February 12, 2008

ASTS Questions & Concerns:
OPTN/UNOS proposed LYFT-based Allocation System

1. LYFT Score, Modeling:
    a. To provide interpretable information for the public, it would be helpful to provide some graphical representation of the overlap of the population by age decile (i.e. what percentage of the 60 year olds would get the same benefit as the population that is 5 years younger?).
    b. The results of the simulations demonstrate that while life years are increased, graft years are not effected. How much of the life years are coming in patients who have returned to dialysis after a failed transplant? What would be the effect of discounting these life years in this population by 0.6 rather than 0.8 as the quality of life after return to dialysis is arguably much worse?
    c. What proportion of the LFYT score is derived from HLA DR typing? That is, all other things being equal, what effect on probability of transplantation does 0, 1, 2 DR matching provide?
    d. We are concerned about the significant bias towards young people because of the significant number of years post-transplant that they will be alive when compared with older patients with competing risks for death. We would argue that life years after transplant should be discounted as is typical in most effectiveness research. Further, the discounting would decrease the uncertainty associated with the tails of the survival distribution and would provide for the ability to test other survival curves than the median survival curve in calculation of LYFT. We think it would inform the community more accurately if you would provide outcome models with standard discounting of life years along with test cases of 50% and 200% of the discount rate. Please discuss the effect discounting will have on outcomes and the ability to use other models. After discounting, please examine models using truncation of survival times at 5, 10, and 15 years.
    e. What was the mean survival time of patients used in the survival models?
    f. Have results of re-transplants, as well as interactions between re-transplantation and other covariates in the LYFT model, been taken into account in the LYFT calculations?
    g. Various contradictory reasons have been given for the exclusion of race from the LYFT calculation. In the past, the committee justified not using ethnicity in the algorithm - saying that there was now little difference in SRTR data. However, recently the rationale was presented that the committee excluded ethnicity as it was too politically charged and the

groups were not well-divided. From SRTR data for live donors and non-ECD deceased donors, there is a difference in outcomes between African Americans and Caucasians. What exactly is driving the exclusion of this predictor from the LYFT score? What is the ethical justification of excluding some variables, race, and including others, diagnosis and age?

h. The variables included in the model include a number where the data was noted as missing. While these variables have significant effects, one wonders about the veracity of the data with inclusion of these variables. Please provide estimates of outcome and uncertainty with and without inclusion of these variables. The list of variables excluded has a number of variables that have significant effect on LYFT, but were excluded for apparently arbitrary reasons, such as "inappropriate for allocation".
i. It is unclear how PRA was used in the model, in particular PRA greater than 80. Please explain how the effect of PRA was estimated.
j. The committee is looking at a model that decreases the benefit of LYFT in order to encourage living donation. Please provide estimates of the decrease in the number of living donor transplantations using estimates from transplant centers with very short waiting times and the pediatric population. Please estimate the outcome with 1.0, 0.8 and 0.6 LYFT.
k. Apparently the LYFT modeling was performed using only SCD kidneys. Please explain what affect this would have on the models as compared to using both SCD and ECD kidneys. What is the effect of having donor factors in both LYFT and DRI?

2. LYFT Score, Statistics:
    a. We understand that LYFT score accurately predicts life years from transplantation. In other words, we understand that for given patient characteristics, the predicted LYFT would be comparable to the average life years gained among historical patients with those characteristics. However, it is likely that some patients with those characteristics had much shorter life years gained than would be predicted by LYFT, and some had much longer life years gained than would be predicted by LYFT. We would like to see this uncertainty about the variability in life years quantified. Please provide a c-statistic and an expression of the residual uncertainty of the outcome and estimates of this uncertainty vs. age. The examination of the variability should also examine the effect of varying the accounting of post transplant years from the median to shorter time intervals such as 25/75, 5 years and 10 years (see below).
    b. Patient survival was censored at the time of re-transplantation. This seems arbitrary. Please explain the advantages and disadvantages of this censoring.
    c. The long term extrapolation of the data necessary appears to be fanciful and would have significant effects on the outcome of LYFT. The use of the latest results to correct the slope of the earlier curves but with the use of the later data to create these slopes would argue against the use of long time frames after transplantation.

d. We are concerned that the LYFT score is based on a dataset that suffers considerably from unmeasured confounding, residual confounding, missing data, and possible misclassification bias. Examples include key missing recipient factors such as cardiovascular disease, and binary comorbidities instead of severity indicators. The appropriateness of risk adjustment based on SRTR data is already a point of controversy in terms of transplant center outcomes monitoring. We are concerned that these same flaws will be perpetuated to something that not only monitors our outcomes but decides which of our patients dies. We understand that LYFT (and the OPTN/SRTR database) is "the best that we can currently do with the data we have." However, we would point out that although the bar for validity and reliability is not as high for research studies, and not even as high for outcomes monitoring, it should be very high when we put our patients' lives on the line.
e. Has an expert panel examined the mathematical and statistical methods used in the development of this allocation policy? If so, was a report made? Were changes in the methods suggested?

3. Donor Profile Index (DPI):
    a. This score comprises a significant portion of the LYFT-based simulations, yet nothing on the SRTR website details its methodology and validation. We would request that simulations from UNOS should be based on either a working paper on the website or a published manuscript. We would hope that the paper describing the DPI would describe the uncertainty of the score. The estimates of uncertainty and those of LFYT should be graphically represented in estimates of the benefit of the system and the age distribution of those receiving transplantation.
    b. We are concerned that the DPI might neglect effect modifiers (interaction terms). In other words, we know clinically that some kidneys will do better in some recipients than in others, and wonder if this kind of donor/recipient matching will be taken into account with the DPI score.

4. Simulations:
    a. Since some patients will be much less likely to receive a transplant (i.e. older patients), it is likely that death on the waiting list will increase. We feel that death on the waiting list should also be a metric taken into consideration when evaluating simulation outcomes. What is the increase in death on the waiting list by age decile from patients left on the list at one and three years as compared to the current system?
    b. After the run of the simulations, what is the resulting distribution of age by DPI by time on waiting list of those patients transplanted? What is the distribution of time on the list and age for patients left on the list?
    c. How would this system affect likelihood and outcomes after re-transplantation?
    d. Decreasing the weighting of LYFT to 80% appears to shift more kidneys to older patients. As the purported benefit of this change was to increase

living donation, has the committee examined delaying the effect of LYFT directly, for example one or two years after listing, rather than decrementing the LYFT weighting?

e. Looking at the older patient, with long waiting times and the highest risk kidneys, what is the benefit of transplantation and the variability around this benefit? Please explain how using LYFT, which looks at survival using SCD kidneys, can be used to examine survival following higher DPI index kidneys.
f. We do not have the outcome of all the simulations available. Given the relatively low bar for the increase in LYFT being considered with the 0.8 LYFT model, have the previous simulations been examined in light of this bar (i.e. has the committee re-reviewed the previous simulations)? Please make available (in a public place such as www.ustransplant.org) the outcome of the previous simulations and the reasons why the committee decided not to use the allocation scheme.
g. Did the simulation examine the effect of policies that would force patients on the waiting list to wait for the best kidney, such as waiting for a better matched kidney or a younger kidney?

5. Logistics:
    a. We are concerned about the cost of HLA typing for continuous PRA. Has this been modeled?
    b. It seems counter-intuitive to decrement the LYFT score for patients with high PRA and then provide a factor to increase transplantation in this population. What are the results of simulation if both factors are removed from the model? What is the sensitivity of the simulations to the factor given to increase transplantation in this population? Looking at the SRTR data on outcome in this population, what is the LYFT tradeoff of providing transplant to this population? What is the justification of provision of priority to this population rather than to another underserved population?
    c. What is the plan for transitioning from the current system to a LYFT-based system? If this is not a smooth transition, we are concerned that patients who have been waiting for many years, but who have low LYFT scores, will be angered and demoralized. One possibility would be to gradually phase in the weight of the LYFT (start with 0.8 waiting time and 0.2 LYFT and then slowly transition to the goal of 0.2 waiting time and 0.8 LYFT).
    d. What will happen to paybacks as this system transitions in, with particular attention to centers that are owed a number of paybacks?
    e. Although the use of dialysis time (instead of waiting time) makes up for nephrologists who fail to refer patients to transplantation in a timely manner (which we acknowledge has been reported many times and amplifies racial, socio-economic, and gender disparities), we are concerned that patients who were offered transplantation but initially refused (and eventually consider it when they get into trouble) will be

treated with the same level of priority (or higher) as those who "did everything the right way from the start."

f. There currently are extreme variations in waiting time until transplantation around the country by DSA. In the LYFT/DPI model, it is likely that these variations in waiting time will translate into variations in age and DPI scores of patients transplanted. Please examine the variation in age at transplantation probability of transplantation by age, and DPI scores across DSA. While the model maybe unstable at the individual DSA level, some estimation of the overall variation can likely be made.

g. Current policy regarding ECD kidneys has a hard cutoff for the DPI that constitutes a higher risk kidney. While we recognize this as an artificial boundary, it does provide for the ability to consent patients up front for the issues regarding outcome for these kidneys. The recent JAMA paper by Merion et al suggests patient characteristics associated with benefit from these kidneys. Given that the OPTN has created a standard of care for kidneys with a DPI of greater than 1.5, how does the committee propose the consent process would work in the future state?