



US 20240266008A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2024/0266008 A1**  
Miller et al. (43) **Pub. Date: Aug. 8, 2024**

(54) **SYSTEMS AND METHODS FOR DESIGNING AUGMENTED RANDOMIZED TRIALS**

(52) **U.S. Cl.**  
CPC ..... *G16H 10/20* (2018.01); *A61B 5/4848* (2013.01); *G16H 50/70* (2018.01)

(71) Applicant: **Unlearn.AI, Inc.**, San Francisco, CA (US)

(57) **ABSTRACT**

(72) Inventors: **David Putnam Miller**, Albany, CA (US); **Jonathan Ryan Walsh**, El Cerrito, CA (US)

Systems and methods for designing random control trials in accordance with embodiments of the invention are illustrated. One embodiment includes a method for designing a target random control trial. The method includes steps for generating a set of prognostic scores for a set of samples. The set of prognostic scores includes prognostic scores at each of several points in time for each sample. The method includes assessing discrimination and bias metrics for the set of generative models based on a set of outcomes for the set of samples that includes outcomes at each of several points in time for each sample. The method includes determining a set of target trial parameters for a randomized control trial (RCT) based on the assessed discrimination and bias metrics, generating result data using the set of generative models, and determining treatment effects for the RCT using the generated result data.

(73) Assignee: **Unlearn.AI, Inc.**, San Francisco, CA (US)

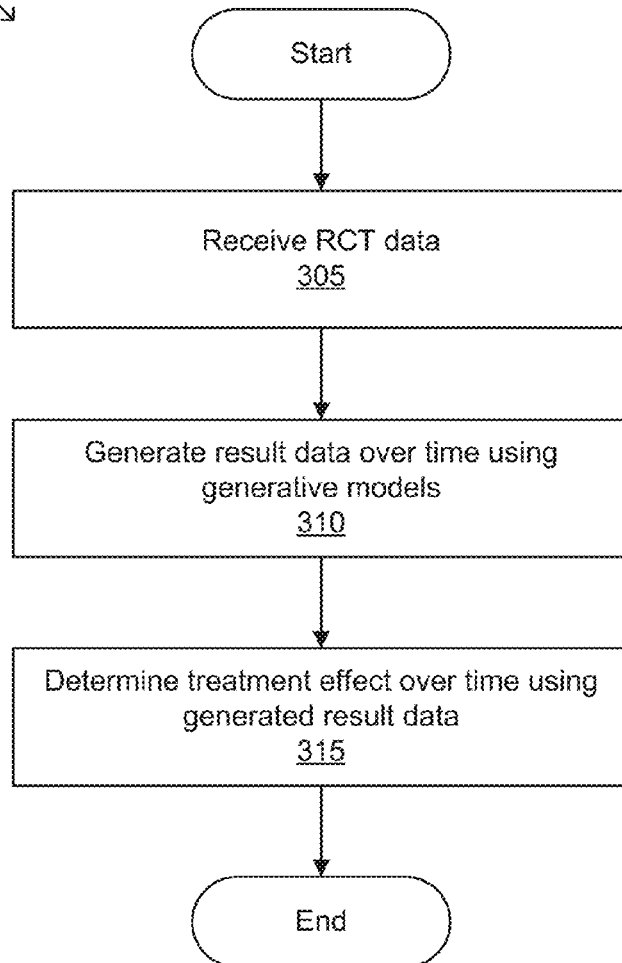
(21) Appl. No.: **18/163,157**

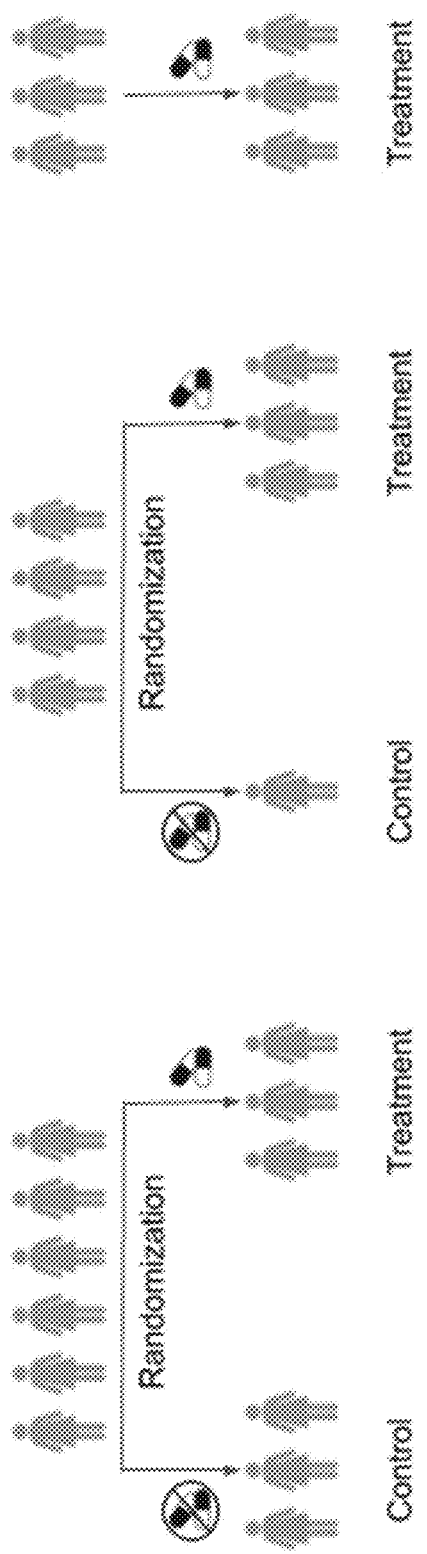
(22) Filed: **Feb. 1, 2023**

**Publication Classification**

(51) **Int. Cl.**  
*G16H 10/20* (2006.01)  
*A61B 5/00* (2006.01)  
*G16H 50/70* (2006.01)

300



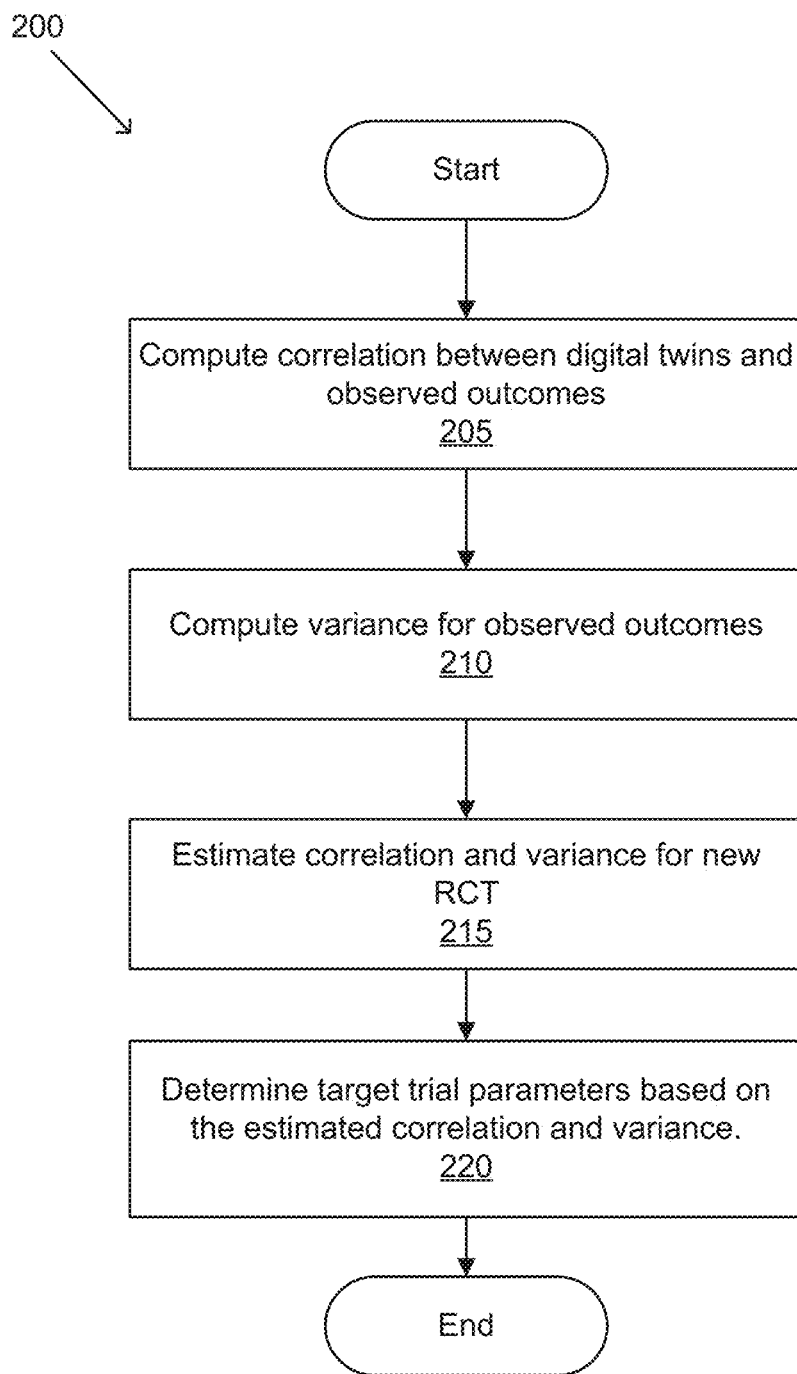


105

110

115

FIG. 1



*FIG. 2*

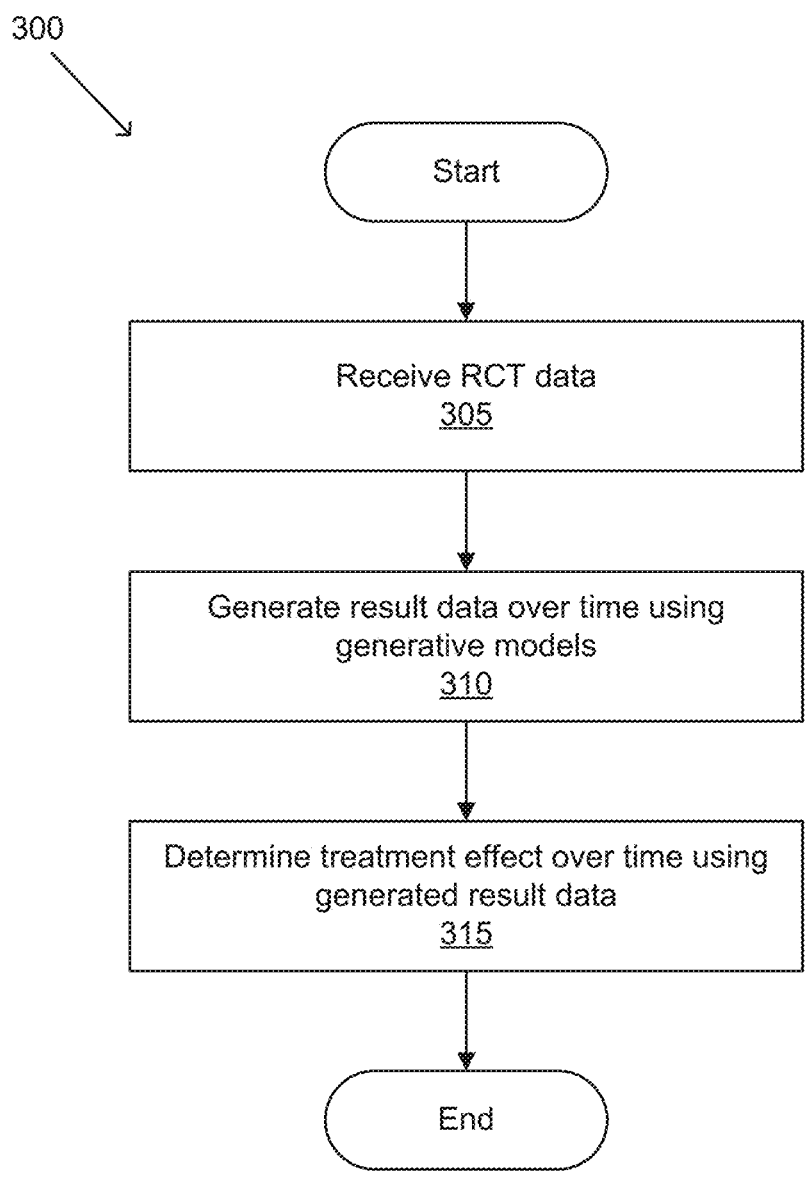


FIG. 3

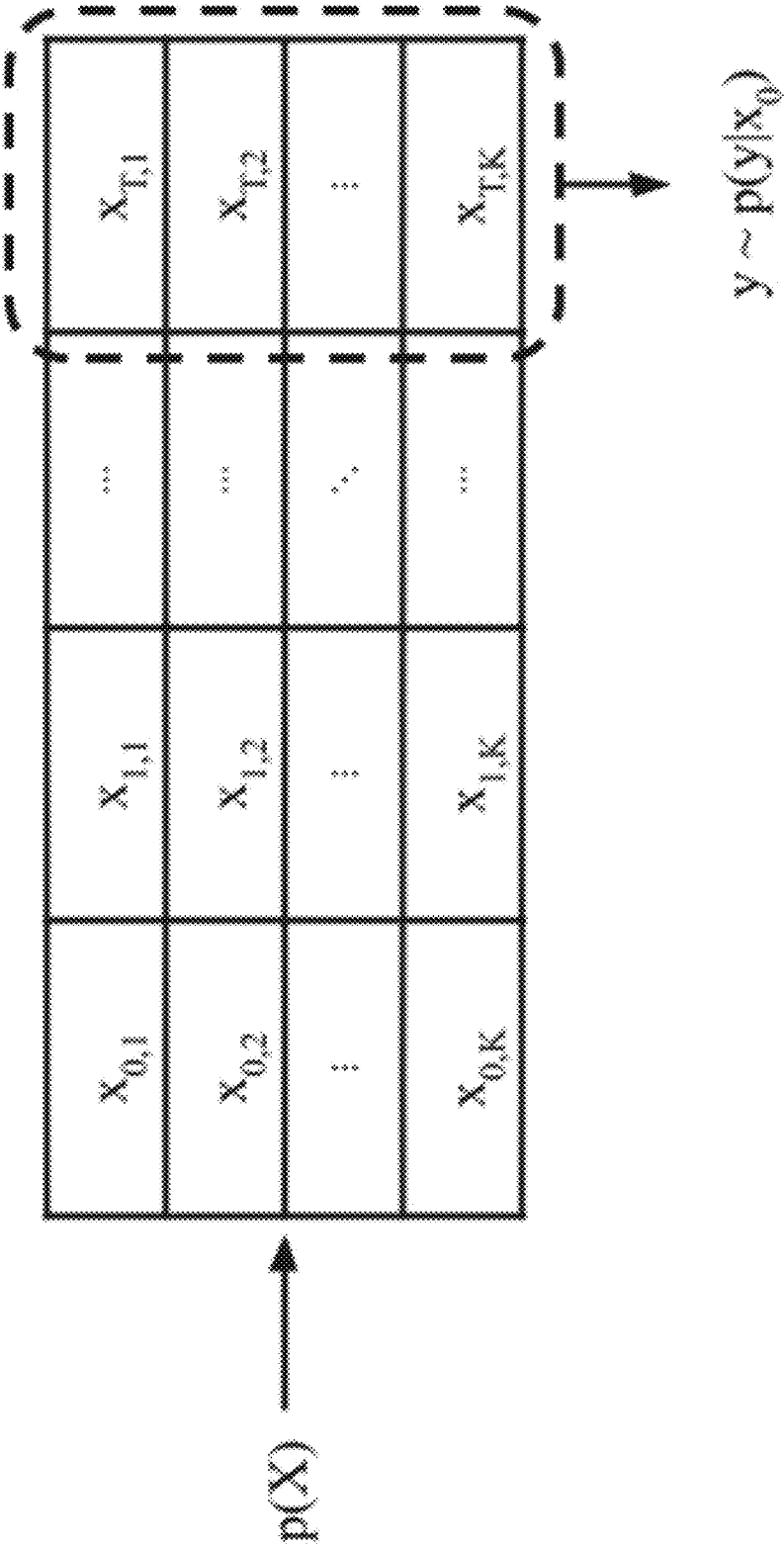


FIG. 4

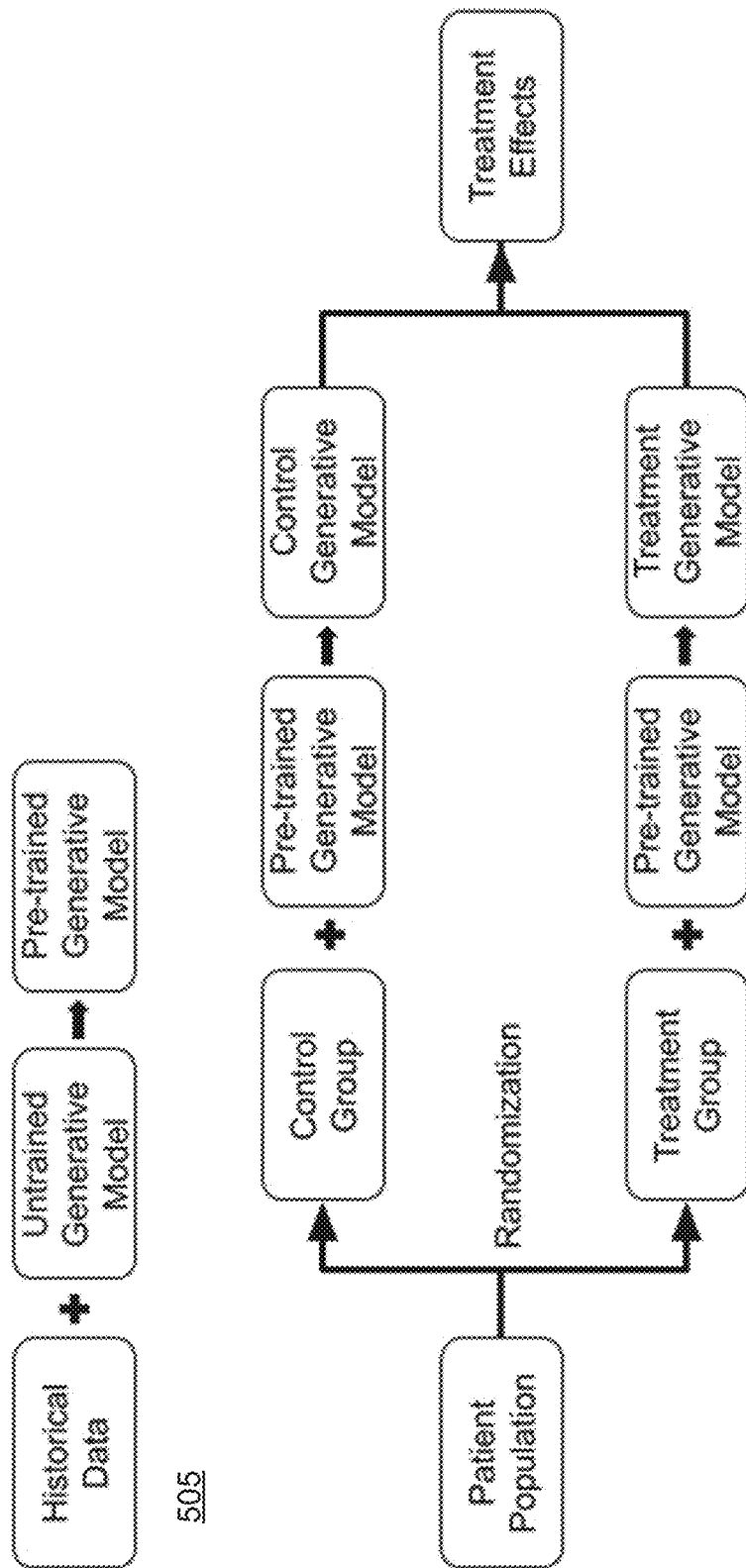


FIG. 5

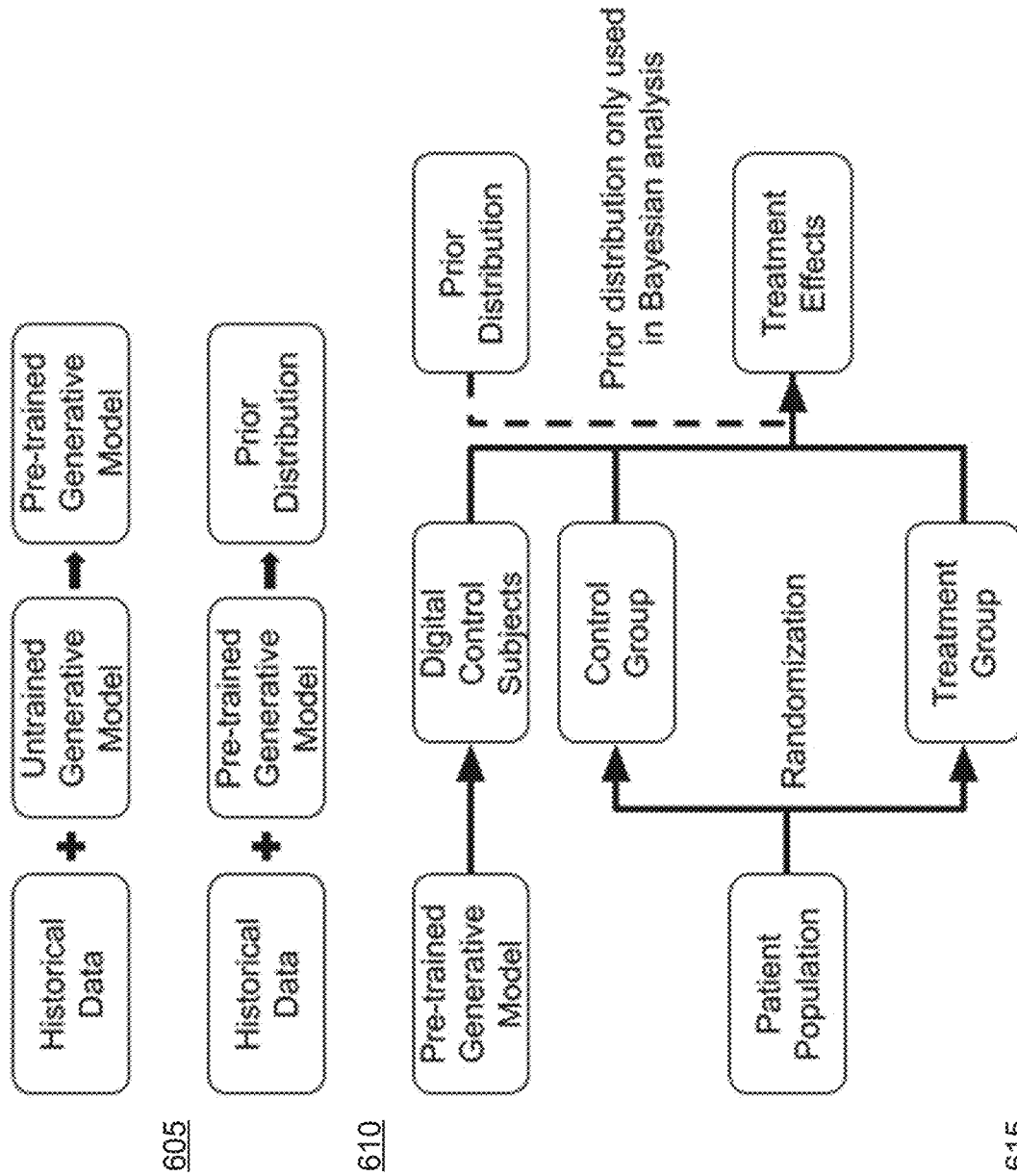


FIG. 6

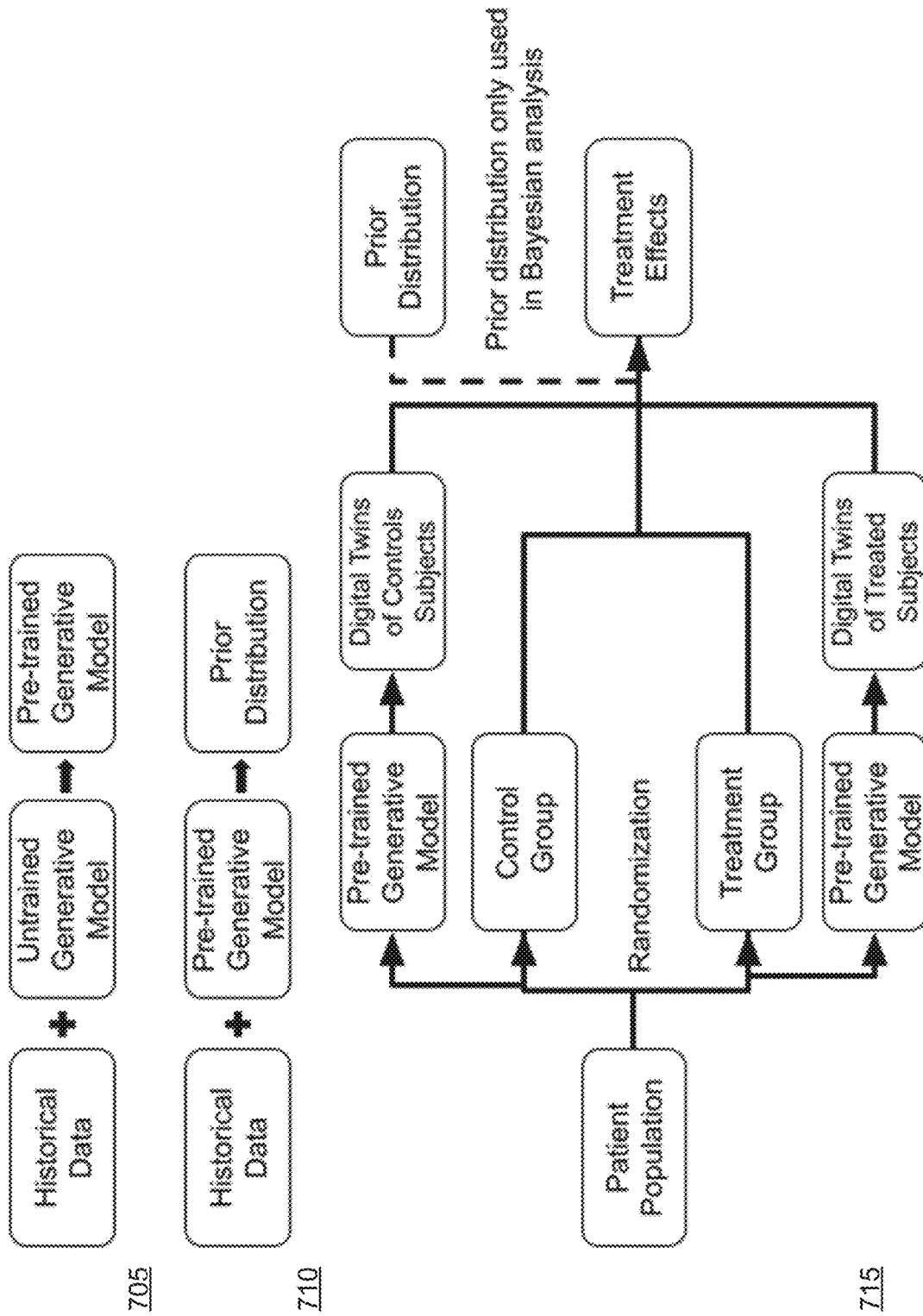


FIG. 7



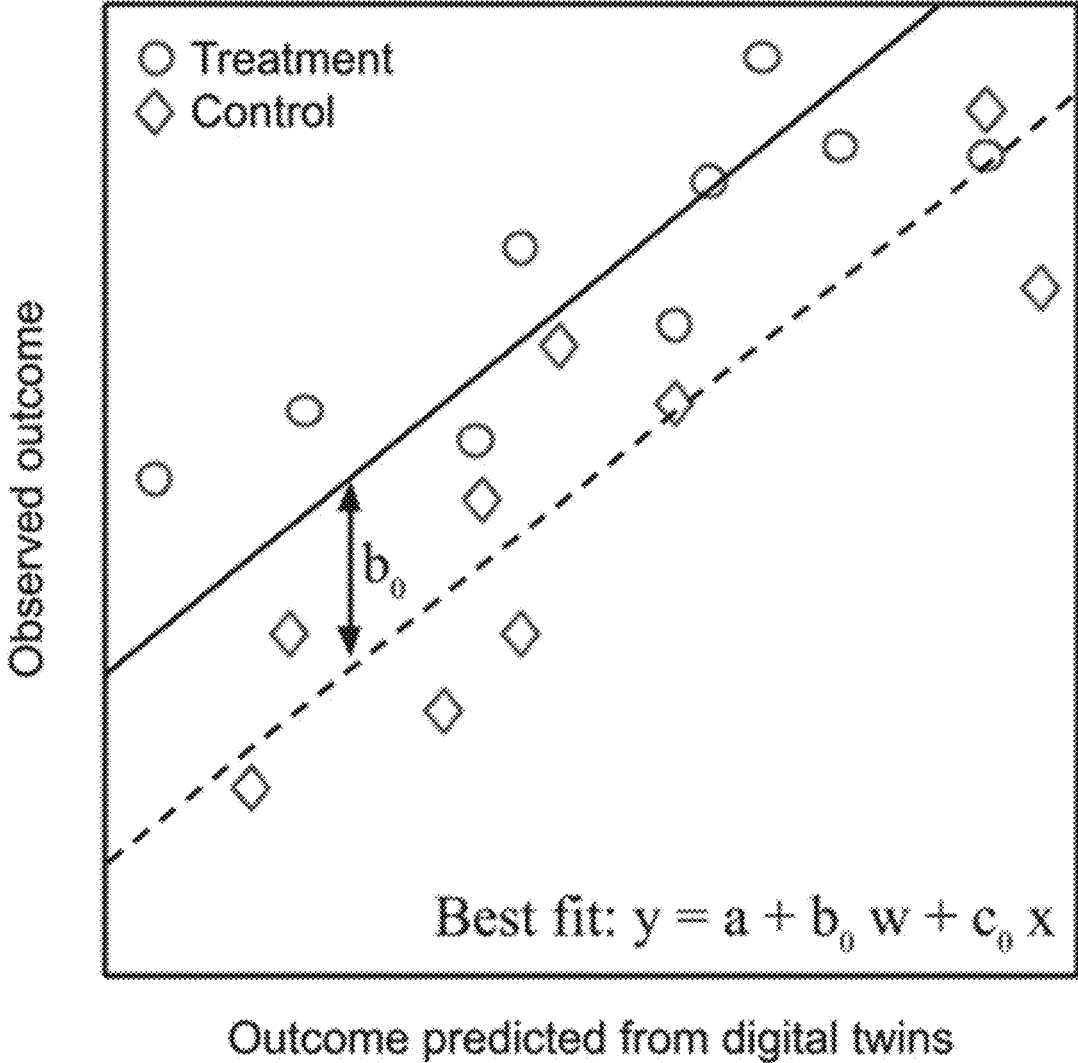


FIG. 8

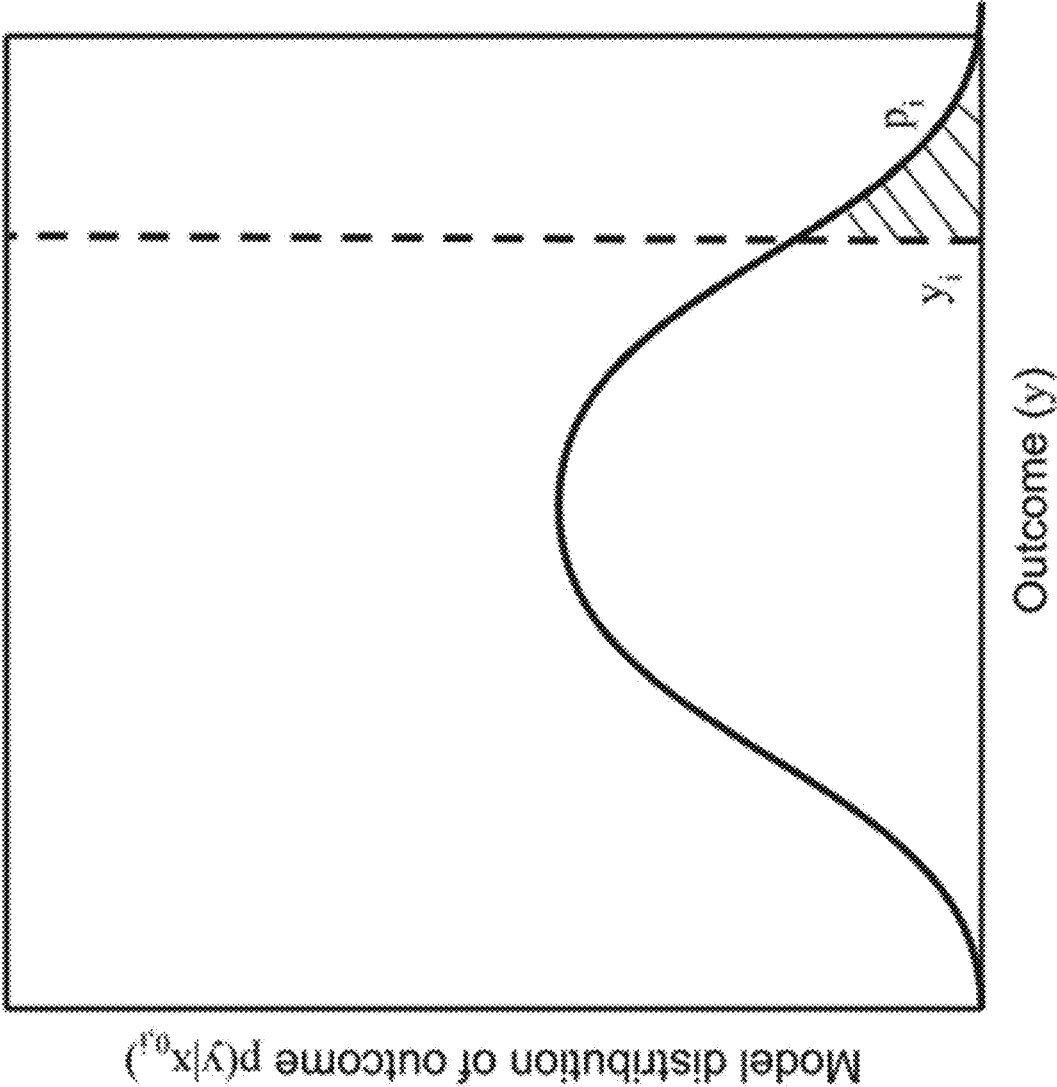


FIG. 9

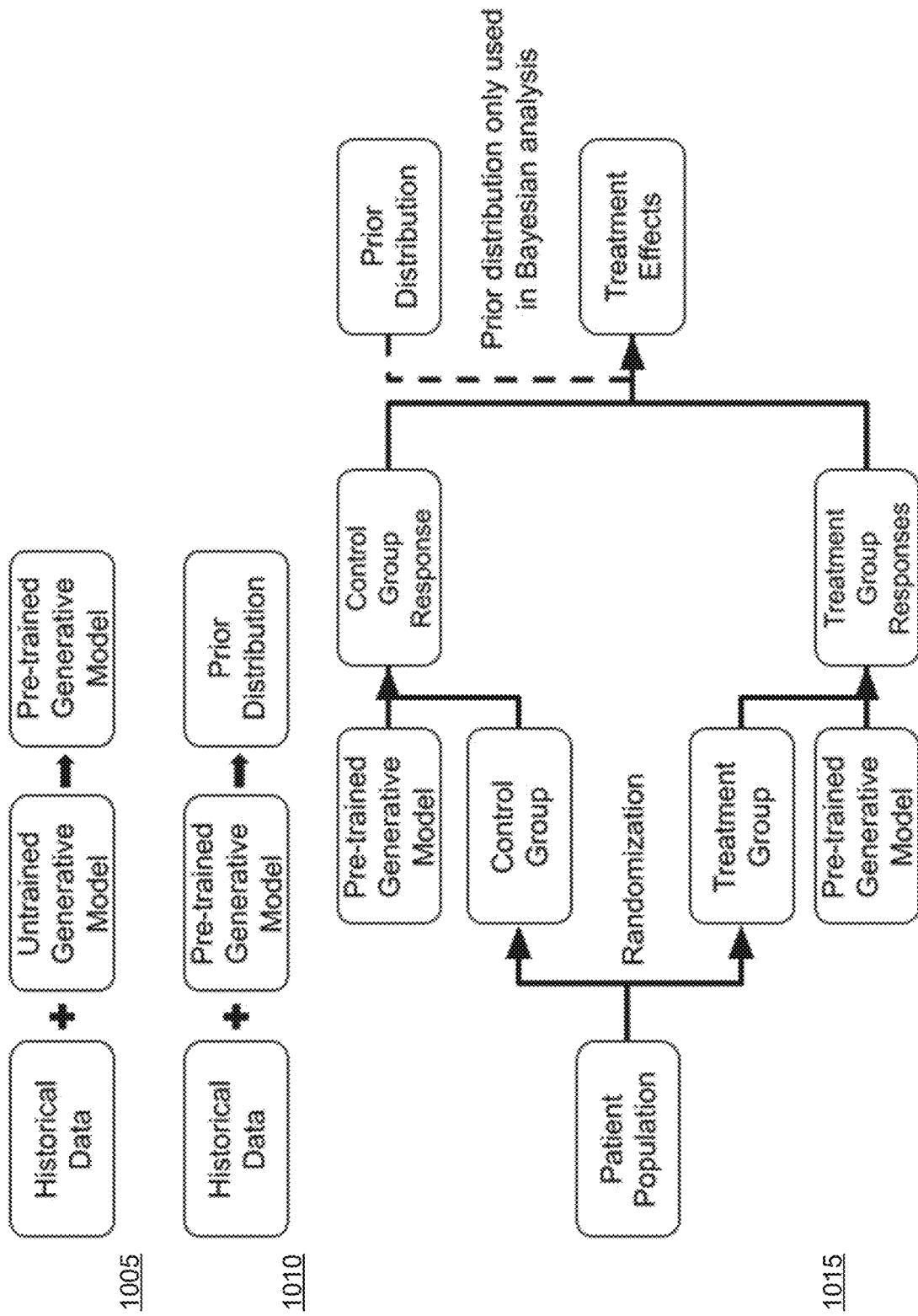


FIG. 10

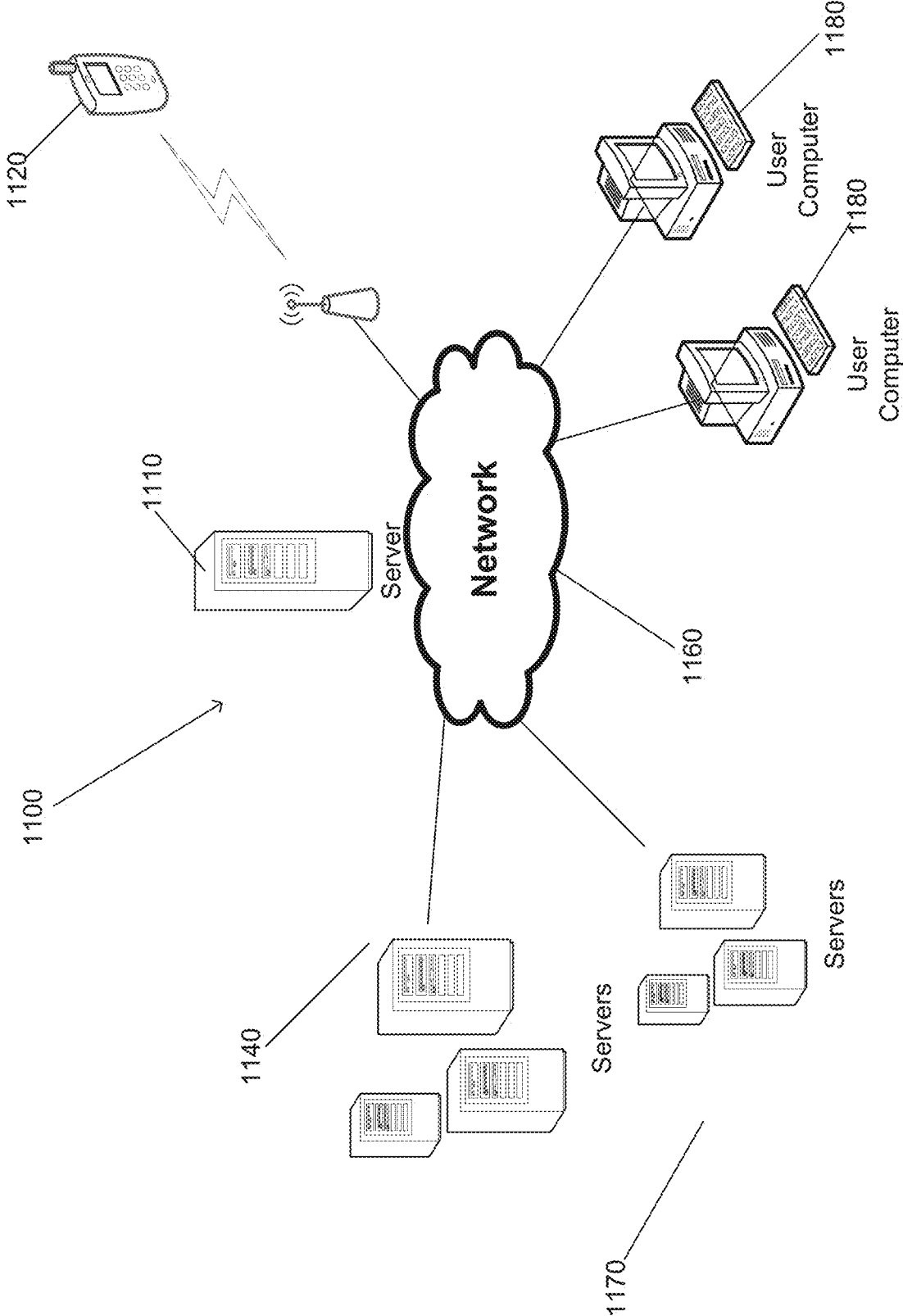


FIG. 11

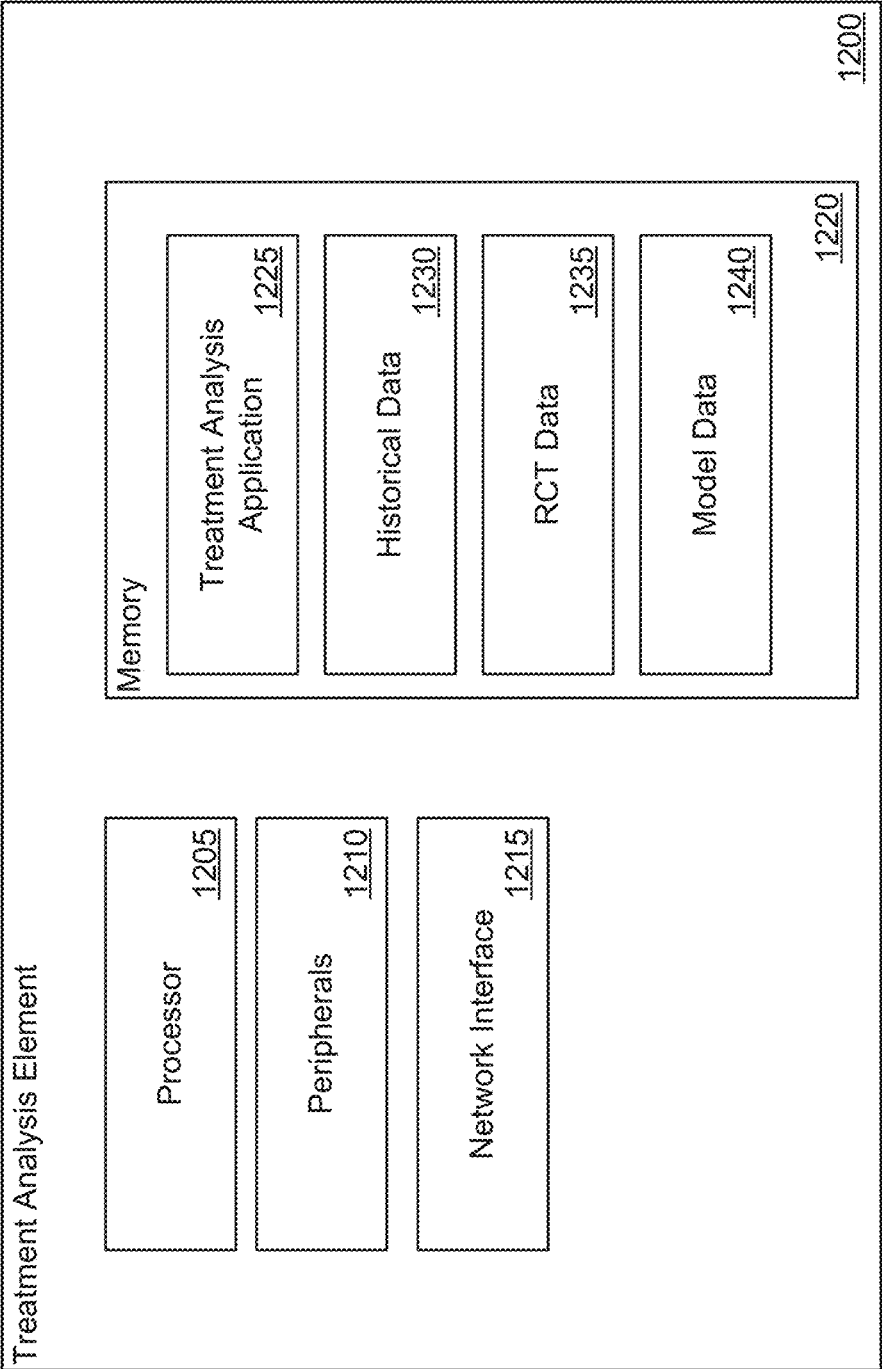


FIG. 12

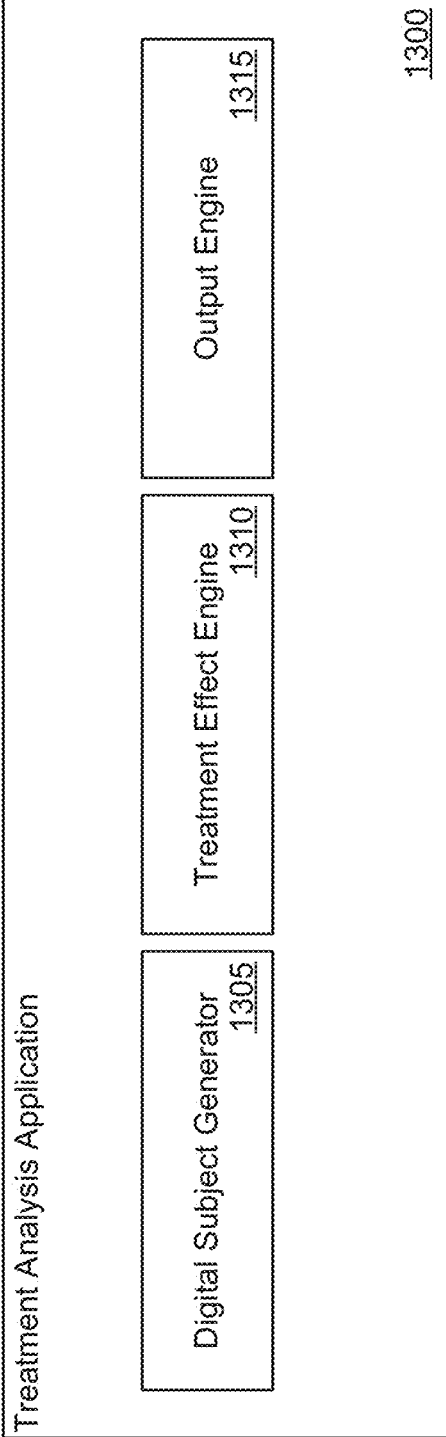


FIG. 13

## SYSTEMS AND METHODS FOR DESIGNING AUGMENTED RANDOMIZED TRIALS

### FIELD OF THE INVENTION

[0001] The present invention generally relates to supplementing data for analysis and, more specifically, to using generative models to supplement data for analysis.

### BACKGROUND

[0002] Randomized Controlled Trials (RCTs) are commonly used to assess the safety and efficacy of new treatments, such as drugs and medical devices. In an RCT, a group of subjects with particular characteristics are randomly assigned to one or more experimental groups receiving new treatments or to a control group receiving a comparative treatment (e.g., a placebo), and the outcomes from these groups are compared in order to assess the safety and efficacy of the new treatments. It is expensive, time consuming and, in some cases, unethical to recruit human subjects to participate in RCTs.

### SUMMARY OF THE INVENTION

[0003] Systems and methods for designing random control trials in accordance with embodiments of the invention are illustrated. One embodiment includes a method for designing a target random control trial. The method includes steps for generating a set of prognostic scores for a set of samples. The set of prognostic scores includes prognostic scores at each of several points in time for each sample. The method includes assessing discrimination and bias metrics for the set of generative models based on a set of outcomes for the set of samples that includes outcomes at each of several points in time for each sample. The method includes determining a set of target trial parameters for a randomized control trial (RCT) based on the assessed discrimination and bias metrics, generating result data using the set of generative models, and determining treatment effects for the RCT using the generated result data. In a further embodiment, the set of prognostic scores are generated based on subjects from a control arm of another trial.

[0004] In still another embodiment, assessing the discrimination and bias metrics includes a Pearson correlation.

[0005] In a still further embodiment, the result data includes panel data from subjects of the RCT and the generated result data includes predicted panel data for a set of one or more digital subjects, wherein the panel data describes the observed values of multiple characteristics at multiple discrete timepoints.

[0006] In yet another embodiment, the predicted panel data for the set of digital subjects is generated based on population statistics of the RCT, and the generated result data is used to supplement control arm data of the RCT data.

[0007] In a yet further embodiment, the predicted panel data for the set of digital subjects is generated based on individual characteristics of the subjects of the RCT.

[0008] In another additional embodiment, determining the treatment effects includes comparing the predicted panel data based on characteristics of a particular subject with the panel data for the particular subject from the RCT data.

[0009] In a further additional embodiment, determining the set of target trial parameters includes minimizing a total number of samples for the target random control trial.

[0010] In another embodiment again, determining the set of target trial parameters includes minimizing a number of samples for the control arm of the target random control trial.

[0011] In a further embodiment again, determining the set of target trial parameters includes minimizing a number of samples for the treatment arm of the target random control trial.

[0012] One embodiment includes a non-transitory machine readable medium containing processor instructions for designing a target random control trial, where execution of the instructions by a processor causes the processor to perform a process that comprises generating a set of prognostic scores for a set of samples, wherein the set of prognostic scores includes prognostic scores at each of several points in time for each sample of the set of samples, assessing discrimination and bias metrics for the set of generative models based on a set of outcomes for the set of samples, wherein the set of outcomes includes outcomes at each of several points in time for each sample of the set of samples, determining a set of target trial parameters for a randomized control trial (RCT) based on the assessed discrimination and bias metrics, generating result data using the set of generative models, and determining treatment effects for the RCT using the generated result data.

[0013] Additional embodiments and features are set forth in part in the description that follows, and in part will become apparent to those skilled in the art upon examination of the specification or may be learned by the practice of the invention. A further understanding of the nature and advantages of the present invention may be realized by reference to the remaining portions of the specification and the drawings, which forms a part of this disclosure.

### BRIEF DESCRIPTION OF THE DRAWINGS

[0014] The description and claims will be more fully understood with reference to the following figures and data graphs, which are presented as exemplary embodiments of the invention and should not be construed as a complete recitation of the scope of the invention.

[0015] FIG. 1 illustrates examples of uses for generative models in the analysis of clinical trials in accordance with various embodiments of the invention.

[0016] FIG. 2 conceptually illustrates an example of a process for designing a random control trial with prognostic effect estimation.

[0017] FIG. 3 conceptually illustrates an example of a process for determining treatment effects of a RCT in accordance with an embodiment of the invention.

[0018] FIG. 4 illustrates an example of generative models for clinical trial panel data in accordance with an embodiment of the invention.

[0019] FIG. 5 illustrates an example of using generative models to estimate treatment effects in accordance with an embodiment of the invention.

[0020] FIG. 6 illustrates an example of borrowing information from digital subjects to estimate treatment effects in accordance with an embodiment of the invention.

[0021] FIG. 7 illustrates an example of borrowing information from digital twins to estimate treatment effects in accordance with an embodiment of the invention.

[0022] FIG. 8 illustrates an example of using generalized linear models and digital twins estimate treatment effects in accordance with an embodiment of the invention.

[0023] FIG. 9 illustrates an example of using a generative model to measure individual treatment responses in accordance with an embodiment of the invention.

[0024] FIG. 10 illustrates an example of estimating treatment effects with individual treatment responses in accordance with an embodiment of the invention.

[0025] FIG. 11 illustrates an example of a treatment analysis system that determines treatment effects in accordance with some embodiments of the invention.

[0026] FIG. 12 illustrates an example of a treatment analysis element that executes instructions to perform processes that determine treatment effects in accordance with various embodiments of the invention.

[0027] FIG. 13 illustrates an example of a treatment analysis application for determining treatment effects in accordance with an embodiment of the invention.

#### DETAILED DESCRIPTION

[0028] Randomized controlled trials (RCTs) for specific indications (i.e., therapeutic areas) have many common elements from one study to the next. In particular, studies collect many of the same fields at baseline (pre-treatment) and collect many of the same outcomes, usually at a common cadence (i.e., visit schedule). The datasets containing the data from these similar RCTs and other supplementary data (e.g., observational studies, summary data from the literature) may then be combined into a single data montage which is suitable for building deep learning models.

[0029] Systems and methods in accordance with some embodiments of the invention can determine treatment effects for a randomized control trial (RCT) using data sampled from a generative model, design RCTs, and/or determine decision rules for treatments. Data sampled from generative models in accordance with some embodiments of the invention may be referred to as 'digital subjects' throughout this description. In many embodiments, digital subjects can be generated to match given statistics of the treatment groups at the beginning of the study. Digital subjects in accordance with numerous embodiments of the invention can be generated for each subject in a study and the generated digital subjects can be used as digital twins for a counterfactual analysis. In various embodiments, generative models can be used to compute a measure of response that is individual to each patient and this response can be used to assess the effect of the treatment. Systems and methods in accordance with several embodiments of the invention can correct for bias that may be introduced by incorporating generated digital subject data.

[0030] In certain embodiments, processes in accordance with a number of embodiments of the invention can improve RCT design by reducing the number of subjects required for different arms of the RCT. Processes in accordance with some embodiments of the invention can improve the ability of a system to accurately determine treatment effects from a RCT by increasing the statistical power of the trial. In many embodiments, the process of conducting a RCT can be improved from the design through the analysis and treatment decisions.

[0031] Examples of uses for generative models in the analysis of clinical trials in accordance with various embodiments of the invention are illustrated in FIG. 1. The first example 115 illustrates that generative models, digital subjects, or digital twins can be used to increase the statistical power of traditional randomized controlled trials. In the

second example 120, generated data is used to decrease the number of subjects required to be enrolled in the control group of a randomized controlled trial. The third example 125 shows that generated can be used as the external comparator arm of a single-arm trial.

[0032] In an RCT, a group of subjects with particular characteristics are randomly assigned to one or more experimental groups receiving new treatments or to a control group receiving a comparative treatment (e.g., a placebo), and the outcomes from these groups can be compared in order to assess the safety and efficacy of the new treatments. Without loss of generality, an RCT can be assumed to include  $i=1, \dots, N$  human subjects. These subjects are often randomly assigned to a control group or to a treatment group such that the probability of being assigned to the treatment group is the same for each subject regardless of any unobserved characteristics. The assignment of subject  $i$  to a group is represented by an indicator variable  $w_i$ . For example, in a study with two groups  $w_i=0$  if subject  $i$  is assigned to the control group and  $w_i=1$  if subject  $i$  is assigned to the treatment group. The number of subjects assigned to the treatment group is  $N_T=\sum_i w_i$  and the number of subjects assigned to the control group is  $N_c=N-N_T$ .

[0033] In various embodiments, each subject  $i$  in an RCT can be described by a vector  $x_i(t)$  of variables  $x_{ij}(t)$  at time  $t$ . In this description, the notation  $X_i=\{X_{ij}(t)\}_{t=1}^T$  denotes the panel of data from subject  $i$  and  $x_{0,i}$  to denote the vector of data taken at time zero. An RCT is often concerned with estimating how a treatment affects an outcome  $y_i=f(X_i)$ . The function  $f(\cdot)$  describes the combination of variables being used to assess the outcome of the treatment. Variables in accordance with a number of embodiments of the invention can include (but is not limited to) simple endpoints based on the value of a single variable at the end of the study, composite scores constructed from the characteristics of a patient at the end of the study, and/or time-dependent outcomes such as rates of range or survival times, among others. Approaches in accordance with various embodiments of the invention as described herein can be applied to analyze the effect of treatments on one or more outcomes (such as (but not limited to) those related to the efficacy and safety of the treatment).

[0034] Each subject has two potential outcomes. If the subject were to be assigned to the control group  $w_i=0$ , then  $\gamma_i^{(0)}$  would be the observed potential outcome. By contrast, if the subject were to be assigned to receive treatment  $w_i=1$ , then  $\gamma_i^{(1)}$  would be the observed potential outcome. In practice, a subject can only be assigned to one of the treatment arms such that the observed outcome is  $Y_i=\gamma_i^{(0)}(1-w_i)+w_i\gamma_i^{(1)}$ . Potential outcomes in accordance with many embodiments of the invention can include various measurements, such as, but not limited to conditional average treatment effect:

$$\tau(x_0) = E[Y | w = 1, x_0] - E[Y | w = 0, x_0] \quad (1)$$

and/or the average treatment effect

$$\tau = E[\tau(x_0)] = E[Y | w = 1] - E[Y | w = 0]. \quad (2)$$



Processes in accordance with several embodiments of the invention can estimate these quantities with high accuracy and precision and/or can determine decision rules for declaring treatments to be effective that have low error rates.

**[0035]** It can be expensive, time consuming and, in some cases, unethical to recruit human subjects to participate in RCTs. As a result, a number of methods have been developed for using external control arms to reduce the number of subjects required for an RCT. These methods typically fall into two buckets referred to as ‘historical borrowing’ or ‘external control’.

**[0036]** Historical borrowing refers to incorporating data from the control arms of previously completed trials into the analysis of a new trial. Typically, historical borrowing applies Bayesian methods using prior distributions derived from the historical dataset. Such methods can be used to increase the power of a randomized controlled trial, to decrease the size of the control arm, or even to replace the control arm with the historical data itself (i.e., an external control arm). Some examples of external control arms include control arms from previously completed clinical trials (also called historical control arms), patient registries, and data collected from patients undergoing routine care (called real world data). Use of these external control arms can have serious drawbacks if the population or design of the current RCT differs from the population or design of the external data sources.

**[0037]** It has recently become possible to apply machine learning methods to create simulated subject records. In addition to data from the RCT, generative models in accordance with several embodiments of the invention can link the baseline characteristics  $x_0$  and the control potential outcome  $\gamma^{(0)}$  through a joint probability distribution  $p_{\theta_j}(\gamma^{(0)}, x_0)$  and a conditional probability distribution  $p_{\theta_c}(\gamma^{(0)} | x_0)$ , in which  $\theta_j$  and  $\theta_c$  are the parameters of the joint and conditional distributions, respectively. Note that a model of the joint distribution will also provide a model of the conditional distribution, but the converse is not true.

**[0038]** In several embodiments, simulated subject records can be sampled from probabilistic generative models that can be trained on various data, such as (but not limited to) one or more of historical, registry, and/or real world data. Such models can allow one to extrapolate to new patient populations and study designs.

**[0039]** In some embodiments, generative models may create data in a specialized format—either directly or indirectly—such as the Study Data Tabulation Model (SDTM) to facilitate seamless integration into standard workflows. In a variety of embodiments, generating entire panels of data can be attractive because many of the trial outcomes (such as primary, secondary, and exploratory endpoints as well as safety information) can be analyzed in a parsimonious way using a single generative model. For simplicity, the notation  $p(\gamma, x_0)$  will be used instead of  $p(X)$  in this description, with the understanding that the former can always be obtained from the latter by generating a panel of data  $X$  and then computing a specific outcome  $\gamma=f(X)$  from the panel.

**[0040]** Systems and methods in accordance with numerous embodiments of the invention can provide various approaches for incorporating data from a probabilistic generative model into the analysis of an RCT. In numerous embodiments, such methods can be viewed as borrowing from a model, as opposed to directly borrowing from a historical dataset. As generative models, from which data

can be borrowed, may be biased (for example, due to incorrect modeling assumptions), systems and methods in accordance with a number of embodiments of the invention can account for these potential biases in the analysis of an RCT. Generative models in accordance with various embodiments of the invention can provide control over the characteristics of each simulated subject at the beginning of the study. For example, processes in accordance with various embodiments of the invention can create one or more digital twins for each human subject in the study. Processes in accordance with certain embodiments of the invention can incorporate digital twins to increase statistical power and can provide more individualized information than traditional study designs, such as study designs that borrow population level information or that use nearest neighbor matches to patients in historical or real world databases.

### Designing Randomized Trials

**[0041]** Systems and methods in accordance with many embodiments of the invention can provide increased statistical power compared to trial designs that do not incorporate generative (or prognostic) models. In certain embodiments, design trials with generative models can be designed to optimize for certain characteristics (e.g., sample size, power, etc.) while maintaining certain desired constraints, such as (but not limited to) a predefined desired power to detect a particular effect size.

**[0042]** An example of a process for designing a random control trial with prognostic effect estimation in accordance with an embodiment of the invention is conceptually illustrated in FIG. 2. Process 200 computes (205) a correlation between prognostic scores (or digital twins) and observed outcomes. In various embodiments, prognostic scores can be generated based on subjects from control arms of other trials. In a variety of embodiments, correlations between prognostic scores can include correlations between a vector of outcomes for each sample and prognostic scores generated for the sample. Correlations in accordance with numerous embodiments of the invention can be computed based on an average difference between observed and predicted outcomes. In many embodiments, observed outcomes can come from control arms of other trials.

**[0043]** In a variety of embodiments, correlations can include discrimination metrics (e.g., Pearson correlations) and/or bias metrics. Correlations for repeated measure outcomes in accordance with many embodiments of the invention can be outcome-specific and time-specific.

**[0044]** Process 200 computes (210) a variance of the observed outcomes. Variances in accordance with various embodiments of the invention can indicate the unexplained variance between the observed outcomes and the prognostic scores. In a variety of embodiments, variances can be computed based on correlations (e.g., discrimination and/or bias metrics) between prognostic scores and outcomes over a range of timepoints.

**[0045]** Process 200 estimates (215) a correlation and a variance for a new RCT. Estimated correlations and/or variances in accordance with a variety of embodiments of the invention can be based on the correlations and variances for the observed outcomes. In certain embodiments, estimated correlations can be higher than the computed correlations while estimated variances are lower than the computed variances. Processes in accordance with some

embodiments of the invention can compute estimated variances based on the computed correlation and/or variance.

**[0046]** Process 200 determines (220) target trial parameters based on the estimated correlation and variance. Target trial parameters in accordance with a number of embodiments of the invention can include (but are not limited to) sample size, control arm size, and/or treatment arm size.

**[0047]** While specific processes for designing random trials are described above, any of a variety of processes can be utilized to design trials as appropriate to the requirements of specific applications. In certain embodiments, steps may be executed or performed in any order or sequence not limited to the order and sequence shown and described. In a number of embodiments, some of the above steps may be executed or performed substantially simultaneously where appropriate or in parallel to reduce latency and processing times. In some embodiments, one or more of the above steps may be omitted.

**[0048]** Designing randomized trials using treatment effect estimators with frequentist and Bayesian approaches in accordance with some embodiments of the invention are described in greater detail below.

#### Designing Randomized Trials Using Treatment Effect Estimators

**[0049]** The design of a randomized trial to estimate the effect of a new intervention on a given outcome can depend on various constraints, such as (but not limited to) the effect size one wishes to reliably detect, the power to detect that effect size, and/or the desired control of the type-I error rate. Of course, there may also be other considerations such as time and cost, and one may be interested in more than one particular outcome. Although many of the examples described herein are directed to optimizing for a single outcome, one skilled in the art will recognize that similar systems and methods can be used to optimize across multiple outcomes without departing from this invention.

**[0050]** Treatment effect estimators (or PROCOVA) in accordance with many embodiments of the invention presume a working model  $Y = \beta_0 + \beta_1 W + \beta_2 M + \epsilon$  where Y, W, and M are a subject's outcome, treatment status, and prognostic score, respectively and  $\epsilon$  is a noise term. This model can be fit via ordinary least-squares and the value of  $\beta_1$  can be taken as the point estimate of the treatment effect,  $\hat{\beta}_1$ . This estimate is unbiased given treatment randomization without any assumptions about the veracity of the working linear model. Similarly, the assumption-free asymptotic sampling variance  $v^2 \equiv \mathbb{V}[\hat{\beta}_1]$  of this estimate is given by:

$$v^2 = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1} + \frac{n_0 n_1}{n_0 + n_1} \left( \frac{\rho_0 \sigma_0}{n_1} + \frac{\rho_1 \sigma_1}{n_0} \right)^2 - 2 \frac{n_0 n_1}{n_0 + n_1} \left( \frac{\rho_0 \sigma_0}{n_1} + \frac{\rho_1 \sigma_1}{n_0} \right) \left( \frac{\rho_0 \sigma_0}{n_0} + \frac{\rho_1 \sigma_1}{n_1} \right) \quad (3)$$

in which  $p_w = \mathbb{C}[Y_w, [M]/\sqrt{\mathbb{V}[M]} \mid \mathbb{V}[Y_w]]$  ( $Y_w$  denote potential outcomes under treatment  $w=1$  and control  $w=0$ ),  $\sigma_w^2 = \mathbb{V}[Y_w]$ ,  $n_0$  and  $n_1$  are the number of enrolled control and treated subjects.

**[0051]** An effect estimate can be declared to be “statistically significant” at level  $\alpha$  if a  $p < \alpha$  where  $p = 2(1 - \Phi(\hat{\beta}_1/v))$  is the two-sided p-value and  $\Phi$  denotes the CDF of the

standard normal density. The probability that  $p < \alpha$  when, in reality, the treatment effect is  $\beta_1$  is given by

$$\text{Power} = \Phi \left( \Phi^{-1}(\alpha/2) + \frac{\beta_1}{v} \right) + \Phi \left( \Phi^{-1}(\alpha/2) - \frac{\beta_1}{v} \right) \quad (4)$$

To power a trial to a given level (e.g. 80%) one must first estimate values for  $\sigma_w^2$  and  $p_w$  using prior data (discussed below) or expert opinion. The power formula (4) can then be composed with the variance formula (3) with  $\sigma_w^2$  and  $p_w$  fixed at their estimates  $\hat{\sigma}_w^2$ , and  $\hat{p}_w$ . The resulting function returns power for any values of  $n_0$  and  $n_1$ .

**[0052]** The goal of a sample size calculation in the design of a clinical trial that uses PROCOVA can be to estimate  $n_0$  and  $n_1$  required to achieve the required power. However, one needs an additional constraint such as (but not limited to) a chosen randomization ratio  $n_0/n_1$ , or minimizing the total trial size  $n_0 + n_1$ . In this example, the randomization ratio is pre-specified, but the same principles can be easily applied to other situations.

**[0053]** In numerous embodiments, processes for designing a trial can be based on a generative (or prognostic) model. Prognostic models in accordance with many embodiments of the invention can be trained (e.g., based on a prior trial) or pre-trained. Processes can then estimate the variances,  $\sigma_w^2$  and correlations,  $p_w$  of the control arm of the trial. One method for obtaining these estimates is to use historical data, such as data from the placebo control arms of previous trials performed on similar populations. In numerous embodiments, estimates can be based on a vector  $Y' = [Y'_1 \dots Y'_m]$  of outcomes for these subjects, gathered during the trials, and their corresponding prognostic scores  $M' = [M'_1 \dots M'_m]$ , calculated with the prognostic model from each subject's vector of baseline covariates X, i.e.  $M'_i = m(X'_i)$ .

**[0054]** In some embodiments, control-arm marginal outcome variance  $\sigma_0^2$  can be estimated with the usual estimator,

$$\hat{\sigma}_0^2 = \frac{1}{m-1} \sum (Y'_i - \bar{Y}')^2$$

where  $\bar{Y}'$  is the sample average. The correlation  $p_0$  between  $M'$  and  $Y'$  can be estimated by  $\hat{p}_0 = \frac{\sum (Y'_i - \bar{Y}') (M'_i - \bar{M}')}{\sqrt{\sum (Y'_i - \bar{Y}')^2 \sum (M'_i - \bar{M}')^2}}$ , the usual sample correlation coefficient. These values may be inflated (for a) or deflated (for  $p_0$ ) in order to provide more conservative estimates of power.

**[0055]** In certain embodiments, an inflation parameter  $\lambda_w$  for the variance and a deflation parameter  $\gamma_w$  for the correlation can be applied to sample size calculation. Inflation and deflation parameters can be used to account for the prognostic model. Define the target effect size  $\beta_1^*$ , the significance threshold  $a$ , the desired power level  $\zeta$ , fraction of subjects to be randomized to the active arm  $\pi$ , and dropout rate  $d$ . Define  $\gamma_w \geq 1$  and  $\lambda_w \in [0,1]$  for  $w=0,1$ . Define the variance of the potential outcome under active treatment  $w$  in the planned trial as  $\gamma_w^2 \hat{\sigma}_w^2$ , so that a large  $\gamma_w$  inflates the estimated variance. Similarly, define the correlation between the potential outcome and the prognostic model under active treatment  $w$  as  $\lambda_w \hat{p}_w$ , so that a small  $\lambda_w$  deflates the estimated correlation. Then  $n$  could be minimized using a numerical optimization algorithm (such as a binary search) such that

$$\zeta \geq \Phi \left( \Phi^{-1} \left( \frac{\alpha}{2} \right) + \frac{\beta_1}{v} \right) + \Phi \left( \Phi^{-1} \left( \frac{\alpha}{2} \right) - \frac{\beta_1}{v} \right), \quad (5)$$

with

$$v^2 = \frac{1}{n} \left( \frac{\gamma_0^2 \sigma_0^2}{1-\pi} + \frac{\gamma_1^2 \sigma_0^2}{\pi} + \frac{\hat{\theta}^2 - 2\hat{\theta}_+ \hat{\theta}}{\pi(1-\pi)} \right),$$

$\hat{\theta} = \hat{p}_0 \hat{\sigma}_0 ((1-\pi)\lambda_0\gamma_0 + \pi\lambda_1\gamma_1)$ , and  $\hat{\theta}_+ = \hat{p}_0 \hat{\sigma}_0 (\pi\lambda_0\gamma_0 + (1-\pi)\lambda_1\gamma_1)$ . The minimum sample size can be estimated to be

$$n_d = \frac{n}{1-d}.$$

**[0056]** Unlike the variances and correlations for a control arm, the corresponding values for the treatment arm can rarely be estimated from data because treatment-arm data for the experimental treatment is likely to be scarce or unavailable. In many embodiments, processes can assume  $\sigma_0^2 = \sigma_1^2$  and  $p_0 = p_1$ , the latter of which holds exactly if the effect of treatment is constant across the population. It may also be prudent (and conservative) to assume a slightly higher value for  $\sigma_1^2$  and a slightly smaller value for  $p_1$  relative to their control-arm counterparts.

**[0057]** With the four parameters  $\sigma_w^2$  and  $p_w$  specified, Equation 4 can be computationally optimized over  $n_0$  and  $n_1$  in the desired randomization ratio  $n_0/n_1$  until the minimum values of  $n_0$  and  $n_1$  are found such that the output power meets or exceeds the desired value (e.g., with a numerical optimization scheme).

**[0058]** In many cases, a trial will aim to assess the effect of the intervention on many different outcomes. Processes in accordance with several embodiments of the invention can use multiple prognostic models (e.g., one to predict each outcome of interest) and/or a multivariate prognostic model. Depending on the variances of the outcomes, and the accuracy with which they can be predicted, sample size calculations on the various outcomes of interest may suggest different required sample sizes. In this case, one could simply choose the smallest sample size that meets the minimum required statistical power on each of the outcomes of interest.

Designing Randomized Trials using Bayesian Treatment Effect Estimators

**[0059]** Bayesian PROCOVA is a generalization of PROCOVA that incorporates additional information about the parameters of the (generalized) linear model. Although an example is described with reference to a joint Normal Inverse-Gamma prior distribution, one skilled in the art will recognize that similar processes may utilize various other choices for prior distributions for these parameters without departing from the invention.

**[0060]** In this example, Bayesian PROCOVA specifies a joint Normal Inverse-Gamma prior distribution for the unknown parameters:

$$\sigma^2 \sim \text{Inverse-Gamma}(\epsilon, \epsilon) \quad (6)$$

$$\frac{1}{\sigma} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \sim N \left( 0, \begin{pmatrix} \lambda^2 & 0 & 0 \\ 0 & 1/\epsilon & 0 \\ 0 & 0 & 1/\epsilon \end{pmatrix} \right) \quad (7)$$

Here,  $\epsilon$  is set to an arbitrary small number (e.g.,  $10^{-4}$ ) to encode diffuse prior distributions over  $\sigma^2$ ,  $\beta_1/\sigma$ , and  $\beta_2/\sigma$ . In certain embodiments, an informative prior can be placed over the ratio  $\beta_0/\sigma$  to express the belief that  $|\beta_0/\sigma| \leq \lambda_B$ . Values for  $\lambda_B$  in accordance with a variety of embodiments of the invention may be elicited by reviewing the predictive performance of the prognostic model on historical trials.

**[0061]** In this formulation,  $\sigma^2$  is the residual variance that isn't explained by the prognostic model. That is, the notation in this section can be linked to the previous through the relation  $\sigma^2 = \sigma_w^2(1-p_w^2)$  for  $w=0,1$ , which assumes that the unexplained variance is the same in the control and treatment groups.

**[0062]** There are many potential methods to elicit  $\lambda_B$  from historical data. In one example, previous clinical trials with similar inclusion criteria to the target trial can be identified. For instance, if the target trial includes only patients with baseline scores on a given diagnostic test that lie within a specified range, past trials where at least P % (e.g.,  $P=90\%$ ) of subjects have baseline scores within that range can be assembled. These past trials can be enumerated as  $j=1, \dots, m$ . Then for each past trial,  $j$ , in the reference set, the pairs  $(Y_{i,j}, M_{i,j})$  can be extracted across all control subjects,  $i$ . Define  $N_j$  as the sample size for each past trial, and set  $E_j = \hat{\beta}_0 / \hat{\sigma}_j$  where:

$$\hat{\beta}_{0,j} = \frac{1}{N_j} \sum_i (Y_{i,j} - M_{i,j}) \quad (8)$$

$$\hat{\sigma}_j^2 = \frac{1}{N_j} \sum_i (Y_{i,j} - M_{i,j} - \hat{\beta}_{0,j})^2 \quad (9)$$

**[0063]** Finally, choose:

$$\lambda_B = \sqrt{\frac{\sum_j E_j^2}{(F_m^{\chi^2})^{-1}(\gamma)}}, \quad (10)$$

in which  $F_m^{\chi^2}$  denotes the cumulative distribution function of a  $X_m^2$  random variable, and  $\gamma \in (0,1)$  is chosen to reflect quantile of the distribution. For example, setting  $\gamma=0.025$  ensures that it's likely that  $|\beta_0/\sigma| < \lambda$ .

**[0064]** Under the linear model used for PROCOVA,  $\beta_1$  can be identified as the treatment effect. It has a Student-t posterior distribution, whose parameters depend on the observed trial data. In numerous embodiments, Bayesian PROCOVA can use a posterior probability-based decision rule to conclude that an effect is "statistically significant" at level  $\alpha$  if the posterior assigns probability exceeding  $(1-\alpha/2)$  to either one of the following events:  $\beta_1 < 0$  or  $\beta_1 > 0$ .

**[0065]** The power of this particular Bayesian decision rule can be given by:

$$\text{Power} = \Phi \left( \Phi^{-1}(\alpha/2) \sqrt{\frac{V_{11}}{\hat{V}} \left( 1 + \frac{(1-p)\beta_0^2}{\sigma^2(n\lambda^2(1-p)+1)} \right)} + \frac{\tau}{\sigma\sqrt{\hat{V}}} \right) + \Phi \left( \Phi^{-1}(\alpha/2) \sqrt{\frac{V_{11}}{\hat{V}} \left( 1 + \frac{(1-p)\beta_0^2}{\sigma^2(n\lambda^2(1-p)+1)} \right)} - \frac{\tau}{\sigma\sqrt{\hat{V}}} \right)$$

in which  $n=n_0+n_1$  is the total sample size,  $p=n_1/n$  is the proportion of subjects assigned to the new intervention, and

$$V_{11} = \frac{n\lambda^2 + 1}{n(n\lambda^2 p(1-p) + p)}, \quad (12)$$

$$\tau = \beta_1 + \left( \frac{1}{n\lambda^2(1-p)+1} \right) \beta_0, \quad (13)$$

$$\hat{V} = \frac{p + (1-p)(n\lambda^2 + 1)^2}{np(n\lambda^2(1-p) + 1)^2}. \quad (14)$$

**[0066]** The goal of a sample size calculation in the design of a clinical trial that uses Bayesian PROCOVA is to estimate  $n_0$  and  $n_1$  required to achieve the required power. However, one needs an additional constraint such as a chosen randomization ratio  $n_0/n_1$ , or minimizing the total trial size  $n_0+n_1$ . For this example, the randomization ratio is pre-specified, but the same principles can be easily applied to other situations.

**[0067]** To perform the sample size calculation, processes in accordance with many embodiments of the invention can estimate the true values of  $\beta_0$  and  $\sigma^2$ , as well as a pre-specified power to detect a given effect size  $\beta_1$ . In principle, both  $\beta_0$  and  $\sigma^2$  can be estimated from the performance of the prognostic model on historical data. Specifically,  $\beta_0$  may be the average difference between the observed and predicted outcomes, and  $\sigma^2 = \sigma(1-p_0^2)$  is the unexplained variance. In many cases, however, processes may set  $\beta_0=0$  when performing a sample size calculation. Finally, given values for  $\beta_0$  and  $\sigma^2$  and the elicited prior distribution, a numerical optimization process can be used to compute the minimum  $n_0$  and  $n_1$  that offer the desired power, subject to a desired constraint on the randomization ratio.

Designing Randomized Trials using Repeated Measures

**[0068]** Generative models in accordance with a number of embodiments of the invention can provide a complete prediction of the multivariate distribution of future outcomes covering the time horizon of an RCT given a complete or partial set of baseline data as inputs. Incorporating the capabilities and outputs of these models at the RCT design and/or analysis stage respectively can yield RCTs that are more statistically powerful and which can therefore be smaller and faster to complete. In certain embodiments, RCTs can be designed using repeated measures.

**[0069]** Let  $X_{ib}$ , where  $i=1$  to  $I$  study subjects and  $b=1$  to  $B$  baseline variables of interest, represent the value of the  $b$ th baseline variable of interest in the  $i$ th study subject. Let  $Y_{ikt}$  represent the value of the  $k$ th outcome of interest in the  $i$ th study subject at a time point  $t$  where  $t=0$  is concurrent with the baseline measurements and  $t>0$  is a post-baseline outcome.

**[0070]** Given a complete or partial set of  $X_{ib}$  and  $Y_{iko}$ , generative models in accordance with a variety of embodiments of the invention can output samples from the multivariate distribution of  $Y_{ikt}$  where  $t>0$ . Prognostic scores for each outcome and each time point may be computed as  $M_{ikt} = E[Y_{ikt} | Y_{i\cdot}, Y_{i\cdot 0}]$  in accordance with certain embodiments of the invention. In some embodiments, prognostic scores and/or other information from the full multivariate distribution may be used for clinical trial planning and analysis.

**[0071]** For the purposes of trial planning, processes in accordance with several embodiments of the invention can begin with model evaluation. Using data not included in the model training set, the discrimination and/or bias of the model may be assessed. These assessments can be outcome-specific and/or time-specific. As before, let  $Y_{ikt}$  represent the value of the  $k$ th outcome of interest in the  $i$ th study subject at a time point  $t$  where  $t>0$  is a post-baseline outcome. Discrimination metrics in accordance with several embodiments of the invention can include the Pearson correlation  $P_{kt}$  between  $Y_{ikt}$  and  $M_{ikt}$  for a given  $k$  and a given  $t$ . Additional functions of  $Y_{ikt}$  and  $M_{ikt}$  may be used to compute bias metrics.

**[0072]** The power to detect a treatment difference in a randomized trial in accordance with a number of embodiments of the invention is a function of the variance of the treatment effect. For studies with repeated measures outcomes using a traditional study design, the variance of the treatment effect is approximately  $\sigma_{okT}^2/(n_0(1-d_0)) + \sigma_{1kT}^2/(n_1(1-d_1))$  where  $n_0$  and  $n_1$  are the baseline control group and treatment group sample sizes respectively,  $d_0$  and  $d_1$  are the control group and treatment group dropout rates through time  $T$  respectively and  $\sigma_{okT}^2$  and  $\sigma_{1kT}^2$  are respectively the control group and treatment group variances of outcome  $k$  at time  $T$ , where  $T$  is the primary timepoint of interest for the primary analysis of the clinical trial.

**[0073]** In several embodiments, designs may use generative model outputs and model evaluation results in a smaller variance of approximately  $(\sigma_0^2/(n_0(1-d_0)) + \sigma_1^2/(n_1(1-d_1))) (1-p_{kT}^2)$ . This formula can be used for frequentist methods with tight type I error control, and can be a critical input for sample size and power formulae, such as those described throughout this specification. In several embodiments, evaluating the formula with and without the value of the generative model output included allows one to compute the attainable sample size reduction or power gain. In many embodiments, even greater sample size reductions can be achieved with Bayesian designs incorporating both correlation and bias metrics from the model evaluation stage.

**[0074]** The traditional repeated measures model for a given outcome,  $k$ , has the form  $Y_{ikt} = \beta_{0k} + \beta_{1k}I_1 + \epsilon_{ikt}$  where  $I_1$  is an indicator for being randomized to treatment and  $\epsilon_{ikt}$  is the residual error where the covariance between  $\epsilon_{ikt}$  and  $\epsilon_{jkt}$  is zero when  $i=j$  and potentially non-zero according to some prespecified structure when  $i \neq j$ .

**[0075]** Incorporating the output of a generative model to the analysis based on a study designed to have greater power, repeated measures models in accordance with a variety of embodiments of the invention can be expanded as  $Y_{ikt} = \beta_{0k} + \beta_{1k}I_1 + \beta_{2k}f(M_{ikt}, t \in [0, \infty)) + \epsilon_{ikt}$ . Gains in power are similar where  $f(M_{ikt})$  is set to  $M_{ikt}$  or  $M_{ikT}$ .

**[0076]** Both the traditional trial design and the enhanced trial design are conservative if  $d_0$  and  $d_1$  are large and the within patient correlations are high; however, such designs

are very rare, as high dropout rates result in generalizability challenges. Thus, processes in accordance with numerous embodiments of the invention are robust to the broad range of dropout patterns and correlation structures that occur in actual practice. In many embodiments, reparameterization of this model to allow  $\beta_{oxz}$  to represent bias provides the general setup for a Bayesian approach to the analysis consistent with greater power increases. Reparameterization of this model to allow  $\beta_{okz}$  to represent bias provides the general setup for a Bayesian approach to the analysis consistent with greater power increases.

#### Determining Treatment Effects

**[0077]** An example of a process for determining treatment effects of a RCT is conceptually illustrated in FIG. 3. Process 300 receives (305) RCT data. RCT data can include panel data collected from subjects of a RCT. RCT data in accordance with a variety of embodiments of the invention can be divided into control and treatment arms based on whether subjects received a treatment. In many embodiments, RCT data can be supplemented with generated subject data. Generated subject data in accordance with a number of embodiments of the invention can include (but is not limited to) digital subject data and/or digital twin data.

**[0078]** In several embodiments, processes can receive historical data that can be used to pre-train generative models and/or to determine a prior distribution for Bayesian analyses. Historical data in accordance with numerous embodiments of the invention can include (but is not limited to) control arms from historical control arms, patient registries, electronic health records, and/or real world data. In several embodiments, historical data can include data where measurements for a set of variables are taken at multiple points in time for each patient.

**[0079]** Process 300 generates (310) digital subject data over time using generative models. Generative models in accordance with certain embodiments of the invention can be trained to generate potential outcome data based on characteristics of an individual and/or a population. Digital subject data in accordance with several embodiments of the invention can include (but is not limited to) panel data, outcome data, etc. In a variety of embodiments, digital subject data can include predicted measurements at multiple points in time. Digital subject data in accordance with various embodiments of the invention can predict the progression of various variables of interest for a given subject over a period of time. In numerous embodiments, generative models can be trained directly on a specific outcome  $p(\gamma|x_0)$ . For example, if a goal of using the generative model is to increase the statistical power for the primary analysis of a randomized controlled trial then it may be sufficient (but not necessary) to only use a model of  $p(\gamma|x_0)$ .

**[0080]** Alternatively, or conjunctively, generative models trained to generate panel data that can be used in the analysis of a clinical trial. Data for a subject in a clinical trial is typically a panel; that is, it describes the observed values of multiple characteristics at multiple discrete timepoints (e.g. visits to the clinical trial site). For example, if a goal of using the generative model is to reduce the number of subjects in the control group of the trial, or as an external comparator for a single arm trial, then generated panel data in accordance with many embodiments of the invention can be used to perform many or all of the analyses of the trial.

**[0081]** In several embodiments, generative models can include (but are not limited to) traditional statistical models, generative adversarial networks, recurrent neural networks, Gaussian processes, autoencoders, autoregressive models, variational autoencoders, and/or other types of probabilistic generative models. For example, processes in accordance with several embodiments of the invention can use sequential models such as (but not limited to) a Conditional Restricted Boltzmann Machine for the full joint distribution of the panel data,  $p(X)$ , from which any outcome can be computed.

**[0082]** An example of generative models for clinical trial panel data in accordance with an embodiment of the invention is illustrated in FIG. 4. Generating panel data in accordance with a variety of embodiments of the invention can enable one to borrow information from the generative model for various analyses in the clinical trial (e.g., primary, secondary, and exploratory endpoints as well as safety information), not just one specific outcome. In addition, digital subjects drawn from the generative model can be of the same form as data obtained from actual subjects in the trial.

**[0083]** Referring back to FIG. 3, process 300 determines (315) treatment effects over time for the RCT using the generated digital subject data. Generative models in accordance with many embodiments of the invention can be incorporated into the analysis of an RCT in a variety of different ways for various applications. In many embodiments, generative models can be used to estimate treatment effect by training separate generative models based on data from the control and treatment arms. Processes in accordance with many embodiments of the invention can use generative models to generate digital subjects to supplement a control arm in an RCT. In certain embodiments, processes can use generative models to generate digital twins for individuals in the control and/or treatment arms. Generative models in accordance with numerous embodiments of the invention used to define individualized responses to treatment. In a variety of embodiments, treatment effects can be determined at each of several points in time to evaluate an RCT over a period of time. Various methods for determining treatment effects in accordance with various embodiments of the invention are described in greater detail herein.

**[0084]** In several embodiments, treatment effects can be determined by fitting generalized linear models (GLMs) to the generated digital subject data and/or the RCT data. In a number of embodiments, multilevel GLMs can be set up so that the parameters (e.g., the treatment effect) can be estimated through maximum likelihood or Bayesian approaches. In a frequentist approach, one can test the null hypothesis  $\beta_0=0$ , whereas the Bayesian approach may focus on the posterior probability  $\text{Prob}(\beta_0 \geq 0 | \text{data, prior})$ .

**[0085]** While specific processes for determining treatment effects in an RCT are described above, any of a variety of processes can be utilized to determine treatment effects as appropriate to the requirements of specific applications. In certain embodiments, steps may be executed or performed in any order or sequence not limited to the order and sequence shown and described. In a number of embodiments, some of the above steps may be executed or performed substantially simultaneously where appropriate or in parallel to reduce latency and processing times. In some embodiments, one or more of the above steps may be omitted.

### Estimating the Treatment Effect Using Generative Models

**[0086]** In many embodiments, processes can estimate treatment effects by training two new generative models: a treatment model using the data from the treatment group,  $p_{\theta_{j_1}}(\gamma^{(1)}, x_0 | w_1)$ , and a control model using the data from the control group,  $p_{\theta_{j_0}}(\gamma^{(0)}, x_0 | w_0)$ . In a variety of embodiments, full panels of data from an RCT can be used to train generative models to create panels of generated data. Such processes can allow for the analysis of many outcomes (including (but not limited to) primary, secondary, and exploratory efficacy endpoints as well as safety information) by comparing the trained treatment models against trained control models. For simplicity, the notation  $p(\gamma, x_0)$  will be used instead of  $p(X)$ , with the understanding that the former can always be obtained from the latter by generating a panel of data  $X$  and then computing a specific outcome  $\gamma=f(X)$  from the panel.

**[0087]** In one embodiment, generative models for the control condition (e.g., a Conditional Restricted Boltzmann Machine) can be trained on historical data from previously completed clinical trials. Then, two new generative models for the control and treatment groups can be obtained by solving minimization problems:

$$\min_{\theta_{j_0}} \left\{ -\sum_i (1-w_i) \log p_{\theta_{j_0}}(Y_i, x_{0,i} | w_0) + \lambda_0 D(p_{\theta_{j_0}}, p_{\theta_{j_1}}) \right\}$$

$$\min_{\theta_{j_1}} \left\{ -\sum_i w_i \log p_{\theta_{j_1}}(Y_i, x_{0,i} | w_1) + \lambda_1 D(p_{\theta_{j_1}}, p_{\theta_{j_0}}) \right\}$$

in which  $\lambda_0$  and  $\lambda_1$  are prior parameters that describe how well pre-trained generative models describe the outcomes in the two arms of the RCT, and  $D(\cdot, \cdot)$  is a measure of the difference between two generative models such as (but not limited to) the Kullback-Leibler divergence. For example, the new generative models may also be Conditional Restricted Boltzmann Machines.

**[0088]** The estimate for the treatment effect can then be computed as

$$\hat{\tau} = \int dy dx y p_{\theta_{j_1}}(y, x | w_1) - \int dy dx y p_{\theta_{j_0}}(y, x | w_0). \quad (15)$$

In several embodiments, treatment effects can be computed by drawing samples from the control and treatment models and comparing the distributions of the samples. Processes in accordance with some embodiments of the invention can further tune the computation of treatment effects by adjusting for the uncertainty in treatment effect estimates. In several embodiments, the uncertainty in treatment effect estimates ( $\lambda_j$ ) can be obtained using a bootstrap by repeatedly resampling the data from the RCT (with replacement), training the updated generative models, and computing an estimate for the treatment effect; the uncertainty is the standard deviation of these estimates. In a number of embodiments, point estimates for the treatment effect and the estimate for its uncertainty can be used to perform a hypothesis test in order to create a decision rule.

**[0089]** In numerous embodiments, processes can begin with a distribution  $\pi(\theta_j)$  for the parameters of the generative model (e.g., obtained from a Bayesian analysis of historical

data). Then, posterior distributions for  $\theta_{j_0}$  and  $\theta_{j_1}$  can be estimated by applying Bayes rule,

$$\log \pi(\theta_{j_0}) = \text{constant} + \sum_i (1-w_i) p_{\theta_{j_0}}(Y_i, x_i | w_0) + \lambda_0 \log \pi(\theta_{j_0}) \quad (16)$$

$$\log \pi(\theta_{j_1}) = \text{constant} + \sum_i w_i p_{\theta_{j_1}}(Y_i, x_i | w_1) + \lambda_1 \log \pi(\theta_{j_1}).$$

In certain embodiments, point estimates for the treatment effect can be calculated as the mean of the posterior distribution

$$\hat{\tau} = \int dy dx d\theta_{j_1} y p_{\theta_{j_1}}(y, x | w_1) \pi(\theta_{j_1}) - \int dy dx d\theta_{j_0} y p_{\theta_{j_0}}(y, x | w_0) \pi(\theta_{j_0}) \quad (17)$$

where the uncertainty is the variance of the posterior distribution,

$$\delta^2 \tau = \int dy dx d\theta_{j_1} y^2 p_{\theta_{j_1}}(y, x | w_1) \pi(\theta_{j_1}) - \left( \int dy dx d\theta_{j_1} y p_{\theta_{j_1}}(y, x | w_1) \pi(\theta_{j_1}) \right)^2 - \left( \int dy dx d\theta_{j_0} y^2 p_{\theta_{j_0}}(y, x | w_0) \pi(\theta_{j_0}) - \left( \int dy dx d\theta_{j_0} y p_{\theta_{j_0}}(y, x | w_0) \pi(\theta_{j_0}) \right)^2 \right) \quad (18)$$

As above, point estimates for the treatment effect and estimates for their uncertainty can be used to perform a hypothesis test in order to create a decision rule in accordance with certain embodiments of the invention. Processes in accordance with a variety of embodiments of the invention can train conditional generative models  $p_{\theta_{j_1}}(\gamma^{(1)} | x_0, w_1)$  and  $p_{\theta_{j_0}}(\gamma^{(0)} | x_0, w_0)$ , as opposed to (or in conjunction with) joint generative models, in order to estimate treatment effects that are conditioned on the baseline covariates  $x_0$ .

**[0090]** It can be difficult to determine the operating characteristics of a decision rule based on these methods. Specifically, extensive simulations can be required in order to estimate the type-I error rate (i.e., the probability that an ineffective treatment would be declared to be effective) and the type-II error rate (i.e., the probability that an effective treatment would be declared ineffective). Well-characterized operating characteristics are required for many applications of RCTs and, as a result, this approach is often impractical. Generative models that rely on modern machine learning techniques are typically computationally expensive to train. As a result, using the bootstrap or Bayesian methods to obtain uncertainties required to formulate reasonable decision rules can be quite challenging.

**[0091]** An example of using generative models to estimate treatment effects in accordance with an embodiment of the invention is illustrated in FIG. 5. In the first stage 505, an untrained generative model of the control condition is trained using historical data, such as (but not limited to), data from previously completed clinical trials, electronic health records, and/or other studies. In the second stage 510, a patient population is randomly divided into a control group and a treatment group as part of a randomized controlled trial. Patients from the population can be randomized into the control and treatment groups with unequal randomization in accordance with a variety of embodiments of the

invention. In this example, two new generative models are trained: one for the control group and one for the treatment group. In certain embodiments, control and treatment generative models can be based on a pre-trained generative model but can be additionally trained to reflect new information from the RCT. Outputs from the control and generative models can then be compared to compute the treatment effects. In several embodiments, Bayesian methods and/or the bootstrap may be used to estimate uncertainties in the treatment effects and decision rules based on p-values and/or posterior probabilities may be applied.

**Borrowing Information from Digital Subjects**

**[0092]** The defining characteristic of a generative model is that one can draw new samples from the model. In several embodiments, each sample from the generative model is a digital subject. Processes in accordance with several embodiments of the invention can draw an initial sample of digital subjects  $(\gamma_i, x_{0,i}) \sim p(\gamma_i, x_{0,i})$  for  $i=1, \dots, M$  such that the moments of the synthetic population match various moments of the actual population in the RCT. Let  $D(\{(\gamma_i, x_{0,i})\}_{i=1}^M)=1$  be a measure of agreement between the moments computed from the digital subject data and the moments computed from the actual population such that the goal is  $D(\{(\gamma_i, x_{0,i})\}_{i=1}^M)=0$ . Processes to generate an initial population in accordance with many embodiments of the invention can choose some  $i$  at random and generate a new sample  $(\gamma'_i, x'_{0,i}) \sim p(\gamma, x_0)$ . Processes can replace digital subject  $i$  with the sample if doing so decreases  $D(\{(\gamma_i, x_{0,i})\}_{i=1}^M)$ . For generative models that use a Markov Chain (e.g., Deep Boltzmann Machines) to generate samples, processes in accordance with a number of embodiments of the invention can compute new samples by taking a one or more steps starting at sample  $i$ .

**[0093]** Define a variable  $s_i=0$  if a given subject is an actual subject from the RCT, and  $s_i=1$  if the subject is a digital subject drawn from the generative model. Data from the subjects in the RCT can be represented as  $(Y_i, x_{0,i}, w_i, s_i=0)$ . Likewise,  $M$  samples  $(Y_i, x_{0,i}, w_i=0, s_i=1) \sim p_{\theta}(\gamma^{(0)}, x_0)$  can be generated using the generative model and  $N_s \leq M$  of the samples can be selected based on the inclusion criteria of the RCT or to match some of the characteristics of the study population, such as (but not limited to) the means and standard deviations of some chosen variables at time zero.

**[0094]** Systems and methods in accordance with certain embodiments of the invention can incorporate digital subjects into an estimate for the treatment effect by fitting a generalized linear model (GLM) given, in its most general form

$$g(E[y_i]) = a + \left( b_0 + \sum_j b_j x_{0,ij} \right) w_i + \left( c_0 + \sum_j c_j x_{0,ij} \right) s_i + \sum_j d_j x_{0,ij} \quad (19)$$

in which  $g(\cdot)$  is a link function. For example,  $g(\mu)=\mu$  corresponds to a linear regression and  $g(\mu)=\log(\mu/(1-\mu))$  corresponds to logistic regression. In various embodiments, this framework can also include Cox proportional hazards models used for survival analysis as a special case. In many embodiments, some of these coefficients may be set to zero to create simpler models.

**[0095]** In the example of equation 19, the terms involving the  $b$  coefficients represent the treatment effect, which may depend on the baseline covariates  $x_0$ . The terms involving the  $c$  coefficients represent potential bias in the generative

model, which may depend on the baseline covariates  $x_0$ . The terms involving the  $d$  coefficients represent potential baseline differences between the treatment and control groups in the trial. The model can be fit using any of a variety of method for fitting GLMs.

**[0096]** In some embodiments, uncertainties in the coefficients can be estimated analytically. Alternatively, or conjunctively, uncertainties in accordance with numerous embodiments of the invention can be estimated using a bootstrap by repeatedly resampling the data (with replacement) and re-fitting the model. Uncertainties in accordance with many embodiments of the invention can be computed as the standard deviations of the coefficients computed by such resampling procedures. Point estimates for the treatment effect and estimates for their uncertainty can be used to perform a hypothesis test in order to create a decision rule in accordance with many embodiments of the invention.

**[0097]** In theory, a perfect generative model will have no bias, with  $c_j=0$  for all  $j$  so that the indicator variable  $s_i$  has no effect. However, machine learning models may not generalize perfectly to data outside of the training set. Typically, the generalization performance of a model is measured by holding out some data from the model training phase so that the held-out data can be used to test the performance of the model. For example, suppose that there are one or more control arms from historical clinical trials in addition to the generative model. Then, the  $c$  coefficients can be estimated in accordance with numerous embodiments of the invention by fitting a reduced GLM on the historical control arm data,

$$g(E[y_i]) = a + \left( c_0 + \sum_j c_j x_{0,ij} \right) s_i. \quad (20)$$

**[0098]** This can be particularly useful in a Bayesian framework, in which a distribution  $\pi(a,c)$  can be estimated for these coefficients using the historical data, where the data-driven prior distribution can be used in a Bayesian analysis (e.g., in Equation 19) in the RCT. Essentially, processes in accordance with some embodiments of the invention can use historical data to determine how well the generative model is likely to generalize to new populations, and then apply this information to the analysis of the RCT. In the limit that  $\pi(c)=\delta(c-0)$ , then the digital subjects become substitutable for the actual control subjects in the RCT. As a result, the better the generative model, the fewer control subjects required in the RCT. In certain embodiments, similar approaches could be used to include prior information on any coefficients that are active when  $w_i=0$ , including the  $d$  coefficients for potential baseline differences between the treatment and control groups.

**[0099]** In certain embodiments, simpler models may be created by setting some of the parameters to zero. To understand the effect of the  $c$  coefficients, a simple case with linear link functions, no interactions, and setting all  $d$  coefficients to zero is described. In this simple case, the GLM above becomes a simple linear regression

$$y_i = a + b_0 w_i + c_0 s_i + \epsilon_i. \quad (21)$$

In addition, suppose a prior distribution that can be expressed by adding the following penalty to the log-likelihood function,

$$\text{penalty}(c_0) = \frac{\lambda}{2} c_0^2, \quad (22)$$

in which  $\lambda$  controls the degree of belief in the quality of the generative model. As mentioned previously, the parameters of the prior distribution could be estimated from historical data and/or specified through some other means. One skilled in the art will recognize that this is not the only choice of prior distribution or means to incorporate prior information, but this provides an example simple enough for analysis to illustrate the properties of processes in accordance with various embodiments of the invention.

**[0100]** In many embodiments, point estimates and uncertainties of the treatment effect can be estimated using a Laplace approximation of the resulting posterior distribution. In practice, various methods including (but not limited to) exact integration, Markov Chain Monte Carlo calculations, and/or variational approximations could be used to obtain a posterior distribution. Using the Laplace approximation (i.e., a series expansion about the maximum of the posterior distribution), it is possible to derive an estimate for the covariance matrix of the posterior distribution of the parameters  $a$ ,  $b_0$ , and  $c_0$ , which can be given by

$$\hat{v} = \frac{1}{\lambda N + \lambda N_S + N N_S} \begin{bmatrix} \lambda + N_S & -(\lambda + N_S) & -N_S \\ -(\lambda + N_S) & N N_S + \lambda(N + N_S) & N_S \\ -N_S & N_S & N + N_S - N_T \end{bmatrix}$$

and the point estimate, which can be given by

$$\begin{bmatrix} \hat{a} \\ \hat{b}_0 \\ \hat{c}_0 \end{bmatrix} = \hat{v} \begin{bmatrix} N_T E[y_i | w_i = 1, s_i = 0] + N_C E[y_i | w_i = 0, s_i = 0] + \\ N_S E[y_i | w_i = 0, s_i = 1] \\ N_T E[y_i | w_i = 1, s_i = 0] \\ N_S E[y_i | w_i = 0, s_i = 1] \end{bmatrix}$$

**[0101]** In the limit that  $\lambda \rightarrow 0$ , point estimate  $\hat{b}_0 = E[\gamma_i | w_i = 1, s_i = 0] E[\gamma_i | w_i = 0, s_i = 0]$  and uncertainty

$$\sigma_{\hat{b}_0}^2 \propto \frac{N}{N_T N_C}.$$

This is the usual frequentist estimate for the treatment effect. Notice that the information from the digital subjects has been completely disregarded because  $\lambda = 0$  expresses the prior belief that the model used to generate the digital subjects is likely to be of poor quality. By contrast, consider the limit  $\lambda \rightarrow \infty$  that expresses the prior belief that the digital subjects generated from the model are statistically indistinguishable from actual control subjects. In this case, the point

$$\hat{b}_0 = \frac{(N_C + N_S) E[y_i | w_i = 1, s_i = 0] - N_C E[y_i | w_i = 0, s_i = 0] - N_S E[y_i | w_i = 0, s_i = 1]}{N_C + N_S}$$

and uncertainty

$$\sigma_{\hat{b}_0}^2 \propto \frac{N + N_S}{N_T (N_C + N_S)}.$$

That is, this treats the digital subject data as if it is exactly substitutable for actual control subject data. Intermediate values of  $\lambda$  borrow intermediate amounts of information from the digital control subjects.

**[0102]** By using Bayesian methods to incorporate digital subject data into a clinical trial, as in the previous example, it is possible to increase the statistical power of the trial and/or to reduce the number of actual subjects required for the control group. If the model used to create the digital subjects can be shown to be accurate by estimating the parameters of the prior distribution using historical data, then it can be possible to create and design trials with attractive operating characteristics (i.e., low type-I and type-II error rates) that do not require large numbers of actual subjects. Operating characteristics in accordance with various embodiments of the invention can be characterized through analytical calculations and/or computer simulations.

**[0103]** An example of borrowing information from digital subjects to estimate treatment effects in accordance with an embodiment of the invention is illustrated in FIG. 6. In the first stage **605**, a generative model of the control condition is trained using historical data, such as (but not limited to) data from previously completed clinical trials, electronic health records, or other studies. In the second stage **610**, if the analysis to be performed is Bayesian, predictions from the generative model are compared to historical data that were not used to train the model in order to obtain a prior distribution capturing how well the predictions generalize to new populations. A frequentist analysis can skip the second stage **610**. In third stage **615**, a randomized controlled trial is conducted (potentially with unequal randomization). The generative model is used to create digital subjects, and all of the data are incorporated into a statistical analysis (including the prior from step **610** if the analysis is Bayesian) to estimate the treatment effects. Bayesian methods, analytical calculations, or the bootstrap may be used to estimate uncertainties in the treatment effects, and decision rules based on p-values or posterior probabilities may be applied. Borrowing Information from Digital Twins

**[0104]** Some methods estimate treatment effects using GLMs while adjusting for covariates. For example, one may perform a regression of the final outcome in the trial against the treatment indicator and a measure of disease severity at the start of the trial. As long as the covariate was measured before the treatment was assigned in a randomized controlled trial, then adjusting for the covariate will not bias the estimate for the treatment effect in a frequentist analysis. When using covariate adjustment, the statistical power is a function of the correlation between the outcome and the covariate being adjusted for; the larger the correlation, the higher the power.

**[0105]** In theory, the covariate that is most correlated with the outcome that one could obtain is an accurate prediction of the outcome. Therefore, another method to incorporate generative models into RCTs in accordance with a variety of embodiments of the invention is to use generative models to predict outcomes and to adjust for the predicted outcomes in a GLM for estimating the treatment effect. Let  $E_p[\gamma_i]$  and



$\text{Var}[\gamma_i]$  denote the expected value and variance of the outcome predicted for subject  $i$  by the generative model, respectively. Depending on the type of generative model, these moments may be computable analytically or, more generally, by drawing samples from the generative model  $p(\gamma_i|x_{0,i})$  and computing Monte Carlo estimates of the moments in accordance with a number of embodiments of the invention. The number of samples used to compute the Monte Carlo estimates can be a parameter selected by the researcher. As above, processes in accordance with several embodiments of the invention can use generative models that generate panel data so that a single generative model may be used for analyses of many outcomes in a given trial (e.g., primary, secondary, and exploratory endpoints as well as safety information). In a number of embodiments, rather than predictions for a given outcome, predictions of multiple outcomes derived from a generative model may all be included in a GLM for a particular outcome. Samples drawn from the generative models in accordance with several embodiments of the invention can be conditioned on the characteristics of a subject at the start of the trial, also referred to as digital twins of that subject.

**[0106]** In many embodiments, digital twins can be incorporated into an RCT in order to estimate the treatment effect by fitting a GLM of the form

$$g(E[y_i]) = a + \left( b_0 + \sum_j b_j x_{0,ij} \right) w_i + \left( c_0 + \sum_j c_j x_{0,ij} \right) g(E_p[y_i]) + \sum_j d_j x_{ij} + \left( z_0 + \sum_j z_j x_{0,ij} \right) w_i g(E_p[y_i]) \quad (23)$$

in which  $g(\cdot)$  is a link function. For example,  $g(\mu) = \mu$  corresponds to a linear regression and  $g(\mu) = \log(\mu/(1-\mu))$  corresponds to logistic regression. This framework in accordance with numerous embodiments of the invention can also include Cox proportional hazards models used for survival analysis as a special case. In many embodiments, some of these coefficients may be set to zero to create simpler models. One skilled in the art will recognize that it is trivial to include other predictions from the generative model as covariates if desired.

**[0107]** The above equation can be generalized to various applications and implementations. The terms involving the  $b$  coefficients represent the treatment effect, which may depend on the baseline covariates  $x_0$ . The terms involving the  $c$  coefficients represent potential bias in the generative model, which may depend on the baseline covariates  $x_0$ . The terms involving the  $d$  coefficients represent potential baseline differences between the treatment and control groups in the trial. The terms involving the  $z$  coefficients reflect that the relationship between the predicted and observed outcomes may be affected by the treatment. The model can be fit using any of a variety of methods for fitting GLMs. In a number of embodiments, uncertainties in the coefficients can be estimated analytically. Alternatively, or conjunctively, processes in accordance with many embodiments of the invention can estimate uncertainties using a bootstrap by repeatedly resampling the data (with replacement) and re-fitting the model; the uncertainties can be the standard deviations of the coefficients computed by this resampling procedure. In some embodiments, point estimates for the

treatment effect and estimates for their uncertainty can be used to perform a hypothesis test in order to create a decision rule.

**[0108]** In some embodiments, variances of the outcomes can be modeled through another GLM that adjusts for the variance of the outcome that is predicted by the generative model. For example, variances in accordance with many embodiments of the invention can be modeled as follows

$$G(\text{Var}[y_i]) = \alpha + \left( \beta_0 + \sum_j \beta_j x_{0,ij} \right) w_i + \left( \gamma_0 + \sum_j \gamma_j x_{0,ij} \right) G(\text{Var}_p[y_i]) + \sum_j \delta_j x_{ij} + \left( \zeta_0 + \sum_j \zeta_j x_{0,ij} \right) w_i G(\text{Var}_p[y_i]) \quad (24)$$

in which  $G(\cdot)$  is a link function that is appropriate for the variance. For example,  $G(\mathbf{94}^2) = \log(\sigma^2)$  can be used for a continuous outcome. In many embodiments, some of these coefficients may be set to zero to create simpler models. One skilled in the art will recognize that other predictions from the generative model can be included as covariates if desired.

**[0109]** Well-trained generative models in accordance with certain embodiments of the invention will have  $g(E[\gamma_i]) \approx g(E_p[\gamma_i])$  and  $G(\text{Var}[\gamma_i]) \approx G(\text{Var}_p[\gamma_i])$  by construction. Therefore, prior knowledge about the coefficients in the GLMs can be used to improve the estimation of the treatment effect. However, machine learning models may not generalize perfectly to data outside of the training set. Typically, the generalization performance of a model is measured by holding out some data from the model training phase so that the held-out data can be used to test the performance of the model. For example, suppose that there are one or more control arms from historical clinical trials in addition to the generative model. Then, the  $c$  coefficients in accordance with various embodiments of the invention can be estimated by fitting a reduced GLM on the historical control arm data,

$$g(E[y_i]) = a + \left( c_0 + \sum_j c_j x_{0,ij} \right) g(E_p[y_i]), \quad (25)$$

for the mean or

$$G(\text{Var}[y_i]) = \alpha + \left( \gamma_0 + \sum_j \gamma_j x_{0,ij} \right) G(\text{Var}_p[y_i]), \quad (26)$$

for the variance. This is particularly useful in a Bayesian framework, in which a distribution  $\pi(a,c)$  or  $\pi(\alpha,\gamma)$  can be estimated for these coefficients using the historical data, where the data-driven prior distribution can be used in a Bayesian analysis of the RCT. Essentially, this uses the historical data to determine how well the generative model is likely to generalize to new populations, and then applies this information to the analysis of the RCT. In the limit that  $\pi(a,c) \rightarrow \delta(a-0)\delta(c-1)$ , then digital twins in accordance with a variety of embodiments of the invention can become substitutable for actual control subjects in the RCT. As a result, the better the generative model, the fewer control subjects required in the RCT. In some embodiments, similar

approaches could be used to include prior information on any coefficients that are active when  $w_i=0$ , including the  $d$  coefficients.

**[0110]** Examples of workflows for frequentist and Bayesian analyses of clinical trials that incorporate digital twins to estimate treatment effects in accordance with various embodiments of the invention are described below. For a frequentist case for a continuous endpoint, consider a simple example

$$E[y_i] = a + b_0 w_i + c_0 E_p[y_i] \quad (27)$$

$$\text{Var}[y_i] = \sigma^2 \quad (28)$$

assuming no interactions and homoscedastic errors. One skilled in the art will recognize how this can be applied to the more general case captured by Equation 23 and Equation 24. In numerous embodiments, simple analyses can lead to results that are more easily interpreted. This model implies a normal likelihood,

$$y_i \sim \mathcal{N}(a + b_0 w_i + c_0 E_p[y_i], \sigma^2) \quad (29)$$

such that the model can be fit (e.g., by maximum likelihood). There are two situations to consider: (1) the design of the trial has already been determined by some method prior to incorporating the digital twins such that the digital twins can be used to increase the statistical power of the trial, or (2) the trial needs to be designed so that it incorporates digital twins to achieve an efficient design with sufficient power. In the case of a continuous endpoint, the statistical power of the trial will depend on the correlation between  $\gamma_i$  and  $E_p[\gamma_i]$ , which can be estimated from historical data, and is a function of the magnitude of the treatment effect. In a variety of embodiments, analytical formulas can be derived in this special case. Alternatively, or conjunctively, computer simulations can be utilized in the general case.

**[0111]** Once the trial is designed, patients are enrolled and followed until their outcome is measured. In some cases, patients may not be able to finish the trial and various methods (such as Last Observation Carried Forward) need to be applied in order to impute outcomes for the patients who have not finished the trial, as in most clinical trials. In a number of embodiments, GLMs can be fit to the data from the trial to obtain point estimates  $\hat{b}_0$  and uncertainties  $\hat{\sigma}_{b_0}$  for the treatment effect. The ratio  $\hat{b}_0/\hat{\sigma}_{b_0}$  follows a Student's  $t$ -distribution which can be used to compute a  $p$ -value  $p_{b_0}$  and the null-hypothesis that there is no treatment effect can be rejected if  $p_{b_0} \leq \mathcal{A}$  in which  $\mathcal{A}$  is the desired control of the type-I error rate. This approach is guaranteed to control the type-I error rate, whereas the realized power will be related to the out-of-sample correlation of  $\gamma_i$  and  $E_p[\gamma_i]$  and the true effect size.

**[0112]** In the Bayesian case for a continuous endpoint with homoscedastic errors, assume a simple analysis,

$$E[y_i] = a + b_0 w_i + c_0 E_p[y_i] \quad (30)$$

$$\text{Var}[y_i] = \sigma^2. \quad (31)$$

In certain embodiments, the simple analysis can lead to results that are more easily interpreted. This model implies a normal likelihood,

$$y_i \sim \mathcal{N}(a + b_0 w_i + c_0 E_p[y_i], \sigma^2), \quad (32)$$

but processes in accordance with various embodiments of the invention can use a Bayesian approach to fit it instead of the method of maximum likelihood. In particular, with historical data representing the condition  $w_i=0$  that was not used to train the generative model, processes in accordance with many embodiments of the invention can fit the model,

$$E[y_i] = a + c_0 E_p[y_i] \quad (33)$$

$$\text{Var}[y_i] = \sigma^2. \quad (34)$$

to the historical data in order to derive prior distributions for the analysis of the RCT. To do so, pick a prior distribution  $\pi_0(a, c_0, \sigma^2)$  such as (but not limited to) a Normal-Inverse-Gamma prior or another appropriate prior distribution. As there are no data to inform the parameters of the prior before analyzing the historical data, processes in accordance with several embodiments of the invention can use a diffuse or default prior. In numerous embodiments, Bayesian updates to the prior distribution can be computed from the historical data to derive a new distribution  $\pi_H(a, c_0, \sigma^2)$ , in which the subscript H can be used to denote that this distribution was obtained from historical data. Processes in accordance with numerous embodiments of the invention can then specify a prior distribution  $\pi_0(b_0)$  for the treatment effect. This could also be derived from data in accordance with many embodiments of the invention if it's available, or a diffuse or default prior could be used. The full prior distribution is now  $\pi_H(a, c_0, \sigma^2)\pi_0(b_0)$ . In various embodiments, such distributions can be used compute the expected sample size in order to design the trial, as in a typical Bayesian trial design. Once the trial is designed, patients can be enrolled and followed until their outcome is measured. In some cases, patients may not be able to finish the trial and various methods (such as Last Observation Carried Forward) can be applied in order to impute outcomes for the patients who have not finished the trial, as in most clinical trials.

**[0113]** In numerous embodiments, GLMs can be fit to obtain a posterior distribution  $\pi_{RCT}(a, b_0, c_0, \sigma^2)$  for the parameters. A point estimate for the treatment effect can be computed by, for example,  $\hat{b}_0 = \int da db_0 dc_0 d\sigma^2 b_0 \pi_{RCT}(a, b_0, c_0, \sigma^2)$ ; though, other Bayesian point estimates could be computed as well. In several embodiments, the posterior probability that the treatment effect is greater than zero can also be computed as  $\text{Prob}(b_0 > 0) = \int da db_0 dc_0 d\sigma^2 \theta(b_0) \pi_{RCT}(a, b_0, c_0, \sigma^2)$ , in which  $\theta(\cdot)$  is a logic function that returns one if the argument is true and zero otherwise. As in a typical Bayesian analysis, the treatment can be declared effective if  $\text{Prob}(b_0 \geq 0)$  exceeds a pre-specified threshold in accordance with a number of embodiments of the invention.

**[0114]** There are two limits to the Bayesian analysis that can be informative. First, in the limit of a flat prior distribution  $\pi_H(a, c_0, \sigma^2)\pi_0(b_0) \propto 1$ , then the point estimate and uncertainty for the treatment effect will converge to give the same results as the maximum likelihood method described previously. Thus, if the generalizability of the digital twin model to the population in the RCT is questionable then the Bayesian analysis will end up being very similar to the frequentist analysis. In contrast to the method used to estimate a treatment effect in a trial including digital subjects, including digital twins in the analysis still leads to a gain in power as long as  $\gamma_i$  is correlated with  $E[\gamma_i]$ . The other instructive limit is  $\pi_H(a, c_0, \sigma^2)\pi_0(b_0) \propto (\alpha-0)\delta(c_0-1)$ . In this limit, the point estimate for the treatment effect converges to  $\hat{b}_0 = N_T^{-1} \sum_i (\gamma_i - E_p[\gamma_i])w_i$ . That is, in some embodiments, the estimate for the treatment effect can be obtained by taking the average of the difference between observed and predicted outcomes for the patients who received the treatment  $w_i=1$ . Notice that this latter prior distribution can lead to a situation in which the data from the patients who received the control treatment  $w_i=0$  can be ignored. Processes in accordance with various embodiments of the invention can run trials without a concurrent control arm.

**[0115]** There are advantages and disadvantages to the frequentist and Bayesian methods that are captured through these simple examples. The frequentist approach to including digital twins in the analysis of an RCT leads to an increase in statistical power while controlling the type-I error rate. If desired, it's also possible to use the theoretical increase in statistical power to decrease the number of subjects required for the concurrent control arm, although this cannot be reduced to zero concurrent control subjects. The Bayesian approach borrows more information about the generalizability of the model used to create the digital twins (e.g., from an analysis of historical data) and, as a result, can increase the power much more than the frequentist approach. In addition, the use of Bayesian methods in accordance with numerous embodiments of the invention can enable one to decrease the size of the concurrent control arm even further. However, the increase in power/decrease in required sample size can come at the cost of an uncontrolled type-I error rate. Therefore, processes in accordance with many embodiments of the invention can perform computer simulations of the Bayesian analysis to estimate the type-I error rate so that the operating characteristics of the trial can be described.

**[0116]** As a final example, it is helpful to consider a simple case in which a GLM is also used for the variance. For example, consider the model

$$E[y_i] = a + b_0 w_i + c_0 E_p[y_i] \quad (35)$$

$$\log \text{Var}[y_i] = \alpha + \beta_0 w_i + \gamma_0 \log_p \text{Var}[y_i], \quad (36)$$

which likelihood

$$y_i \sim \mathcal{N}(a + b_0 w_i + c_0 E_p[y_i], e^{\alpha + \beta_0 w_i + \gamma_0 \log \text{Var}[y_i]}). \quad (37)$$

**[0117]** Models in accordance with a number of embodiments of the invention can allow for heteroscedasticity in which the variance of the outcome is correlated with the variance predicted by the digital twin model, and in which

the variance may be affected by the treatment. In several embodiments, a system of GLMs can be fit (e.g., using maximum likelihood, Bayesian approaches, etc.), as was the case for the simpler model. One skilled in the art will clearly recognize that one could also include the interaction or other terms in order to model more complex relationships if necessary. In addition, one skilled in the art will also recognize that including interactions can lead to estimates of conditional average treatment effects in addition to average treatment effects.

**[0118]** An example of borrowing information from digital twins to estimate treatment effects in accordance with an embodiment of the invention is illustrated in FIG. 7. In the first part **705**, a generative model of the control condition is trained using historical data from previously completed clinical trials, electronic health records, or other studies. In the second part **710**, if the analysis to be performed is Bayesian, predictions from the generative model are compared to historical data that were not used to train the model in order to obtain a prior distribution capturing how well the predictions generalize to new populations. A frequentist analysis does not need to obtain a prior distribution. In the third part **715**, a randomized controlled trial is conducted (potentially with unequal randomization), digital twins are generated for each subject in the trial, and all of the data are incorporated into a statistical analysis (including the prior from step **710** if the analysis is Bayesian) to estimate the treatment effects. Bayesian methods, analytical calculations, or the bootstrap may be used to estimate uncertainties in the treatment effects, and decision rules based on p-values or posterior probabilities may be applied.

**[0119]** An example of using generalized linear models and digital twins estimate treatment effects in accordance with an embodiment of the invention is illustrated in FIG. 8. This drawing illustrates the concept using a simple analysis of a continuous outcome. The x-axis represents the prediction for the outcome from the digital twins, and the y-axis represents the observed outcome of the subjects in the RCT. A linear model is fit to the data from the RCT, adjusting for the outcome predicted from the digital twins. If no interactions are included, then two parallel lines are fit to the data: one to the control group and one to the treatment group. The distance between these lines is an estimate for the treatment effect. Both frequentist and Bayesian methods may be used to analyze the generalized linear model.

Using Generative Models to Define Response and Estimate Treatment Effects

**[0120]** In the previous sections, a response is defined in the units of the outcome,  $\gamma$ . For example, an analysis with a survival outcome would produce a treatment effect with the units of time. For example, one may find that the treatment improves survival by 6 months, on average. This is typically useful for aiding interpretation of the treatment effect. However, defining the treatment effect in terms of its natural units may mask important characteristics of the treatment in a population that is heterogeneous. That is, it may be beneficial to measure treatment effects in different ways in order to understand how specific individuals respond to a treatment.

**[0121]** In numerous embodiments, generative models can be used to define individualized responses to treatment. In particular, consider the tail probability of a continuous outcome

$$p_i = \int_{\gamma_i}^{\infty} dy p(y | x_{0,i}) \quad (38)$$

which defines the probability of observing an outcome greater than or equal to  $\gamma_i$  under a generative model of the control condition. That is, there is probability  $p_i$  that subject  $i$  would have an outcome better than  $\gamma_i$  if they had received the control. Note that this example assumes that larger  $\gamma_i$  is better, but it is trivial to consider the opposite case by changing the lower limit of integration to  $-\infty$  and the upper limit to  $\gamma_i$ . In several embodiments,  $p_i$  can be computed using a Monte Carlo estimate by creating  $N$  digital twins with  $\gamma_j \sim p(\gamma | x_i)$  for  $j=1, \dots, N$  and approximating

$$p_i \approx \frac{1}{N} \sum_i \theta(y_j \geq \gamma_i)$$

in which  $\theta(\cdot)$  is a logical function equal to one if the argument is true and zero otherwise.

**[0122]** An example of using a generative model to measure individual treatment responses in accordance with an embodiment of the invention is illustrated in FIG. 9. A generative model can be used to compute a tail-area probability for the observed response. This describes how likely it would be for the patient to demonstrate a better response under the control condition than what was observed in the trial.

**[0123]** In the case of a single subject,  $p_i$  can be interpreted as a measure of evidence against the null hypothesis that the data were drawn from the generative model. Consider the case of a subject in the control arm with  $w_i=0$ . There are two reasons why a small  $p_i$  may be observed in this case: first, it may simply be due to random chance; second, it may indicate that the generative model of the control condition is biased. This could reflect an improperly trained model or a mismatch between the training data and the concurrent control arm of the RCT. Suppose, however, that  $p_i$  is generally large for those subject with  $w_i=0$ . Then, this indicates that the generative model of the control condition is consistent with the data from the concurrent control arm of the RCT. In this case, if a subject with  $w_i=1$  has a small  $p_i$  then this could result from either random chance or a response to the treatment.

**[0124]** The intuition described above suggests two additional uses of generative models of the control condition in the analysis of RCT. First, a generative model can be used to measure the discrepancy between a concurrent control arm in an RCT and its expected behavior from historical data. For example, the average surprise  $\mathcal{S} = -\sum_i (1-w_i) \log p_i$  is a measure of this discrepancy. A large surprise  $\mathcal{S}_c$  either indicates a problem with the generative model or a problem with the control group. There are many other ways to combine the  $p_i$  for the control group into a score to measure discrepancy. Although this analysis cannot definitively determine the cause of the discrepancy, it can flag potential problems that may merit further investigation by clinical trial sponsors or regulatory authorities.

**[0125]** Processes in accordance with numerous embodiments of the invention can fit a linear model to the data from the RCT using  $p_i$  as the measure of response in order to estimate a treatment effect. Processes in accordance with a

variety of embodiments of the invention can use a normalizing transform to define the response as  $\Phi^{-1}(p_i)$  and fit the linear model

$$\Phi^{-1}(p_i) = a + \left( b_0 + \sum_j b_j x_{0,j} \right) w_i + \sum_j d_j x_{ij} + \epsilon_i \quad (39)$$

in which  $\Phi^{-1}(\cdot)$  can be the inverse cumulative distribution function of the standard normal distribution. The parameter  $\alpha$  accounts for the bias of the generative model. As described in other sections, an informative prior for  $\alpha$  can be derived from a regression against historical data in accordance with certain embodiments of the invention. In this case,  $\hat{\alpha} = E[(\Phi^{-1}(p_i))]$  will be equal to 0 if the generative model is unbiased. In several embodiments, analyses can proceed as described throughout for the frequentist and Bayesian cases with this redefined response. The resulting estimate for the treatment effect has the advantage of being individualized by construction, it describes how much better a particular patient responded compared to their distribution of predicted outcomes under the control condition. However, this also has a disadvantage of being impossible to interpret without referring to the generative model. Nevertheless, such approaches may be particularly useful for treatments with heterogeneous effects.

**[0126]** An example of estimating treatment effects with individual treatment responses in accordance with an embodiment of the invention is illustrated in FIG. 10. In part 1005, a generative model of the control condition is trained using historical data from previously completed clinical trials, electronic health records, or other studies. In part 1010, if the analysis to be performed is Bayesian, predictions from the generative model are compared to historical data that were not used to train the model in order to obtain a prior distribution capturing how well the predictions generalize to new populations. A frequentist analysis can skip part 1010. In part 1015, a randomized controlled trial is conducted (potentially with unequal randomization), the generative model is used to define responses for each subject based on transformed tail-area probabilities, and all of the data are incorporated into a statistical analysis (including the prior from step 1010 if the analysis is Bayesian) to estimate the treatment effects. Bayesian methods, analytical calculations, or the bootstrap may be used to estimate uncertainties in the treatment effects, and decision rules based on p-values or posterior probabilities may be applied.

## Systems for Determining Treatment Effects

### Treatment Analysis System

**[0127]** An example of a treatment analysis system that determines treatment effects in accordance with some embodiments of the invention is illustrated in FIG. 11. Network 1100 includes a communications network 1160. The communications network 1160 is a network such as the Internet that allows devices connected to the network 1160 to communicate with other connected devices. Server systems 1110, 1140, and 1170 are connected to the network 1160. Each of the server systems 1110, 1140, and 1170 is a group of one or more servers communicatively connected to one another via internal networks that execute processes that provide cloud services to users over the network 1160. One

skilled in the art will recognize that a treatment analysis system may exclude certain components and/or include other components that are omitted for brevity without departing from this invention.

[0128] For purposes of this discussion, cloud services are one or more applications that are executed by one or more server systems to provide data and/or executable applications to devices over a network. The server systems **1110**, **1140**, and **1170** are shown each having three servers in the internal network. However, the server systems **1110**, **1140** and **1170** may include any number of servers and any additional number of server systems may be connected to the network **1160** to provide cloud services. In accordance with various embodiments of this invention, treatment analysis systems in accordance with various embodiments of the invention may be provided by a process being executed on a single server system and/or a group of server systems communicating over network **1160**.

[0129] Users may use personal devices **1180** and **1120** that connect to the network **1160** to perform processes that determine treatment effects in accordance with various embodiments of the invention. In the shown embodiment, the personal devices **1180** are shown as desktop computers that are connected via a conventional “wired” connection to the network **1160**. However, the personal device **1180** may be a desktop computer, a laptop computer, a smart television, an entertainment gaming console, or any other device that connects to the network **1160** via a “wired” connection. The mobile device **1120** connects to network **1160** using a wireless connection. A wireless connection is a connection that uses Radio Frequency (RF) signals, Infrared signals, or any other form of wireless signaling to connect to the network **1160**. In FIG. 11, the mobile device **1120** is a mobile telephone. However, mobile device **1120** may be a mobile phone, Personal Digital Assistant (PDA), a tablet, a smartphone, or any other type of device that connects to network **1160** via wireless connection without departing from this invention.

[0130] As can readily be appreciated the specific computing system used to determine treatment effects is largely dependent upon the requirements of a given application and should not be considered as limited to any specific computing system(s) implementation.

#### Treatment Analysis Element

[0131] An example of a treatment analysis element that executes instructions to perform processes that determine treatment effects in accordance with various embodiments of the invention is illustrated in FIG. 12. Treatment analysis elements in accordance with many embodiments of the invention can include (but are not limited to) one or more of mobile devices, cloud services, and/or computers. Treatment analysis element **1200** includes processor **1205**, peripherals **1210**, network interface **1215**, and memory **1220**. One skilled in the art will recognize that a treatment analysis element may exclude certain components and/or include other components that are omitted for brevity without departing from this invention.

[0132] The processor **1205** can include (but is not limited to) a processor, microprocessor, controller, or a combination of processors, microprocessor, and/or controllers that performs instructions stored in the memory **1220** to manipulate data stored in the memory. Processor instructions can con-

figure the processor **1205** to perform processes in accordance with certain embodiments of the invention.

[0133] Peripherals **1210** can include any of a variety of components for capturing data, such as (but not limited to) cameras, displays, and/or sensors. In a variety of embodiments, peripherals can be used to gather inputs and/or provide outputs. Treatment analysis element **1200** can utilize network interface **1215** to transmit and receive data over a network based upon the instructions performed by processor **1205**. Peripherals and/or network interfaces in accordance with many embodiments of the invention can be used to gather data that can be used to determine treatment effects.

[0134] Memory **1220** includes a treatment analysis application **1225**, historical data **1230**, RCT data **1235**, and model data **1240**. Treatment analysis applications in accordance with several embodiments of the invention can be used to determine treatment effects of an RCT, to design an RCT, and/or determine decision rules for treatments.

[0135] Historical data in accordance with many embodiments of the invention can be used to pre-train generative models to generate potential outcomes for digital subjects and/or digital twins. In numerous embodiments, historical data can include (but is not limited to) control arms from historical control arms, patient registries, electronic health records, and/or real world data. In many embodiments, predictions from the generative model can be compared to historical data that were not used to train the model in order to obtain a prior distribution capturing how well the predictions generalize to new populations.

[0136] In some embodiments, RCT data can include panel data collected from subjects of a RCT. RCT data in accordance with a variety of embodiments of the invention can be divided into control and treatment arms based on whether subjects received a treatment. In many embodiments, RCT data can be supplemented with generated subject data. Generated subject data in accordance with a number of embodiments of the invention can include (but is not limited to) digital subject data and/or digital twin data.

[0137] In several embodiments, model data can store various parameters and/or weights for generative models. Model data in accordance with many embodiments of the invention can include data for models trained on historical data and/or trained on RCT data. In several embodiments, pre-trained models can be updated based on RCT data to generate digital subjects.

[0138] Although a specific example of a treatment analysis element **1200** is illustrated in this figure, any of a variety of treatment analysis elements can be utilized to perform processes for determining treatment effects similar to those described herein as appropriate to the requirements of specific applications in accordance with embodiments of the invention.

#### Treatment Analysis Application

[0139] An example of a treatment analysis application for determining treatment effects in accordance with an embodiment of the invention is illustrated in FIG. 13. Treatment analysis application **1300** includes digital subject generator **1305**, treatment effect engine **1310**, and output engine **1315**. One skilled in the art will recognize that a treatment analysis application may exclude certain components and/or include other components that are omitted for brevity without departing from this invention.

**[0140]** Digital subject generators in accordance with various embodiments of the invention can include generative models that can generate digital subject and/or digital twin data. Generative models in accordance with certain embodiments of the invention can be trained to generate potential outcome data based on characteristics of an individual and/or a population. Digital subject data in accordance with several embodiments of the invention can include (but is not limited to) panel data, outcome data, etc. In several embodiments, generative models can include (but are not limited to) traditional statistical models, generative adversarial networks, recurrent neural networks, Gaussian processes, auto-encoders, autoregressive models, variational autoencoders, and/or other types of probabilistic generative models.

**[0141]** In various embodiments, treatment effect engines can be used to determine treatment effects based on generated digital subject data and/or data from a RCT. In some embodiments, treatment effect engines can use digital subject data from digital subject generators to determine a treatment effect in a variety of different applications, such as, but not limited to, comparing separate generative models based on data from the control and treatment arms of a RCT, supplementing a control arm in an RCT, comparing predicted potential control outcomes with actual treatment outcomes, etc. Treatment effects engines in accordance with some embodiments of the invention can be used to determine individualized responses to treatment. In certain embodiments, treatment effect engines can determine biases of generative models of the digital subject generator and incorporate the biases (or corrections for the biases) in the treatment effect analyses.

**[0142]** Output engines in accordance with several embodiments of the invention can provide a variety of outputs to a user, including (but not limited to) decision rules, treatment effects, generative model biases, recommended RCT designs, etc. In numerous embodiments, output engines can provide feedback when the results of generative models of a digital subject generator diverge from the RCT population. For example, output engines in accordance with certain embodiments of the invention can provide a notification when a difference between generated control outcomes for digital twins of subjects from a control arm and their actual control outcomes exceeds a threshold.

**[0143]** Although a specific example of a treatment analysis application is illustrated in this figure, any of a variety of Treatment analysis applications can be utilized to perform processes for determining treatment effects similar to those described herein as appropriate to the requirements of specific applications in accordance with embodiments of the invention.

**[0144]** Although specific methods of determining treatment effects are discussed above, many different methods of treatment analysis can be implemented in accordance with many different embodiments of the invention. It is therefore to be understood that the present invention may be practiced in ways other than specifically described, without departing from the scope and spirit of the present invention. Thus, embodiments of the present invention should be considered in all respects as illustrative and not restrictive. Accordingly, the scope of the invention should be determined not by the embodiments illustrated, but by the appended claims and their equivalents.

1. A method for designing a target random control trial, the method comprising:

generating a set of prognostic scores for a set of samples, wherein the set of prognostic scores comprises prognostic scores at each of a plurality of points in time for each sample of the set of samples;

assessing metrics for a set of generative models based on a set of outcomes for the set of samples;

determining a set of target trial parameters for a randomized control trial (RCT) based on the assessed metrics;

generating result data using the set of generative models; and

determining treatment effects for the RCT using the generated result data.

2. The method of claim 1, wherein the set of prognostic scores are generated based on subjects from a control arm of another trial.

3. The method of claim 1, wherein assessing the metrics comprise a correlation.

4. The method of claim 1, wherein the result data comprises panel data from subjects of the RCT and the generated result data comprises predicted panel data for a set of one or more digital subjects, wherein the panel data describes the observed values of multiple characteristics at multiple discrete timepoints.

5. The method of claim 4, wherein:

the predicted panel data for the set of digital subjects is generated based on population statistics of the RCT; and

the generated result data is used to supplement control arm data of the RCT data.

6. The method of claim 4, wherein the predicted panel data for the set of digital subjects is generated based on individual characteristics of the subjects of the RCT.

7. The method of claim 6, wherein determining the treatment effects comprises comparing the predicted panel data based on characteristics of a particular subject with the panel data for the particular subject from the RCT data.

8. The method of claim 1, wherein determining the set of target trial parameters comprises minimizing a total number of samples for the target random control trial.

9. The method of claim 1, wherein determining the set of target trial parameters comprises minimizing a number of samples for the control arm of the target random control trial.

10. The method of claim 1, wherein determining the set of target trial parameters comprises minimizing a number of samples for the treatment arm of the target random control trial.

11. A non-transitory machine readable medium containing processor instructions for designing a target random control trial, where execution of the instructions by a processor causes the processor to perform a process that comprises:

generating a set of prognostic scores for a set of samples, wherein the set of prognostic scores comprises prognostic scores at each of a plurality of points in time for each sample of the set of samples;

assessing metrics for a set of generative models based on a set of outcomes for the set of samples;

determining a set of target trial parameters for a randomized control trial (RCT) based on the assessed bias metrics;

generating result data using the set of generative models; and

determining treatment effects for the RCT using the generated result data.

**12.** The non-transitory machine readable medium of claim **11**, wherein the set of prognostic scores are generated based on subjects from a control arm of another trial.

**13.** The non-transitory machine readable medium of claim **11**, wherein assessing the metrics comprise a correlation.

**14.** The non-transitory machine readable medium of claim **11**, wherein the result data comprises panel data from subjects of the RCT and the generated result data comprises predicted panel data for a set of one or more digital subjects, wherein the panel data describes the observed values of multiple characteristics at multiple discrete timepoints.

**15.** The non-transitory machine readable medium of claim **14**, wherein:

the predicted panel data for the set of digital subjects is generated based on population statistics of the RCT; and

the generated result data is used to supplement control arm data of the RCT data.

**16.** The non-transitory machine readable medium of claim **14**, wherein the predicted panel data for the set of digital subjects is generated based on individual characteristics of the subjects of the RCT.

**17.** The non-transitory machine readable medium of claim **16**, wherein determining the treatment effects comprises comparing the predicted panel data based on characteristics of a particular subject with the panel data for the particular subject from the RCT data.

**18.** The non-transitory machine readable medium of claim **11**, wherein determining the set of target trial parameters comprises minimizing a total number of samples for the target random control trial.

**19.** The non-transitory machine readable medium of claim **11**, wherein determining the set of target trial parameters comprises minimizing a number of samples for the control arm of the target random control trial.

**20.** The non-transitory machine readable medium of claim **11**, wherein determining the set of target trial parameters comprises minimizing a number of samples for the treatment arm of the target random control trial.

\* \* \* \* \*