February 11, 2022

Predicting counterfactual risks under hypothetical treatment strategies: an application to HIV

Barbra A. Dickerman, PhD,[1,2] Issa J. Dahabreh, MD, ScD,[1,2,3] Krystal V. Cantos, PhD,[1,2] Roger W. Logan, PhD,[1,2] Sara Lodi, PhD,[1,2,4] Christopher T. Rentsch, PhD,[5,6,7] Amy C. Justice, MD, PhD,[6,7,8] Miguel A. Hernán, MD, DrPH[1,2,3,9]

Affiliations:
[1] CAUSALab, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, US
[2] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, US
[3] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, US
[4] Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, US
[5] Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK
[6] Veterans Affairs Connecticut Healthcare System, West Haven, Connecticut, US
[7] Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, US
[8] Center for Interdisciplinary Research on AIDS, Yale School of Public Health, New Haven, Connecticut, US
[9] Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts, US

Correspondence: Barbra A. Dickerman, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, 8th Floor, Department of Epidemiology, Boston, MA, 02115. Email: bad788@mail.harvard.edu.

Declarations

Funding: This research was supported by National Institutes of Health grants K99 CA248335 (B.A.D.) and R37 AI02634 and Providence/Boston Center for AIDS Research grant P30 AI042853 (S.L.).

Conflicts of interest/competing interests: None.
Word counts: Abstract, 195; Text, 4,378

ABSTRACT

The accuracy of a prediction algorithm depends on contextual factors that may vary across deployment settings. To address this inherent limitation of prediction, we propose an approach to counterfactual prediction based on the g-formula to predict risk across populations that differ in their distribution of treatment strategies. We apply this to predict 5-year risk of mortality among persons receiving care for HIV in the U.S. Veterans Health Administration under different hypothetical treatment strategies. First, we implement a conventional approach to develop a prediction algorithm in the observed data and show how the algorithm may fail when transported to new populations with different treatment strategies. Second, we generate counterfactual data under different treatment strategies and use it to assess the robustness of the original algorithm's performance to these differences and to develop counterfactual prediction algorithms. We discuss how estimating counterfactual risks under a particular treatment strategy is more challenging than conventional prediction as it requires the same data, methods, and unverifiable assumptions as causal inference. However, this may be required when the alternative assumption of constant treatment patterns across deployment settings is unlikely to hold and new data is not yet available to retrain the algorithm.

Keywords: counterfactual prediction, causal inference, dataset shift, machine learning, parametric g-formula, transportability

Introduction

The performance of a prediction algorithm depends on contextual factors that may vary across populations or time periods [1-4]. Therefore, an algorithm that performs well when developed in a particular database may underperform when applied to individuals from a different population or time period.

As an example, suppose that we were interested in predicting the 5-year risk of death among individuals with HIV in a health system. To do so, we developed a high-performance (e.g., well-calibrated) prediction algorithm using historical data from a sample of health system members. However, if in our healthcare system antiretroviral therapy was reserved for the most severely immunosuppressed individuals during the time period from which historical data were available, then there is no guarantee that the prediction algorithm will retain its high performance if, in the future, antiretroviral therapy becomes available to all individuals with HIV. For example, our prediction algorithm might rely heavily on an initial (baseline) measure of viral load that predicts subsequent 5-year mortality strongly if treatment access is restricted but weakly if there is universal access to treatment (because immediate treatment initiation quickly reduces viral load and thus the baseline value of viral load becomes less relevant). That is, changes in the treatment strategies over time bring about different patterns of associations between the predictors and outcome, which in turn can affect the performance of the prediction algorithm [3, 4].

In the causal inference literature, the extension of inferences from one population to another is referred to as transportability [5, 6]. Adopting this terminology, the above problem is a transportability problem for prediction under a different treatment strategy (but in the same target population). Such problems – sometimes referred to as "dataset shift" or "domain adaptation" problems in computer science – have been recently cited as a major driver of failure of so-called artificial intelligence systems in medical settings [7]. The ideal way to address differences in treatment strategies is to collect updated data to retrain an algorithm to predict 5-year mortality after clinical practice changes [7]. However, we would have to wait at least 5 years to obtain these data, leaving open the question of how to predict risk in the meantime. One option is to use the existing data by reformulating prediction tasks as counterfactual prediction tasks [1-4, 8-10].

Counterfactual prediction requires answering a "what if" question before the prediction takes place [1]. In our example, we would first answer the question "*what if* treatment were available to all individuals with HIV?" by simulating a counterfactual population with the same baseline characteristics as our study population but in which treatment is universally available after baseline, and then develop the prediction algorithm using

those counterfactual data. Sound counterfactual prediction requires the same methods, data, and assumptions as causal inference, which is the task defined by the contrast of two or more counterfactual predictions [11].

Here we describe an application of the g-formula to transport prediction algorithms across populations that differ only in their distribution of treatment strategies. We predict the 5-year mortality risk under different treatment strategies among individuals with HIV in the U.S. Veterans Health Administration, the largest provider of HIV care in the U.S. [12]. In the following sections, we first describe a conventional approach to develop a prediction algorithm in the observed data and discuss the limitations of using such an algorithm under changes in treatment strategy. We then generate counterfactual data under different treatment strategies and use them to assess the robustness of the original algorithm's performance to changes in treatment strategies and to develop counterfactual prediction algorithms.

Conventional (factual) prediction

A (factual) prediction algorithm is a mapping from the candidate predictors V to one or more outcomes Y. In our example, Y is a binary indicator for death (1: dead, 0; alive) within 5 years and V are the characteristics at the start of the 5-year period (baseline) shown in **Table 1**.

The (factual) prediction task is to use the observed data on Y and V to predict the conditional risks $\Pr[Y = 1|V]$ in the Veterans Aging Cohort Study (VACS), which includes individuals living with HIV and accessing care in the Veterans Health Administration. Data on inpatient and outpatient diagnoses, laboratory test results, and dispensed medications (to ascertain treatments) are recorded in electronic medical records during routine clinical care. Deaths are ascertained using inpatient medical records and the VA Beneficiary Identification Records Locator Subsystem mortality database [13].

We focused on 6,707 individuals who were aged 35 years or older with moderate-to-severe immunosuppression (defined as CD4 cell count ≤500 cells/μL), no history of AIDS, no previous use of antiretroviral therapy, and who were receiving care (defined as having CD4 cell count and plasma HIV-RNA viral load measurements within the past 3 months) between January 2000 and August 2012. Most individuals were male (98%), the mean age was 50 years (standard deviation, 8.6 years), the mean CD4 cell count was 223 cells/μL (standard deviation, 147 cells/μL), and the median viral load was 58,887 copies/mL (interquartile range, 15,280-168,000 copies/mL). Over the 5-year

follow-up, 994 individuals died and 2,364 were censored after 12 months without a CD4 cell count or viral load measurement.

We restricted the analysis to uncensored individuals (who are, by definition, the only ones with known Y), under the assumption that a predictive algorithm developed in the uncensored would apply to the entire study population. Below we will relax this assumption. As a first step, we randomly split the observed data into a training set (80%) to develop this algorithm and a test set (20%) to evaluate its performance. The training set included 3,474 individuals (804 deaths) and the test set included 869 individuals (190 deaths). Both sets had an almost identical distribution of baseline characteristics (**Appendix Table 1**).

To accomplish the factual prediction task, we fit a LASSO-regularized logistic regression model to the training set [14]. The LASSO combines coefficient shrinkage with predictor selection (because some coefficients are shrunk to zero) [15]. The optimal regularization parameter λ was selected using 10-fold cross-validation. We used the fitted model to predict the 5-year risk of death for each individual in the test set. Note that this model was developed for illustrative purposes rather than for clinical use; it is distinct from the established predictive index known as the VACS Index, which has been shown to have good discrimination and calibration across various samples of people receiving antiretroviral therapy for HIV infection [16, 17].

We then examined measures of calibration and discrimination in the test set [18]. Calibration refers to how closely the predicted risks agree with the observed outcomes. We evaluated this graphically by generating a local regression-smoothed calibration plot (the 45-degree line in the plot indicates perfect calibration) [18, 19]. Discrimination refers to how well a prediction model can distinguish individuals who do vs. do not experience the outcome of interest. We assessed this by calculating the c-statistic, which is the area under the receiver-operating characteristic curve for binary outcomes and the probability that the model will assign a higher predicted risk to a randomly selected person who experiences the outcome than a randomly selected person who does not (a c-statistic value of 1 indicates perfect discrimination, while a value of 0.5 indicates no discriminatory ability). All analyses in this paper were conducted using R version 4.0.4 and SAS version 9.4 (SAS Institute, Inc., Cary, NC, USA).

LASSO predictor selection resulted in a model that included the baseline predictors age, CD4 cell count, viral load, history of infection with hepatitis C virus, and HIV diagnosis in the past year (**Appendix Table 2**). This prediction algorithm showed good discrimination (c-statistic 0.71) and calibration in the test set (**Figure 1a**).

The limits of factual prediction

Our prediction algorithm is expected to perform differently in populations in which the relations between the predictors V and Y differ from those in our training data. A shift in the V-Y relation across populations or datasets may be due to several reasons. To see this, consider the simplified setting with no censoring represented by the causal directed acyclic graph (DAG) in **Figure 2**.

The node Y denotes death status at 5 years. The nodes P_0 and P_1 denote predictors of mortality measured at the start of baseline month 0 (when the predictive algorithm is developed) and of month 1, respectively. The nodes A_0 and A_1 denote treatment (antiretroviral therapy) initiation by the end of month 0 and of month 1, respectively. The nodes L_0 and L_1 denote predictors of both treatment and mortality measured at the start of baseline month 0 and of month 1, respectively. The causal DAG also includes unmeasured variables U, which are therefore not usable in our analysis. For simplicity of presentation, our causal DAG only includes the predictors and treatments at baseline and one month later (in practice, they are measured in each of the 60 months of follow-up), and it assumes no censoring throughout the follow-up and no deaths in the first two months.

The baseline predictors V used to develop the predictive algorithm are selected from the union of L_0 and P_0. The predictive algorithm cannot include the variables A_0, L_1, A_1, and P_1 (because they are not yet measured at the start of month 0 when the algorithm is developed), but the strength and direction of the associations between the V variables and the outcome Y depend on the variables A_0, L_1, A_1, and P_1 (because they are part of the causal structure that links V and Y). A predictive algorithm that is optimal in one population may be suboptimal in another population if the two populations differ in the magnitude or direction of any of 3 types of associations between V and Y: 1) those that involve only baseline variables, including the direct causal effects of V on Y (represented by the direct arrows from L_0 and P_0 to Y) and the effect of U (a common cause of L_0 and Y), 2) those that involve post-baseline variables except for treatment, that is, the indirect causal effects of V on Y in the absence of treatment (represented by the causal pathways from L_0 and P_0 to Y through L_1 and P_1 but not through A_0 and A_1), and 3) those that involve post-baseline treatment, that is, the indirect causal effects of V on Y through treatment (represented by the causal pathways from L_0 to Y through A_0 and A_1 but not through P_1 and L_1).

It follows that, even when considering two populations with identical baseline characteristics L_0 and P_0 (and therefore the potential subset V) and with an identical causal structure linking V to the outcome Y, the associations between V and Y are

expected to differ depending on the treatment strategies applied to each population. This is precisely the problem considered in this paper. That is, when developing an algorithm to be deployed in a healthcare system over time, we will need to adapt or retrain the algorithm when the distribution of treatment strategies in the system shifts. Because outcome data will not be immediately available after such treatment changes, a timely retraining of the algorithm is not possible. Therefore, we propose an approach to generate counterfactual data under hypothetical treatment strategies.

Counterfactual prediction

The data under different treatment strategies could be generated by conducting a trial in which individuals sampled from the target population are randomly allocated to one of several treatment strategies. Conceptualizing an observational analysis as the emulation of a target trial may help reduce bias when estimating counterfactual risks [20]. The protocol components of our target trial (including the same eligibility criteria, outcome, and follow-up described previously) are described in **Table 2**.

We considered three treatment strategies g: (1) immediate initiation of antiretroviral therapy at baseline and continuation over follow-up; (2) initiation upon an AIDS diagnosis or a CD4 cell count <350 cells/μL and continuation over follow-up; and (3) no initiation over follow-up. There is a 3-month grace period for initiating treatment under strategies 1 and 2 so that individuals who have not yet initiated treatment by the end of the grace period would start at that time. Strategy 3 is not a clinically realistic treatment strategy but is included here as an extreme example of shift. In this target trial, the 5-year mortality risk had individuals adhered to their assigned treatment strategy g is an unbiased estimate of the counterfactual 5-year mortality risk had all individuals in the population adhered to that strategy $\Pr[Y^g = 1]$ [21]. All treatment strategies discussed here implicitly include an intervention to eliminate censoring during follow-up.

We emulated each component of the protocol of this target trial using the observational VACS data to obtain estimates of $\Pr[Y^g = 1]$ under each strategy g. A correct emulation requires data on the eligibility criteria and, at each time k, the outcome Y_k, censoring C_k, treatment A_k, and confounders L_k under the assumption of no unmeasured confounding given L_k (encoded in the causal DAG above through the absence of direct arrows from the unmeasured variables U to the treatment variables A_k) and the assumption that censored and uncensored individuals in the original population had the same risk conditional on the measured variables L_k [22]. That is, these assumptions are required when generating counterfactual data under any treatment strategy g. We used the g-formula [23, 24], a generalization of standardization to time-varying treatments that appropriately handles treatment-confounder feedback, that is, a setting like ours in

which the measured confounders (e.g., CD4 cell count) are affected by prior treatment (antiretroviral therapy), to generate counterfactual data under these assumptions.

The g-formula risk under strategy g is a weighted average (integral) of the mortality risk conditional on covariate history and weighted by the distribution of covariate histories compatible with the conditional distribution of covariates under g. We implemented a parametric g-formula algorithm in two stages, as previously described [25]. First, we fit parametric models for the distribution of outcome and covariates over time using maximum likelihood methods. See **Appendix Table 3** and **Appendix 1** for details on the variables and models used in our g-formula analysis of the VACS data. Second, we used a Monte Carlo simulation to approximate the weighted average under each strategy g. By generating indicators for death for individuals, a byproduct of the Monte Carlo simulation is a dataset comprised of individuals' trajectories had everyone followed treatment strategy g. We repeated the process under each of the three treatment strategies to obtain three datasets, each representing the counterfactual trajectory of 6,707 individuals following each treatment strategy. Each dataset had an identical distribution of baseline predictors. Increasing the Monte Carlo simulation sample size to 100,000 did not affect the results described below.

Assessing the performance of the original prediction model under treatment shifts

We applied the original prediction model to the three counterfactual datasets simulated under hypothetical treatment strategies to predict individuals' 5-year risk of death. We calculated the measures of calibration and discrimination in the same 20% of the data in each hypothetical setting, so that any differences in performance across settings could be solely attributed to differences in treatment strategies.

We found that the original prediction model's discriminatory ability was largely preserved (c-statistics ranging from 0.70-0.73) but its predictions were miscalibrated under all hypothetical treatment strategies (**Figure 3**). Risk was overpredicted under strategies that started treatment sooner (**Figure 3a-b**). This occurred because these treatment strategies disrupted the associations between the predictors and death, such that higher CD4 cell count was associated with a lower risk of death under strategies that start treatment sooner, which was not captured by the original prediction model (**Appendix Figure 1**). Risk was underpredicted under the never treat strategy (**Figure 3c**). This occurred because, under the never treat strategy, the associations between key predictors such as CD4 cell count and death were strongest, which again was not captured by the original prediction model developed in a setting where the observed treatment pattern had dampened these associations (**Appendix Figure 1**).

Developing counterfactual prediction models

The simulated counterfactual trajectories also provide an opportunity to develop counterfactual prediction models for use in settings where the original prediction model is expected to perform poorly but new data are unavailable to adapt or retrain the model. A counterfactual prediction model uses simulated values of the counterfactual outcome Y^g and observed data on baseline predictors V to predict a given individual's counterfactual risk under a given treatment strategy, $\Pr[Y^g = 1|V]$. Developing such a model involves the same approach used to develop the original prediction model, except now we use the counterfactual outcomes under hypothetical treatment strategies.

Specifically, we fit new LASSO-regularized logistic regression models to the training sets simulated under each strategy to predict the 5-year risk of death had everyone received treatment under that strategy, $\Pr[Y^g = 1|V]$. This model considered information on the same candidate predictors as the original prediction model, counterfactual outcomes under the new strategies, and no information on postbaseline treatment initiation. LASSO predictor selection resulted in models that included the same predictors as in the original prediction model, except for diagnosis of HIV in the past year (**Appendix Table 2**).

To assess the performance of the counterfactual prediction models, we applied them to test sets simulated under each strategy (each again using the same 20% of the data) to predict 5-year risk of death and calculated the same measures of calibration and discrimination, as previously described. Unlike the original prediction model, the counterfactual prediction showed good calibration under all the treatment strategies we considered (**Figure 3d-f**). Note that our primary analysis involved fitting g-formula models to the entire observed dataset to generate counterfactual data which were then split into train and test sets, which ensures that the train and test sets have the same exact data generating process at the potential expense of overfitting if the sample is not large. Results were similar when we estimated g-formula models separately in the training and test sets.

Assessing the performance of the original prediction algorithm under the natural course

To informally assess the adequacy of our procedure, we also simulated a dataset under the natural course, that is, under the same distribution of treatment strategies as in the observed data and under no censoring. If our counterfactual prediction methodology is correct, we expect that the original algorithm will have a similar performance in this simulated data as in the observed data.

A comparison of the performance of the algorithm in the observed data and in the data simulated under the natural course requires consideration of how censoring by loss to follow-up was handled, because approximately one-third of individuals were lost to follow-up in the observed data and therefore had no known value of Y.

A common approach in the machine learning literature for addressing censoring by loss to follow-up is to restrict the analysis to individuals with known value of Y. In keeping with common practice, this is the approach that we used to develop the original prediction model. However, this approach makes the often-unrealistic assumption that censored and uncensored individuals are marginally (unconditionally) exchangeable in terms of outcome risk. By contrast, when generating the natural course data without censoring, we made the weaker assumption that censored and uncensored individuals are only exchangeable within strata defined by covariate history at each time point.

A better approach would be to use inverse-probability weighting [22] to "eliminate" censoring in the observed data before the prediction algorithm is developed. This approach relies on the same conditional exchangeability assumption made when generating the natural course data under no censoring, permitting a direct comparison of algorithm performance across datasets with the same distribution of treatment strategies and also with no censoring.

In this case, results were similar regardless of whether the original prediction model was developed with or without inverse-probability weighting in the observed data (**Appendix Figure 2**). This similarity suggests that the original algorithm was not sensitive to the choice between censoring assumptions in this particular application. Also, the original prediction model showed similar performance in the observed data (without inverse probability weighting) and in the natural course data (similar c-statistics and calibration curves, as shown in **Figure 1**). This similarity supports the adequacy of our simulation procedure for counterfactual prediction.

DISCUSSION

In an application of counterfactual prediction to mortality risk among individuals receiving care for HIV in the largest integrated healthcare system in the U.S., we found evidence that a conventional (factual) prediction model would yield systematically miscalibrated predictions when applied to new populations with different treatment patterns. The miscalibration was most pronounced under strategies with extreme differences in access to treatment (e.g., no access) but also evident under clinically

realistic treatment strategies that depend on evolving risk factors, largely because treatment disrupted the associations between those risk factors and the outcome. By contrast, a counterfactual prediction model showed good calibration under all treatment strategies considered.

A common goal is to transport a prediction model developed in one population to a new population with a different data distribution. This may involve populations that have (1) different distributions of treatment strategies over time and different distributions of covariates at baseline, in which case both differences need to be accounted for, or (2) different distributions of treatment strategies but the same distributions of covariates at baseline. Here we consider the second case to introduce methodology for counterfactual prediction under hypothetical treatment strategies.

Regardless of which components of the joint distribution of the variables differ across populations, the most natural solution to this transportability problem is to use data from the new population to adapt or retrain a factual prediction model [7, 26-28]. However, this requires data from the new population. If those data are not available, then we need to recast the prediction question as a counterfactual prediction question, which fundamentally changes the data analysis. In fact, counterfactual prediction requires the same data (eligibility criteria, outcome, censoring, time-varying treatments, time-varying confounders), assumptions (exchangeability, positivity, consistency [22]) and methods as those required for causal inference [11].

In this application, we focused on counterfactual prediction to address treatment strategy shift within one healthcare system over time. We first took advantage of parametric g-formula procedures to generate counterfactual datasets had everyone been treated under one of three treatment strategies. We then applied a predictive algorithm (LASSO-regularized logistic regression in our example) to predict the counterfactual risk under a given treatment strategy. Therefore, via parametric g-formula calculations, we estimated the joint distribution of the observed data and used it to simulate counterfactual data that differed only in their distribution of treatment strategies, assuming that other components of the joint distribution of the variables remained the same. A similar approach can be used in the presence of other types of differences in the joint distribution or “dataset shift”.

Counterfactual prediction to address changes in treatment strategies requires rich data on time-varying treatments and confounders, like the VACS data. When these data are available, our methodology has several strengths.

First, our approach makes it explicit that the validity of counterfactual prediction under different treatment strategies relies on the same conditions as the validity of extending causal inferences for treatment effects across populations. These conditions include exchangeability between treatment groups and positivity of treatment assignment within a given population, as well as exchangeability between, and positivity of inclusion across, different populations [6]. In this particular application, we focus on changes in treatment strategies within a single population over time and so we are not concerned with transportability between populations or positivity of being in the data from the target population, but these may be important considerations in other applications. In any case, counterfactual prediction tasks require the application of well-established methods for causal inference. Causal directed acyclic graphs can be used to make assumptions explicit [4, 29-31], and g-methods can be used to carry out the data analysis under those assumptions [32].

Second, we used the g-formula to appropriately handle treatment-confounder feedback, that is, the setting in which the decision to treat is based on patients' values of evolving risk factors that are also affected by previous treatment. Treatment-confounder feedback has not been explicitly considered by previously proposed approaches to counterfactual prediction [29, 30]. Inverse-probability weighting, which has formed the basis of some previous applications [2, 8, 9], can also appropriately handle treatment-confounder feedback. It does not involve the Monte Carlo simulation procedure that we leveraged here to generate counterfactual individual trajectories.

Third, the parametric g-formula is a flexible g-method to generate counterfactual data under any type of treatment strategy. Our application involved sustained treatment strategies, in which treatment initiation was either static (start immediately or never) or dynamic (start upon AIDS diagnosis or when CD4 cell count drops below a certain level), but the same methodology can accommodate arbitrarily complex strategies (e.g., involving multiple treatments).

When using counterfactual prediction to address differences in treatment patterns, the possibility of different versions of treatment must be considered [33]. Here we used a treatment variable which reflects the distribution of antiretroviral therapies available during the study period to estimate counterfactual risks after shifts in treatment strategies. That is, we considered two time periods over which the rules for determining treatment assignment are different but the versions of treatment itself are assumed to be the same (or different in ways that do not matter) [34]. If different versions of treatment are deployed across settings and those versions have different effects on the counterfactual outcomes of interest, then our counterfactual predictions will be incorrect unless they incorporate information on different versions of treatment. If these different

versions involve a treatment not currently available in the data, then counterfactual prediction will not be generally possible.

In summary, we propose methodology for counterfactual prediction as an interim mitigation strategy for shifts in treatment patterns while awaiting new data to retrain a prediction algorithm. Specifically, our proposed approach to counterfactual prediction permits the generation of counterfactual data under hypothetical treatment strategies, which can be used to assess the robustness of a prediction model to particular shifts, and, if needed, develop counterfactual prediction models. Estimating counterfactual risks under a particular treatment strategy is more challenging than conventional prediction as it requires the same data, methods, and unverifiable assumptions as causal inference. However, this may be necessary when the alternative assumption of constant treatment patterns across deployment settings is unlikely to hold and a factual prediction model cannot be assumed to retain good performance when transported to new patient populations.

References

1. Dickerman BA, Hernán MA. Counterfactual prediction is not only for causal inference. European Journal of Epidemiology. 2020;35(7):615-7. doi:10.1007/s10654-020-00659-8
2. van Geloven N, Swanson SA, Ramspek CL, et al. Prediction meets causal inference: the role of treatment in clinical prediction models. European Journal of Epidemiology. 2020;35:619-30.
3. Schulam P, Saria S. Reliable decision support using counterfactual models. Advances in Neural Information Processing Systems. 2017. pp. 1697–708.
4. Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. Biostatistics. 2020;21(2):345-52. doi:10.1093/biostatistics/kxz041
5. Dahabreh IJ, Hernan MA. Extending inferences from a randomized trial to a target population. Eur J Epidemiol. 2019;34(8):719-22. doi:10.1007/s10654-019-00533-2
6. Dahabreh IJ, Robertson SE, Steingrimsson JA, Stuart EA, Hernan MA. Extending inferences from a randomized trial to a new target population. Stat Med. 2020;39(14):1999-2014. doi:10.1002/sim.8426
7. Finlayson SG, Subbaswamy A, Singh K, et al. The Clinician and Dataset Shift in Artificial Intelligence. N Engl J Med. 2021;385(3):283-6. doi:10.1056/NEJMc2104626
8. Sperrin M, Martin GP, Pate A, Van Staa T, Peek N, Buchan I. Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. Stat Med. 2018;37(28):4142-54. doi:10.1002/sim.7913
9. Pajouheshnia R, Peelen LM, Moons KGM, Reitsma JB, Groenwold RHH. Accounting for treatment use when validating a prognostic model: a simulation study. BMC Med Res Methodol. 2017;17(1):103. doi:10.1186/s12874-017-0375-8
10. Lin L, Sperrin M, Jenkins DA, Martin GP, Peek N. A scoping review of causal methods enabling predictions under hypothetical interventions. Diagn Progn Res. 2021;5(1):3. doi:10.1186/s41512-021-00092-9
11. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. Chance. 2019;32(1):42-9.
12. U.S. Department of Veterans Affairs. Veterans Affairs HIV Program Fact Sheet. 2020. https://www.hiv.va.gov/pdf/HIV-program-factsheet.pdf.
13. Justice AC, Dombrowski E, Conigliaro J, et al. Veterans Aging Cohort Study (VACS): Overview and description. Med Care. 2006;44(8 Suppl 2):S13-24. doi:10.1097/01.mlr.0000223741.02074.66
14. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological). 1996;58(1):267-88.

15. Steyerberg EW, Eijkemans MJC, Habbema JDF. Application of shrinkage techniques in logistic regression analysis: a case study. Stat Neerl. 2001;55(1):76-88. doi:10.1111/1467-9574.00157
16. Tate JP, Justice AC, Hughes MD, et al. An internationally generalizable risk index for mortality after one year of antiretroviral therapy. AIDS. 2013;27(4):563-72. doi:10.1097/QAD.0b013e32835b8c7f
17. Tate JP, Sterne JAC, Justice AC, Veterans Aging Cohort Study and the Antiretroviral Therapy Cohort Collaboration. Albumin, white blood cell count, and body mass index improve discrimination of mortality in HIV-positive individuals. AIDS. 2019;33(5):903-12. doi:10.1097/QAD.0000000000002140
18. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Annals of Internal Medicine. 2015;162(1):W1-W73. doi:10.7326/M14-0698
19. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med. 2014;33(3):517-35. doi:10.1002/sim.5941
20. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183(8):758-64. doi:10.1093/aje/kwv254
21. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med. 2017;377(14):1391-8. doi:10.1056/NEJMsm1605385
22. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2020.
23. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period — application to the healthy worker survivor effect [published errata appear in Mathl Modelling 1987;14:917-21]. Mathematical Modelling. 1986;7:1393-512.
24. Taubman SL, Robins JM, Mittleman MA, Hernán MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. Int J Epidemiol. 2009;38(6):1599-611. doi:10.1093/ije/dyp192
25. Young JG, Cain LE, Robins JM, O'Reilly EJ, Hernan MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. Stat Biosci. 2011;3(1):119-43. doi:10.1007/s12561-011-9040-7
26. Sugiyama M, Krauledat M, Müller KM. Covariate shift adaptation by importance weighted cross validation. Journal of Machine Learning Research. 2007;8:985-1005.
27. Gretton A, Smola A, Huang J, Schmittfull M, Borgwardt K, Schölkopf B. Covariate shift by kernel mean matching. In: Quiñonero-Candela J, Sugiyama, M.,

Schwaighofer, A. and Lawrence, N. D., editor. Dataset Shift in Machine Learning. Cambridge, MA: The MIT Press. p. 131-60.
28. Steingrimsson JA, Gatsonis C, Dahabreh IJ. Transporting a prediction model for use in a new target population. 2021;https://arxiv.org/abs/2101.11182v2.
29. Subbaswamy A, Saria S. Counterfactual normalization: Proactively addressing dataset shfit and improving reliability using causal mechanisms. Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence. 2018. pp. 947-57.
30. Subbaswamy A, Schulam P, Saria S. Preventing failures due to dataset shift: Learning predictive models that transport. Artificial Intelligence and Statistics (AISTATS). 2019.
31. Dahabreh IJ, Robins JM, Haneuse S, Hernán MA. Generalizing causal inferences from randomized trials: counterfactual and graphical identification. 2019;https://arxiv.org/abs/1906.10792v1.
32. Robins J, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. Stat Med. 2008;27(23):4678-721. doi:10.1002/sim.3301
33. Hernan MA, VanderWeele TJ. Compound treatments and transportability of causal inference. Epidemiology. 2011;22(3):368-77. doi:10.1097/EDE.0b013e3182109296
34. VanderWeele TJ, Hernan MA. Causal Inference Under Multiple Versions of Treatment. J Causal Inference. 2013;1(1):1-20. doi:10.1515/jci-2012-0002

Table 1. Baseline characteristics of 6,707 eligible individuals, Veterans Aging Cohort Study, 2000-2012.[a]

Characteristic, %	
Age (years)	
<45	29
45-49.9	22
50-54.9	21
55-59.9	14
≥60	13
Male	98
CD4 count (cells/μL)	
<100	27
100-199	19
200-299	20
300-399	19
≥400	15
HIV-RNA viral load (copies/mL)	
<10,000	20
10,000-49,999	27
50,000-99,999	17
100,000-199,999	15
≥200,000	22
History of infection with hepatitis C virus	11
Diagnosis of HIV in the past year	13
History of prophylaxis for opportunistic infections[b]	29

Percentages may not sum to 100% due to rounding.

[a] Baseline ranges from January 2000 to August 2012. Eligible individuals had complete data on all variables.

[b] Includes treatment with atovaquone, dapsone, pentamidine isethionate, or combinations of sulfonamides and trimethoprim, including derivatives.

Table 2. Specification and emulation of a target trial of antiretroviral therapy initiation strategies and risk of death using observational data from the Veterans Aging Cohort Study.

Protocol component	Target trial specification	Target trial emulation
Eligibility criteria	• Aged ≥35 • Diagnosis of HIV-1 infection • Moderate-to-severe immunosuppression (defined as CD4 count ≤500 cells/μL) • No history of AIDS • No previous antiretroviral therapy • Receiving care (defined as CD4 count and HIV-RNA viral load measurements within the past 3 months) between January 2000 and August 2012 Baseline is defined as the first month in which all eligibility criteria are met.	Same as for the target trial.
Treatment strategies	(1) Initiation of cART immediately, defined as within 3 months of baseline and irrespective of CD4 count, and continuation over follow-up (2) Initiation of cART within 3 months of an AIDS diagnosis or a CD4 count <350 cells/μL and continuation over follow-up (3) No initiation of cART over follow-up cART is defined as treatment with (1) boosted PI + ≥2 NRTI, (2) NNRTI + ≥2 NRTI, (3) INSTI + ≥2 NRTI, (4) entry inhibitors + ≥2 NRTI, or (5) dual therapy (coformulated medications including dolutegravir + rilpivirine (Juluca), dolutegravir + 3TC (Dovato), raltegravir + 3TC (Dutrebis); any combination with boosted PI + 3TC; boosted PI + INSTI). Individuals must have a clinical visit at least once every 12 months to assess prognostic factors (CD4 count, HIV-RNA viral load) associated with adherence and loss to follow-up.	Same as for the target trial. We considered cART to be continuous after initiation. AIDS diagnosis and CD4 count were ascertained by the treating physicians.
Treatment assignment	All individuals are randomly assigned to a strategy at baseline, and individuals and their treating physicians will be aware of the assigned treatment strategy.	Same as for the target trial with adjustment for baseline confounders in an attempt to emulate randomization (within covariate strata).
Outcomes	All-cause mortality.	Same as for the target trial. Deaths were ascertained using inpatient medical records and the VA Beneficiary Identification Records Locator Subsystem mortality database.
Follow-up	Starts at baseline and ends at the month of death, incomplete follow-up (12 months after the last recorded prognostic factors), 5 years after baseline, or administrative end of follow-up (September 2017), whichever happens first.	Same as for the target trial.
Statistical analysis	Parametric g-formula to estimate 5-year risk of death under each treatment strategy, with adjustment for pre- and post-baseline prognostic factors associated with adherence and loss to follow-up.	Same as for the target trial with adjustment for baseline confounders.

Abbreviations: cART, combined antiretroviral therapy; INSTI, integrase strand transfer inhibitor; NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside/nucleotide reverse transcriptase inhibitor; PI, protease inhibitor.

Supplementary Appendix to:

Predicting counterfactual risks under hypothetical treatment strategies: an application to HIV

Barbra A. Dickerman, PhD, Issa J. Dahabreh, MD, ScD, Krystal V. Cantos, PhD, Roger W. Logan, PhD, Sara Lodi, PhD, Christopher T. Rentsch, PhD, Amy C. Justice, MD, PhD, Miguel A. Hernán, MD, DrPH

TABLE OF CONTENTS

Appendix Table 1. Baseline characteristics of individuals in the training and test sets, Veterans Aging Cohort Study, 2000-2012.[a]

Characteristic, %	**Training set** (N=3,474)	**Test set** (N=869)
Age (years)		
<45	25	26
45-49.9	22	20
50-54.9	22	24
55-59.9	15	13
≥60	16	17
Male	98	98
CD4 count (cells/μL)		
<100	30	28
100-199	19	21
200-299	19	19
300-399	19	19
≥400	13	13
HIV-RNA viral load (copies/mL)		
<10,000	17	19
10,000-49,999	25	25
50,000-99,999	18	16
100,000-199,999	16	15
≥200,000	25	24
History of infection with hepatitis C virus	11	12
Diagnosis of HIV in the past year	11	9
History of prophylaxis for opportunistic infections[b]	31	31

Percentages may not sum to 100% due to rounding.

[a] Baseline ranges from January 2000 to August 2012.

[b] Includes treatment with atovaquone, dapsone, pentamidine isethionate, or combinations of sulfonamides and trimethoprim, including derivatives.

Appendix Table 2. LASSO-regularized logistic regression model coefficients of the factual and counterfactual prediction models, Veterans Aging Cohort Study, 2000-2017.[a]

	Factual prediction model	Counterfactual prediction models		
Predictors		**Start antiretroviral therapy immediately**	**Start antiretroviral therapy upon AIDS diagnosis or CD4 count <350 cells/μL**	**Never start antiretroviral therapy**
Intercept	-1.4912	-1.9247	-1.8016	-0.8081
Age	0.3194	0.3533	0.3275	0.4450
Sex	--	--	--	--
CD4 count (cells/μL)	-0.3457	-0.3841	-0.2955	-0.4868
HIV-RNA viral load (copies/mL)	0.0149	0.0106	0.0263	0.0287
History of infection with hepatitis C virus	1.3207	1.1706	1.1789	1.2408
Diagnosis of HIV in the past year	0.3851	--	--	--
History of prophylaxis for opportunistic infections	--	--	--	--

[a] HIV-RNA viral load was *ln*-transformed and CD4 count was *sqrt*-transformed. Continuous predictors (age, CD4 count, HIV-RNA viral load) were then standardized to mean = 0, standard deviation = 1 based on the training set distributions.

Ap_BDpendix Table 3. Variables used to model 5-year risk of death among individuals with HIV, Veterans Aging Cohort Study, 2000-2017.

A. Time-fixed variables	Functional form as predictor	Variable name	Categories
Age	Linear	age_0	N/A
Sex	Indicator	sex	Female/Male
CD4 count (*sqrt*-transformed)	Linear	sqrtcd40	N/A
HIV-RNA viral load (*ln*-transformed)	Linear	lnrna0	N/A
History of infection with hepatitis C virus	Indicator	hcv_pos_ab_v	Yes/No

B. Time-varying variables	Modeling as dependent	Variable name	Functional form as predictor
Month of follow-up	Not predicted	month2	Restricted cubic splines, 3 knots at 5, 28, and 53 months
Visit to measure CD4 count	Logistic[a]	visit_cd4	Linear (months since the last measurement)
CD4 count (*sqrt*-transformed)	Linear[b]	sqrtcd4	Indicators (current and lagged values)
Visit to measure HIV-RNA viral load	Logistic[a]	visit_rna	Linear (months since the last measurement)
HIV-RNA viral load (*ln*-transformed)	Logistic, then log-linear[c]	lnrna	Indicators (current and lagged values)
Combined antiretroviral therapy	Logistic to failure[d]	everhaart	Indicator (combined antiretroviral therapy), Linear (months since its initiation)
Diagnosis of an AIDS-defining illness	Logistic to failure[d]	aids	Indicator (AIDS-defining illness), Linear (months since diagnosis)

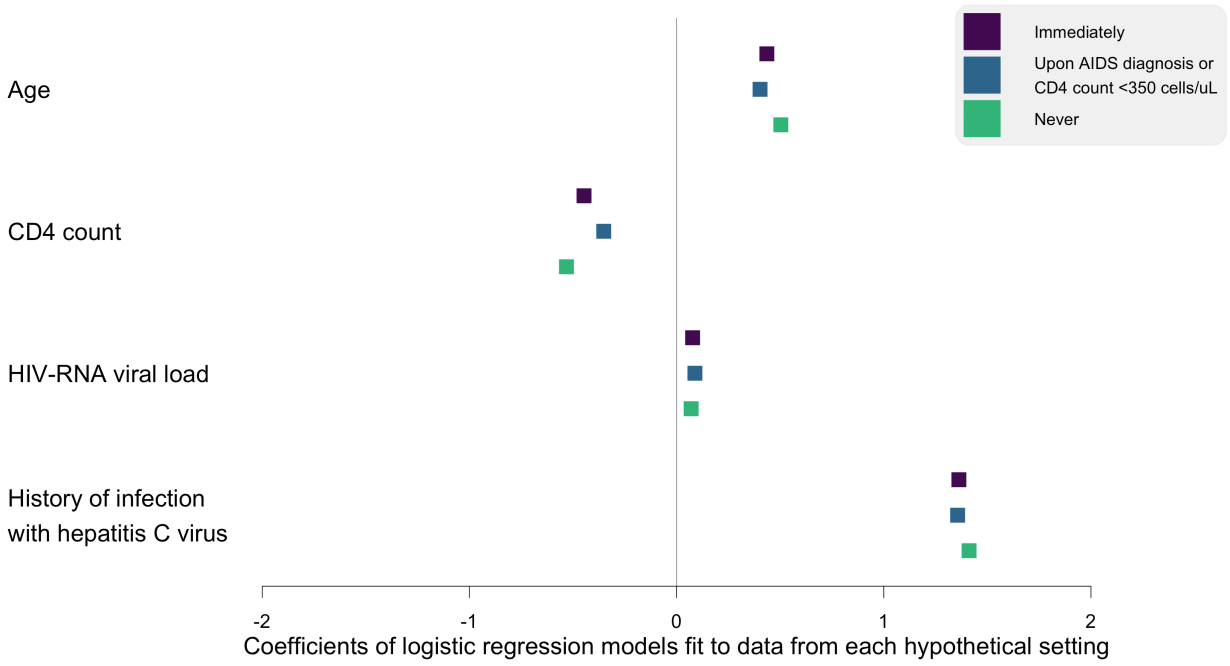
[a] Fits logistic model for an indicator of the visit process, which is assumed to be an additional time-varying confounder. For the simulation, a visit indicator is generated based on the logistic model parameters. When the simulated indicator is 1, the measurement value is simulated as described below. When the simulated indicator is 0, the measurement value is carried forward from the last simulated value. Values are carried forward for up to 12 months, after which the simulated visit indicator is set to 1.

[b] Fits linear model to records with a simulated visit, described above. For the simulation, variables predicted by a linear model were assigned a value equal to the predicted value plus the standard error multiplied by a random number from a Normal (0,1) distribution. Therefore, two subjects with the same risk factor history were not necessarily predicted to have exactly the same risk factor value at the next time point. Simulated values of continuous risk factors were truncated so that they did not fall outside of the observed range.

[c] Variables with many zero values were predicted in two stages. First, we fit a logistic model for an indicator that the variable is nonzero. Second, we fit a linear model for the natural log of the nonzero values. Simulated values were truncated so that they did not fall outside of the observed range.

[d] Fits logistic model only to records where the first lagged value of the variable equals zero. For the simulation, variables predicted by the logistic model were assigned a value of 1 if the predicted probability was greater than a random number from a uniform distribution. After the first 1 is generated, the value was set to 1 thereafter.

Appendix Figure 1. Associations between key predictors and death in hypothetical settings defined by different antiretroviral therapy initiation strategies, Veterans Aging Cohort Study. Logistic regression models were fit to the same training set observations in each hypothetical setting and included the same predictors for direct comparability. HIV-RNA viral load was *ln*-transformed and CD4 count was *sqrt*-transformed. Continuous predictors (age, CD4 count, HIV-RNA viral load) were then standardized to mean = 0, standard deviation = 1.



Appendix 1 Models used in the parametric g-formula

Model 1 Outcome (death)

The LOGISTIC Procedure

Model Information

Data Set	WORK.PARAM
Response Variable	event
Number of Response Levels	2
Weight Variable	_weight_
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	286167
Number of Observations Used	283803
Sum of Weights Read	286167
Sum of Weights Used	283803

Response Profile

Ordered Value	event	Total Frequency	Total Weight
1	1	994	994.00
2	0	282809	282809.00

Probability modeled is event=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.6780	0.3572	462.0592	<.0001
age_0	1	0.0530	0.00358	218.9970	<.0001
sex	1	-0.3273	0.2926	1.2514	0.2633
sqrtcd40	1	0.0561	0.00879	40.7161	<.0001
lnrna0	1	0.0796	0.0222	12.8405	0.0003
hcv_pos_ab_v	1	0.9952	0.0753	174.6497	<.0001
month2	1	-0.00615	0.00543	1.2849	0.2570
month2_spl1	1	0.0148	0.00722	4.2051	0.0403
sqrtcd4_l1	1	0.0521	0.0195	7.1545	0.0075
sqrtcd4	1	-0.2136	0.0194	120.9885	<.0001
ts_last_sqrtcd4	1	-0.0541	0.0290	3.4830	0.0620
lnrna	1	0.0280	0.00912	9.4430	0.0021
ts_last_lnrna	1	0.1005	0.0284	12.5317	0.0004
everhaart	1	-0.6132	0.0988	38.5058	<.0001
tseverhaart_inter	1	0.00631	0.00417	2.2851	0.1306
aids	1	1.4808	0.1194	153.7881	<.0001
tsaids_inter	1	-0.0277	0.00617	20.2034	<.0001

Model 2 Visit to measure CD4 count

The LOGISTIC Procedure

Model Information

Data Set	WORK.PARAM
Response Variable	visit_cd4
Number of Response Levels	2
Weight Variable	_weight_
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	279220
Number of Observations Used	279220
Sum of Weights Read	279220
Sum of Weights Used	279220

Response Profile

Ordered Value	visit_cd4	Total Frequency	Total Weight
1	1	67328	67328.00
2	0	211892	211892.00

Probability modeled is visit_cd4=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	308471.78	303766.83
SC	308482.32	303946.00
-2 Log L	308469.78	303732.83

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4736.9498	16	<.0001
Score	4781.0516	16	<.0001
Wald	4680.2110	16	<.0001

The SAS System

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.1690	0.0508	1824.8048	<.0001
age_0	1	0.00805	0.000527	232.8895	<.0001
sex	1	-0.00298	0.0295	0.0102	0.9194
sqrtcd40	1	-0.00404	0.00119	11.4459	0.0007
lnrna0	1	0.0212	0.00301	49.7930	<.0001
hcv_pos_ab_v	1	-0.0632	0.0160	15.6887	<.0001
month2	1	-0.00598	0.000836	51.0817	<.0001
month2_spl1	1	0.0161	0.000996	260.8463	<.0001
sqrtcd4_l2	1	0.0597	0.00330	327.3119	<.0001
sqrtcd4_l1	1	-0.0630	0.00329	366.3845	<.0001
ts_last_sqrtcd4_l1	1	0.0337	0.00437	59.4472	<.0001
lnrna_l1	1	0.0138	0.00131	111.7310	<.0001
ts_last_lnrna_l1	1	0.0550	0.00443	154.2067	<.0001
everhaart_l1	1	0.6493	0.0157	1717.4627	<.0001
tseverhaart_l1_inter	1	-0.0139	0.000576	586.1415	<.0001
aids_l1	1	0.2542	0.0327	60.3638	<.0001
tsaids_l1_inter	1	-0.00646	0.00122	27.9804	<.0001

Model 3 CD4 count

The REG Procedure
Model: MODEL1
Dependent Variable: sqrtcd4

Number of Observations Read 67568
Number of Observations Used 67568

Root MSE	2.77103	R-Square	0.7832
Dependent Mean	18.41451	Adj R-Sq	0.7831
Coeff Var	15.04807		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.17939	0.12102	18.01	<.0001
age_0	1	-0.01235	0.00125	-9.85	<.0001
sex	1	0.15957	0.07061	2.26	0.0238
sqrtcd40	1	0.10083	0.00280	35.96	<.0001
lnrna0	1	0.10370	0.00711	14.59	<.0001
hcv_pos_ab_v	1	-0.21224	0.03825	-5.55	<.0001
month2	1	-0.02882	0.00198	-14.58	<.0001
month2_spl1	1	0.03866	0.00235	16.43	<.0001
sqrtcd4_l2	1	0.27967	0.01016	27.53	<.0001
sqrtcd4_l1	1	0.53070	0.01013	52.37	<.0001
ts_last_sqrtcd4_l1	1	0.04433	0.01153	3.85	0.0001
lnrna_l1	1	-0.05814	0.00307	-18.95	<.0001
ts_last_lnrna_l1	1	-0.08017	0.01169	-6.85	<.0001
everhaart_l1	1	0.86122	0.03703	23.26	<.0001
tseverhaart_l1_inter	1	-0.00656	0.00138	-4.77	<.0001
aids_l1	1	-0.43022	0.07479	-5.75	<.0001
tsaids_l1_inter	1	0.00384	0.00282	1.36	0.1736

Model 4 Visit to measure HIV-RNA viral load

The LOGISTIC Procedure

Model Information

Data Set	WORK.PARAM
Response Variable	visit_rna
Number of Response Levels	2
Weight Variable	_weight_
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	279236
Number of Observations Used	279236
Sum of Weights Read	279236
Sum of Weights Used	279236

Response Profile

Ordered Value	visit_rna	Total Frequency	Total Weight
1	1	69147	69147.00
2	0	210089	210089.00

Probability modeled is visit_rna=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	312587.48	107192.93
SC	312598.02	107382.65
-2 Log L	312585.48	107156.93

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	205428.549	17	<.0001
Score	149909.420	17	<.0001
Wald	66184.8312	17	<.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.4530	0.0912	24.6538	<.0001
age_0	1	0.00414	0.000955	18.8014	<.0001
sex	1	0.0468	0.0540	0.7516	0.3860
sqrtcd40	1	-0.00879	0.00218	16.2475	<.0001
lnrna0	1	0.00662	0.00544	1.4804	0.2237
hcv_pos_ab_v	1	-0.0830	0.0291	8.1576	0.0043
month2	1	-0.0114	0.00143	64.1313	<.0001
month2_spl1	1	0.0202	0.00178	129.3813	<.0001
sqrtcd4_l2	1	0.0337	0.00409	67.9847	<.0001
sqrtcd4_l1	1	-0.0928	0.00552	282.9134	<.0001
ts_last_sqrtcd4	1	-2.3031	0.0103	49973.7290	<.0001
lnrna_l1	1	0.0507	0.00234	468.3945	<.0001
ts_last_lnrna_l1	1	0.9612	0.00675	20273.4216	<.0001
everhaart_l1	1	0.6448	0.0269	573.5337	<.0001
tseverhaart_l1_inter	1	-0.0116	0.00104	125.7539	<.0001
aids_l1	1	0.0577	0.0553	1.0872	0.2971
tsaids_l1_inter	1	-0.00393	0.00216	3.3063	0.0690
sqrtcd4	1	0.0625	0.00437	204.4490	<.0001

Models 5, 6 HIV-RNA viral load models

The LOGISTIC Procedure

Model Information

Data Set	WORK.PARAM
Response Variable	zlnrna
Number of Response Levels	2
Weight Variable	_weight_
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	69371
Number of Observations Used	69371
Sum of Weights Read	69371
Sum of Weights Used	69371

Response Profile

Ordered Value	zlnrna	Total Frequency	Total Weight
1	1	38303	38303.000
2	0	31068	31068.000

Probability modeled is zlnrna=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.9176	0.1173	61.1971	<.0001
age_0	1	-0.00618	0.00120	26.5277	<.0001
sex	1	0.1913	0.0698	7.5218	0.0061
sqrtcd40	1	-0.00393	0.00266	2.1959	0.1384
lnrna0	1	0.0129	0.00678	3.6018	0.0577
hcv_pos_ab_v	1	0.1455	0.0367	15.7016	<.0001
month2	1	0.0335	0.00210	255.4762	<.0001
month2_spl1	1	-0.0373	0.00231	260.0867	<.0001
sqrtcd4_l2	1	0.00653	0.0104	0.3955	0.5294
sqrtcd4_l1	1	0.0610	0.0106	32.9589	<.0001
ts_last_sqrtcd4	1	-0.0519	0.0109	22.8529	<.0001
lnrna_l1	1	0.4459	0.00352	16043.7095	<.0001
ts_last_lnrna_l1	1	-0.0170	0.00560	9.1851	0.0024
everhaart_l1	1	-0.7678	0.0424	327.6621	<.0001
tseverhaart_l1_inter	1	0.00293	0.00137	4.5885	0.0322
aids_l1	1	0.2390	0.0765	9.7524	0.0018
tsaids_l1_inter	1	-0.00158	0.00275	0.3298	0.5658
sqrtcd4	1	-0.0805	0.00397	409.9896	<.0001

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: llnrna

Number of Observations Read	38303
Number of Observations Used	38303

Root MSE	0.26880	R-Square	0.4765
Dependent Mean	1.90385	Adj R-Sq	0.4763
Coeff Var	14.11869		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.86245	0.01574	118.36	<.0001
age_0	1	-0.00315	0.00016391	-19.19	<.0001
sex	1	0.02041	0.00853	2.39	0.0168
sqrtcd40	1	0.01083	0.00037796	28.66	<.0001
lnrna0	1	0.01242	0.00091590	13.56	<.0001
hcv_pos_ab_v	1	-0.03044	0.00481	-6.33	<.0001
month2	1	0.00153	0.00023643	6.49	<.0001
month2_spl1	1	-0.00455	0.00030560	-14.87	<.0001
sqrtcd4_l2	1	0.00335	0.00114	2.93	0.0034
sqrtcd4_l1	1	0.01208	0.00118	10.26	<.0001
ts_last_sqrtcd4	1	-0.00225	0.00148	-1.53	0.1270
lnrna_l1	1	0.04406	0.00046800	94.15	<.0001

ts_last_lnrna_l1	1	0.01561	0.00068627	22.75	<.0001
everhaart_l1	1	-0.25648	0.00423	-60.65	<.0001
tseverhaart_l1_inter	1	0.00360	0.00017097	21.05	<.0001
aids_l1	1	0.01024	0.00890	1.15	0.2498
tsaids_l1_inter	1	-0.00018530	0.00035672	-0.52	0.6035
sqrtcd4	1	-0.03245	0.00049821	-65.13	<.0001

Model 7 Combined antiretroviral therapy

The LOGISTIC Procedure

Model Information

Data Set	WORK.PARAM
Response Variable	everhaart
Number of Response Levels	2
Weight Variable	_weight_
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	61239
Number of Observations Used	61239
Sum of Weights Read	61239
Sum of Weights Used	61239

Response Profile

Ordered Value	everhaart	Total Frequency	Total Weight
1	1	3744	3744.000
2	0	57495	57495.000

Probability modeled is everhaart=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.4464	0.1988	151.4150	<.0001
age_0	1	0.0151	0.00211	51.0855	<.0001
sex	1	0.2021	0.1132	3.1843	0.0744
sqrtcd40	1	0.0586	0.00849	47.5716	<.0001
lnrna0	1	0.0296	0.0140	4.4560	0.0348
hcv_pos_ab_v	1	-0.4206	0.0619	46.1860	<.0001
month2	1	-0.0921	0.00353	679.7947	<.0001
month2_spl1	1	0.1121	0.00594	356.1836	<.0001
sqrtcd4_l2	1	0.0142	0.0145	0.9558	0.3282
sqrtcd4_l1	1	-0.0566	0.0164	11.8398	0.0006
ts_last_sqrtcd4	1	-0.2170	0.0290	56.1688	<.0001
lnrna_l1	1	0.0214	0.0138	2.4147	0.1202
ts_last_lnrna	1	-0.2647	0.0288	84.4924	<.0001
aids_l1	1	-0.1760	0.1437	1.5013	0.2205
tsaids_l1_inter	1	0.00386	0.00994	0.1509	0.6977
sqrtcd4	1	-0.0836	0.0111	57.1759	<.0001
lnrna	1	0.1268	0.0127	99.0249	<.0001

Model 8 Diagnosis of an AIDS-defining illness

The LOGISTIC Procedure

Model Information

Data Set	WORK.PARAM
Response Variable	aids
Number of Response Levels	2
Weight Variable	_weight_
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	264252
Number of Observations Used	264252
Sum of Weights Read	264252
Sum of Weights Used	264252

Response Profile

Ordered Value	aids	Total Frequency	Total Weight
1	1	536	536.00
2	0	263716	263716.00

Probability modeled is aids=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.3722	0.5226	69.9983	<.0001
age_0	1	0.00559	0.00534	1.0973	0.2949
sex	1	0.3339	0.2760	1.4637	0.2263
sqrtcd40	1	0.0452	0.0141	10.3141	0.0013
lnrna0	1	0.1482	0.0344	18.5283	<.0001
hcv_pos_ab_v	1	-0.1702	0.1494	1.2968	0.2548
month2	1	-0.0450	0.00837	28.8932	<.0001
month2_spl1	1	0.0503	0.0110	21.0638	<.0001
sqrtcd4_l2	1	0.0360	0.0263	1.8776	0.1706
sqrtcd4_l1	1	-0.0507	0.0321	2.4984	0.1140
ts_last_sqrtcd4	1	-0.4732	0.0528	80.2139	<.0001
lnrna_l1	1	0.0477	0.0212	5.0559	0.0245
ts_last_lnrna	1	0.0570	0.0440	1.6772	0.1953
everhaart	1	-0.1112	0.1175	0.8952	0.3441
tseverhaart_inter	1	0.00355	0.00691	0.2644	0.6071
sqrtcd4	1	-0.2364	0.0227	108.6731	<.0001
lnrna	1	0.0223	0.0197	1.2758	0.2587