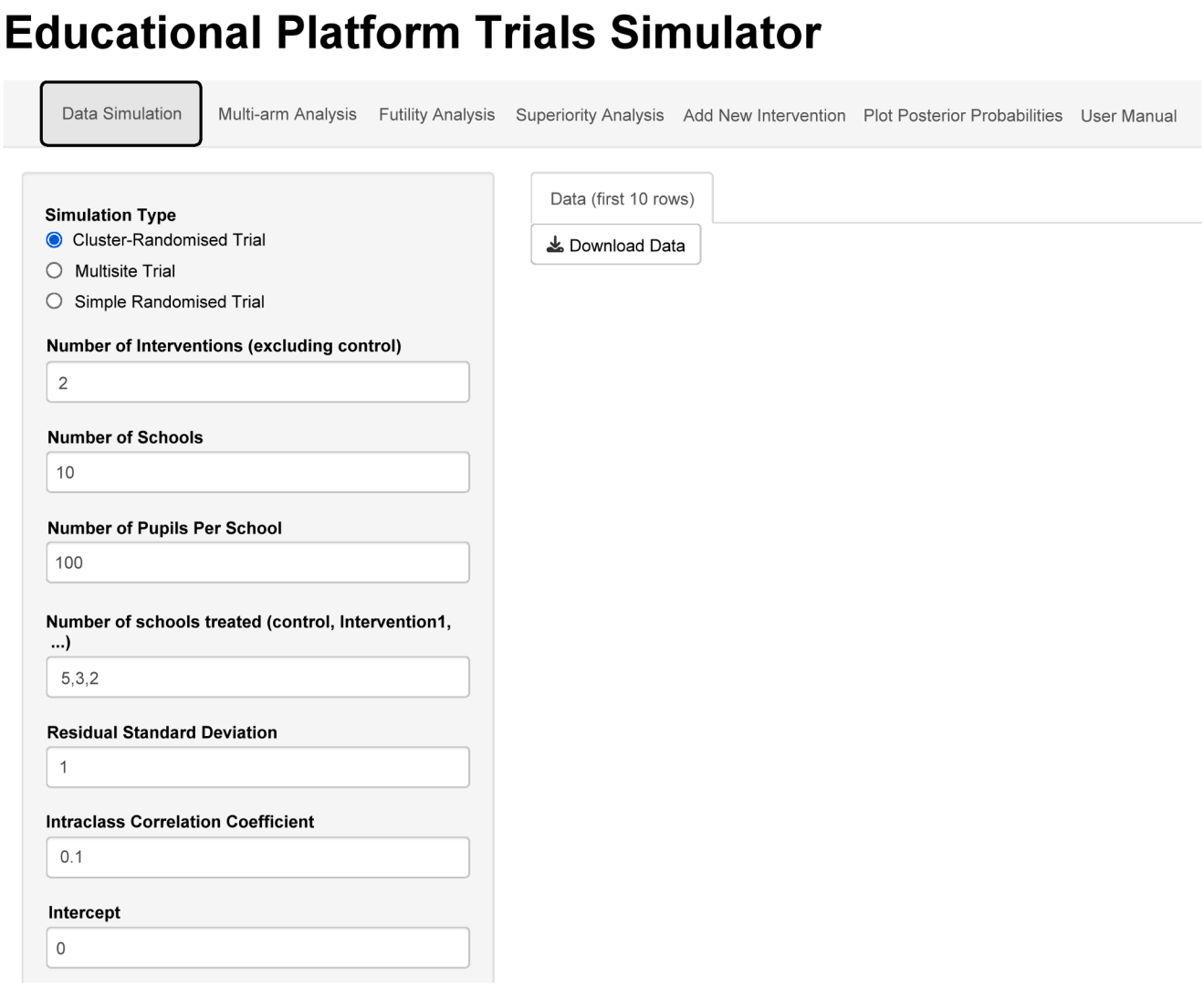


**Fig. 1.** The initial window of the EPTS software (note that the input bar on the left hand side is truncated).

$$Post_{ij} = \begin{cases} \beta_0 + \beta_1 Pre_{ij} + \sum_{k=1}^{K} \beta_{2.k} T_{ijk} + \varepsilon_{ij} \ \textit{for SRT} \\ \beta_0 + \beta_1 Pre_{ij} + \sum_{k=1}^{K} \beta_{2.k} T_{ijk} + b_{j0} + \varepsilon_{ij} \ \textit{for CR} \\ \beta_0 + \beta_1 Pre_{ij} + \sum_{k=1}^{K} \beta_{2.k} T_{ijk} + b_{j0} + \sum_{k=1}^{K} b_{jk} T_{ijk} + \varepsilon_{ij} \ \textit{for MST}, \end{cases}$$

where $\beta_0$ is the intercept, $\beta_1$ is effect of pre-test, $\beta_{2.k}$ is the effect of intervention $k$, and $\varepsilon_{ij}$ denotes the residual for pupil $i$ in school $j$. The $b_{j0}$ is a school-specific random intercept and $b_{jk}$ is a random effect for school-by- $k^{th}$ intervention interactions.

In order to conduct simulations, the user must first specify the type of trial: CRT, MST or SRT. The user specifies the following components in the input panel: number of interventions, standard deviation of residuals, intercept, effect size, random seed, and attrition rates. For CRT and MST, the user must also specify the number of pupils per school, number of schools, and for CRT, the percentage of intervention schools as well as the intraclass correlation coefficient (ICC). For MST, the user specifies the percentage of pupils in intervention groups, standard deviation of random intercept and the standard deviation of random slope. For SRT, the user specifies the total number of participants and the percentage of participants in intervention groups.

Optionally, the user can define covariates to include in the simulation model. Covariate names must begin with a letter and may contain letters, numbers, periods (.), and underscores (_). Names cannot start with a digit or underscore, nor include spaces or special characters. Each covariate must be specified as either continuous or categorical. For continuous covariates, the user defines the standard deviation and regression coefficient. For categorical covariates, the user selects a reference category, assigns probabilities to each category, and specifies coefficients for all non-reference categories.

The simulated data comprises pupil ID (or participant ID for SRT), school ID (for CRT and MST), multi-arm interventions, and their corresponding covariates and post-test scores. For CRT and MST, the dataset includes unique school IDs to represent clustering, while for SRT, the dataset includes participant IDs only. The first ten rows of the simulated data are displayed on the right side of the panel. The simulated data can be saved in CSV format.

### 2.2.2. *Multi-arm analysis*

The Multi-arm Analysis tab allows the user to analyse the CRT, MST and SRT datasets using Frequentist and Bayesian models. First, the user must specify the type of analysis as follows: crtBayes: Bayesian analysis of cluster randomised trials using vague priors. crtFREQ: Analysis of cluster randomised trials using multilevel model under a frequentist setting. mstBayes: Bayesian analysis of multisite randomised trials using vague priors. mstFREQ: Analysis of multisite randomised trials using

multilevel model under a frequentist setting. srtBayes: Bayesian analysis of simple randomised trials using vague priors. srtFREQ: Analysis of simple randomised trials under a frequentist setting.

In the input panel, after using simulated data or uploading the dataset and specifying the variables, the user can configure the settings for Bayesian analysis by specifying the number of Markov Chain Monte Carlo (MCMC) iterations per chain and the threshold for estimating the Bayesian posterior probability. Vague priors are used as recommended by [20] and implemented in [22]. For frequentist analysis, the user can choose the Analytical (default), Permutation or Bootstrap option to compute effect size confidence intervals. The bootstrapping method includes case re-sampling at the student level, case re-sampling at the school level, case re-sampling at both levels and residual bootstrapping. The functions crtBayes, crtFREQ, mstBayes, mstFREQ, srtBayes and srtFREQ from the eefAnalytics package were utilised for multi-arm analysis 22. The output will be a forest plot. The plot displays the effect sizes using within and total variances. Users can customise the plot by enabling the Plot Customization checkbox. The plot can be downloaded in TIFF, PDF, SVG, or EPS formats.

#### 2.2.3. Futility analysis

The Futility Analysis tab allows users to assess futility in platform trials. Futility is defined using a minimum expected effect size. We will estimate Bayesian posterior probabilities that the effect size for any intervention arm will be beyond a specific relevant threshold (i.e., minimum expected effect size) utilising trial outcome data from the pre- and post- intervention stage [20]. The intervention is considered futile if this probability is less than a pre-specified probability threshold [7,12,23]:

$$\mathrm{P}(\text{Effect size} \geq \text{Threshold} \mid \text{Data}) < \text{Futility Threshold}.$$

The functions crtBayes, mstBayes and srtBayes from eefAnalytics package were used to check futility for CRT, MST and SRT data, respectively. The user can adjust settings by specifying the number of MCMC iterations per chain, and the threshold for estimating the Bayesian posterior probabilities, along with the probability threshold for futility analysis. The results of the futility analysis will indicate whether each intervention is futile or not. The results can be saved in CSV format.

#### 2.2.4. Superiority analysis

Superiority is generally defined as the probability for an intervention being better than a designated reference intervention, sometimes also referred to as a "common control" [7,24,25]. The Superiority Analysis tab evaluates whether interventions are statistically superior to the reference intervention. Using Bayesian posterior probabilities, the analysis determines whether the probability of an intervention's effect size (relative to the reference intervention) exceeding a specified threshold is greater than a user-defined superiority threshold:

$$\mathrm{P}(\text{Effect size} \geq \text{Threshold} \mid \text{Data}) > \text{Superiority Threshold}.$$

Hence, in difference to the Futility Analysis which is carried out in relation to the actual control, this Superiority Analysis tab enables direct comparisons among interventions.

Users can configure parameters such as the number of MCMC iterations per chain, reference intervention, the effect size threshold, and the superiority threshold for posterior probabilities. The results of the superiority analysis will indicate which interventions, if any, are superior to the reference intervention. The result can be exported in CSV format.

#### 2.2.5. Add new intervention

The Add New Intervention tab allows users to expand an existing platform trial by adding a new intervention arm. This feature is crucial for adaptive platform trials, where new interventions can be introduced and evaluated alongside ongoing interventions. Using pre-existing trial data from pre- and post-intervention stages, users can add a new intervention group to the dataset with customisable parameters.

To generate the new intervention group, users need to specify several key parameters, including the number of new schools or clusters, the number of pupils per school, the expected effect size, and the attrition rate for the new intervention. Users have the flexibility to incorporate multiple covariates into the new intervention group. When adding a new intervention, users can specify any number of covariates that exist in the current dataset. The simulation process uses the mean and standard deviation of continuous covariates and the observed distributions of binary and categorical covariates from the existing dataset to generate corresponding values for the new intervention group, ensuring alignment with the original dataset's characteristics. For CRT, the new intervention group will be assigned to specific clusters, and for MST, the user can control the proportion of pupils receiving the new intervention within each new school. The resulting dataset can be downloaded for further analysis.

#### 2.2.6. Plot posterior probabilities

The Plot Posterior Probabilities tab visualises the posterior probabilities estimate from multiple intervention arms across different thresholds, allowing for comparison to check futility. The user can modify settings by specifying the number of iterations per chain for MCMC and the range of thresholds for estimating the Bayesian posterior probabilities. The user can also add a vertical line to represent the pre-specified threshold for estimating Bayesian posterior probability (Threshold) and a horizontal line to indicate the threshold of Bayesian posterior probability (ProbThreshold). By enabling the Plot Customization checkbox, users can personalize their plot with a range of options. These include setting a custom title, as well as custom labels for the x- and y-axes and specifying tick intervals. Users can also change the colour of the vertical and horizontal line, rename each intervention group, and choose specific colours for each intervention. Additionally, they can define the plot's width and height in inches to match their preferred output dimensions. The plot is created using the ggplot2 package and can be downloaded in TIFF, PDF, SVG, or EPS formats [26].

### 2.3. Performance

Computational time required to run analyses depends on several key factors. The number of pupils (or total number of individuals in study population) plays a major role, as larger sample sizes generally lead to longer computation times, especially in complex models such as multi-arm analyses. Similarly, the number of schools or clusters contributes to the complexity of hierarchical models like CRTs and MSTs, increasing the time needed for simulation and estimation. The inclusion of multiple intervention arms increases the computational workload, as each arm introduces additional parameters and comparisons. The type of model being used significantly affects processing time as well; for instance, Bayesian models typically require more time to fit compared to frequentist ones, and CRTs and MSTs tend to be slower than SRTs due to their nested structure. In Bayesian analyses, the number of MCMC iterations is a crucial determinant of computational time, with more iterations leading to longer runtimes. In frequentist approaches, the use of resampling techniques such as permutations and bootstraps similarly increase processing time, especially when a high number of iterations is required for stable estimates.

To estimate the time needed to perform various analyses within the Shiny application, we measured and recorded the computation time for each analysis. We used simulated data, including a four-arm design with 10 schools and 100 pupils per school for the CRT (Appendix A), MST designs (Appendix B), and a four-arm SRT (Appendix C) with 1000 individuals. In addition, the simulation inputs for CRT, MST, and SRT are provided in Appendix D (Tables D1–D3, respectively), along with an estimated computation time for respective approach (Table D4). The analytic (default) methods generally required the least amount of time, while bootstrap and permutation-based approaches were substantially

more time-consuming.

In the Shiny web application version, the server connection will time out after 15 min of inactivity, which may interrupt long-running computations such as analyses of large datasets, Bayesian analysis with a high number of MCMC iterations, or frequentist analysis involving extensive permutations or bootstrap resampling. For computationally intensive tasks that may exceed this time limit, we strongly recommend users to either run the equivalent R functions provided in the *epts* package or use the offline version of the application, which can be launched locally via the `runEPTS()` function.

## 3. Illustrative example

We present an example to illustrate the usage of the app. First, we simulate a CRT dataset with three interventions, ten schools, and 100 pupils per school. The number of schools assigned are 2 for the control group, 3 for Intervention 1, 2 for Intervention 2, and 3 for Intervention 3. The complete simulation inputs are provided in Appendix D (Table D1). Fig. 2 shows an example of simulated data with three interventions and three covariates. The simulated data can be downloaded and saved as a CSV file for future use in other tabs. This simulated data set is available in Appendix A. For multi-arm analysis, click on the "Multi-arm analysis" tab. Under the "Data Source" section, either select "Use Simulated Data" or upload a dataset by clicking the "Browse" button under "Choose CSV File". After choosing the "crtBayes" option via radio buttons, we input the variables and set the number of simulations (MCMC iterations per chain) to 10,000, with a 0.05 threshold of effect size for estimating Bayesian posterior probability. Fig. 3 displays the effect sizes as a forest plot. Next, we select the "Futility analysis" tab and set the number of simulations to the default (10,000), with a 0.05 threshold of effect size for estimating Bayesian posterior probability and a 0.8 probability threshold for futility analysis. The threshold of effect size beyond 0.05 is often considered meaningful in educational trials, as an effect beyond 0.05 is converted to one month of additional progress, while an effect size below 0.05 is interpreted as 0 months of progress [27]. The idea of considering a 0.80 probability threshold comes from the concept of statistical power in educational trials, where 0.80 is considered a meaningful and acceptable value [28]. The screenshot of the futility analysis results is presented in Fig. 4A. Then, to perform the superiority analysis, we should navigate to the "Superiority Analysis" tab. We use the same default values as in the futility analysis and set intervention 2 as the reference intervention. The screenshot of the superiority analysis results is shown in Fig. 4B To illustrate how to introduce a new intervention arm, we navigate to the 'Add New Intervention' tab. After uploading the existing dataset, we proceed to specify the parameters for the new intervention. In this example, we add 2 new schools with 100 pupils each, set an expected effect size of 0.4, and an attrition rate of 0.1. Once configured, clicking "Add New Intervention" integrates the new intervention group into the dataset. The dataset, which includes the new intervention arm, is provided in Appendix E. Finally, we select the "Plot Posterior Probabilities" tab and input the parameters similar to the "Futility tab". After selecting "Add a vertical line" and "Add a horizontal line", we specified the "Value for Vertical Line" as 0.05 and the "Value for Horizontal Line" as 0.8. Fig. 5 shows the posterior probability plot. In Fig. 5, with a threshold of 0.05 and a ProbThreshold of 0.8, interventions 1 and 2 were found to be futile. A step-by-step video tutorial is also provided in Appendix F to demonstrate the entire process outlined above.

To validate the performance of the EPTS, we conducted an additional simulation study based on a real-world educational trial. Specifically, we used summary estimates from the Lexia trial, a multisite education trial, as input parameters for the simulation [29]. Key trial characteristics, including the number of interventions, sample size per school, standard deviations, and coefficients, were entered into the simulator. Detailed input parameters are provided in Appendix D (Table D5). After running the simulation, we compared the resulting effect size and coefficients to those reported in the original Lexia trial. The simulation and actual trial data showed good agreement, with only minor deviations. For example, the effect size obtained from the simulated data was 0.10 compared to the reported effect size of 0.11, and the pre-test coefficients were similarly close (simulated: 4.15; reported: 4.14). As part of a sensitivity analysis, we also varied the simulated effect sizes while considering all the other parameters with similar values from Lexia trial. When simulating an effect size of 0.15, the resulting estimate was 0.13; similarly, a simulated effect size of 0.10 yielded a result of 0.09. These findings further support the model's consistency and its capacity to approximate real-world outcomes within a reasonable margin. Adjustments to the model parameters could further improve the accuracy, but the current simulation already offers valuable insights.

## 4. Impact

The presented R-Shiny app provides software with a graphical user interface for simulating platform trials of various design types which are

Data (first 10 rows)

| pupils | schools | interventions | pretest | gender | ethnicity | posttest |
|---|---|---|---|---|---|---|
| 1 | 1 | 1.00 | 0.98 | 1.00 | 2.00 | 4.04 |
| 2 | 1 | 1.00 | -0.62 | 1.00 | 2.00 | 0.31 |
| 3 | 1 | 1.00 | -0.73 | 0.00 | 0.00 | 1.05 |
| 4 | 1 | 1.00 | -0.52 | 1.00 | 2.00 | -0.32 |
| 5 | 1 | 1.00 | -1.75 | 1.00 | 2.00 | -2.26 |
| 6 | 1 | 1.00 | 0.88 | 0.00 | 2.00 | 3.76 |
| 7 | 1 | 1.00 | 1.37 | 0.00 | 0.00 | 3.02 |
| 8 | 1 | 1.00 | -1.69 | 0.00 | 2.00 | -1.25 |
| 9 | 1 | 1.00 | -0.63 | 1.00 | 0.00 | 2.93 |
| 10 | 1 | 1.00 | 0.02 | 1.00 | 0.00 | NA |

**Fig. 2.** Simulated cluster randomised data with three interventions.

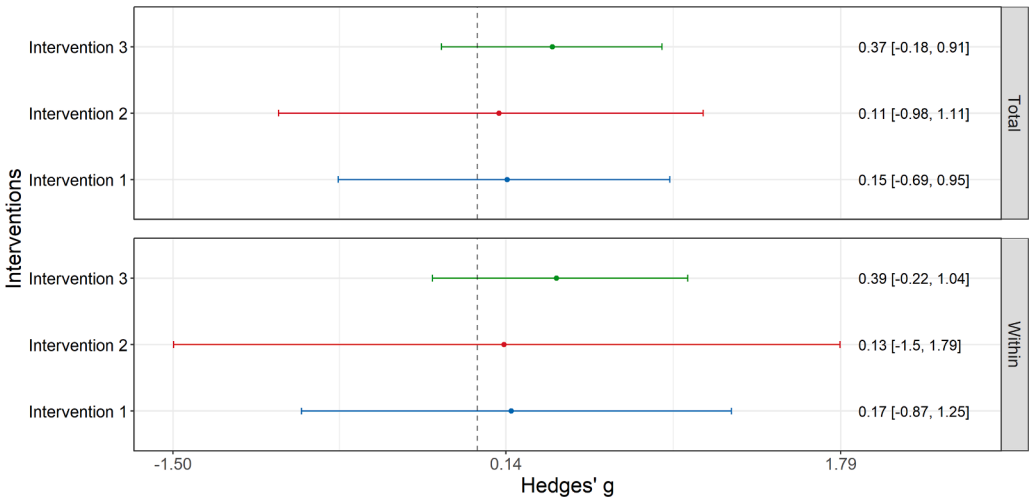


**Fig. 3.** Forest plot of comparison of effect sizes (Top: Estimated using total variance, Bottom: Estimated using within-cluster variance).

A

| Intervention | P(Effect size > Threshold) | Futility |
|---|---|---|
| 1 | 0.62 | Intervention 1 is futile. |
| 2 | 0.55 | Intervention 2 is futile. |
| 3 | 0.93 | Intervention 3 is not futile. |

B

| Intervention | P(Effect size > Threshold) | Superiority |
|---|---|---|
| 1 | 0.49 | Not Superior to the Reference Intervention |
| 2 | NA | Reference |
| 3 | 0.75 | Not Superior to the Reference Intervention |

**Fig. 4.** Screenshot of results A: Futility analysis, B: Superiority analysis.

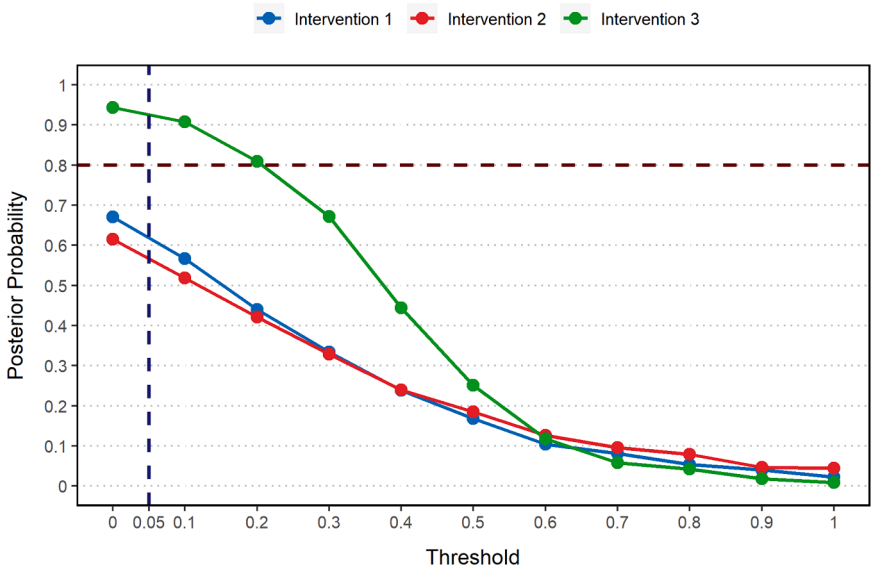


**Fig. 5.** Posterior probability plot for a four-arm cluster randomised trial.

of particular relevance in the educational sciences. This app is filling a critical gap since it is often very hard to get access to real data sets in education, requiring lengthy approval processes such as in the UK via the Integrated Data Service (IDS). Hence, there is a general lack of available educational data for development and testing of analytic methods as well as for training purposes. Furthermore, the tool may be used to understand the impact of setting and changing certain parameters on trial outcomes. Such results provide important insights for the design of new trials and can be used to discard certain parameter settings or trial configurations without running actual, costly, trials.

The EPTS allows users without programming expertise to simulate CRT, MST, as well as SRT data, set key parameters of the simulation, and simulate platform trials in education and other social science disciplines. Additionally, it can be used to simulate traditional two-arm trials, making it a versatile tool for trial design and analysis. To the best of our knowledge, this is the first free, open-source, web-based software based on the R environment enabling the simulation of data from CRT and MST designs.

## 5. Conclusion

In this paper, we have introduced an R-Shiny app for simulating CRTs, MSTs and SRTs. The app is an open-source, browser-based simulator for planning platform trials in the context of education. The app allows the user to set key features of the simulation and create various scenarios. The software also includes various graphical outputs to aid in interpreting futility and superiority. With its user-friendly interface and powerful simulation capabilities, EPTS is a valuable tool for researchers engaged in studies based on CRT, MST or SRT design, providing an accessible yet comprehensive approach to trial simulation. While the EPTS and this paper have been motivated by an unmet need in the educational sciences, the tool can in principle be used also in other sciences, assuming that researchers are able to match conventions and notations accordingly. For instance, a multi-centre clinical trial could correspond to either a multi-site trial or a cluster-randomised trial in education. While it is beyond the scope of this paper to draw comprehensively all such connections, we encourage the wider use of this EPTS, and welcome feedback, by users from various sciences.

## Data availability

The data simulated for the illustrative example section is available in Appendix A. The source code is available in GitHub repository (https://github.com/Mohammad-sayari/Educational-Platform-Trials-Simulator).

## CRediT authorship contribution statement

**Mohammad Sayari:** Writing – review & editing, Writing – original draft, Visualization, Software. **Akansha Singh:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology. **Germaine Uwimpuhwe:** Writing – review & editing, Validation, Software, Methodology. **Nasima Akhter:** Writing – review & editing. **Tahani Coolen-Maturi:** Writing – review & editing. **Rashmika Gupta:** Writing – review & editing. **Jochen Einbeck:** Writing – review & editing, Validation, Supervision, Methodology.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Mohammad Sayari reports financial support was provided by Education Endowment Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*Funding*

This research was funded by a grant to Durham University from the Education Endowment Foundation.

*Acknowledgements*

The authors would like to thank Dr Andrea Simkus for her input into conceptual questions concerning the simulation of platform trials.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.softx.2025.102214.

## References

[1] Park JJ, Harari O, Dron L, Lester RT, Thorlund K, Mills EJ. An overview of platform trials with a checklist for clinical readers. J Clin Epidemiol 2020;125:1–8. https://doi.org/10.1016/j.jclinepi.2020.04.025.
[2] Berry SM, Connor JT, Lewis RJ. The platform trial: an efficient strategy for evaluating multiple treatments. JAMA 2015;313(16):1619–20. https://doi.org/10.1001/jama.2015.2316.
[3] Burki T. Platform trials: the future of medical research? Lancet Respir Med 2023;11(3):232–3. https://doi.org/10.1016/S2213-2600(23)00052-8.
[4] Thorlund K, Haggstrom J, Park JJ, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians. BMJ 2018:360. https://doi.org/10.1136/bmj.k698.
[5] Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006;5(1):27–36. https://doi.org/10.1038/nrd1927.
[6] Hummel, J., S. Wang, and J. Kirkpatrick, Using simulation to optimize adaptive trial designs: applications in learning and confirmatory phase trials. 2015. p. 401–13. https://doi.org/10.4155/CLI.15.14.
[7] Saville BR, Connor JT, Ayers GD, Alvarez J. The utility of bayesian predictive probabilities for interim monitoring of clinical trials. Clinical Trials 2014;11(4):485–93. https://doi.org/10.1177/1740774514531352.
[8] Lachin JM. A review of methods for futility stopping based on conditional power. Stat Med 2005;24(18):2747–64. https://doi.org/10.1002/sim.2151.
[9] Saville BR, Berry SM. Efficiencies of platform clinical trials: a vision of the future. Clinical Trials 2016;13(3):358–66. https://doi.org/10.1177/1740774515626362.
[10] Tarran B. Education trials and tribulations. Significance 2019;16(6):8–9. https://doi.org/10.1111/j.1740-9713.2019.01331.x.
[11] Walton AE, Collins LM, Klasnja P, Nahum-Shani I, Rabbi M, Walton MA, Murphy SA. arXiv preprint. 2020. https://doi.org/10.48550/arXiv.2005.05880.
[12] Meyer EL, Mesenbrink P, Mielke T, Parke T, Evans D, König F. & EU-PEARL (EU Patient-cEntric clinicAl tRial pLatforms) Consortium., systematic review of available software for multi-arm multi-stage and platform clinical trial design. Trials 2021;22:1–14. https://doi.org/10.1186/s13063-021-05130-x.
[13] Meyer EL, Mielke T, Parke T, Jacko P, Koenig F. SIMPLE—A modular tool for simulating complex platform trials. SoftwareX 2023;23:101515. https://doi.org/10.1016/j.softx.2023.101515.
[14] Wathen JK. Octopus-optimize clinical trials on platforms using simulation 2020. https://github.com/kwathen/OCTOPUS.
[15] Krotka P, Hees K, Jacko P, Magirr D, Posch M, Roig MB. NCC: an R-package for analysis and simulation of platform trials with non-concurrent controls. SoftwareX 2023;23:101437. https://doi.org/10.1016/j.softx.2023.101437.
[16] Thorlund K, Golchi S, Haggstrom J, Mills E. Highly efficient Clinical Trials Simulator (HECT): software application for planning and simulating platform adaptive trials. Gates Open Res 2019;3:780. https://doi.org/10.12688/gatesopenres.12912.2.
[17] Jaki T, Pallmann P, Magirr D. The R package MAMS for designing multi-arm multi-stage clinical trials. J Stat Softw 2019;88:1–25. https://doi.org/10.18637/jss.v088.i04.
[18] Development Team FACTS. FACTS: fixed and adaptive clinical trial simulator [Computer software]. TX, USA: Berry Consultants LLC, A.; 2025. https://www.berryconsultants.com/software/facts/.
[19] Sayari, M., Singh, A., Uwimpuhwe, G., Einbeck, J., epts: educational Platform Trials Simulator. 2025. CRAN. https://CRAN.R-project.org/package=epts.
[20] Uwimpuhwe G, Singh A, Higgins S, Kasim A. Application of bayesian posterior probabilistic inference in educational trials. Int J Res Method Edu 2021;44(5):533–54. https://doi.org/10.1080/1743727X.2020.1856067.
[21] Singh A, Uwimpuhwe G, Li M, Einbeck J, Higgins S, Kasim A. Multisite educational trials: estimating the effect size and its confidence intervals. Int J Res Method Edu 2022;45(1):18–38. https://doi.org/10.1080/1743727X.2021.1882416.
[22] Uwimpuhwe G, Singh A, Higgins S, Xiao Z, Troyer ED, Kasim A. eefAnalytics: robust Analytical methods for evaluating educational interventions using Randomised Controlled Trials Designs, R software package. CRAN 2025. https://cran.r-project.org/package=eefAnalytics.
[23] Meyer EL, Mesenbrink P, Di Prospero NA, Pericàs JM, Glimm E, Ratziu V, Sena E, König F, Investigators EU-PEARL NASH. Designing an exploratory phase 2b

platform trial in NASH with correlated, co-primary binary endpoints. PLoS One 2023;18(3):e0281674. https://doi.org/10.1371/journal.pone.0281674.
[24] Granholm A, Kaas-Hansen BS, Lange T, Schjørring OL, Andersen LW, Perner A, Jensen AKG, Møller MH. An overview of methodological considerations regarding adaptive stopping, arm dropping, and randomization in clinical trials. J Clin Epidemiol 2023;153:45–54. https://doi.org/10.1016/j.jclinepi.2022.11.002.
[25] Viele K, Saville BR, McGlothlin A, Broglio K. Comparison of response adaptive randomization features in multiarm clinical trials with control. Pharm Stat 2020;19 (5):602–12. https://doi.org/10.1002/pst.2015.
[26] Wickham H, Chang W, Wickham MH. Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics. Version 2016;2(1):1–189. https://doi.org/10.32614/CRAN.package.ggplot2.
[27] Education Endowment Foundation. The education endowment foundation toolkit guide. London: EEF; 2023. https://d2tic4wvo1iusb.cloudfront.net/production/documents/toolkit/Toolkit_guide_v1.2_-_2023.pdf?v=1740582004.
[28] Hedges LV, Rhoads C. Statistical power analysis in education research. National Center for Special Education Research; 2010. https://ies.ed.gov/ncser/pubs/20103006/pdf/20103006.pdf.
[29] Tracey, L., Elliott, L., Fairhurst, C., Mandefield, L., Fountain, I., & Ellison, S. (2021). Lexia Reading Core5®. https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Lexia-report-unconditional-effect-sizes.pdf?v=1745835609.