SRC Analytical Methods Subcommittee Meeting Minutes

Analytical Methods Subcommittee Teleconference

August 2, 2022, 11:00 AM – 1:30 PM CDT

Voting Members:
David Vock, PhD (Co-chair)
Shu-Xia Li, PhD
Brent Logan, PhD
Katherine Panageas, DrPH
Not in attendance:
Andrew Schaefer, PhD

New Voting Members:
Erika Helgeson, PhD
Megan Neely, PhD
Not in attendance:
William (Bill) Irish, PhD

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

HRSA:
Adriana Martinez
Shannon Dunne, JD

SRTR Staff:
Ajay Israni, MD, MS
Ryutaro Hirose, MD
Jon Miller, PhD
David Zaun, MS
Grace Lyden, PhD
Larry Hunsicker, MD
Nicholas Wood, PhD
Not in attendance:
Josh Pyke, PhD

Welcome and opening remarks

Dr. Jon Snyder called the Analytical Methods Subcommittee (AMS) meeting to order. Dr. David Vock introduced new voting members:

- Erika Helgeson, PhD, Assistant Professor of Biostatistics, University of Minnesota, works in kidney transplant recipient and kidney donor outcomes
- William (Bill) Irish, PhD, East Carolina University, kidney transplant work
- Megan Neely, PhD, Duke University, posttransplant lung outcomes research

Dr. Vock reviewed the agenda, and Dr. Snyder went over conflict of interest management. Dr. Vock proceeded with the first item.

Relaxed LASSO in the program-specific reports

Dr. Grace Lyden described the purpose and process of the Scientific Registry of Transplant Recipients (SRTR) fit-and-build cycle. One of the responsibilities of SRTR is to evaluate transplant center performance, which is done by comparing observed counts of outcomes to what would be expected based on a national model (includes pretransplant, offer acceptance, and posttransplant outcomes). Currently, SRTR uses a model "build" and a model "fit" cycle. Variables to be included in

each model during each 6-month evaluation are chosen during the build cycle, which historically has been performed every few years.

During the build cycle, risk factors are chosen using the least absolute shrinkage and selection operator (LASSO). The LASSO identifies coefficients that are nonzero on average across 10 multiply-imputed datasets. During the fit stage, which happens during each 6-month evaluation cycle, the selected variables are used to fit the model to the current cohort. The LASSO is used similarly to in the build process, but cross-validation is not used to get the optimal lambda value in the fit. Instead, the optimal lambda identified by the build is used to fit one penalized model.

In June 2021, the AMS heard a proposal to move to a more frequent build, or constructing the model based on all available data every 6 months. The subcommittee discussed the possibility of implementing a relaxed LASSO, which uses LASSO purely for variable selection but not coefficient estimation. The proposed implementation involves fitting the LASSO model and selecting the optimal lambda through cross-validation. Predictors with nonzero coefficients are selected, and a second, unpenalized model uses only the selected predictors.

Dr. Lyden said a reason why SRTR thought it should replace the current process of building every few years with building every 6 months was that a lot of variable selection happens at the build stage, leading to variable selection based on old data. This may not be optimal, as variables become more or less predictive over time. Also, the current approach can be inflexible when new data elements become available. It would be beneficial to include new data elements with current factors that were not available at the build stage but are now available for the entire evaluation cohort. Additional reasons the AMS supported building more frequently and using the relaxed LASSO included the need to consider all available predictors each cycle and a relaxed LASSO reducing downward bias induced by the LASSO penalty.

In contrast, a reason not to build every 6 months is the short 3-week period available to prepare the initial draft report release and the 1-month period to prepare the final reports. During these short preparation windows, running a build cycle again if an issue occurred may hinder SRTR's ability to meet reporting deadlines. Another concern is the possibility of an increased burden on transplant programs, if the set of risk-adjustment variables was changing more frequently than it currently does.

Dr. Lyden said the aims of this presentation were to describe considerations for running a build at each program-specific report (PSR) cycle, hear from the subcommittee on how to make this process more feasible, and compare the current fit process (taking variables from the build and fitting LASSO with one specific lambda to current data) to the fit with a relaxed LASSO (fitting LASSO with one specific lambda followed by an unpenalized model with the predictors that had nonzero coefficients).

Dr. Jon Miller explained that one of the program reports was run to introduce the model without the LASSO penalty and compare hazard ratios (HRs), as well as Membership and Professional Standards Committee (MPSC) flags, from the last PSR round. The process of producing these PSRs includes data cleaning and fitting intercept-only models (model of the outcomes without any predictors). At the point of the draft release in March, centers are not provided with any expected values or HRs but can check the data cleaning process; so, in theory, the model fitting process could be stopped at

the intercept-only models for the draft release.

However, the other reason for running a draft release is to test the multiple imputation and LASSO models. For example, if a variable in the multiple imputation step is too correlated to other variables, it can be identified early and predictors can be changed so imputations can run. A 10-times multiple imputation on the dataset is run, which takes significant time since the data are at the level of the individual patient. When predictions are made, the least favorable value is given to individual patients with missing data, and all patients remain in the dataset the whole time. The cumulative hazard from the LASSO is used to calculate expected counts for each patient, and counts are summed within centers. The sum at a center is used as the expected count, and HRs are calculated from observed (O) and expected (E) event counts using a Bayes shrinkage from a gamma (O+2, E+2) posterior.

Dr. Miller presented data from a build and fit from the July 2022 PSR cycle. He said there was the problem of running the build in sufficient time to meet the deadline for program reporting. He gave a comparison of the adult heart 1-year graft survival. The cohort contained 6,500 patients, being one of the 56 models fit for the PSR. Adult heart is one of the moderate cohort sizes, with kidney being bigger and intestine smaller. It is considered an average representation for build compared to fit.

The LASSO is used for both the build and fit stages, but in the build stage, the penalty factor (lambda) is selected by cross-validation. In the fit stage, the LASSO uses the optimal lambda from the build to penalize coefficient estimates, resulting in additional variable selection after the build. An advantage of reselecting at the fit cycle is that the continuous variables are constructed with linear splines, and the LASSO can re-estimate the nonlinearity in continuous predictors. Dozens of splines are considered in the fit process for use in overall variable selection and for selecting the form of the nonlinearity for continuous predictors. For the heart 1-year graft survival in the build, this was 101 variables. Expanding these out to the categorical including the linear splines, the design matrix totaled 652 individual predictors.

The fit cycle does not consider the variables that are not selected in the build process. Forty-eight predictors totaled in the heart graft survival fit with a design matrix of 572 columns. Lambda selection done through cross-validation makes a build cycle take longer. At least 300 different lambdas are selected for best cross-validated fit consideration, which takes 2.5 hours for heart and 5 hours for kidney deceased donor. By comparison, the fit process has 55 lambdas, and uses the lambda that was selected in the build process. It does not do a cross-validation and takes only seconds to estimate.

Running a build cycle as currently done, with each model being estimated in sequence, would add 4 to 6 days to PSR production, which would cause problems for meeting the deadline if anything needed recalculation. Options to get the build to run faster include reducing the length of the LASSO's lambda sequence or only using the first multiply-imputed dataset to choose lambda and then using the chosen lambda to estimate coefficients across all 10 multiply-imputed datasets. Reducing the length of the lambda sequence resulted in minimal time savings; using only the first multiply-imputed dataset to select lambda reduced time but may be suboptimal. Another method considered was running kidney, heart, liver, lung, and intestine in parallel rather than sequentially. This reduced the multiple-implementation process from 3 days to 1. Additionally, if SRTR converts to using a period-prevalent approach, this would eliminate the need to fit separate models for 1-year

and 3-year outcomes. Under the forthcoming period-prevalent approach, the number of models would therefore be reduced by at least half.

Dr. Miller compared the current fit with the relaxed LASSO. The summer PSR was calculated, followed by using a model derived using the relaxed LASSO. Dr. Miller presented comparisons of model coefficients and derived center-level HRs.

In coefficient comparisons for the lung 1-year adult graft survival model, the coefficients from unpenalized models are much further away from the null. Some of the more substantial changes are in the continuous variables that are fit with the linear splines. Interpretation is difficult just from analyzing the coefficients for splines.

Graphs for 90-day graft survival and 1-year, conditional on surviving to 90-day, graft survival demonstrated high correlation between the current approach and the relaxed LASSO. Few programs moved substantially. One fewer program was flagged for review with the unpenalized model.

Dr. Lyden said a caveat to keep in mind was that the initial variable selection and cross-validation of the LASSO models was still occurring only in a build stage using prior data, and Dr. Miller did not run a new build on current data for this analysis.

For this analysis, there were a total of three models: 1) LASSO with cross-validation on all data (build – not rerun for this analysis), 2) LASSO with no cross-validation on current data (fit), and 3) unpenalized Cox proportional hazards model on current data (fit). She asked the subcommittee for their thoughts on reducing build timing and if it was necessary to confirm relaxed LASSO was more "correct" (under simulation) or just as good as the current approach.

Dr. Vock asked for clarification on the goals of variable selection, specifically what they were hoping to get out of doing variable selection. Dr. Miller noted that the goal was to achieve the best prediction of an individual patient's outcome, and Dr. Lyden added the best prediction for a patient if they were to receive care someone like them (with their risk profile) receives on average across all centers. This is why models used for expected counts do not include center and are fit on a national level. Dr. Ryutaro Hirose said that building calculators that consider center-specific results can help patients and providers make decisions, although it is important to be transparent about selected variables for which an outcome is risk adjusted. Dr. Vock suggested the Minnesota Supercomputing Institute for reducing build time—buying and maintaining servers through the institute could help increase capability to run build cycles in parallel.

Dr. Shu-Xia Li brought up the possibility of clinicians giving input on what variables should be included. Dr. Lyden thought it would be a challenge implementing the practice with the PSRs and might be subjective if SRTR clinical staff weighed in on which factors to force in the models. Dr. Snyder said that SRTR had attempted a process of bringing variable lists to OPTN organ-specific committees for feedback (little feedback was received) and noted at least one instance where a variable was forced into models due to a new development in the field. He offered his opinion that a documented, reproducible process would be preferable to a panel of clinical experts recommending variables, while noting that SRTR has a process of bringing recommended model changes to the SRTR Review Committee (SRC) for consideration. Dr. Brent Logan said his organization (Center for

International Blood and Marrow Transplant Research [CIBMTR]) allows the community to provide variable feedback every 2 to 3 years on what should be incorporated into the models. Dr. Snyder thought it was possible SRTR could implement a similar process. Dr. Hirose expressed concern that variables could change significantly from build to build, and asked if model stability (in terms of which variables are included) could be assessed under the new approach. Drs. Lyden and Miller agreed analyses could be conducted to apply the build cycle to each PSR and assess changes to the set of selected variables across cycles.

Dr. Vock questioned if a uniform variable list should be data driven or determined by expert agreement. Dr. Miller sided with data driven, as the purpose of the PSR process is achieving the best predictions as determined by cross-validated prediction error, not applying causal interpretation to coefficients. A data-driven approach produces the best possible prediction. Dr. Snyder said SRTR erred on the side of creating accurate predictions, rather than relying on subjective face validity to drive variable selection. Dr. Vock thought a middle ground might be possible. Dr. Hirose said they should always try to get the most accurate predictors of outcome, and that it is impossible to add all variables clinicians request since they are not always collected in a reliable fashion. Dr. Lyden said that the period-prevalent method would ensure uniformity in variable selection across the 1- and 3-year posttransplant models.

Dr. Miller asked if it was a worthwhile to do simulations, as opposed to relying on published literature supporting the use of the relaxed LASSO. Dr. Megan Neely was comfortable with relying on the literature and asked if bias was the only metric SRTR was considering or if SRTR was also considering the variability or noise in the predictions that result from the relaxed versus standard LASSO, as the standard LASSO might be slightly biased but with less variability. Dr. Lyden said they were not proposing fitting an unpenalized model and choosing lambda based on the estimate of prediction error for the two-step process. She was not aware of any literature saying that prediction error would be smaller if SRTR implemented the version currently under discussion. Dr. Neely also asked if SRTR put confidence bands on the scores given to programs, and Dr. Miller said HRs have 95% Bayesian credible intervals.

Dr. Logan suggested doing bootstrap resampling on the current data set, and compare bootstrap sampling-based estimates of predictive performance. For each bootstrap sample, apply either a relaxed LASSO or current approach, and use it to get a bias-corrected prediction performance. Accurate predictions for independent patient profiles is what should be driving decisions, and bootstrap sampling could be used to determine how prediction performance fairs with these two approaches. Dr. Li agreed.

Dr. Vock pointed out, in the context of how the LASSO was planned to be used in the refitted/relaxed approach, that it may be beneficial to use fewer linear spline terms and consider basis functions that would be less correlated and still allow for a lot of nonlinearity. Dr. Katherine Panageas asked if the intention behind running simulations was to publish a manuscript with a comprehensive simulation study, with confirmatory data to appease concerns about the relaxed LASSO moving forward. Drs. Lyden and Miller agreed that simulations would be for confirmatory purposes.

Dr. Vock thanked the committee for the robust discussion and encouraged Drs. Miller and Lyden to continue the development process and report back to the committee.

A race-free kidney donor profile index

Dr. Miller presented an SRTR study that investigated kidney donor profile index (KDPI) being calculated without the Black race factor. The current study is in preparation for peer-reviewed publication. Estimated glomerular filtration rate (eGFR), an estimate of kidney function used for determining when a kidney candidate becomes eligible to accrue waiting time, was previously calculated from serum creatinine, age, and a Black race predictor, which has recently been removed from policy due to concerns it induces bias and may contribute to racial disparities in access to kidney transplant.

Other algorithms using a Black race predictor are now being scrutinized, including the kidney donor risk index (KDRI), which, when converted to a percentile score (ie, the Kidney Donor Profile Index [KDPI]), currently affects kidney allocation. SRTR aimed to recreate the original KDRI analyses as closely as possible, using the results as a comparison to what would happen if the Black race variable was removed from the model. The analysis included looking at strengths of the coefficients and predictive value changes. For the refitted KDRI graft failure model, the donor, transplant, and stratifying factors that were included in the proportional hazards model matched the functional forms of the ones in the original Rao paper. Because the paper did not describe which recipient variables were selected or their functional forms, recipient variable selection was recreated through a backwards selection process. After producing the closest re-creation of the original model, the Black donor race variable was removed and estimations were made for the KDRI and KDPI. The predictive power of the model was also estimated from concordance and integrated Brier score in the evaluation dataset.

For KDPI conversion, a newer cohort (2015-2021) was used. Dr. Lyden added a prediction of kidney nonuse (organ recovered for transplant but not transplanted) under race-free KDPI (kidneys from deceased donors 2015-2021 cohort), with logistic regression for nonuse conditional on the closest KDPI re-creation as well as all other underlying KDRI variables. This model includes a direct effect of the KDPI or “labelling effect” (effect of KDPI above and beyond the underlying variables), which was used to predict what would happen if Black donor race was not included in the original KDRI model.

Dr. Miller showed that removing the Black race indicator from calculation of KDRI did not meaningfully change the predictive power of the model. He noted that other model coefficients changed, some substantially (eg, the effect of serum creatinine), when removing the Black race coefficient.

Under the current KDRI/KDPI formulation, 31% of Black donors are classified as highest risk (KDPI over 85). When removing the Black race variable from the calculation, 17% of Black donors are classified into the highest risk category. By contrast, non-Black donors classified as high risk increased from 13% to 16%. This happens because the KDPI is a percentile rank. If some donors have their KDPI lowered, others must have their KDPI increase. Approximately an equal number of Black donors were moved out of that highest risk as there were non-Black donors that were moved into the highest risk, such that the fraction of all donors that were classified as highest risk remained at 16%. Predicted nonutilization remained at about 20% for each formulation of KDRI.

Dr. Miller concluded there was no substantive impact on the model due to removing the Black donor race variable, and there are no indications of negative effects (ie, increased nonutilization) if it were

removed. Improved equity for Black recipients could potentially be better promoted by constructing a new paradigm for calculating kidney donor risk. One possibility for overhauling KDRI/KDPI is moving away from the relative measurement of percentile based on each year (KDPI) to a measure of absolute risk.

Dr. Vock asked if the modeling of nonuse was based on an assumption that the way physicians react to an offer would remain constant under race-free KDRI. Dr. Miller confirmed this, saying that is how the model and logistic regression approach would behave. When simulations for policy development are done, that assumption is generally made. Dr. Lyden added that because the analysis is retrospective in design, it might not be best to go so far to say they are assuming future behaviors. Simulations of changes to allocation policy that use past data are not applied to predict what would happen in the future. It is also unclear from the nonuse modeling how things would play out if KDPI were different in allocation since there are considerations for who the organ is offered to and whether it would be accepted given acceptance practices that might depend on the labeling. Dr. Lyden said it was a dynamic system and did not consider the logistic regression to be an adequate model to assess if different KDPI would change the breakdown of waiting time for Black transplant candidates, which organs they end up getting, and how the organ quality affects their posttransplant outcomes. She suggested taking the nonuse model a step further and simulating the impact on Black candidates, not just the impact on the organs themselves.

Dr. Hirose said donor families preferred the term “nonutilized” (“nonused”) organs as opposed to discards. He also added it was important to establish the purpose of the metric before figuring out what to include and what not to include. Dr. Larry Hunsicker said it was suboptimal to use a somewhat poorly documented risk stratification from 10 years ago. As the study belongs to OPTN, any discussion of it should be done in conjunction with OPTN. Dr. Snyder suggested producing a better donor risk index, as well as presenting the race issue before the kidney committee. Dr. Hunsicker said it was better to focus on the issue in terms of social consequences, as eliminating or keeping race in KDPI or KDRI appears to make little difference. Dr. Vock added that future work should assess if these changes would positively affect access for Black kidney transplant candidates.

Closing business

With no other business being heard, the meeting concluded. The next meeting date is to be determined.