Variable selection in individual patient data meta-analysis

Michael Seo1, Orestis Efthimiou1

1Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

Abstract

*…to write last*

# Introduction

Individual patient data (IPD) meta-analysis (MA) of Randomized Clinical trials (RCTs) is considered to be the gold standard in evidence synthesis.1 Despite being more resource-intensive than the standard aggregate data (AD) MA, IPD MA achieves higher power to detect differential treatment than the AD MA since it models the individual risk across hundreds of patients as opposed to few studies. Furthermore, IPD MA is less prone to ecological bias as within trial information can be directly used to estimate how patient level characteristics modify treatment effect.1–3  
 There are two ways of performing IPD MA (one- and two-stage approaches). A two-stage approach first analyzes each study separately and then uses standard meta-analysis to pool the aggregate term of interest, such as treatment effect estimate and its standard error. A one-stage approach simultaneously models the individual participant data from all studies while accounting for the study level clustering of individuals. Similar to the comparison between IPD MA and AD MA, one-stage approach offers greater flexibility than the two stage approach to distinguish differences between patients both within and across studies.1–3   
 A common goal of IPD MA is to estimate the relative treatment effect, accounting for possible differences in the distribution of covariates among trials. Another goal would be to identify possible treatment covariate interactions (i.e. effect modification).2 If important interactions are present, there may be important clinical implications, i.e. on whether to treat a particular patient subgroup. For both goals, it is important to identify correct set of covariates.   
 However, studies often collect too many covariates, some of which might be unrelated to the outcome of interest.4 Including all covariates in a model will lead to more complicated models and possibly lead to overfitting. On the other hand, including too few covariates runs the risk of missing important covariates that moderate the effect.  
 In this paper, we are interested in exploring methods for variable selection in IPD meta-analysis. Variable selection has a long history in statistics.5 We can use simple methods that select variables based on a threshold. Stepwise selection is a very popular approach, but often criticized.6 Relatively more recent techniques such as LASSO have gained ground.6,7 Best subset approach has recently received more attention due to development in a faster algorithm.8 Bayesian methods, such as Bayesian model averaging, have been developed and widely used.9,10  
 Currently, it is unclear how these methods are going to perform in an IPD meta-analysis, where the aim is not prediction of patient-level effects, but estimation of treatment effects and interactions. We compare different methods ranging from naïve models that pool all clinical trials into one dataset and random effects models that preserve the clustering of patients within studies. By comparison, we want to explore whether we need to preserve the clustering of patients within studies when performing variable selection. We performed simulations to answer this question and applied these methods to two real datasets, from cardiology and psychiatry.

# Real datasets

## Drug-eluting or bare-metal stents for percutaneous coronary intervention

The dataset comprises of IPD from randomized clinical trials (RCTs) in patients who have undergone percutaneous coronary intervention for coronary artery disease. Eliminating studies that have systematically missing covariates, we will use reduced number of clinical trials, with 8 studies and 11133 patients that compared the effects of using drug-eluting or bare metal stents while going through percutaneous coronary intervention. Outcome we focus on is cardiac death or myocardial infarction at a 1-year landmark. The dataset contains information on a number of patient-level covariates. These include one continuous variable, age, one count variable, the number of implanted stents, and seven binary covariates: gender, diabetes, clinical presentation at the time of percutaneous coronary intervention, multivessel disease, stent placement in the left anterior descending artery, overlapping stents, and mean stent diameter greater than 3.11

## Antidepressant treatment of major depression

The dataset comprises of IPD from RCTs in patients receiving antidepressant treatment for major depression. We have 4 placebo-controlled trials and 1261 patients in the acute phase treatment of major depression. Outcome of interest is depression severity between the antidepressants and placebo at week 6 or 8. Patient level covariates include two binary variables, sex and episode frequency dichotomized at greater than or equal to three episodes and 9 continuous variables: baseline severity, age, age at onset, episode duration, and 17-item Hamilton Rating Scale for depression constituting five subscales of anhedonia, guilt, bodily symptoms, appetite, and insomnia.12

# Available methods for variable selection in individual patient data meta-analysis

In this section, we outline the various methods that can be used for selecting variables in IPD meta-analysis. We start from simpler methods and then discuss more advanced approaches. In the models’ description, we assume that all available studies have collected information on all covariates of interest, and that there are no missing outcomes or covariate data from all patients. For issues related to missing outcome data, see Section 3.8.

## Notation and general considerations

We will use to denote a patient randomized in study to receive treatment (where can be 0 or 1). For this patient we have information on a vector of patient level covariates . Without loss of generality, we will assume that all continuous covariates are centered on zero. We also have information on an outcome of interest, which we will denote as . In this paper, we focus on the case where is either continuous or binary.  
 We split the patient-level covariates in three categories: covariates that have no effect on the outcome of interest (‘nuisance covariates’), covariates that affect the outcome but do not interact with the treatment (‘prognostic effects’), and covariates that affect the outcome and have an interaction with treatment (‘effect modifiers’). For example, if patients’ age is not related to *y* then age is a nuisance parameter. If age is related to *y* but do not have an interaction with treatment, then age is a prognostic effect. If age is related to *y* and related to the treatment, then age is an effect modifier.  
 Let us focus to the generalized linear model:

Where is a parameter of interest such as probability of an event when is binary, is a link function, is the effect of prognostic effect, is the coefficient for effect modifier, and is the treatment effect. The goal of an IPD meta-analysis is to estimate the average treatment effect (i.e. the relative treatment effect for ) and to identify important treatment-covariate interactions, also known as effect modifications. To that aim, variable selection methods can be utilized to select the correct set of covariates that yield accurate average treatment effect and effect modifications.

## Simple null and simple glm models - not accounting for the study (Simple-null and Simple-glm)

Following this method, one would aggregate data from all studies into a single dataset. These simple models serve as baseline models of comparison. Simple null model is a simple linear (logistic) regression that only fits the response to the treatment effect without including any covariates. Simple glm model is a multiple linear (logistic) regression that incorporates all available covariates.

## Stepwise variable selection - not accounting for the study (STEP-naïve)

Using this method, there are three choices, “forward”, “backward” and “bidirectional”. Forward stepwise regression starts with a small model (i.e. with just an intercept), considers all one variable expansion and adds the variable that has the lowest AIC. This process continues until the AIC stops improving. Backward stepwise regression starts from the full model and eliminates variable according to the AIC. Bidirectional stepwise regression will consider both adding and removing one variable at each step, and take the best option according to the AIC.5,13

## LASSO - not accounting for study (LASSO-naïve)

LASSO regression is a simple technique to reduce model complexity and prevent overfitting, which may result from fitting multiple linear regression. The penalty term () regularizes the coefficients by using L1 lasso penalty in the optimization function. There are two common ways of selecting the optimal lambda value: lambda-min and lambda-1se. Lambda-min rule selects the best model that minimizes the cross validation error (i.e. mean squared error for the continuous outcome and misclassification rate for binary outcome). Lambda-1se rule selects the most parsimonious model whose error is no more than one standard error above the error of the model using lambda-min. The aim of the lambda-1se rule is to choose the simplest model that has accuracy comparable with the model using lambda-min.6,14,15

## Generalized linear mixed effects model using LASSO (GLMM-LASSO)

Following this method, one would include the clustering of patients within studies and assume random effects structure on the treatment effect. This approach fits a generalized linear mixed model including an L1-penality term that enforces variable selection and shrinkage. Optimal lambda value can be chosen based on lowest AIC or BIC on a set of lambda values, or through cross validation.16

## Bayesian LASSO with mixed effects (Bayes-LASSO)

Park and Casella (2008) introduced the Bayesian LASSO that uses Laplacian double exponential prior on the covariate effect (). Using such a prior, one can obtain shrinkage estimate of the covariate effect. An advantage of using Bayesian LASSO is that one can obtain standard error bound, which is not available in classical LASSO unless bootstrap is used. The degree of sparseness is controlled by , which can be given non-informative distribution or estimated using cross validation.17

## Stochastic search variable selection (SSVS)

This Bayesian model introduces indicator variables ( to select covariates in each step of the MCMC iterations. A mixture prior on the covariate effect is used: , where the first density is centred around zero and has a small variance.9,18 Meuwissen and Goddard (2004) introduced a variant of SSVS where was assumed random and estimated in the model with own prior and g fixed at 100.19

## Considerations regarding missing data

For frequentist methods, multiple imputation is a standard way to approach missing data.20 For Bayesian methods, fully model based approaches can be used. One would write down the statistical model for full data and sample from the distribution in each iteration of MCMC.21 However, for simplicity and for fair comparison between Bayesian and frequentist methods, one can rely on analyzing based on complete data. Covariates that are systematically missing can be dropped (i.e. missing completely for a certain trial). And, remaining observations with missing data in either the outcome or covariates can be omitted.

# Simulations

## Overview

In this section we describe a simulation study we performed to compare the various methods. We explored dichotomous-continuous outcomes and different scenarios regarding the covariates. For each scenario, we performed 100 simulations.

## Data generating mechanism

For the case of a continuous outcome, we used the following model

where refers to the outcome of the patient, refers to mean outcome of the patient, refers to standard deviation of the outcome, refers to study specific baseline risk, refers to coefficient of the prognostic factor, refers to relative treatment effect of second treatment compared to the first, refers to coefficient of the effect modifier, refers to average treatment effect of second treatment compared to the first, and refers to the heterogeneity of the treatment effect across studies.

For the dichotomous outcomes the data generating mechanism is as follows

Where notations are the same except now the outcome follows Bernoulli distribution with logit link function. We explored four scenarios, described in Table 1.

In order to generate the data, we follow the next steps

1. We generate the rows of the predictor matrix from , where has entry equal to for continuous covariates and for discrete covariates. We generate treatment indicator from . For each study, we generate a total number of patients for each study from
2. We generate the treatment effect of each study from where the average treatment effect is fixed at a value of 1 and is drawn from
3. We generate a study baseline risk ( from and assume that the risk is independent across studies.
4. We include a random error component, denoted as , sampled from for continuous outcome and for discrete outcome.
5. Lastly, we generate from for discrete outcome and for continuous outcome.

## Measure of performance

A common goal of IPD is to obtain accurate estimation of treatment effect and effect modifiers. As a measure of performance, we calculate how much these parameter estimates deviates from the true value using mean squared error (MSE). We report the mean squared error for the treatment effect, false effect modifier and true effect modifier. False effect modifiers are variables that are not effect modifiers, but are falsely selected as effect modifiers by the model. MSE of these would simply be squared sum of these variable effect estimates since the true value of these values are 0. Similarly, true effect modifier MSE would measure how accurately the model has identified the effect modifiers. We also compare standard error of the treatment effect between different models. Standard error should be comparable across different models or else having lower MSE for the treatment effect for a certain model would not be as significant.

## Methods compared and fitting details

For the stepwise regression, we use bidirectional selection model using the core R package. We limited the lower bound of the scope to include at least the treatment effect. LASSO regression is fitted using R package glmnet and we chose to use lambda-min since lambda-1se seems more suitable for prediction purposes and we are interested in the coefficient estimates (i.e. treatment effect). Similarly, we chose not to penalize the treatment effect, thus allowing the treatment effect to be always selected. Generalized linear mixed effect model using LASSO is fitted using glmmLasso R package. The optimal lambda value is chosen by comparing BIC values from set of lambda values from 0 to 500. Again, we did not penalize the treatment effect and the baseline risk. Bayesian LASSO with mixed effects and SSVS are fitted using JAGS software through R framework utilizing parallel computation. Shrinkage priors are placed on both prognostic effect and effect modifiers, but we did not put any shrinkage priors on the treatment effect or baseline risk. For SSVS, we used the variant of SSVS that Meuwissen and Goddard (2004) proposed where was assumed random. For Bayesian methods, 3 chains of 10,000 iterations each has been run with 1000 iteration burn-in. Code for these in JAGS are given in the appendix. In regards to the missing data for the stent and depression datasets, only the complete data has been used after omitting systematically missing covariates and remaining observations with missing values in covariates or outcomes. For the simulations, we did not introduce any missing data in the data generating mechanism.

## Simulation results

In table 2, we compare measure of performance across different models.

# Using real datasets

## Stents dataset

Results are shown in Table 3. After controlling for the covariates, we find that the treatment effect has wider standard error. Some marginally significant effect modifier we find is the stent placement in the left anterior descending artery (ladtreated). It was selected in stepwise and SSVS model and in the Bayesian LASSO model, its coefficient remains relatively larger compared to other coefficients. Although, it was not significant for all the models, we can say that this variable acts as a possible effect modifier

We present estimated relative treatment effect at different patient subgroups. For instance, if we choose to find treatment effect of patients who received stent placement in the left anterior descending artery and five implanted stents, we get a large effect of the treatment. For instance, running full generalized linear model, patients in this subgroup have odds ratio of 0.48 with 95% CI of 0.19 to 1.27. This means that treated patients in this subgroup will see about 52% decrease in the odds of dying. On the other hand, if we choose subgroup of patients who did not receive stent placement in the left anterior descending artery and only one implanted stents, we get a smaller treatment effect with odds ratio 0.92 with 95% CI (0.76, 1.11). Here we only see about 8% decrease in the odds of dying from cardiac death or myocardial infarction. Drug-eluting stents are much more expensive technique than bare-metal stents. If the treatment effect is not as high as in this subgroup, patients can safely use bare-metal stents instead of drug-eluting stents. Further results for different subgroups and models are shown in Table 4. Do similar analysis for SSVS?

# Discussion

Among the models we have considered, we have not considered best subset regression. The best subset regression is a model selection approach that consists of testing all possible combination of the predictor variables, and then selects the best model according to some statistical measures, such as adjusted R-squared or Mallow’s Cp. One caveat with this approach is the high computational cost as all combination has to be considered. Recently, Bertsimas et al. (2016) showed that the best subset selection problem can be formulated as a mixed integer optimization problem and demonstrated that this can be solved at even larger problem sets.8 However, currently the R packages , best-subset, that implements the optimization, only allows continuous outcomes.22 Other traditional approaches to best subset regression, such as the R package leaps, cannot run for high number of covariates.

# Bibliography

1. Thomas D, Radji S, Benedetti A. Systematic review of methods for individual patient data meta- analysis with binary outcomes. *BMC Med Res Methodol*. 2014;14:79. doi:10.1186/1471-2288-14-79

2. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340:c221. doi:10.1136/bmj.c221

3. Debray TPA, Moons KGM, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res Synth Methods*. 2015;6(4):293-309. doi:10.1002/jrsm.1160

4. Heinze G, Wallisch C, Dunkler D. Variable selection – A review and recommendations for the practicing statistician. *Biom J Biom Z*. 2018;60(3):431-449. doi:10.1002/bimj.201700067

5. Miller AJ. Selection of Subsets of Regression Variables. *J R Stat Soc Ser Gen*. 1984;147(3):389-425. doi:10.2307/2981576

6. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Ser B Methodol*. 1996;58(1):267-288.

7. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *Ann Stat*. 2004;32(2):407-499. doi:10.1214/009053604000000067

8. Bertsimas D, King A, Mazumder R. Best subset selection via a modern optimization lens. *Ann Stat*. 2016;44(2):813-852. doi:10.1214/15-AOS1388

9. O’Hara RB, Sillanpää MJ. A review of Bayesian variable selection methods: what, how and which. *Bayesian Anal*. 2009;4(1):85-117. doi:10.1214/09-BA403

10. Boulet S, Ursino M, Thall P, Jannot A-S, Zohar S. Bayesian variable selection based on clinical relevance weights in small sample studies—Application to colon cancer. *Stat Med*. 2019;38(12):2228-2247. doi:10.1002/sim.8107

11. Piccolo R, Bonaa KH, Efthimiou O, et al. Drug-eluting or bare-metal stents for percutaneous coronary intervention: a systematic review and individual patient data meta-analysis of randomised clinical trials. *The Lancet*. 2019;393(10190):2503-2510. doi:10.1016/S0140-6736(19)30474-X

12. Noma H, Furukawa TA, Maruo K, et al. Exploratory analyses of effect modifiers in the antidepressant treatment of major depression: Individual-participant data meta-analysis of 2803 participants in seven placebo-controlled randomized trials. *J Affect Disord*. 2019;250:419-424. doi:10.1016/j.jad.2019.03.031

13. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning : With Applications in R*. New York: Springer; 2013.

14. Hastie T, Tibshirani R, Friedman, Jerome. *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. New York: Springer; 2001.

15. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1-22.

16. Groll A, Tutz G. Variable selection for generalized linear mixed models by L1-penalized estimation. *Stat Comput*. 2014;24(2):137-154. doi:10.1007/s11222-012-9359-z

17. Park T, Casella G. The Bayesian Lasso. *J Am Stat Assoc*. 2008;103(482):681-686.

18. George EI, McCulloch RE. Variable Selection via Gibbs Sampling. *J Am Stat Assoc*. 1993;88(423):881-889. doi:10.1080/01621459.1993.10476353

19. Meuwissen THE, Goddard ME. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet Sel Evol GSE*. 2004;36(3):261-279. doi:10.1051/gse:2004001

20. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45(3):67.

21. Daniels MJ, Hogan JW. *Missing Data In Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Champman & Hall; 2008.

22. Hastie T, Tibshirani R, Tibshirani RJ. Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *ArXiv Prepr ArXiv170708692*. 2017.

**Table 1:** Overview of the scenarios we explored in our simulations

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Scenario** | **Type of outcome** | **# of studies** | **# of patients per treatment arm** | **# of covariates** | **# of prognostic factors** | **True values, prognostic factors** | **# of effect modifiers** | **True values, effect modifiers1** | **# of nuisance covariates** |
| 1 | Continuous | 5 | 30-300 | 10 | 2 continuous  1 dichotomous | 0.1, -0.1, 0.2 | 1 continuous  1 dichotomous | 0.2 (0.2), 0.3 (-0.2) | 3 continuous  2 dichotomous |
| 2 | Continuous | 10 | 30-300 | 20 | 4 continuous  2 dichotomous | 0.1, -0.1, 0.2, 0.2, -0.2, 0.3 | 2 continuous  2 dichotomous | 0.2 (0.1), 0.3 (0.1), -0.1 (-0.1), -0.2 (-0.2) | 6 continuous  4 dichotomous |
| 3 | Binary | 5 | 30-300 | 10 | 2 continuous  1 dichotomous | 0.1, -0.1, 0.2 | 1 continuous  1 dichotomous | 0.2 (0.2), 0.3 (-0.2) | 3 continuous  2 dichotomous |
| 4 | Binary | 10 | 30-300 | 20 | 4 continuous  2 dichotomous | 0.1, -0.1, 0.2, 0.2, -0.2, 0.3 | 2 continuous  2 dichotomous | 0.2 (0.1), 0.3 (0.1), -0.1 (-0.1), -0.2 (-0.2) | 6 continuous  4 dichotomous |
|  |  |  |  |  |  |  |  |  |  |

**1** For each covariate, we have the main effect on the outcome and in the brackets, we report the interaction with the treatment (i.e. effect modification)

**Table 2:** Results from the simulations. Model abbreviations as per Section 3. MSE: mean squared error

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Scenario** | **model** | **False effect modifier MSE** | **True effect modifier MSE** | **Treatment MSE** | **Treatment effect standard error** |
| 1 | Simple-null | 0.000 | 0.065 | 0.038 | 0.091 |
| Simple-glm | 0.016 | 0.020 | 0.048 | 0.20 |
| Step-naïve | 0.0088 | 0.029 | 0.039 | 0.12 |
| LASSO-naïve | 0.000 | 0.060 | 0.037 | NA |
| GLMM-LASSO | 0.0029 | 0.042 | 0.043 | NA |
| Bayes-LASSO |  |  |  |  |
| SSVS | 0.0014 | 0.031 | 0.026 | 0.23 |
| 2 | Simple-null | 0.000 | 0.045 | 0.033 | 0.070 |
| Simple-glm | 0.0090 | 0.010 | 0.037 | 0.19 |
| Step-naïve | 0.0053 | 0.012 | 0.031 | 0.10 |
| LASSO-naïve | 0.000 | 0.015 | 0.029 | NA |
| GLMM-LASSO | 0.0046 | 0.010 | 0.13 | NA |
| Bayes-LASSO |  |  |  |  |
| SSVS | 0.0014 | 0.0067 | 0.013 | 0.15 |
| 3 | Simple-null | 0.000 | 0.065 | 0.030 | 0.15 |
| Simple-glm | 0.061 | 0.059 | 0.15 | 0.36 |
| Step-naïve | 0.035 | 0.071 | 0.091 | 0.20 |
| LASSO-naïve | 0.0012 | 0.063 | 0.033 | NA |
| GLMM-LASSO | 0.000 | 0.065 | 0.036 | NA |
| Bayes-LASSO |  |  |  |  |
| SSVS | 0.0032 | 0.049 | 0.048 | 0.38 |
| 4 | Simple-null | 0.000 | 0.045 | 0.10 | 0.10 |
| Simple-glm | 0.028 | 0.037 | 0.14 | 0.33 |
| Step-naïve | 0.017 | 0.037 | 0.12 | 0.18 |
| LASSO-naïve | 0.000 | 0.024 | 0.085 | NA |
| GLMM-LASSO | 0.000 | 0.039 | 0.078 | NA |
| Bayes-LASSO |  |  |  |  |
| SSVS | 0.0040 | 0.011 | 0.027 | 0.25 |

**Table 3:** Results from fitting various models in the stents dataset. Parameter abbreviations as per Section 2.1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **Simple null (Std. Err)** | **Simple glm (Std. Err)** | **Step-naïve**  **(Std. Err)** | **LASSO-naïve** | **GLMM-LASSO** | **Bayes-LASSO**  **(Std. Err)** | **SSVS**  **(Std. Err/% selected)** |
| Average treatment effect | -0.21 (0.085) | -0.13 (0.47) | 0.25 (0.18) | -0.21 | -0.112 | -0.054 (0.44) | 0.008 (0.28/ 100) |
| Heterogeneity (τ) |  |  |  |  | 0.105 | 0.043 | 0.019 |
| age | 0 | 0.84 (0.083) | 0.81 (0.055) | 0 | 0.331 | 0.67 (0.083) | 0.66 (0.080 / 100) |
| gender | 0 | 0.012 (0.13) | 0 | 0 | 0 | -0.024 (0.10) | -0.006 (0.065 / 21.3) |
| diabetes | 0 | 0.54 (0.13) | 0.51 (0.092) | 0 | 0.081 | 0.40 (0.12) | 0.43 (0.10 / 99.4) |
| stable\_cad | 0 | -0.53 (0.15) | -0.47 (0.10) | 0 | -0.131 | -0.46 (0.13) | -0.48 (0.12 / 99.1) |
| multivessel | 0 | 0.25 (0.13) | 0.16 (0.093) | 0 | 0 | 0.20 (0.12) | 0.19 (0.14 / 69) |
| ladtreated | 0 | 0.23 (0.13) | 0.24 (0.13) | 0 | 0 | 0.086 (0.11) | 0.059 (0.10 / 35.4) |
| overlap | 0 | 0.48 (0.18) | 0.49 (0.13) | 0 | 0.024 | 0.29 (0.14) | 0.34 (0.15 / 88.7) |
| m\_dia\_above\_3 | 0 | -0.43 (0.26) | 0 | 0 | 0 | -0.10 (0.18) | -0.042 (0.15 / 35.5) |
| num\_stent | 0 | 0.038 (0.064) | 0.054 (0.057) | 0 | 0 | 0.039 (0.053) | 0.017 (0.043 / 17.4) |
| age:treat | 0 | -0.056 (0.11) | 0 | 0 | 0.172 | -0.054 (0.10) | -0.045 (0.085 / 29.0) |
| gender:treat | 0 | 0.031 (0.19) | 0 | 0 | 0 | 0.047 (0.13) | 0.018 (0.087 / 25.4) |
| diabetes:treat | 0 | -0.067 (0.19) | 0 | 0 | 0 | -0.001 (0.13) | -0.017 (0.094 / 30.3) |
| stable\_cad:treat | 0 | 0.113 (0.20) | 0 | 0 | 0 | 0.013 (0.15) | 0.006 (0.10 / 30.8) |
| multivessel:treat | 0 | -0.184 (0.19) | 0 | 0 | 0 | -0.074 (0.14) | -0.078 (0.14 / 41.2) |
| ladtreated:treat | 0 | -0.32 (0.18) | -0.34 (0.18) | 0 | 0 | -0.20 (0.15) | -0.16 (0.17 / 56.2) |
| overlap:treat | 0 | 0.008 (0.25) | 0 | 0 | 0 | 0.016 (0.16) | -0.025 (0.13 / 36) |
| m\_dia\_above\_3:treat | 0 | 0.445 (0.41) | 0 | 0 | 0 | 0.23 (0.33) | 0.14 (0.21 / 49.6) |
| num\_stent:treat | 0 | -0.082 (0.095) | -0.11 (0.074) | 0 | 0 | -0.075 (0.074) | -0.046 (0.058 / 25.9) |

% selected in SSVS shows how many times a given variable is selected throughout the iteration

Abbreviation: num\_stents, number of implanted stents; stable\_cad, clinical presentation at the time of percutaneous coronary intervention; ladtreated, stent placement in the left anterior descending artery; m\_dia\_above\_3, mean diameter greater than 3

**Table 4:** Estimated treatment effect (and 95% CI) for different subgroup population in Stent dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Scenarios** | **Simple null TE (95% CI)** | **Simple glm TE (Std. Err)** | **Step-naïve TE**  **(Std. Err)** | **LASSO-naïve TE** | **GLMM-LASSO TE** | **Bayes-LASSO TE**  **(Std. Err)** | **SSVS TE**  **(Std. Err)** |
| ladtreated + 5 num\_stents |  |  |  |  |  |  |  |
| not ladtreated + 1 num\_stents |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Abbreviation: num\_stents, number of implanted stents; ladtreated, stent placement in the left anterior descending artery; TE, treatment effect; CI, confidence interval