**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 outcome across hundreds or thousands of patients as opposed to usually few available 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 general ways to perform an IPD MA usually termed one-stage and two-stage approaches. A two-stage approach first analyzes each study separately and then uses standard meta-analysis methods to pool the aggregate term of interest, such as the estimate of treatment effect and its standard error. A one-stage approach simultaneously models the individual participant data from all studies while keeping intact the randomization of each study, i.e. accounting for the clustering of individual patients. One-stage approaches are usually thought to offer greater flexibility than two-stage approaches to distinguish differences between patients both within and across studies.1–3

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 is 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 include in the analysis relevant patient-level covariates. However, studies often collect large number of covariates, some of which might be unrelated to the outcome of interest.4 Including all covariates in a model will give more complicated models and may lead to overfitting. On the other hand, including too few covariates runs the risk of missing important covariates that moderate the effect. Thus, selecting which variables to include in an IPD MA is an important question.

Variable selection has a long history in statistics.5 Simple methods select variables based on a selection criterion such as AIC. For example, stepwise selection is a very popular approach, but has been often criticized.6 Relatively more recent techniques such as LASSO have gained ground. Through efficient algorithm such as least angle regression, LASSO shrinks large coefficients to reduce overfitting and selects variables by forcing certain variables to zero.6–8 Since the development of LASSO, there has been many extensions. One particular model applies LASSO to generalized linear mixed effects model, which is often used in IPD MA.3,9 Bayesian methods have also been developed. Bayesian LASSO has an advantage of obtaining standard error bounds, which is not reported in classical LASSO unless bootstrap is used.10 Bayesian model averaging methods, such as stochastic search variable selection (SSVS) has been developed to find robust estimation of effect size.11,12

It is currently unclear how the aforementioned methods perform in an IPD meta-analysis, where the aim is estimation of treatment effects and interactions. Hereby 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 8 RCTs in 11133 patients who have undergone percutaneous coronary intervention for coronary artery disease. The RCTs compared the effects of using drug-eluting versus bare metal stents. The outcome we focus on in our analysis is composite, i.e. 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 (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).13

## Antidepressant treatment of major depression

The dataset comprises IPD from four placebo-controlled trials on 1261 patients. The RCTs explored four placebo-controlled trials on 1261 patients. The RCTs explored the effects of antidepressant treatment for acute major depression. The outcome of interest is depression severity on a continuous scale 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).14

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

In this section, we outline several methods that can be used for selecting variables in IPD MA. 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 ~~data generating mechanisms~~ general model framework (?)

We 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 range 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 factors’), 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 does not have an interaction with treatment, then age is a prognostic factor. If age is related to *y* and interacts with the treatment, then age is an effect modifier. We will denote the prognostic factors of patient as and the effect modifiers as .

We model our data through generalized linear mixed effects model (GLMM). The systematic component (i.e. linear predictor) is given below

where is a parameter of interest (e.g. probability of an event when is binary), is a link function (e.g. log-odds for binary outcomes), vector includes the regression coefficients of the prognostic factors, is the vector of coefficients for effect modifiers, and is the treatment effect. The model is a mixed effects model since the treatment effect is randomly distributed as below

where is the average treatment effect and is the heterogeneity parameter. The goal of an IPD MA is to estimate the average treatment effect and to identify important treatment-covariate interactions, i.e. effect modification. Variable selection methods aim to select the correct set of covariates that yield accurate average treatment effect and effect modifications.

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

Stepwise variable selection has been extensively used in the past. There are three different flavors of stepwise selection, depending on the directionality of the selection procedure: “forward”, “backward” and “bidirectional”. Forward stepwise regression starts with a small model (i.e. with only the intercept), considers all one-variable expansions of the model, 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.5,15

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

LASSO regression is a technique aimed at reducing model complexity and preventing overfitting.6 The model uses a L1 penalty term in the optimization function, controlled by a penalty parameter . The inclusion of the penalty term leads to a shrinkage of the regression coefficients. Some of the coefficients may shrink to zero, and the corresponding covariates are excluded from the model. Thus, different values of correspond to different models, and variable selection is achieved. For continuous outcome, the objective is to minimize

Similarly, for dichotomous outcome, the objective is to minimize

The exact value of is usually determined by k-fold cross validation.16 Lambda value that minimizes the cross validation error (i.e. mean squared error for the continuous outcome and misclassification rate for binary outcome) is selected. For the purpose of IPD MA, one could use LASSO naively, i.e. not accounting for the clustering of patients in the different studies. In classical LASSO, no simple formula for standard errors exists. Calculating standard error using bootstrap has been suggested6, but there doesn’t seem to be a consensus on whether the method is statistically valid.17

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

This method directly generalizes the naïve LASSO. Here we account for 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 the cross validation error similarly to the naïve LASSO.9 One approach to fit GLMM is through penalized quasi-likelihood (PQL), suggested by Breslow and Clayton (1993).18 Then, for given heterogeneity, the objective to minimize

where the negative log likelihood is same as given previously in naïve LASSO in section 3.3 and heterogeneity estimate for the random effects can be updated from an approximate EM algorithm.9

## 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 standard errors are automatically calculated. The degree of sparseness is controlled by , which can be given diffuse hyper-prior.10 The model is same generalized linear mixed effects model described in (1), but the prior on the covariate effects is now steeper around zero, therefore allowing shrinkage.

where is a diffuse hyper-prior from gamma distribution. By placing a Laplacian double exponential prior, the posterior density for the continuous outcome is given by

The maximum a posterior estimates (i.e. values of and that maximizes the posterior density) are found by similar optimization problem as that of naïve LASSO.

## Stochastic search variable selection (SSVS)

Introduced in the paper George and McCulloch (1993), this Bayesian model introduces indicator variables to select covariates in each step of the MCMC iterations. The model framework is the same as described in (1), but in this model, a mixture prior on the covariate effect is used

Similarly,

where the first density is centered around zero and has a small variance.11,19 Meuwissen and Goddard introduced a variant of SSVS where was assumed random and estimated in the model with own prior and fixed at 100.20

## Considerations regarding missing data

For frequentist methods, multiple imputation is a standard way to approach missing data.21 For Bayesian methods, fully model 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.22 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

In this section we describe a simulation study we performed to compare the various methods for variable selection in IPD MA. We explored dichotomous and continuous outcomes and different scenarios regarding the covariates. For each scenario, we performed 1000 simulations. In what follows we describe the various scenarios we explored, the approaches we compared, and the methods we used to compare their performance.

## Data generating mechanism

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

where denotes the outcome of patient randomized in study ; denotes the expected outcome of the patient ; refers to standard deviation of the outcome; denotes the study-specific baseline outcome; denotes the coefficients of the prognostic factors; denotes the treatment effect of second treatment compared to the first; denotes the coefficients of the effect modifiers; denotes the average treatment effect of second treatment compared to the first; and denotes the heterogeneity of the treatment effect across studies.

For the dichotomous outcomes, the data generating mechanism we used is as follows:

Here the outcome follows Bernoulli distribution with logit link function. We explored XXX scenarios, described in Table 1.

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

1. For each study, we start by determining the total number of patients for each treatment arm, by sampling from and rounding. Then, we generate the patient-level covariates . For continuous covariates, we sample from where is total number of covariates and has entry equal to . The correlation coefficient ( is set to be 0.3. For discrete covariates, we sample from a . We generate treatment indicator by sampling from a .
2. We generate the treatment effect of each study by drawing where the average treatment effect and depend on the scenario.
3. We generate a study baseline effect ( from and assume that the effect is independent across studies.
4. Lastly, we generate from for discrete outcome and for continuous outcome.

We generated data for all scenarios using R.23 The code we used is available at <https://github.com/MikeJSeo/phd/blob/master/varselect/helpful.functions.R>.

## Models compared

We use naïve models with no variable selection as a reference. First, we fit genearlized linear mixed effects model (glmm) with only the treatment effect. Another reference model we use is glmm model with fully specified covariates. With these reference models, we compare the variable selection models described in section 3. The models include: STEP-naïve, LASSO-naïve, GLMM-LASSO, BAYES-LASSO, and SSVS

## Measure of performance

The usual goal of an IPD MA is to estimate average treatment effects, to identify effect modifiers and to estimate their impact on the outcome of interest. Thus, as a measure of performance, we assessed how much the corresponding parameter estimates deviated from their true values, using the mean squared error (MSE). We report the MSE 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. Among the models that report standard error, we compare the 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.

## Fitting details

For fitting generalized linear mixed effects model, we use lmer package. (citation)

For the stepwise regression, we use bidirectional selection model using the core R package.23 Bidirectional selection model is often a default approach and it is preferred because it has a possibility to delete a variable at each step. We start from a base model that has treatment effect included and add additional covariates at each stage.

LASSO regression is fitted using R package glmnet and lambda value is chosen that minimizes the 10-fold cross validation.8 Similarly, we chose not to penalize the treatment effect, thus allowing the treatment effect to be always selected in the model.

Generalized linear mixed effect model using LASSO is fitted using glmmLasso R package.9 The optimal lambda value is chosen by cross validation. 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. Shrinkage priors were placed on both prognostic effect and effect modifiers, but we did not put any shrinkage priors on the treatment effect or on baseline effect.

For SSVS, we used the variant of SSVS that Meuwissen and Goddard20 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 available at <https://github.com/MikeJSeo/phd/tree/master/varselect>

## Simulation results

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

# Application in real datasets

## Drug-eluting vs. bare-metal stents

Here we use the data described in [Section 2.1](#Section21). Results are shown in Table 3. After controlling for the covariates, we find that the treatment effect has wider standard error. Overall, there was little evidence of effect modification in most covariates. The covariate with the strongest effect was the indicator on whether stent placement was performed in the left anterior descending artery (ladtreated in Table 3). This covariate was selected in stepwise and SSVS model and in the Bayesian LASSO model, while its coefficient remains relatively larger compared to other coefficients.

Our readers should note that in real-life applications, results like the ones presented in Table 3 may not be very informative for clinical practice. What would arguably be more useful for clinical doctors, are the estimated relative treatment effect at clinically relevant patient subgroups. For instance, if we estimate treatment effect of patients who received stent placement in the left anterior descending artery and had five implanted stents, we get XXXX. 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 XXXX. Note that drug-eluting stents are a much more expensive technique than bare-metal stents, and in settings with limited resources, it might be worth prioritizing…

## Antidepressants

Here we use the data described in [Section 2.2](#Section22)….

# 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.24 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.24 However, currently the R packages that implements the optimization, only allows continuous outcomes.25 Other traditional approaches to best subset regression, such as the R package leaps, cannot run for high number of covariates. Thus, although promising, the best subset method may be infeasible to use in practice.

**References**

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

9. 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

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

11. 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

12. 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

13. 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

14. 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

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

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

17. Kyung M, Gill J, Ghosh M, Casella G. Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal*. 2010;5(2):369-411. doi:10.1214/10-BA607

18. Breslow NE, Clayton DG. Approximate Inference in Generalized Linear Mixed Models. *J Am Stat Assoc*. 1993;88(421):9-25. doi:10.2307/2290687

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

20. 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

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

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

23. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018. https://www.R-project.org/.

24. 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

25. 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 covariates** | **# of prognostic factors** | **True values, prognostic factors** | **# of effect modifiers** | **True values, effect modifiers1** | **# of nuisance covariates** |
| 1 | Continuous | 5 | 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 | 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 | 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 | 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 | GLMM-null | 0.000 | 0.065 | 0.038 | 0.091 |
| GLMM-full | 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 | GLMM-null | 0.000 | 0.045 | 0.033 | 0.070 |
| GLMM-full | 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 | GLMM-null | 0.000 | 0.065 | 0.030 | 0.15 |
| GLMM-full | 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 | GLMM-null | 0.000 | 0.045 | 0.10 | 0.10 |
| GLMM-full | 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** | **GLMM-null (Std. Err)** | **GLMM-full  (Std. Err)** | **Step-naïve**  **(Std. Err)** | **LASSO-naïve** | **GLMM-LASSO** | **Bayes-LASSO**  **(Std. Err)** | **SSVS**  **(Std. Err/% selected)** |
| Average treatment effect (log-odds ratio) | -0.11 (0.087) | -0.13 (0.47) | 0.25 (0.18) | -0.21 | -0.112 | -0.054 (0.44) | 0.008 (0.28/ 100) |
| Heterogeneity (τ) | 0 | 0 |  |  | 0.105 | 0.043 | 0.019 |
| age | 0 | 0.70 (0.089) | 0.81 (0.055) | 0 | 0.331 | 0.67 (0.083) | 0.66 (0.080 / 100) |
| gender | 0 | -0.022 (0.14) | 0 | 0 | 0 | -0.024 (0.10) | -0.006 (0.065 / 21.3) |
| diabetes | 0 | 0.48 (0.13) | 0.51 (0.092) | 0 | 0.081 | 0.40 (0.12) | 0.43 (0.10 / 99.4) |
| stable\_cad | 0 | -0.56 (0.15) | -0.47 (0.10) | 0 | -0.131 | -0.46 (0.13) | -0.48 (0.12 / 99.1) |
| multivessel | 0 | 0.30 (0.14) | 0.16 (0.093) | 0 | 0 | 0.20 (0.12) | 0.19 (0.14 / 69) |
| ladtreated | 0 | 0.20 (0.13) | 0.24 (0.13) | 0 | 0 | 0.086 (0.11) | 0.059 (0.10 / 35.4) |
| overlap | 0 | 0.41 (0.18) | 0.49 (0.13) | 0 | 0.024 | 0.29 (0.14) | 0.34 (0.15 / 88.7) |
| m\_dia\_above\_3 | 0 | -0.29 (0.26) | 0 | 0 | 0 | -0.10 (0.18) | -0.042 (0.15 / 35.5) |
| num\_stent | 0 | 0.017 (0.064) | 0.054 (0.057) | 0 | 0 | 0.039 (0.053) | 0.017 (0.043 / 17.4) |
| age:treat | 0 | -0.085 (0.11) | 0 | 0 | 0.172 | -0.054 (0.10) | -0.045 (0.085 / 29.0) |
| gender:treat | 0 | 0.049 (0.19) | 0 | 0 | 0 | 0.047 (0.13) | 0.018 (0.087 / 25.4) |
| diabetes:treat | 0 | -0.080 (0.19) | 0 | 0 | 0 | -0.001 (0.13) | -0.017 (0.094 / 30.3) |
| stable\_cad:treat | 0 | 0.11 (0.20) | 0 | 0 | 0 | 0.013 (0.15) | 0.006 (0.10 / 30.8) |
| multivessel:treat | 0 | -0.18 (0.19) | 0 | 0 | 0 | -0.074 (0.14) | -0.078 (0.14 / 41.2) |
| ladtreated:treat | 0 | -0.37 (0.18) | -0.34 (0.18) | 0 | 0 | -0.20 (0.15) | -0.16 (0.17 / 56.2) |
| overlap:treat | 0 | -0.043 (0.25) | 0 | 0 | 0 | 0.016 (0.16) | -0.025 (0.13 / 36) |
| m\_dia\_above\_3:treat | 0 | 0.53 (0.41) | 0 | 0 | 0 | 0.23 (0.33) | 0.14 (0.21 / 49.6) |
| num\_stent:treat | 0 | -0.077 (0.098) | -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