

SRC-AMS Meeting Minutes

Analytical Methods Subcommittee Teleconference

July 11, 2024, 12:00 PM – 2:30 PM CDT

Voting Members:

David Vock, PhD (Co-chair)
Joel Adler, MD, MPH
Syed Ali Husain, MD, MPH, MA, FASN
Erika Helgeson, PhD
William (Bill) Irish, PhD
Megan Neely, PhD
William Parker, MD, MSCP, PhD

Not in Attendance:

Brent Logan, PhD
Andrew Schaefer, PhD

Ex-Officio Members:

Jon Snyder, PhD (Co-chair)

HRSA:

Kayla Rochelle

Not in Attendance:

Shannon Dunne, JD
Adriana Alvarez, MS

SRTR Staff:

Larry Hunsicker, MD
Ajay Israni, MD, MS
Grace Lyden, PhD
Jon Miller, PhD
Nicholas Wood, PhD
Dave Zaun, MS

Welcome and opening remarks

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

SRTR's 5-tier metric summaries

SRTR produces 5-tier summaries of key transplant program performance metrics, as recommended by participants in SRTR's 2012 Consensus Conference on *Transplant Program Quality and Surveillance*. Dr. Snyder reviewed the SRTR search results page for transplant centers, which present these 5-tier classifications, updated semiannually, for survival on the waiting list, getting a deceased donor transplant faster, and first year graft survival.

Dr. Snyder reviewed the methodology for converting the Bayesian posterior distribution of the program's rate ratio to a score ranging from 0 to 1, which is then assigned to 1 of 5 performance tiers. Dr. Grace Lyden described a few example histograms that illustrated applying the score function to 10,000 sampled rate ratios from the posterior distribution of the program's rate ratio. The method was published in the journal *Health Services Research*¹ in 2018. SRTR is currently exploring refining the system such that each metric uses the same scoring function. This was

¹ Wey A, Salkowski N, Kasiske BL, Israni AK, Snyder JJ. A Five-Tier System for Improving the Categorization of Transplant Program Performance. *Health Serv Res*. 2018 Jun;53(3):1979-1991. doi: 10.1111/1475-6773.12726. Epub 2017 Jun 13. PMID: 28608369; PMCID: PMC5980219.

brought before the AMS at their March 2024 meeting and reviewed by the SRC at their April 2024 meeting. During the April 2024 meeting of the SRC, members reviewed the methodology and expressed the desire for the AMS to make the decision on refining the 5-tier methodology. Some SRC members supported using a consistent score function, while others supported the current approach allowing different cut points of the tiering function and/or score function. Dr. Vock pointed out in the SRC meeting that the score function is monotonic and rank preserving, so SRTR can achieve goals by either changing the score function or the tiering cut points. Dr. Vock added it would be helpful to have sentiments expressing clinically meaningful distinctions between tiers.

Dr. Snyder showed graphs highlighting how different score functions affect the tier assignments. Dr. Snyder said exponents in the score function that are closer to zero result in a shallower slope. Shallower slopes result in fewer programs in the extreme tiers, and often result in pulling programs toward that middle tier. Dr. Snyder posited SRTR should adopt a standardized shape parameter ± 4.82 , which assigns similar weights to rate ratios less than 0.33 or greater than 3.0.

Dr. Vock was concerned there would be questions of whether the rate ratios of 3 and 0.33 would have the same meaning across different metrics, and if the tiering had clinically relevant thresholds. Dr. Snyder reviewed a web-based application SRTR has developed to visualize the impact of different scoring functions on tier assignments across various performance metrics. He pointed out the need to explain the difference between Tier 3 and Tier 5 to the general public, and Dr. William Irish questioned what “average” means, and how the word is interpreted. Dr. William Parker said more direct feedback from the SRC may be needed on whether to display absolute rates rather than relative risks/rates. Dr. Irish suggested having a qualitative descriptor of the program that the average patient could interpret.

The AMS continued to debate whether to adopt a standardized score function across all metrics. Dr. Snyder suggested the subcommittee review the other material on the agenda, which may better inform their decision. He suggested the attendees first review Dr. Jon Miller’s work on presenting absolute risks to better contextualize the tier assignments. The committee agreed to hear Dr. Miller’s presentation.

Dr. Miller first reviewed components of the public SRTR website that provide predicted absolute rates for waitlist mortality, transplant rate, and first-year graft survival by tier for an “average risk” patient to provide context to the tiers. In addition to the predicted outcome within each tier, the user can access a “Learn More” page, which also provides the minimum and maximum predicted outcomes within each tier, standardized to an average-risk patient. The average patient’s expected outcome is obtained from the national Kaplan-Meier estimate.

Dr. Miller said that one proposal for integrating the more accessible metric of 1-year survival into the tiers would be to use the minimum and maximum to anchor the ends of the gauges, if moving towards a gauge (ie, speedometer-type) icon. Dr. Larry Hunsicker noted that while there is not much absolute difference across tiers for 1-year graft survival, there is a much larger difference for transplant rates, and this should be emphasized and made clear to patients. Dr. Irish agreed. Dr. Vock said it may be helpful to have tiers constructed such that, for example, Tier 5 centers are twice as good as the national standard. Dr. Hunsicker noted this could be done in terms of relative risk, but does not solve the absolute risk issue.

Dr. Snyder asked members if SRTR should remove the Learn More link a user needs to click on to see the range within each tier, and instead add the ranges on the main site. He also asked members about endpoints, such as anchoring the ends of the gauge icon. If both the minimum and maximum are used in absolute predicted survival, the median could go inside each tier range. Dr. Lyden suggested an infographic at the top of the page showing the median in each tier. Dr. Hunsicker said these questions should be reviewed by the Patient and Family Affairs Subcommittee (PFAS) and the Human Centered Design Subcommittee (HCDS).

Members were comfortable with how SRTR comes up with the average patient survival at a program within the tier. They discussed how to display ranges, with Dr. Hunsicker suggesting a midpoint of each tier and Dr. Vock advising against placing ranges earlier on the website because of a potential for making the information more cluttered and confusing. Dr. Joel Adler added that there was value in transparency in making these ranges obvious on the website, from the standpoint of communicating perspective and uncertainty around program placement. Ms. Amy Ketterer added her experience of talking with patients in her role and thought that ranges would be confusing for patients. The subcommittee agreed that SRTR should explore a key graphic near the search results that makes absolute risk within each tier assignment more obvious to the user and continue to have a "Learn More" link if a user wishes to see more information about expected outcomes within each tier.

Dr. Snyder then introduced Dr. Lyden to review various methodologies under consideration for displaying uncertainty in tier assignments.

Dr. Lyden's presentation posed two questions to the subcommittee:

- 1) Do members agree on the proposed methodology to calculate credible intervals?
- 2) What are additional thoughts on other methods of conveying uncertainty in a program's tier assignment?

Dr. Lyden discussed the motivation behind credible intervals for program scores. Program performance can be on the line between tiers, there is not 100% certainty behind programs' underlying performance, and the SRC strongly supported the idea of adding credible intervals around program score to show it may extend over multiple tiers. She said the most logical choice for credible interval is 2.5th and 97.5th percentiles of distribution that would give a standard two-sided 95% credible interval for the score.

Dr. Lyden showed an example of a midsized program with 10 observed events and 12 expected, resulting in a rate ratio of 0.86. There were 10,000 draws sampled from the posterior distribution of the rate ratio, and fed through a score function with uniform K and the 2.5th and 97.5th percentiles of the distribution were taken as the limits of the 95% credible interval. She then noted that there is no need for sampling or to mathematically derive the distribution of the score because the score function is strictly increasing or strictly decreasing. Therefore, the 2.5th and 97.5th percentiles of the posterior distribution of the rate ratio can simply be put through the scoring function to derive the 2.5th and 97.5th limits of the credible interval of the score. Dr. Lyden noted that using a 95% credible interval can result in wide intervals for the score.

Dr. Lyden proposed three possible solutions. For wide intervals, SRTR could reduce the nominal coverage probability, from, for example, 95% to 50%. However, she noted this may be nonstandard since 95% credible intervals are a generally accepted standard. The second solution would be to add shading on the icon to indicate the density of the score's distribution. The last solution is showing the probability of being in each tier, or higher or lower than the assigned tier, instead of using credible intervals.

Dr. Lyden demonstrated these solutions using examples of tier probabilities. A very small program had some chance of being in any tier, while another example showed quite a bit of uncertainty and no probability that it is a Tier 1 program, with a 60% chance of being in Tier 4 or Tier 5. Another example was with a Tier 1 program that had a 23% chance of being in Tier 2.

Dr. Snyder suggested for the five tier icons, showing a bar chart with the heights of the bars determined by the probability of being in each tier. Dr. Adler suggested giving a range of where the absolute performance was. Dr. Hunsicker said there would be fewer problems using the median rather than the mean, which are different because the score distribution is not symmetric. Dr. Hunsicker pointed out that for solution two, adding numbers would confuse patients. Dr. Ajay Israni agreed and suggested testing it with different audiences.

Members voted unanimously in support of adding a measure of uncertainty to the tier assignment. Members then voted for their preferred method of displaying this uncertainty by rank ordering the possible solutions. Members preferred showing the probability of being in each tier for professionals while supporting shading to convey the most likely location of the tier assignment for patients and the general public, but supported presenting the idea to SRTR's PFAS to gather their feedback.

The members then revisited the two earlier questions regarding conveying range and making it more visible on the site. There was no objection to this, or to standardizing the score functions across the metrics. Dr. Parker suggested showing the distribution of tiers and absolute risks in each tier for different score functions to the SRC or its other subcommittees. Once this was generated, it would be revisited by the subcommittee. The final recommendation was for SRTR to continue to explore various options, bringing them to the other subcommittees, and bringing this information back to the AMS subcommittee for future consideration.

Closing business

Dr. Snyder said the next quarterly meeting would be scheduled soon. A call for nominees will be announced for replacing those rolling off AMS at the end of the year. With no other business being heard, the meeting concluded.