Biometrical Journal **52** (2022) 61, zzz–zzz / DOI: 10.1002/bimj.200100000

CITIES: Clinical Trials With Intercurrent Events Simulator

Ahmad Hakeem Abdul Wahab[*1], **Yongming Qu** [2], **Hege Michiels** [3], **Junxiang Luo** [4], **Run Zhuang** [5], **Dominique McDaniel** [6], **Dong Xi** [7], **Elena Polverejan** [1], **Steven Gilbert** [8], **Stephen Ruberg** [9], and **Arman Sabbaghi** [5]

[1] Janssen Pharmaceuticals *(Statistics and Decision Sciences, 08560)*
[2] Eli Lilly and Company *(Department of Statistics, Data and Analytics, 46285)*
[3] Ghent University *(Department of Applied Mathematics, Computer Science and Statistics, 9000)*
[4] Moderna *(Department of Biostatistics and Programming, 02139)*
[5] Purdue University *(Department of Statistics, 47907)*
[6] Drexel University *(Department of Epidemiology and Biostatistics, 19121)*
[7] Gilead Sciences *(Department of Biostatistics, 94404)*
[8] Pfizer *(Global Product Development, 02139)*
[9] Analytix Thinking LLC *(Indianapolis, 46236)*

Although clinical trials are often designed with randomization and well-controlled protocols, complications will inevitably arise in the presence of intercurrent events (ICEs) such as treatment discontinuation. These can lead to missing outcome data and possibly confounding causal inference when the missingness is a function of a latent stratification of patients defined by intermediate outcomes. The pharmaceutical industry has been focused on developing new methods that can yield pertinent causal inferences in trials with ICEs. However, it is difficult to compare the properties of different methods developed in this endeavor as real-life clinical trial data cannot be easily shared to provide benchmark datasets. Further, different methods consider distinct assumptions for the underlying data generating mechanisms, and simulation studies often are customized to specific situations or methods. We develop a novel, general simulation model and corresponding Shiny application in R for clinical trials with ICEs, aptly named the Clinical Trials With Intercurrent Events Simulator (CITIES). It is formulated under the Rubin Causal Model where the considered treatment effects account for ICEs in clinical trials with repeated measures. CITIES facilitates the effective generation of data that resemble real-life clinical trials with respect to their reported summary statistics, without requiring the use of the original trial data. We illustrate the utility of CITIES via two case studies involving real-life clinical trials that demonstrate how CITIES provides a comprehensive tool for practitioners in the pharmaceutical industry to compare methods for the analysis of clinical trials with ICEs on identical, benchmark settings that resemble real-life trials.

Key words: Causal Inference; Estimands; Intercurrent Events; Repeated Measures.

1 Introduction

Randomized controlled clinical trials (RCTs) begin as designed experiments that use treatment randomization to balance baseline covariates (in expectation) and create the basis for valid statistical inferences (Fisher, 1971), thereby enabling unbiased and unambiguous causal inferences on the direct effects of a treatment or intervention (Hariton and Locascio, 2018). During the course of a clinical trial, especially for larger and longer trials, some patients inevitably discontinue their (randomly) assigned treatment due to adverse effects (AEs) of the treatment, lack of efficacy (LoE), personal or administrative reasons, or excess efficacy (EE) (Akacha et al., 2017a). Disruptions to the planned clinical trial protocol and discontinuations of trial medications can confound causal inferences on the effects of the receipt of treatment for the trial subjects (Frangakis and Rubin, 2002; Carpenter et al., 2014). While the first three reasons for discontinuing a study treatment are well recognized, the latter can be illustrated by an example in obesity studies.

*Corresponding author: aabdulw1@its.jnj.com

Overweight participants may reach their weight target and not want to have further weight loss. They think the current weight can be maintained through diet and exercise and thereby discontinue their study medication (Davidson et al., 1999). This is a pervasive and important problem in pharmaceutical drug development, as well as in academic and government-funded clinical trials. Such disruptions are referred to as intercurrent events (ICEs) in the recent International Council of Harmonization (ICH) Guidelines on this matter, namely the ICH E9 (R1) Addendum (ICH, 2019). More formally, ICEs are "events that occur after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest".

A natural thought is to only include patients who adhere to the study medication in the analysis without protocol deviations, which is akin to a "per-protocol" (PP) analysis. However, this conditions on the actual post-randomization outcomes within each arm. This breaks the randomization and ultimately does not correspond to a causal estimand since the subpopulations of interest may be inherently different (Imbens and Rubin, 2015; Lipkovich et al., 2020). A strategy to overcome this is to account for the different latent strata of patients, defined based on whether patients experience ICEs under treatment or control, in order to obtain valid inferences on the causal effects of treatment for the target latent subgroup of subjects. An example would be the population of patients who will not have ICEs and who adhere to treatment assignment, be it on both arms or at least the experimental arm. Disruptions due to ICEs, and their consequences on the possibility of obtaining valid and unbiased causal inferences, are especially amplified in clinical trials with repeated measurements, as they provide more chances for ICEs to occur.

ICH E9 (R1) suggests that principal stratification can be an important and relevant strategy to handle ICEs. A consequence of these new guidelines was an increased dedication of efforts and resources by statisticians in the pharmaceutical industry to develop new causal inference methodologies that can assess direct treatment effects without being confounded by treatment discontinuations or ICEs. Such new methods are arguably more useful for treatment evaluations in the pharmaceutical industry. One example of such a methodology is the Tripartite Estimand Approach (TEA) of Akacha et al. (2017b). This framework was proposed to disentangle the pure causal effects of a treatment intervention in the presence of ICEs. It consists of three estimands that are relevant and meaningful not only to patients, prescribers, and payers, but also to sponsors and regulators. The TEA estimands are the probability of discontinuation due to AEs, the probability of discontinuation due to LoE, and the direct treatment effect (Pearl, 2013; Qu and Lipkovich, 2021) in patients who can adhere to the treatment of interest under either treatment or control. An application of TEA can be found in Qu et al. (2021).

However, comparing and contrasting the properties of competing methods developed in this endeavor is difficult. One difficulty is that real-life clinical trial data are not easily accessed and shared across different teams and organizations to establish benchmarks. Another difficulty in evaluating different methods is that they involve distinct assumptions for the underlying data generating mechanisms. In addition, novel methods may not be as readily testable on clinical trials from other companies. These difficulties have led to the unfortunate current situation in which publications describing new methodologies for the pharmaceutical industry only include summary information and performance metrics of the new models, and which do not provide information on the relative advantages and disadvantages of the new models with respect to other models, trials, or data generating mechanisms. The lack of a platform to evaluate competing methods ultimately impedes the development and identification of effective methods for clinical trials with ICEs.

One resolution of this dilemma is the use of clinical trial simulators to evaluate competing methods (Boulesteix et al., 2018). Comprehensive software platforms such as FACTS, Certara®, and Cytel® can simulate clinical trials. However, these industry platforms do not incorporate ICEs in a causal setting, and often require prior programming experience and extensive training. Furthermore, they are fairly expensive, compelling researchers in academic institutions with limited resources to use statistical packages that are free, or code entire simulators themselves. Paux and Dmitrieniko (2019) developed the `mediana` R package to simulate, model, and evaluate clinical trials with multiple endpoints based on the Clinical Scenario Evaluation (CSE) approach developed by Benda et al. (2010) and refined by Friede et al. (2010). Although this software package can simulate data from many distributions, such as the Negative Binomial

and truncated distributions, it does not provide the flexibility to integrate different functions of discontinuation from varying sources. Sofrygin et al. (2017) developed the `simcausal` R package for specification and simulation of complex longitudinal data structures using a nonparametric structural equations model (NPSEM) that can be represented using directed acyclic graphs (DAG, Pearl, 1995). This package permits the integration of correlated outcomes using copulas with accessible syntax and direct outputting of causal graphs. However, it does not enable users to directly incorporate different functions of treatment discontinuation. Ultimately, the programming difficulty for these software platforms impedes their usage by researchers. Existing software packages also contribute to the perception of a black box simulator, which is not helpful in facilitating meaningful methodological discussions between statisticians, clinicians, and medical providers.

In this paper we present a clinical trial simulator that incorporates realistic scenarios for ICEs or treatment discontinuation. Our simulator is called the Clinical Trials with Intercurrent Events Simulator (CITIES). It is executed as a Shiny application in R, and can run on any web browser. In contrast to existing R Shiny applications for simulating clinical trials, such as those developed by Thorlund et al. (2019), Wojciechowski et al. (2015), Grayling and Wason (2020), and Karanevich et al. (2021), CITIES is unique in that it simulates potential outcomes under the Rubin Causal Model (RCM, Rubin, 1974; Holland, 1986), and summarizes causal effects in the presence of ICEs in a transparent and intuitive manner. CITIES is effectively a direct application of potential outcomes using TEA as defined in Akacha et al. (2017b), which forms the genesis of this work. An advantage of CITIES is that it can incorporate and control different sources of patient discontinuation with varying functional trends across treatment and time cleanly and directly with graphical representation, thereby demystifying the existing perception of a black box simulator in this context.

We proceed in Section 2 to outline how the potential outcomes are generated, how the varying sources of discontinuation are integrated, and how the causal effects and treatment discontinuation summaries are calculated in CITIES. We demonstrate applications of CITIES for simulating data resembling two real-life clinical trials with ICEs in Section 3. Finally, we provide concluding remarks on clinical trial simulators with ICE and potential extensions of CITIES in Section 4.

2 The CITIES Tool

2.1 The Clinical Trial Setting in CITIES

CITIES builds a realistic model describing patient behaviors and decisions as to whether they continue to take their medication and follow the trial protocol, or whether they discontinue their study treatment during the trial. As with any model, assumptions must be made about how patients will respond to the treatments during the trial. CITIES involves a modest number of user-defined input parameters that characterize the assumptions and functional relationships underlying the model. They are entered via four tabs in CITIES that users can easily interact with: Mean Treatment Response Settings, LoE & EE, Admin & AE, and Average Treatment Effects. Once the input values are specified, simulations can be executed to produce multiple datasets that resemble real-life clinical trial results. Different methodologies can then be applied and compared in multiple ways (e.g., with respect to bias, coverage probabilities, and Type I error rates) under different realistic clinical trial scenarios by means of the multiple simulated datasets. We describe the different input parameters, tabs, and model assumptions in the remainder of this section. Outputs of CITIES include visualizations that help the user understand how the inputs are used in the data generating model.

2.2 Mean Treatment Response Settings

CITIES considers longitudinal clinical measures for both continuous and binary outcomes in patients across time. We represent such measures using potential outcomes, and model continuous measures across

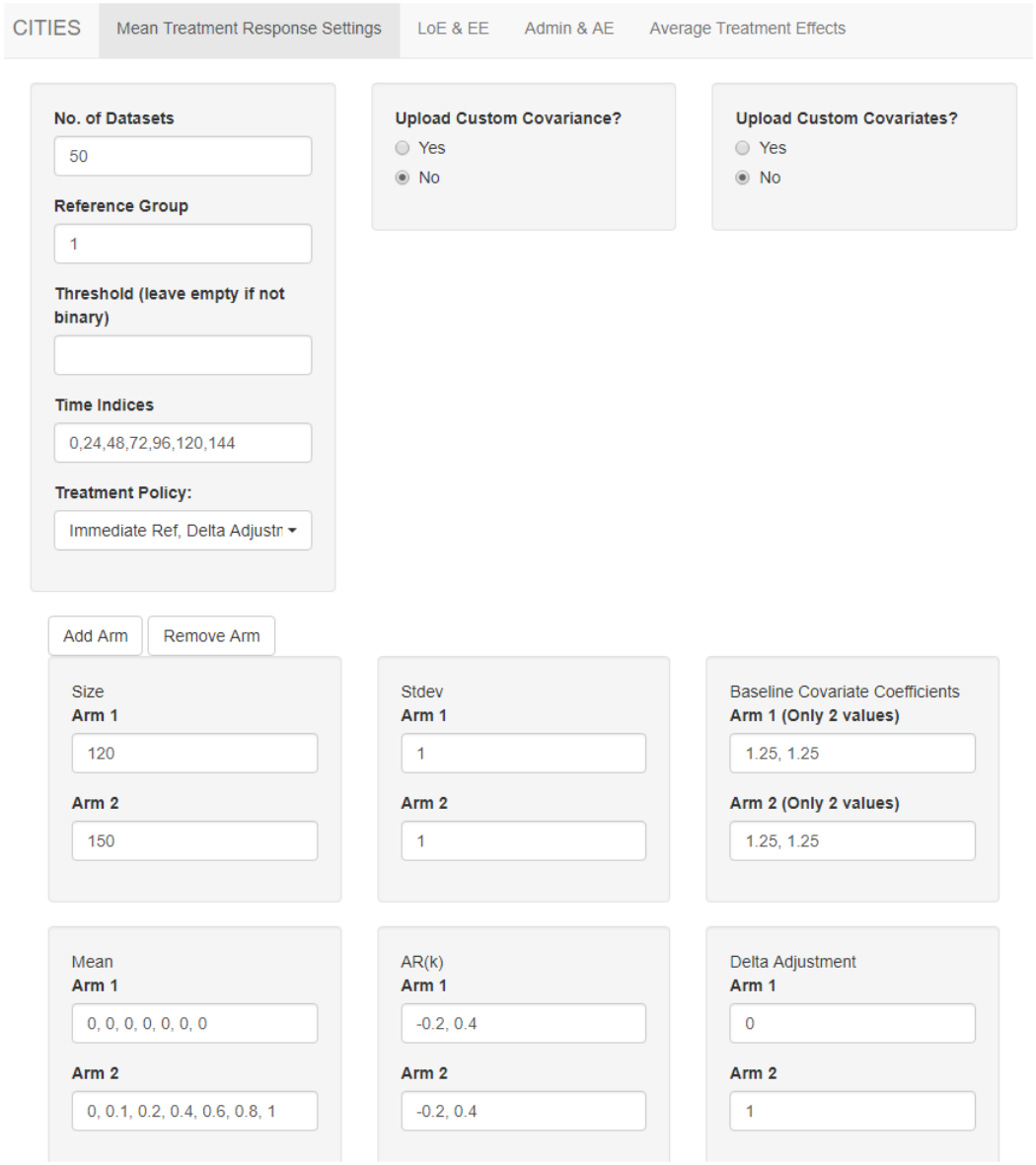

Figure 1 Illustration of the Mean Treatment Response Settings tab in CITIES.

time using a multivariate Normal distribution with a default specified autocorrelation structure or user provided structure in its covariance matrix.

The potential outcomes representation in CITIES utilizes the Stable Unit-Treatment Value Assumption (SUTVA, Imbens and Rubin, 2015, p. 9–13). Under SUTVA, there are no lurking variations in treatment that give rise to different potential outcomes, and there is no interference amongst subjects. In this case we denote the potential outcome for subject $i = 1, \ldots, I$ at time $j = 1, \ldots, J$ under treatment $t = 1, \ldots, T$ by $Y_{ij}(t)$. The observed outcome Y_{ij}^{obs} for subject i at time j is a function of the treatment assignment indicator T_i and their potential outcomes, namely, $Y_{ij}^{\text{obs}} = \sum_{t=1}^{T} I\{t = T_i\} Y_{ij}(t)$. Causal effects under the Rubin Causal Model (RCM) are defined in terms of comparisons of potential outcomes for a set of

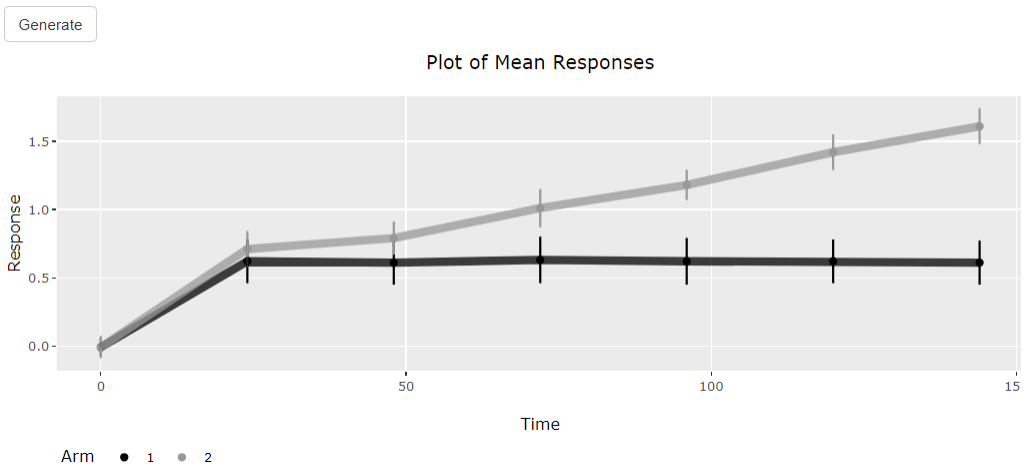

Figure 2 Mean Profile of the Mean Treatment Response Settings tab in CITIES, without discontinuation of study treatment (i.e., under no ICE). Generic names such as arms 1 and 2 are used, since CITIES can generate trials with more than 2 arms.

subjects (Imbens and Rubin, 2015, p. 5–7). For example, relative to some reference treatment arm or group T_i', the standard individual treatment effect for subject i at time j is $\tau_{ij}(t) = Y_{ij}(t) - Y_{ij}(T_i')$, which is consistent with how treatment effect is defined in ICH E9(R1): "how the outcome of a [test] treatment compares to what would have happened to the same patient with no treatment or an alternative treatment [control]". Ultimately, the corresponding finite-population average treatment effect $\tau_{\text{ATE},j}(t)$ at time j is the average of all the individual treatment effects at time j, i.e., $\tau_{\text{ATE},j}(t) = \sum_{i=1}^{I} \tau_{ij}(t)/I$.

We let Y_{i0} denote patient i's baseline measurement and $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^T$ be a vector of length p containing baseline covariates, taken prior to their treatment assignment. Since the emphasis of this paper is on randomized clinical trials, these measurements and baseline covariates are independent of treatment assignment, i.e., they have the same distribution across all treatment arms before treatment is assigned. The vector consisting of the baseline measure Y_{i0} and all potential outcomes $\boldsymbol{Y}_i(t) = (Y_{i1}(t), \ldots, Y_{iJ}(t))^{\mathsf{T}}$ under treatment t for a single subject i with the accompanying vector of covariate coefficients $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_p(t))^T$ of length p, is assumed to follow a multivariate Normal distribution with mean vector $(\mu_0, \boldsymbol{\mu}(t) + \boldsymbol{X}_i^T \boldsymbol{\beta}(t))^{\mathsf{T}}$ and covariance matrix $\boldsymbol{\Sigma}(t)$. The specified structure for $\boldsymbol{\Sigma}(t)$ in CITIES follows either an autoregressive process of order k scaled by a parameter $\sigma^2(t)$ or a custom covariance supplied by the user, both of which may vary across treatment assignment. In other words, we entertain the possibility of heterogeneous treatment effects. More formally,

$$\begin{pmatrix} Y_{i0} \\ \boldsymbol{Y}_i(t) \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} \mu_0 \\ \boldsymbol{\mu}(t) + \boldsymbol{X}_i^T \boldsymbol{\beta}(t) \end{pmatrix}, \quad \boldsymbol{\Sigma}(t) = \begin{pmatrix} \Sigma_{11}(t) & \Sigma_{12}(t) \\ \Sigma_{21}(t) & \Sigma_{22}(t) \end{pmatrix} \right),$$

where $\boldsymbol{\mu}(t) = (\mu_1(t), \ldots, \mu_J(t))^{\mathsf{T}}$ and $\boldsymbol{\Sigma}(t) = \sigma^2(t)\text{AR}_t(k)$ by default or what was user supplied. Here, $\text{AR}_t(k)$ is defined as the correlation matrix for the $\text{AR}(k)$ process on the outcomes for treatment t. Specifically, when $k = 1$, the (i, j)th element of the $\text{AR}(1)$ correlation matrix is defined as $\rho(t)^{|i-j|}$, for an autocorrelation parameter $\rho(t) \in (0, 1)$, $1 \leq i, j \leq J$. For $1 < k < J - 1$, the correlation matrix is populated from a vector of partial autocorrelations using the recursive relation in (Joe, 2006, p. 2178). Conditional on the baseline measurement Y_{i0} and baseline covariates $\boldsymbol{X}_{i0}$, CITIES generates the potential outcomes for each subject across both treatments according to

$$[\boldsymbol{Y}_i(t) \mid Y_{i0}, \boldsymbol{X}_i] \sim \text{N}\left(\boldsymbol{\mu}(t) + \boldsymbol{X}_i^T \boldsymbol{\beta}(t) + \Sigma_{21}(t)\Sigma_{11}^{-1}(t)(Y_{i0} - \mu_0), \Sigma_{22}(t) - \Sigma_{21}(t)\Sigma_{11}^{-1}(t)\Sigma_{12}(t)\right),$$

with $\boldsymbol{Y}_i(0)$ independent of $\boldsymbol{Y}_i(1)$ conditional on Y_{i0}. Thus each patient arrives at the clinical trial with a baseline clinical observation of $\boldsymbol{Y}_i(0)$, and subsequent potential outcomes for that patient emanate from that baseline according to the user defined treatment mean response and covariance structure.

CITIES generates clinical trials based on the specified number of data sets, subjects per treatment, mean vectors for the corresponding treatments, the supplied covariance structure, baseline covariates and associated coefficients $\boldsymbol{\beta}(t)$. Further, users will also specify meaningful time points for each measurement under the "Time Indices", such as weeks or days. Users can have multiple arms by clicking the "Add Arm" button and remove arms by clicking the "Remove Arm" button. As such, users will have to supply a reference group for subsequent pairwise analyses.

As regulatory agencies often request post-treatment follow-up, users can request estimands based on treatment policy assumptions in CITIES, which include Immediate Reference and Delta Adjustment. The former assumes that patients who discontinue treatment will immediately drop back to their reference arm's potential outcomes while the latter takes in a delta adjustment value per arm and decrements the potential outcomes by the supplied delta after treatment discontinuation. Note that the potential outcomes adjusted using delta adjustment will never exceed the corresponding potential outcomes on the reference arm, i.e. they will at most revert back to their potential outcomes in the reference group.

Figure 1 and 2 illustrate the Mean Treatment Response Settings tab in CITIES where higher values are favorable. Note on the top right that users can choose to upload custom covariances and covariates instead of using the default options. Regardless of whether the user intends to work on the original scale or change from baseline measurements, the first input in both the vector of treatment and control means need to be the same. In addition, users can also specify a threshold value to dichotomize the data. For example, success might be defined as an HbA1c being under 7% or a 90% reduction in Psoriasis Area and Severity Index relative to baseline (PASI)(Puig, 2015). Based on the supplied number of datasets, the "Generate" button then simulates the datasets and visualizes the means and standard errors for both treatment arms on the panel on the right. Here the standard errors are defined as the standard deviation of the means over the number of generated datasets. The error bars in this panel correspond to the means plus and minus one standard error.

2.3 Modelling LoE and EE

CITIES models treatment discontinuation due to either LoE or EE by means of discontinuation probabilities that are a function of the potential outcomes $\mathbf{Y}_i(t)$ for the assigned treatment. If a patient's response to treatment exhibits little improvement from baseline, or is even worse than baseline, then the patient or their attending physician is more likely to discontinue the study treatment. The discontinuation probabilities are specified to reflect this LoE. The opposite holds if a patient's response to treatment exhibits a great deal of improvement from baseline, which is EE. For treatment discontinuation due to EE, we envision realistic scenarios in which patients decide to stop taking their study treatment because they perceive that they are doing so well that there is no longer a need to remain on the study treatment. This can occur in trials of psychiatric disorders such as depression and schizophrenia in which patients can feel better during the trial and then decide to stop taking their medication. There may also be situations in which the effect of the treatment may be so large that it causes a patient to discontinue, as in cases of an excessive decrease in blood pressure or blood glucose levels that result in hypotension or hypoglycemia, respectively.

To formally describe the LoE curve, let $\Pr\{\mathrm{DC}_{\mathrm{LoE},ij}(t)\}$ denote the probability of treatment discontinuation due to LoE for a particular patient i at time j under treatment t, $p_{\mathrm{LoE,max}} \in [0,1]$ be the specified maximum value of the probability of discontinuation due to LoE, $\Delta Y_{ij}(t) := Y_{ij}(t) - Y_{i0}$ be the change from baseline to time point j in potential outcome $Y_{ij}(t)$, and L_{LoE} and U_{LoE} denote lower and upper thresholds on $\Delta Y_{ij}(t)$, respectively. Then, we use a piecewise linear function to model the probability of

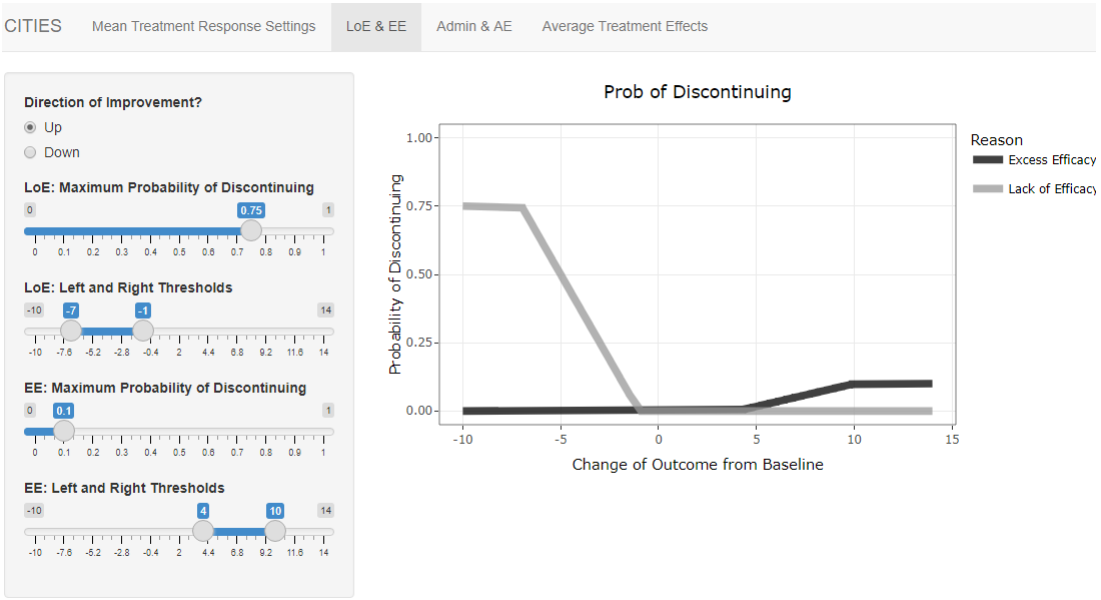

Figure 3 Illustration of the LoE & EE tab in CITIES. In this example, higher outcome values indicate better responses, $p_{\mathrm{LoE,max}} = 0.75$, $y_{\mathrm{L,LoE}} = -7$, $y_{\mathrm{U,LoE}} = -1$, $p_{\mathrm{EE,max}} = 0.1$, $y_{\mathrm{L,EE}} = 4$, $y_{\mathrm{U,EE}} = 10$. The "Prob of Discontinuing" plot visualizes the treatment discontinuation probabilities due to LoE and EE on the y-axis as a function of the change from baseline on the x-axis under these values of the input parameters.

discontinuation due to LoE as follows:

$$\Pr\{\mathrm{DC}_{\mathrm{LoE},ij}(t)\} = \begin{cases} p_{\mathrm{LoE,max}} & \text{if } \Delta Y_{ij}(t) \leq L_{\mathrm{LoE}}, \\ \frac{p_{\mathrm{LoE,max}}\{U_{\mathrm{LoE}} - \Delta Y_{ij}(t)\}}{(U_{\mathrm{LoE}} - L_{\mathrm{LoE}})} & \text{if } L_{\mathrm{LoE}} < \Delta Y_{ij}(t) \leq U_{\mathrm{LoE}}, \\ 0 & \text{if } \Delta Y_{ij}(t) > U_{\mathrm{LoE}}, \end{cases} \tag{1}$$

The same is done to describe the EE curve, but in the opposite direction of improvement. This describes a Missing Not At Random (MNAR) pattern since the change in baseline measurement used to calculate the $\Pr\{\mathrm{DC}_{\mathrm{LoE},ij}(t)\}$ is also missing or unobserved.

In the LoE & EE tab in CITIES, users first specify whether higher outcome values indicate better responses. The direction of improvement can also be changed based on the nature of the study. In diabetes lower blood glucose is desirable, and is often measured by reductions in glycated hemoglobin or HbA1c (EMA, 2018b; Karges et al., 2014). In contrast, trials on Alzheimer's disease can utilize cognition rating scales to measure a patient's disease status, and higher cognition scores are of interest for such scales (EMA, 2018a; Verdile et al., 2004). The LoE and EE curves in CITIES will reflect the specified direction dynamically. For example, when higher outcome values indicate better responses, $\Pr\{\mathrm{DC}_{\mathrm{LoE},ij}(t)\}$ will be a decreasing function while $\Pr\{\mathrm{DC}_{\mathrm{EE},ij}(t)\}$ will be an increasing function of $d_{ij}(t)$. Figure 3 illustrates the LoE & EE tab in CITIES.

2.4 Modelling Admin and AE

In addition to LoE and EE, patients can discontinue their study treatment due to perceived or actual adverse reactions. Such adverse reactions can range from headaches or nausea to more severe or serious complications such as cardiac arrhythmia, inflammation of the brain, or even death. Although the occurrence of

adverse reactions and a patient's or physician's decision to discontinue study treatment is complex, in order to balance simplicity with the complexity of real clinical trials, CITIES models discontinuation due to AEs in two parts. First is the occurrence of an AE at a particular time under an assigned treatment, and second is the probability of deciding to discontinue treatment due to the AE at that time. It is worth noting that AEs can be recurrent events and in clinical trial reports, only the proportion of patients with at least one AE and subsequently discontinuations due to AE are generally reported. To that end, let K be a poisson random variable that describes the number of AEs. Then the probability of observing k AEs on a particular patient given treatment arm t is $\Pr\{K = k|T = t\} = (\lambda_t d)^k \exp(-\lambda_t d)/k!$, where d is the duration of the study and λ_t is the event rate per unit of time on arm t. Define $p_{\text{DC}|AE,t}$ to be the probability of discontinuing if an AE does occur, i.e. $\Pr(\text{DC} = 1|\text{AE} = 1, T = t)$. Then, given a patient has k AEs, the probability of not discontinuing due to AE during the study on treatment arm t is $(1-p_{\text{DC}|AE,t})^k$. Then the cumulative rate of treatment discontinuation due to AE up to time d on treatment arm t is $r_{DC|AE,t} = 1-\exp(-\lambda d p_{\text{DC}|\text{AE,t}})$.

To simulate AE and the subsequent discontinuations, we first divide the trial periods into intervals $\tau_0 < \tau_1 < \ldots < \tau_{J-1}$, which need not be equally spaced. For each interval $(\tau_{j-1}, \tau_j]$, $j = 1, \ldots, J$, we generate K number of AEs at time j from a Poisson distribution with rate $\lambda(\tau_j - \tau_{j-1})$. Should the subject experience k AEs during the j-th interval, the treatment discontinuation indicator variable can be generated from a Bernoulli distribution with probability of $1 - (1 - p_{\text{DC}|\text{AE,t}})^k$.

All input parameters for the AE generator are available from clinical trial summary results and publications. From clinical trial results, it is standard protocol to list the proportion of patients that did not experience any AE on treatment arm t, i.e. $\hat{\Pr}\{K = 0|T = t\}$. Then we have $\Pr\{K = 0|T = t\} = \exp(-\lambda_t d)$, which can provide an estimate for λ_t, i.e. $\hat{\lambda_t} = -d^{-1} \log\left(\hat{\Pr}\{K = 0|T = t\}\right)$. In assuming that the AEs follow a Poisson process, each event is independent of the occurrence of the previous event. Given a patient has k events, the probability of not discontinuing due to AEs during the study is $(1 - p_{\text{DC}|\text{AE,t}})^k$ on that treatment arm. Then, on average, the corresponding discontinuation rate due to AE during the study period is $r_{\text{DC}|\text{AE,t}} = 1 - \exp(-\lambda_t d \times p_{\text{DC}|\text{AE,t}})$. Let $\hat{r}_{\text{DC}|\text{AE,t}}$ be the estimator for $r_{\text{DC}|\text{AE,t}}$. Then $\hat{p}_{\text{DC}|\text{AE,t}} = \log\left(1 - \hat{r}_{\text{DC}|\text{AE,t}}\right) / \log\left(\hat{\Pr}(K = 0|T = t)\right)$.

Figure 4 Illustration of the Admin & AE tab in CITIES. In this example, the probability of discontinuing due to administrative reasons are 0.02 on both arms respectively, $\hat{\Pr}(K > 0, T = 1) = 0.7$, $\hat{r}_{\text{DC}|\text{AE,1}} = 0.1$, $\hat{\Pr}(K > 0, T = 2) = 0.9$ and $\hat{r}_{\text{DC}|\text{AE,2}} = 0.1$.

Patients may also discontinue treatment due to administrative reasons. Examples include personal events that preclude further participation in a clinical trial (e.g., moving to a new location, pregnancy, and change in marital status), patients' unwillingness to continue participating in frequent doctor visits, or other burdens related to clinical evaluations of the patient. Such discontinuations are typically considered

to be unrelated to study treatment and not within the control of the clinicians running the trial. As such, in CITIES the probability of discontinuation due to administrative reasons is a flat probability that is the same at each time point that can vary across the experimental and control arms. In the Admin & AE tab in CITIES, users need to specify the the probability of discontinuing due to administrative reasons, the probability of observing at least one AE and the proportion of discontinuing due to AE on the control and experimental arms (Figure 4)

2.5 Summary of the Data Generation Mechanism in CITIES

CITIES incorporates all of the previously described components in its data simulation (Figure 5). First, the baseline responses Y_{i0} for the patients are generated independently of treatment according to the specified Normal distribution. Second, all potential outcomes $\mathbf{Y}_i(0)$ and $\mathbf{Y}_i(1)$ for each patient i are generated based on the user's specification for the multivariate Normal distribution under treatments conditional on the baseline response. These potential outcomes correspond to how each patient would respond to each treatment when there are no ICEs. This mode of generating data in CITIES is consistent with the ICH E9(R1) definition and is compatible with the potential outcomes framework as defined by Splawa-Neyman et al. (1990), Splawa-Neyman (1935), and Rubin (1974), and used in the causal inference literature (Holland, 1986). Third, CITIES calculates for each patient i at each time j the probabilities that they discontinue due to the various types of ICEs described previously in this section, and generates discontinuation statuses for the patients according to Bernoulli random variables with the calculated probabilities. Once a subject discontinues treatment, all subsequent visits are missing, unless one selects a treatment policy option for generating outcome data post discontinuation of a patient's randomized study treatment. Furthermore, a patient can have multiple reasons for discontinuation at a given visit and CITIES records an appropriate flag variable for each reason separately.

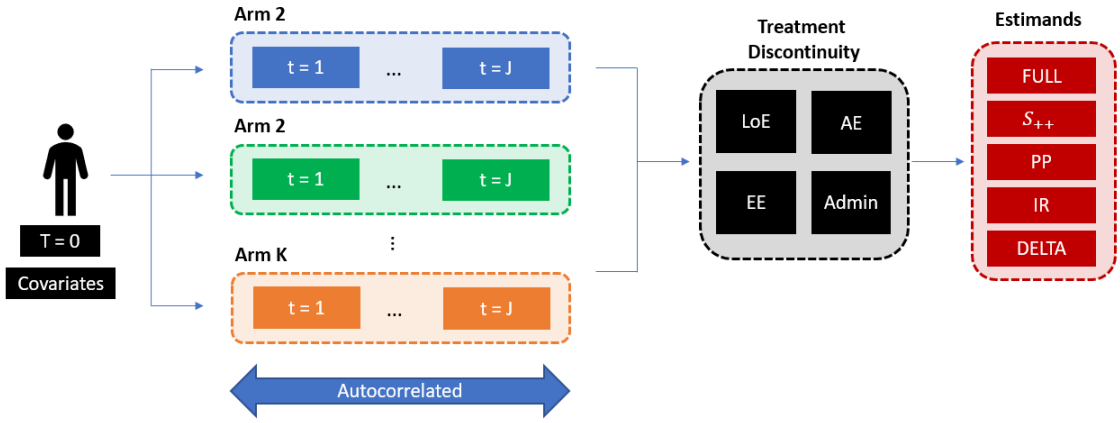

Figure 5 Illustration of the simulation process in CITIES.

The combination of the assigned treatment and the consequent outcomes constitute the *realized* data for the patient. These data would be observed in the hypothetical case of no ICEs or treatment discontinuations in the trial. The realized data are distinct from the *observed* data, which correspond to the outcomes that are actually observed in the trial in the presence of ICEs or treatment discontinuations. In case the user is interested in knowing everything that hypothetically would have happened to the patient under the treatment (i.e., the realized data), CITIES has the capability of saving the simulated potential outcomes under the study treatment in the user's local drive. This feature of generating and saving the potential outcomes for ICEs on each treatment for each patient distinguishes CITIES from existing clinical trial simulators. It

yields a more complete picture of the outcomes under both treatments, and enables assessments of the true treatment effects under different ICE patterns.

2.6 Mean Treatment Profiles in CITIES

There are five different summaries of the mean treatment profile for the data simulated by CITIES. First are the mean treatment response profiles for the potential outcomes $Y_{ij}(t)$, which we refer to as the "Full" mean profile. This summary does not consider treatment discontinuation, i.e., it is a potential profile under no ICE. This summary can serve as a reference in comparing the other treatment response profiles that account for ICEs.

The second set of summaries are the mean treatment response profiles at each time point as they would be observed in the clinical trial. We refer to these summaries as the per-protocol (PP) mean profiles. It is important to recognize that these mean profiles are generated from non-randomized groups of patients. Statisticians often model these data via the Mixed Model with Repeat Measurements (MMRM) and report least square means (LSMeans) so that the visual displays of the treatment profiles are more meaningful.

The third summary of the treatment profiles compares the potential outcomes across treatment arms for those patients that are able to adhere to either of the treatment assignments for the entire study duration. This is a natural causal estimand in clinical trials with ICEs that captures the direct effect of the treatment. We adopt the notation from Qu et al. (2020) and refer to this estimand as the adherers average causal effect for the S_{++} principal stratum, denoted by AdACE(S_{++}). Qu et al. (2021) provide consistent estimators under specific assumptions for the treatment effect in this principal stratum. CITIES adopts the same definition of a causal estimand as that given by the ICH (2019), which is "estimated by comparing the outcomes in a group of subjects on the treatment to those in a similar group of subjects on the control". Estimating the treatment effect in this principal stratum corresponds to answering the question: "For those patients who can adhere to both treatment and control, does the experimental treatment yield greater efficacy compared to the control treatment that is already on the market?".

The fourth and fifth estimands fall under treatment policy assumptions that can be used to produce a complete response profile for each patient, should the user opt for the Immediate Reference or Delta Adjustment, as illustrated in Section 2.2.

Once the "Submit" action button is clicked, a progress bar is provided in the bottom right of the window to indicate the iteration of the simulator. Figure 6 illustrates the five types of estimands in the Average Treatment Effects tab in CITIES and table 1 summarizes their estimates for the corresponding input parameters. At the final time of 144 (bottom row) in this simulated trial, estimates for the treatment policy estimands (IR & Delta) are 0.72 and 0.87, which are understandably lower than the PP estimand due to their conservative assumption of post study treatment discontinuation data. We note that the estimates of the causal estimands are at least 0.1 higher that the PP estimand with tighter standard errors. Although this trial was simulated under conditions where simulation settings were the same for both arms, barring the treatment means and probability of observing at least one AE in each arm, the estimands yielded different estimates without directly perturbing the means within each principal strata.

The final visualization provided in CITIES summarizes the percentage of treatment discontinuation at each time point and the corresponding ICEs behind the discontinuations. An illustration is provided in Figure 7. This visualization can be particularly useful when identifying input parameter values that yield simulated data resembling real clinical trials when the original data are not available.

3 Case Studies Involving Real-Life Clinical Trials

3.1 Diabetes Clinical Trial

We first demonstrate the application of CITIES to simulate data resembling the results of a Phase III Diabetes clinical trial described in Stenlöf et al. (2013). This clinical trial had a duration of 26 weeks, involved

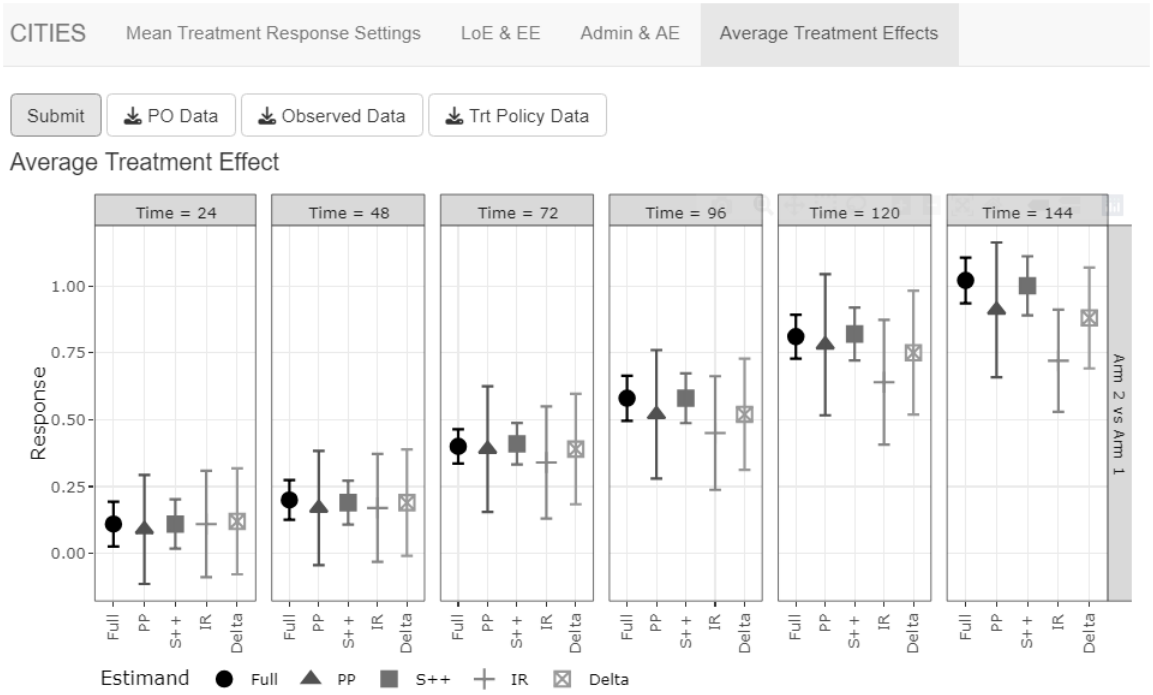

Figure 6 Treatment effect profiles for the five estimands in CITIES based on Figures 1-4.

Table 1 Treatment effect estimates for all five estimands with their accompanying standard errors across time as obtained by CITIES based on input parameters in Figures 1-4. Here, a delta value of 1 is used for Arm 2.

Time	FULL	PP	AdACE(S_{++})	IR	DELTA
24	0.11 (0.08)	0.1 (0.22)	0.11 (0.09)	0.12 (0.21)	0.13 (0.21)
48	0.2 (0.07)	0.17 (0.19)	0.19 (0.08)	0.16 (0.19)	0.19 (0.19)
72	0.4 (0.06)	0.38 (0.21)	0.41 (0.08)	0.34 (0.21)	0.39 (0.21)
96	0.58 (0.08)	0.52 (021)	0.58 (0.09)	0.45 (0.19)	0.53 (0.19)
120	0.81 (0.08)	0.76 (0.23)	0.82 (0.1)	0.62 (0.23)	0.73 (0.22)
144	1.02 (0.09)	0.9 (0.23)	1 (0.11)	0.72 (0.19)	0.87 (0.18)

double-blind, randomized treatment assignment. Each of the 584 subjects in the trial received either 100 mg or 300 mg of the experimental treatment, or placebo once daily. In this particular demonstration of CITIES we consider simulating clinical trials that only involve the 100 mg dose and placebo. The primary endpoint in the clinical trial was the change from baseline in HbA1c at week 26.

We specified input parameters for the Mean Treatment Response Settings tab in CITIES (Figure S1) based on the raw means and standard errors of the response that were given in the reported Figure 2 of Stenlöf et al. (2013) using the xyscan software (Ullrich, 2022), a tool to extract data points, such as numeric values, from a plot. We specified an AR(1) parameter $\rho = 0.5$ as a starting point, since these values are rarely published. The input parameter values for both the LoE & EE (Figure S3) and the probability of discontinuing due to administrative reasons were selected based on the reported study flow diagram, or Figure 1, in (Stenlöf et al., 2013), while the parameters for the AE Poisson process was based on the reported Table 4 in the published trial (summarized in Figure S4). The probability of discontinuing due to

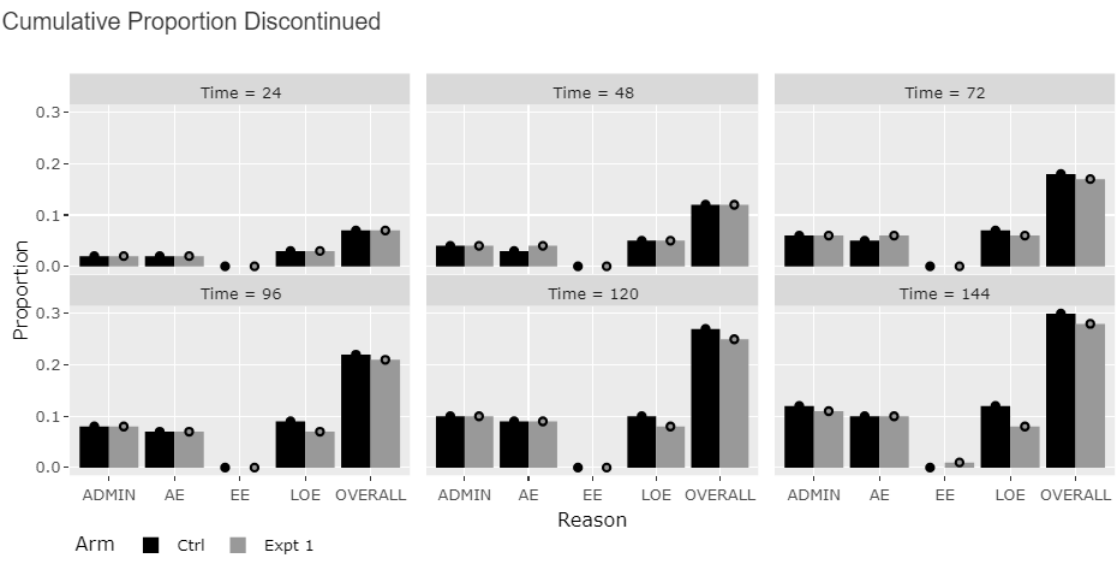

Figure 7 Percentages of treatment discontinuations, and the corresponding ICEs in CITIES based on Figures 1, 3, 4.

EE is set to 0 in this case because no ICEs due to EE were observed in this clinical study. In addition, we compared the percentage of observed treatment discontinuation by arm from our simulated data (Figure S6 and Table 2) with the reported Figure 1 in (Stenlöf et al., 2013) to validate the appropriateness of our specified input parameter values.

Based on the datasets simulated via CITIES under our specified input parameter values, estimates for all four estimands are around -0.78 (Table 3). These estimates are similar to the reported treatment effect given in (Stenlöf et al., 2013), which is -0.77 as can be seen from the reported Graph A in Figure 2. The trends of the endpoints provided by CITIES, and illustrated in Figure S5, resembles results for the mITT population from the original study. This case study serves to demonstrate how CITIES facilitates the simulation of data resembling real-life clinical trials, and how the resulting simulated data can be used to understand estimands and evaluate estimators more effectively.

Table 2 Summary information for both the Diabetes trial and simulated data via CITIES.

	Arm 1 (Control)		Arm 2 (Experimental)	
	Reported Values	CITIES	Reported Values	CITIES
Number of subjects enrolled in trial	192	192	195	195
Proportion discontinue due to AE	0.01	0.01	0.02	0.02
Proportion discontinue due to LoE	0.02	0.03	0.005	0.01
Proportion discontinue due to EE	0	0	0	0
Proportion discontinue due to Admin	0.135	0.12	0.1	0.08
Total discontinuations	0.165	0.16	0.125	0.11

3.2 Alzheimer's Clinical Trial

We next demonstrate the application of CITIES to simulate data resembling the reported results of a Phase II clinical trial in early-stage Alzheimer's disease (Mintun et al., 2021). This trial had 257 enrolled patients,

Table 3 Treatment effect estimates for all four estimands with their accompanying standard errors across time as obtained by CITIES based on input parameters in Figures S1, S3, S4.

Time	FULL	PP	AdACE(S_{++})
6	-0.56 (0.05)	-0.56 (0.08)	-0.55 (0.05)
12	-0.73 (0.06)	-0.73 (0.09)	-0.71 (0.06)
18	-0.76 (0.06)	-0.77 (0.09)	-0.75 (0.06)
26	-0.78 (0.07)	-0.78 (0.09)	-0.77 (0.07)

with 131 assigned to receive the active treatment and the remaining 126 assigned to the placebo group. All patients had early symptomatic Alzheimer's disease, with amyloid and tau deposits on their Positron-Emission Tomography (PET). The active treatment was administered intravenously according to a carefully designed titration scheme, with 700 mg for the first three doses and 1400 mg for the remaining doses, with an allowance for downward titration or switching to placebo if amyloid plaques achieved sufficiently low levels in the brain. The primary endpoint was the change in the Integrated Alzheimer's Disease Rating Scale (iADRS) from baseline. The iADRS is a score that ranges from 0 to 144. Lower values of iADRS indicate greater functional and cognitive impairment.

We specified input parameter values for the Mean Treatment Response Settings tab in CITIES as in Figure S7 based on the reported values in Figure 2A from (Mintun et al., 2021), with the baseline measurement for all patients on both arms set to 106. As before, we used the xyscan software (Ullrich, 2022) to estimate the LS mean changes from baseline for the placebo and experimental treatment at weeks 12, 24, 36, 52 and 76, and used the estimates as input parameters. The input parameter values for the LoE & EE and Admin & AE tabs (Figures S9 and S10, respectively) were specified based on the values given in the reported Figure 1 in (Mintun et al., 2021). As in the previous case study, we compared the percentage of treatment discontinuation from our simulated data (Figure S12 and Table 4) with the reported Figure 1 in Mintun et al. (2021) to validate the appropriateness of our specified input parameter values. Probability of discontinuing due to administrative reasons was extracted from the reported Figure 1 in Mintun et al. (2021) and the parameters for the AE poisson process was populated using the reported Table 2 in the published trial.

We conclude from the simulated datasets in CITIES that treatment effect estimates for all four estimands are approximately 3.2 at week 76 (Table 5). This is similar to the reported treatment effect of 3.2 in the original analysis in (Mintun et al., 2021), which was obtained from a mixed model for repeated measures (MMRM) analysis on the mITT population consisting of all randomized subjects who received at least one dose of the study treatment. Details of the MMRM analysis can be found in the paper. This case study further demonstrates how CITIES provides an interactive and transparent platform in which users can simulate realistic clinical trials with ICEs based on summary statistics of real-life clinical trials to understand different estimands and estimators.

Table 4 Summary information for both the Alzheimer's trial and simulated data via CITIES.

	Arm 1 (Control)		Arm 2 (Experimental)	
	Reported Values	CITIES	Reported Values	CITIES
Number of subjects enrolled in trial	126	126	131	131
Proportion discontinue due to AE	0.072	0.07	0.305	0.3
Proportion discontinue due to LoE	0	0	0	0
Proportion discontinue due to EE	0	0	0	0
Proportion discontinue due to Admin	0.198	0.17	0.069	0.06
Total discontinuations	0.27	0.24	0.374	0.36

Table 5 Treatment effect estimates for all four estimands with their accompanying standard errors across time as obtained by CITIES based on input parameters in Figures S7, S9, S10.

Time	FULL	PP	AdACE(S_{++})
12	0.47 (0.84)	0.33 (1.23)	0.48 (0.9)
24	1.23 (0.96)	1.14 (1.53)	1.24 (1.05)
36	2.47 (0.97)	2.42 (1.53)	2.45 (1.09)
52	3.71 (0.85)	3.65 (1.28)	3.68 (1.08)
64	3.29 (0.82)	3.22 (1.45)	3.25 (1.11)
76	3.14 (0.88)	3.14 (1.57)	3.14 (1.16)

4 Discussion

As is true with all fields of science, communicating ideas and sharing information that are embodied in a model can be challenging. CITIES reduces this challenge by allowing users to share and present information dynamically with ease and transparency. CITIES has four innovations and major advantages over existing clinical trial simulator software. First, it generates efficacy data under the potential outcomes framework with baseline covariates that are consistent with the ICH E9(R1) definition of a treatment effect. Second, it generates potential outcomes for ICEs. Third, it incorporates a set of realistic study treatment discontinuation models for the primary set of ICEs that are frequently observed in real-life clinical trials, ultimately generating missing data through an MNAR pattern as the missingness is a function of the unobserved outcomes themselves (Section 2.3). Fourth, CITIES can handle more realistic longitudinal settings where patients can discontinue at different visits during the trial. This combination of the four unique elements can ultimately better serve to simulate realistic clinical trial datasets for the purpose of developing, implementing, and comparing different inferential methods for different estimands of interest in RCTs with ICEs.

While CITIES has considerable novelty and flexibility in this first version, we plan to further update CITIES in the future so that the subsequent versions would allow even greater flexibility but retain a controlled environment in which methodologies can be compared under common scenarios for their utilities. Specifically, future versions will provide functionality in simulating longitudinal binary data (not simply dichotomized continuous outcomes) and time to event outcomes. In addition, we intend to integrate the baseline covariate and outcome measurements into modelling the AEs.

It is important to recognize that CITIES does not invoke any monotonicity assumptions (Frangakis and Rubin, 2002) for the generation of the potential outcomes and ICEs. For example, a simulated data set may contain patients that discontinue due to an AE on the treatment but complete the trial under the placebo, patients that discontinue due to LoE under placebo but complete the trial under treatment, as well as patients that discontinue due to AE on treatment and discontinue due to LoE on placebo. All of these types of patients are realistic scenarios in a randomized clinical trial, and can be simulated in CITIES.

Further, we do not believe that one should think of CITIES as covering one type of estimand versus another. Instead, CITIES seeks to generate clinical trial data that is realistic enough to be meaningful for statistical methodological research related to any estimand. We are using the potential outcomes framework to simulate data for each patient on both treatments. The observed data that are the output of any simulation then depend on which treatment the patient is randomized to and their adherence, or lack thereof, based on DC reasons governed by the user input parameters.

As for principal stratification (PS), we understand there are apprehensions in being adopted as part of the mainstream for drug development and regulatory approval. There are many reasons for this, including a lack of understanding about PS, lack of easily accessible software for performing the necessary causal inferences on the PS estimands, and a lack of evidence as to their utility. It is the third issue that we seek to address with our clinical trial simulation engine. By simulating data under the potential outcomes

framework, which to the best of our knowledge is currently not done by any other clinical trial simulator, we can enable the user to assess if, when, and how the PS estimands could be useful, and to explore different estimators and their properties. We believe that the principal stratum of adherers is of fundamental interest to all stakeholders and that inferences for causal effects in this stratum can help to answer the important question: "What can I expect to happen when I take this medication as prescribed?" That question forms the basis of the adherers' average causal effect (abbreviated as "AdACE" by Qu et al. (2021)) of a treatment. CITIES enables a meaningful approach to simulate data to research such a causal estimand.

A summary that is a source of much debate is one that compares the potential outcomes for the principal stratum of patients who can adhere to the experimental treatment without regard of their adherence to the control treatment (Permutt, 2018). This estimand is denoted by AdACE(S_{*+}) (Qu et al., 2020). Sponsors have expressed interest in this estimand in the past. The difficulty with this estimand is that some patients who can adhere to the experimental treatment for the duration of the study may fail to adhere to the control treatment. In the datasets that are simulated by CITIES, a patient may be observed to adhere to the experimental treatment but could have an ICE on the control treatment, for example due to lack of efficacy (if the control is a placebo) or an adverse event (if the control is itself another active treatment). CITIES generates potential outcomes for both efficacy or potential outcomes $Y_{ij}(t)$ and adherence $A_{ij}(t)$ on both the experimental treatment and control arm as depicted in Table 6.

Table 6 Example of a subject with their potential outcomes Y and adherence A, where $A = 1$ indicates adherence and $A = 0$ indicates discontinuation of the study treatment.

Treatment	Outcome	Baseline	Visit 1	Visit 2	Visit 3	Visit 4
Experimental	Y	0	1	2	3	4
	A	1	1	1	1	1
Control	Y	0	0	~~0.5~~	~~0.5~~	~~1~~
	A	1	1	0	0	0

When $A = 0$, the patient is no longer adherent to their study medication, and therefore the response to that study treatment does not exist and that visit and subsequent visits, as indicated by the strike-through text in the table. For the value of this estimand to be defined one must define how patients who adhere to experimental but not control would behave in the presence of their non-adherence when randomised to control treatment. In the table above, what is the "true" value to simulate for Visits 2-4 on the Control arm?

One can make assumptions about the patient response on "no treatment," which is a different treatment than placebo treatment in this case, or assumptions about their response if they receive rescue medication, or assumptions about their response if they pursue any other intervention (e.g., surgery, diet and exercise, behavioral therapy, etc.). But those responses are NOT responses to the randomized study treatment. Such assumptions relate to a particular estimand using a "treatment policy" approach and measuring patient response regardless of their adherence to their randomized study treatment. But that is an estimand for the effect of initiation of study treatment (i.e., what is often referred to as an Intent-to-Treat estimand) and not an estimand for the effect of taking the study treatment (i.e., the direct treatment effect). Consequently, this version of CITIES does not summarize the mean profiles for the S_{*+} principal stratum.

Ultimately, CITIES can enable new types of insightful research to be conducted for the problem of estimating treatment effects in clinical trials in the presence of ICEs that result in discontinuation of study treatments.

Conflict of Interest *The authors have declared no conflict of interest.*

Acknowledgements The authors thank the Editors, the Guest Editors, the Associate Editor, and the referees for their very constructive comments and valuable suggestions that substantially improved this paper and the CITIES simulator.

References

Akacha, M., Bretz, F., Ohlssen, D., Rosenkranz, G., and Schmidli, H. (2017a). Estimands and their role in clinical trials. *Statistics in Biopharmaceutical Research*, 9(3):268–271.

Akacha, M., Bretz, F., and Ruberg, S. (2017b). Estimands in clinical trials – broadening the perspective. *Statistics in Medicine*, 36(1):5–19.

Benda, N., Branson, M., Maurer, W., and Friede, T. (2010). Aspects of modernizing drug development using clinical scenario planning and evaluation. *Therapeutic Innovation & Regulatory Science*, 44(3):299–315.

Boulesteix, A., Binder, H., Abrahamowicz, M., and Sauerbrei, W. (2018). On the necessity and design of studies comparing statistical methods. *Biometrical Journal*, 60(1):216–218.

Carpenter, J. R., Roger, J. H., Cro, S., and Kenward, M. G. (2014). Response to comments by seaman et al. on "analysis of longitudinal trials with protocol deviation: A framework for relevant, accessible assumptions, and inference via multiple imputation," Journal of Biopharmaceutical Statistics 23:1352-1371. *Journal of Biopharmaceutical Statistics*, 24(6):1363–1369.

Davidson, M. H., Hauptman, J., DiGirolamo, M., Foreyt, J. P., Halsted, C. H., Heber, D., Heimburger, D. C., Lucas, C. P., Robbins, D. C., Chung, J., and Heymsfield, S. B. (1999). Weight control and risk factor reduction in obese subjects treated for 2 years with orlistat: A randomized controlled trial. *JAMA : the journal of the American Medical Association*, 281(3):235–242.

EMA (2018a). Clinical investigation of medicines for the treatment of alzheimer's disease. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicines-treatment-alzheimers-disease-revision-2_en.pdf.

EMA (2018b). Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-clinical-investigation-medicinal-products-treatment-prevention-diabetes-mellitus_en.pdf.

Fisher, R. A. (1971). *The Design of Experiments*. Hafner Publishing Company, 9 edition.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

Friede, T., Nicholas, R., Stallard, N., Todd, S., Parsons, N., Valdés-Márquez, E., and Chataway, J. (2010). Refinement of the clinical scenario evaluation framework for assessment of competing development strategies with an application to multiple sclerosis. *Therapeutic Innovation & Regulatory Science*, 44(6):713–718.

Grayling, M. J. and Wason, J. M. (2020). A web application for the design of multi-arm clinical trials. *BMC Cancer*, 20(1):80–80.

Hariton, E. and Locascio, J. J. (2018). Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. *BJOG : an international journal of obstetrics and gynaecology*, 125(13):1716–1716.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

ICH (2019). E9(r1) statistical principles for clinical trials: Addendum: Estimands and sensitivity analysis in clinical trials. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 1 edition.

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97(10):2177–2189.

Karanevich, A., Meier, R., Graw, S., McGlothlin, A., and Gajewski, B. (2021). Optimizing sample size allocation and power in a bayesian two-stage drop-the-losers design. *The American Statistician*, 75(1):66–75.

Karges, B., Rosenbauer, J., Kapellen, T., Wagner, V. M., Schober, E., Karges, W., and Holl, R. W. (2014). Hemoglobin a1c levels and risk of severe hypoglycemia in children and young adults with type 1 diabetes from germany and austria: A trend analysis in a cohort of 37,539 patients between 1995 and 2012. *PLoS Medicine*, 11(10):e1001742–e1001742.

Lipkovich, I., Ratitch, B., and Mallinckrodt, C. H. (2020). Causal inference and estimands in clinical trials. *Statistics in Biopharmaceutical rRsearch*, 12(1):54–67.

Mintun, M. A., Lo, A. C., Duggan Evans, C., Wessels, A. M., Ardayfio, P. A., Andersen, S. W., Shcherbinin, S., Sparks, J., Sims, J. R., Brys, M., Apostolova, L. G., Salloway, S. P., and Skovronsky, D. M. (2021). Donanemab in early alzheimer's disease. *The New England Journal of Medicine*, 384(18):1691–1704.

Paux, G. and Dmitrieniko, A. (2019). *Mediana: An R Package for Clinical Trial simulations*.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Pearl, J. (2013). Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*.

Permutt, T. (2018). Effects in adherent subjects. *Statistics in Biopharmaceutical Research*, 10(3):233–235.

Puig, L. (2015). Pasi90 response: the new standard in therapeutic efficacy for psoriasis. *Journal of the European Academy of Dermatology and Venereology*, 29(4):645–648.

Qu, Y., Fu, H., Luo, J., and Ruberg, S. J. (2020). A general framework for treatment effect estimators considering patient adherence. *Statistics in Biopharmaceutical Research*, 12(1):1–18.

Qu, Y. and Lipkovich, I. (2021). Implementation of ICH E9 (R1): A few points learned during the COVID-19 pandemic. *Therapeutic Innovation & Regulatory Science*, 55(5):984–988.

Qu, Y., Luo, J., and Ruberg, S. J. (2021). Implementation of tripartite estimands using adherence causal estimators under the causal inference framework. *Pharmaceutical Statistics : The Journal of The Pharmaceutical Industry*, 20(1):55–67.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Sofrygin, O., van der Laan, M. J., and Neugebauer, R. (2017). simcausal R package: Conducting transparent and reproducible simulation studies of causal effect estimation with complex longitudinal data. *Journal of Statistical Software*, 81(2):1–47.

Splawa-Neyman, J. (1935). Statistical problems in agricultural experimentation (with discussion). *Journal of the Royal Statistical Society, Series B*, pages 107 – 180.

Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465 – 472.

Stenlöf, K., Cefalu, W. T., Kim, K.-A., Alba, M., Usiskin, K., Tong, C., Canovatchel, W., and Meininger, G. (2013). Efficacy and safety of canagliflozin monotherapy in subjects with type 2 diabetes mellitus inadequately controlled with diet and exercise. *Diabetes, Obesity & Metabolism*, 15(4):372–382.

Thorlund, K., Golchi, S., Haggstrom, J., and Mills, E. (2019). Highly efficient clinical trials simulator (hect): Software application for planning and simulating platform adaptive trials. *Gates Open Research*, 3:780–780.

Ullrich, T. (2022). xyscan. https://www.star.bnl.gov/public/comp/vis/Thomas/xyscanDistributionPage/. Accessed: 2022-02-24.

Verdile, G., Fuller, S., Atwood, C. S., Laws, S. M., Gandy, S. E., and Martins, R. N. (2004). The role of beta amyloid in alzheimer's disease: still a cause of everything or the only one who got caught? *Pharmacological Research*, 50(4):397–409.

Wojciechowski, J., Hopkins, A., and Upton, R. (2015). Interactive pharmacometric applications using r and the shiny package: Interactive pharmacometric applications with shiny. *CPT: Pharmacometrics and Systems Pharmacology*, 4(3):146–159.

