Analytical Methods Subcommittee Meeting Minutes

Teleconference

March 7, 2023, 10:30 AM – 1:00 PM CST

Voting Members:	**HRSA:**	**SRTR Staff:**
David Vock, PhD (Co-Chair)	Shannon Dunne, JD	Ajay Israni, MD, MS
Erika Helgeson, PhD	Adriana Martinez	Grace Lyden, PhD
William (Bill) Irish, PhD		Jon Miller, PhD
Shu-Xia Li, PhD		Nicholas Wood, PhD
Brent Logan, PhD		David Zaun, MS
Katherine Panageas, PhD		
Andrew Schaefer, PhD		

Voting Members Not in Attendance:
Megan Neely, PhD

Ex-Officio Members:
Jon Snyder, PhD (SRTR Co-Chair)

Welcome and opening remarks

Drs. Jon Snyder and David Vock called the Analytical Methods Subcommittee (AMS) meeting to order. Dr. Snyder reviewed the agenda and conflict of interest management. Dr. David Vock proceeded with the first item.

Build every time and relaxed LASSO in the program-specific reports

Dr. Grace Lyden introduced the new proposal for building risk-adjustment models used in the production of the Scientific Registry of Transplant Recipients' (SRTR's) program-specific reports (PSRs). The proposal would make two significant changes. First, SRTR would build the risk models from scratch every 6 months (ie, reselect variables for inclusion in the proportional hazards model build every time), rather than simply refitting or re-estimating coefficients from a previously selected set of covariates every 6 months with less frequent build cycles. Second, during every build, SRTR would investigate a relaxed least absolute shrinkage and selection operator (LASSO) to arrive at final parameter estimates. She explained that SRTR creates performance metrics presented in the PSRs by fitting models to national data, which are used to estimate the expected outcomes at each transplant program. Those expected outcomes are then compared with what is observed during the evaluation time period. This is done for posttransplant outcomes, pretransplant outcomes, and offer acceptance. If the proposal is approved by the committee, SRTR will implement the build every time with the option to relax the LASSO for posttransplant outcomes first, incorporating other metrics as programming progresses.

Historically, SRTR has used the LASSO in the build and fit cycle to get these expected models. Every few years, a build is done with all available predictors, using cross-validation to select the optimal lambda for penalization. With the LASSO, SRTR considers a range of lambdas that control the amount of penalization. A fit is done every 6 months, in which the selected lambda from the most recent build is used, along with that set of predictors that had nonzero coefficients in the build. These are used for a new LASSO fit with the current data. The single penalized model from that LASSO is used to get the expected outcome for a particular metric.

At the previous AMS meeting, the committee discussed transitioning from the build and fit cycle to a build every 6 months, in which SRTR could consider the full range of possible risk predictors and do cross-validation to select the optimal lambda. Dr. Lyden explained this would increase flexibility in terms of incorporating new data elements that the transplant system has started to collect, which would expand the scope of risk adjustment and result in being more responsive to national trends over time. Dr. Lyden also proposed transitioning to a relaxed LASSO approach, in which a penalized model would do variable selection to determine which variables are included but coefficients for variables that are selected are estimated without penalty (ie, relaxed). This would be explored to determine if it improves cross-validated prediction error. Dr. Lyden explained this process would reduce the bias of the LASSO estimator toward zero and increase the magnitude of model coefficients, leading to more risk adjustment and potentially better performance. In the previous AMS meeting, members supported using this process if empirical results showed better performance.

Dr. Lyden then described the current approach of the build and fit cycle. The build is done every few years. If the evaluation cohort contains at least 10 events, SRTR attempts to develop a risk-adjustment model. If there are fewer than 10 events, an intercept-only or unadjusted model is developed. Ten multiply-imputed data sets are developed to deal with missingness, and each undergoes a cross-validated LASSO fitting process. With each imputed data set, a range of lambdas is considered, and each lambda gets an estimate of the prediction error for the penalized regression at that lambda. This is followed by another check for if it should be an intercept-only model. For each of the multiply-imputed data sets, SRTR analyzes the cross-validated error at lambda 1 standard error (SE), or the lambda that results in the most parsimonious model with the cross-validated error within 1 SE of the minimum cross-validated error. If nine or more of the multiply-imputed data sets have variables selected at lambda 1 SE, the process moves forward with an adjusted model. Continuing the process, one value of lambda is selected and applied to all 10 multiply-imputed data sets—that is, the value of lambda that minimized the median cross-validated error across the 10 imputed data sets. Given the one value of lambda, a fit is obtained in each of the multiply-imputed data sets and final coefficients are averaged across the 10 multiply-imputed fits.

Dr. Lyden transitioned to the proposed approach of building every time with the option to relax the LASSO. She and Dr. Jon Miller devised a method to only relax the LASSO if doing so improves the model, which is defined by reducing cross-validated prediction error. The method is essentially a two-dimensional tuning parameter that 1) selects the optimal lambda and 2) decides if relaxing improves prediction error. Whatever combination of these tuning parameters results in the lowest cross-validated prediction error is the desired combination with which to move forward.

Dr. Brent Logan suggested potential alternatives, such as using a grouped LASSO to fit the imputed data sets in one model. The data sets could be stacked, and grouped penalization could guarantee consistent variable selection across imputed data sets. Dr. Katherine Panageas asked why the work would be done every 6 months, and if 10 or more events was adequate. Dr. Snyder noted that SRTR runs this process every 6 months since SRTR is obligated to release evaluations every 6 months by contract. Dr. William Irish asked what would change within the 6-month time frame, and whether this would change how a variable affects outcome. Dr. Lyden explained that the baseline hazard of outcomes can change over time, as risks vary over time. Dr. Miller added that they would not necessarily expect a lot of change in 6 months; however, building every 6 months rather than every few years allows better capture of temporal changes. Dr. Panageas asked if SRTR was comfortable with potentially dealing with a big change in 6 months within the transplant system. Dr. Miller noted that large fluctuations do not tend to occur, with hazard ratios (HRs) not changing substantially between the build and the fit. Dr. Lyden pointed out the importance of doing the expected model fitting in tandem with assessing what was observed, in the case of changing underlying risk and the potential to overestimate or underestimate expected outcomes.

Dr. Vock asked how important it was to have model stability of all beta coefficients if covariates change dramatically from build to build, which was a user trust issue. Dr. Logan added that if the ultimate goal is to look at center outcomes, 11 events is not enough to be doing that. Because the data would be across a lot of centers, a higher threshold would be better to do this reliably. He said it was unclear if we should seek risk adjustments with so few events. Dr. Logan believed that everything would end up shrunk to the mean in the presence of so few events.

Members discussed what might be a better threshold. Dr. Lyden said that, for the builds, it is not uncommon for many parameters to be estimated, particularly when continuous spline basis expansions are included, which can add up to 20 parameters per continuous variable. Dr. Miller added that the largest models currently have approximately 80 parameters selected. Dr. Vock asked how many events were needed before selecting more than one factor. Covariates might not be selected until there are 15 events, for example. SRTR did not have these data readily available but can look into this question.

Dr. Nicholas Wood gave his perspective regarding the comment of whether the 6-month cadence impacts estimates. Running a parallel process described here for offer acceptance which uses a 1-year cohort, half of the data get turned over each PSR cycle. He said that offer acceptance behavior can react to changes in allocation policy. The coefficient estimates could change dramatically depending on how that allocation policy works, such as who is getting what offers, and how clinicians respond to the offers.

Dr. Lyden asked Drs. Logan and Panageas if it was important that multiply-imputed data sets select the same variables, and if there were strong reasons to move toward this. Dr. Logan said an advantage of this was it provided some stability to the variable selection process, however, it may backfire from a prediction performance standpoint. He did not have a good sense of how it would end up performing. Dr. Panageas agreed.

Next, Dr. Miller presented a comparison of posttransplant evaluations from the published January 2023 PSRs to the evaluation using the proposed build and relaxed LASSO process. He reminded the committee that a major concern is finding a process that can be run within the given time frame. The draft PSR release has about 15-20 days from the time that the data are available until results need to be published. This left roughly 3 weeks to run 56 different models (for posttransplant outcomes alone). The new proposed build process ends up with 37 out the 56 models having any risk adjustment, including eight of the 14 1-year graft survival models. The current fit process had 25 of the 56 models having any risk adjustment, including seven of 14 1-year graft survival models.

Dr. Miller explained the reasons the intercept-only model was chosen in the build process. In 11 of the 19 cases, it was too few events (fewer than 10 events), with these cases generally happening in pediatric and living donor models. He noted that there are some cases where the program's overall evaluation may be based on a risk-adjusted deceased donor model in combination with an intercept-only living donor model. In eight of the 19 cases, it was too few of the imputations selecting predictors (ie, more than one imputation had zero predictors selected). These were generally cohorts that had few events to begin with.

Comparing the build process to the fit with in-sample concordance (the in-sample C statistic) demonstrated that model discrimination was improved in every case where there was a risk-adjusted model. The number of variables selected was almost always more in the new process. There was also strong overlap in the predictors that were selected by the new process as compared to the current process.

Dr. Miller moved on to a comparison of HRs resulting from the new process as compared to those of the current process. In comparing 90-day adult HRs between the new process and the current process, HRs were very similar, with very few programs changing in terms of whether they are flagged for review by the Membership and Professional Standards Committee (MPSC). Programs that did flip were close to the boundary (HR=1.75). For the 1-year conditional adult HR, there are a couple more that are flagged under the build. For the pediatric HR comparisons, there were no flagging changes. Dr. Miller said there were a handful of programs that move between the tiers, and most programs stay within their tier. There were no programs that moved two or more tiers when the build and fit processes were compared for adult recipients. Most pediatric programs also do not move at all, and the four that do only move one tier.

Dr. Miller said that assuming SRTR goes ahead with this for the January 2024 PSR, SRTR will explore ways to expand the lambda sequence fed to the relaxed LASSO. Timewise, the relaxed LASSO is the limiting factor, and SRTR is looking at a few other algorithms to expand the number of relaxed models that can be explored. In a few cases, there were problems with convergence, in which cases the process reverts to the penalized model.

Dr. Ajay Israni brought up the issue of the variables in the model changing every 6 months, as some centers use SRTR-provided workbooks to populate them with recent or upcoming transplants. So, for that function, the workbooks would not work as well potentially. New variables added could also add burden on the center. Dr. Israni said they need to be very transparent with the community about these changes, perhaps pursuing a publication in the *American Journal of Transplantation*.

Dr. Logan expressed concern about the C statistic. The C statistics shown are the in-sample C statistics. The C statistic should be expected to go up in-sample because the process is being optimized based on in-sample data. He cautioned against overinterpreting this result. Dr. Miller suggested exploring different measures of discrimination and predictive accuracy. Dr. Logan said to get a bias-corrected version via cross-validation or bootstrapping of the full data set, which is a time-intensive process. Dr. Vock agreed with Dr. Logan's concerns about interpretation. He also wondered if the broader community was comfortable with adding more predictors into these risk adjustments, which he was fine with.

Dr. Lyden thought that, philosophically, the community was generally in favor of more risk adjustment. Dr. Irish agreed, as there is always a concern that what is being included is not enough. Dr. Snyder also agreed with this, and also stated that it seemed relaxing the LASSO did not have a huge impact, as it was only chosen in 12% of the adjusted models. Dr. Lyden suggested expanding computational capacity to look at more lambdas for relaxing. Dr. Snyder thought if it could be done computationally and, if more variables were selected, it would be viewed favorably. He suspected programs would welcome more risk adjustment, and SRTR should be expanding its model-building strategy to try to improve risk adjustment for centers.

As the committee did not have major concerns, SRTR plans to continue with this project, anticipating it will be launched in January 2024. Drs. Lyden and Miller will look into the suggestions, and the committee will be updated on progress in the next meeting where there will be opportunity to discuss any new concerns or insights.

Core risk adjustment for intercept-only LASSO models

Dr. Lyden said the LASSO process can result in an intercept-only model, even when there are 10 or more events. If more than one imputed data set does not select anything, the model is intercept only, with nothing selected. This happened in eight of the 56 posttransplant models. Risk adjustment is the cornerstone of program evaluation, and it becomes difficult to argue that an intercept-only model is providing risk adjustment, even when it is the optimal model on the LASSO path, because there are no risk factors in the model.

Because of this, the idea of forcing in a core set of risk-adjustment factors has come up. The idea of a core set would lean on variables that are identified *a priori*. For example, age at transplant is always related to patient survival. Similarly, primary cause of end-organ failure is generally accepted as a factor that relates to posttransplant risk. Dr. Lyden proposed consideration of a small set (two to three) of risk-adjustment factors in a post-LASSO processing step if the LASSO leads to an intercept-only model.

She explained there has been resistance to this idea historically, due to the belief that if the LASSO does not select any risk-adjustment factors, the best possible model is an intercept-only model. Her opinion is that this is not necessarily true because the LASSO only considers the models that are implied by the path of lambdas given to it. Each result in a penalized model, which identifies a certain set of predictors. Every possible set of predictors is not considered because an exhaustive search like that is computationally prohibitive, which is why LASSO or relaxed LASSO are used instead of best-subset selection. While the relaxed LASSO is a good approximation of the subset

selection, only a few lambdas are considered for relaxing in the SRTR approach. Dr. Lyden said this indicates there may be good models out there that currently are not being considered in the process, especially when there are few events to identify the best predictors from 250 used as inputs. An example of this can be found in the paper "False discoveries occur early on the LASSO path" in the *Annals of Statistics*, where the authors demonstrate that even when effects are strong, true and false predictors are always going to be interspersed on the path. So, it makes sense that early in the path, where values of lambda are large, the few predictors that come into the model might be a mix of true and false predictors.

Dr. Lyden proposed considering what the *a priori* good model might be when the LASSO process results in an intercept-only model. Supposing there is sufficient evidence for modeling and there was an intercept-only model despite attempts to identify adjusters, the following process would be followed. First, using the same folds as in the LASSO, we would obtain the cross-validated error for an intercept-only model. Then, we would obtain the cross-validated error for an unpenalized model that includes the two to three *a priori* selected core risk factors for that organ and outcome. Between these two models, the model with lower cross-validated error would be chosen.

Dr. Lyden noted input would be sought from SRTR senior staff about which small set of predictors should be explored for each organ and outcome being modeled.

Dr. Lyden asked the committee if this approach seemed reasonable. She said if it resulted in more adjusted models, it may enhance face validity and possibly improve estimation of expected. Dr. Vock asked if they considered something in between the two options of 1) forcing in a small subset of covariates and 2) just letting the variable selection run on its own. Another option is to set a lower penalty for some variables during the variable selection process. He said an argument could be made that the five core predictors, for example, should be penalized less.

Dr. Snyder suggested taking an informal survey of SRTR senior staff to determine what variables to adjust for; Dr. Vock said there probably would be selection bias. Dr. Lyden thought adjusting for age would be a good idea, since there was a distribution across transplant centers.

There was no major opposition to Dr. Lyden's proposal. Dr. Lyden said it had yet to be tested. Dr. Vock suggested looking at the number of events of the intercept-only models from the recent analysis. The committee supported further exploration of this option.

Closing business

With no other business being heard, the meeting concluded. The next meeting will be scheduled for mid-summer 2023.

References

1. Su W, Bogdan M, Candès E. False discoveries occur early on the LASSO path. *Ann Stat*. 2017;45(5):2133–2150.