***At-a-Glance***

**Proposal to Implement Pre-Transplant Performance Review by the Membership and Professional Standards Committee**

- **Affected/Proposed Policy:** OPTN Bylaws, Appendix D. 10: *Additional Transplant Program Requirements for Transplant Hospitals and Transplant Programs* and Appendix M. *Definitions*

- **Membership and Professional Standards Committee**

  Currently, transplant program performance monitoring relies almost exclusively on risk-adjusted graft and patient survival rates among recipients. The overemphasis on post-transplant metrics may result in risk-aversion and decreased transplant volumes, and may not be in the best interest of waitlisted patients. Further, post-transplant outcomes may not identify structural problems (e.g., understaffing) that prevent a program from keeping up with the needs of its waitlist population. As such, a more holistic approach to performance monitoring is necessary.

  The purpose of this proposal is to provide the MPSC with a tool, the Composite Pre-transplant Metric (CPM), for identifying kidney and liver programs that may be in need of review based on outlying performance in accepting deceased donor organ offers, transplanting waitlisted patients, and/or mitigating waitlist mortality. The CPM is an aggregate, pre-transplant performance metric that combines programs' acceptance rate, geography-adjusted transplant rate, and waitlist mortality rate observed-to-expected (O/E) ratios into a single number for prioritizing programs for potential review.

- **Affected Groups**
  Transplant Administrators
  Transplant Data Coordinators
  Transplant Physicians/Surgeons
  Transplant Program Directors

- **Number of Potential Candidates Affected**
  All patients registered on either the liver or kidney waitlist could be affected due to increased attention on pre-transplant performance metrics. As of August 8, 2014, there were 15,778 registered liver candidates and 101,056 registered kidney candidates.

- **Compliance with OPTN Strategic Plan and Final Rule**
  This proposal is consistent with the OPTN Final Rule, which stresses the importance of reviewing inter-transplant program variability in waitlist mortality. In addition, the proposal addresses the OPTN key goal of increasing access to transplants.

- **Specific Requests for Comment**
  Should transplant program performance monitoring become more comprehensive by including pre-transplant (i.e. waiting list management) performance in addition to post-transplant outcomes? Is the CPM a reasonable method for creating a more balanced performance assessment and identifying programs that need further inquiry by the MPSC? Readers are encouraged to provide feedback on these particular questions as well as comments on all aspects of the proposal.

**Proposal to Implement Pre-Transplant Performance Review by the Membership and Professional Standards Committee**

**Affected/Proposed Policy: OPTN Bylaws, Appendix D. 10: Additional Transplant Program Requirements for Transplant Hospitals and Transplant Programs and Appendix M. Definitions**

**Membership and Professional Standards Committee**

**Public comment response period: September 29, 2014 – December 5, 2014**

**Summary and Goals of the Proposal:**

Currently, transplant program performance monitoring relies almost exclusively on risk-adjusted graft and patient survival rates among recipients. The overemphasis on post-transplant metrics may result in risk-aversion and decreased transplant volumes,[1,2] and may not be in the best interest of waitlisted patients. Further, post-transplant outcomes may not identify structural problems (e.g., understaffing) that prevent a program from keeping up with the needs of its waitlist population. As such, a more holistic approach to performance monitoring is necessary.

The purpose of this proposal is to provide the MPSC with a tool, the Composite Pre-transplant Metric (CPM), for identifying kidney and liver programs that may be in need of review based on outlying performance in accepting deceased donor organ offers, transplanting waitlisted patients, and/or mitigating waitlist mortality. The CPM is an aggregate, pre-transplant performance metric that combines programs' acceptance rate, geography-adjusted transplant rate, and waitlist mortality rate observed-to-expected (O/E) ratios into a single number for prioritizing programs for potential review.

**Background and Significance of the Proposal:**

Since 1994, the OPTN has reviewed risk-adjusted patient and graft survival outcomes to monitor transplant program performance. The intent of this oversight has been and continues to be to identify opportunities for process improvement that lead to improved patient outcomes. In recent years, members of the MPSC have questioned whether the review of only post-transplant outcomes is broad enough to fully assess whether transplant programs are serving the needs of their patients.

This overly narrow definition of patient outcomes was most evident in several high profile cases of waitlist mismanagement in the early 2000's. In one case, a transplant program did not have a full-time surgeon on-site and was, in turn, unable to keep up with the needs of its waitlisted patients. Deceased donor transplant offers were frequently turned down, transplant volumes decreased, and waitlisted patients were dying at a higher than expected rate. In another example, a newly established transplant program was insufficiently staffed to handle the immediate influx of thousands of patients, substantially affecting patients' access to transplantation. In both of these cases, graft and patient survival rates among recipients were not extraordinary and thus were insufficient for uncovering these systemic cases of waitlist mismanagement. The MPSC's "functional inactivity" thresholds, which trigger a program for review if they have performed zero

[1] Schold JD, Buccini LD, Srinivas TR, et al. The association of center performance evaluations and kidney transplant volume in the United States. Am J Transplant 2013;13:67-75.

[2] Cameron, Andrew M., and Brigitte E. Sullivan. "Regulatory Oversight in Transplantation: There and Back Again." *JAMA surgery* 148.11 (2013): 997-998.

transplants over a specified time frame (e.g., 3 consecutive months for liver, heart, and kidney programs) also did not identify these cases.

In April 2008, the United States General Accountability Office (GAO) issued a report[3] that highlighted these cases and the need to develop and implement "a set of activity-level indicators to detect problems that prolong the time patients wait for transplants." The report emphasized the utility of waitlist activity measures such as transplant rates and organ offer acceptance rates. These "pre-transplant" metrics – risk-adjusted (i.e., "case-mix" adjusted) acceptance rates, transplant rates, as well as waitlist mortality rates – are produced by the SRTR contractor. HRSA charged the OPTN to find ways to use these metrics and perhaps other measures of waitlist activity to expand the suite of performance metrics used to oversee transplant programs.

In 2006, the OPTN's Joint Board-MPSC Process Improvement Working Group began evaluating the usefulness of organ offer acceptance rates and other measures of pre-transplant activity. They observed that while risk-adjusted acceptance rates and transplant rates were correlated (programs with high acceptance rates tended to have high transplant rates), they were not *so* highly correlated as to make either of the two redundant and irrelevant in light of the other. The group concluded that both metrics add value and could be useful for monitoring transplant programs' pre-transplant activity. They also concluded that acceptance rates, though potentially a very powerful metric for identifying programs with waitlist management problems, should be used but *not* as a stand-alone metric.

**The Development of a New Metric:**

With this mandate to develop an approach for monitoring pre-transplant performance, coupled with the GAO report and the Joint Board-MPSC working group recommendations to use acceptance rates but not as a stand-alone metric, the OPTN contractor developed the Composite Pre-transplant Metric (CPM) for MPSC to consider as a potential approach for identifying programs in need of review. The CPM is a weighted average of the following case-mix adjusted, pre-transplant observed-to-expected (O/E) ratios produced by the SRTR contractor: waitlist mortality rates (liver only), *geography-adjusted* transplant rates, and organ offer acceptance rates. To account for widely varying sample sizes across institutions, the O/E ratios are first attenuated ("shrunken" closer to 1.0) depending on the strength of evidence underlying each program's ratios, using an approximation to the Empirical Bayes[4] method.

The CPM can be interpreted as an "aggregate, pre-transplant O/E ratio," with an average value of around 1.0. Unusually high CPM values - typically associated with high mortality, low transplant, and low acceptance rates - are generally around 1.5 to 2.5, or even higher. Low outlier values, which generally reflect increased waitlist activity and lower mortality rates, tend to be between about 0.50 and 0.75. The CPM is intended to identify programs that *may* have a need for improvement in waitlist management; it is not a definitive indication that a problem actually exists.

**Figure 1** shows that the CPM distribution for liver programs is centered around 1.0, with a few programs having values as low as about 0.5, and others having values near or above 2.0. CPM is intended to identify only a small number of programs with highly aberrant pre-transplant performance metrics for further review. In the July 2012 cohort, just 6 (5%) of 130 liver programs had CPM exceeding 1.5, the proposed threshold for triggering MPSC review.

[3] GAO Report on Organ Transplant Programs to the Ranking Member, Committee on Finance, U.S. Senate, April 2008.
[4] Efron, Bradley, and Carl N. Morris. *Stein's paradox in statistics*. WH Freeman, 1977.

The distribution of CPM among liver programs tends to look like this:

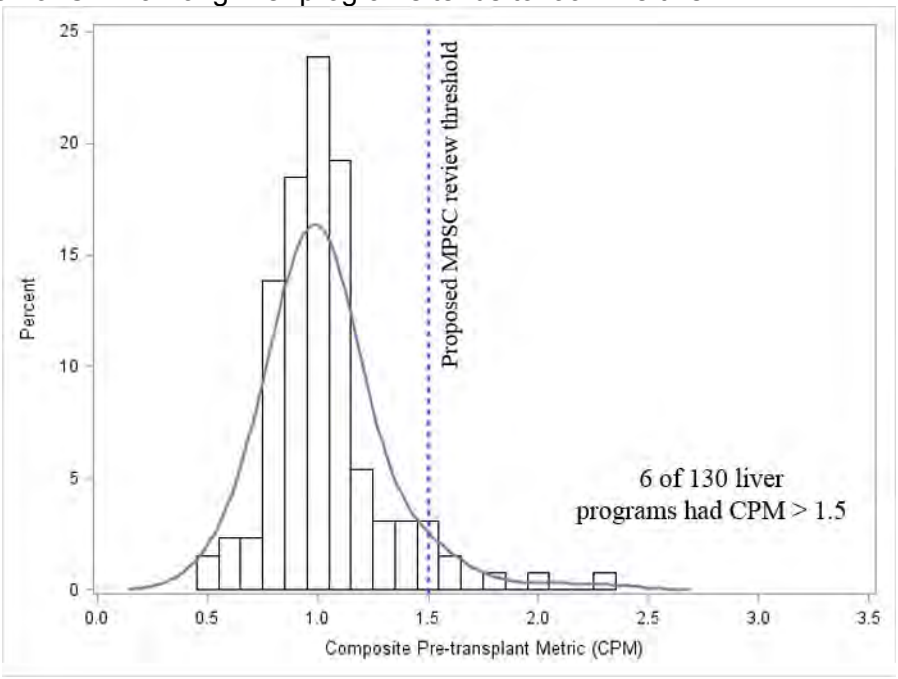


**Figure 1.** CPM distribution for liver programs, July 2012 PSR cohort (calendar year 2011 data).

The distribution of CPM among kidney programs tends to look like this:

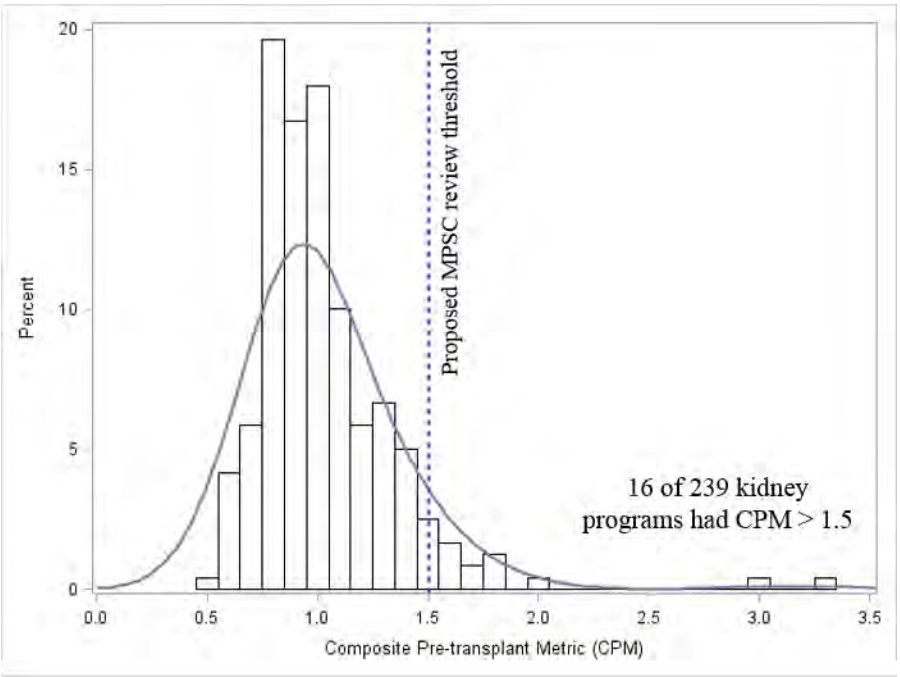


**Figure 2.** CPM distribution for kidney programs, July 2012 PSR cohort (calendar yr. 2011 data).

**Figure 2** reveals a similar CPM distribution for kidney programs compared to liver programs (Figure 1), except for the presence of two programs with extreme values near or above 3.0. In

the July 2012 cohort, just 16 (7%) of 239 kidney programs had CPM exceeding 1.5, the proposed threshold for triggering MPSC review.

**Profiles of Programs with Unusually High CPM Values**

The liver program with the highest CPM of 2.29 was accepting deceased donor liver offers at a rate only 10% of expected, based on national data and adjusting for donor characteristics (e.g., age, DCD) as well as candidate characteristics (e.g., age, MELD score). In other words, for every ten similar offers accepted by an average program, this program accepted just one. In turn, this program was only transplanting patients at a rate 62% of expected. Also, waitlist mortality was 67% higher than expected, although this rate was not statistically different from expected.

*Statistical profiles of liver programs*

Pre-transplant metrics for liver program with CPM of 2.29.

- Acceptance rate: O/E = 0.10 (p<0.01) 74 fewer accepted offers than expected.
- Transplant rate O/E = 0.62 (p<0.01) 23 fewer transplants than expected.
- WL mortality rate O/E = 1.67 (p=0.37) 2 more deaths than expected.

Pre-transplant metrics for liver program with the second highest CPM of 2.03:

- Acceptance rate O/E = 0.25 (p<0.01) 44 fewer accepted offers than expected.
- Transplant rate O/E = 0.27 (p<0.01) 41 fewer transplants than expected.
- WL mortality rate O/E = 1.26 (p=0.21) 7 more deaths than expected.

*Statistical profiles of kidney programs*

Pre-transplant metrics for kidney program with CPM of 3.26.

- Acceptance rate: no offers received
- Transplant rate O/E = 0.06 (p<0.01) 31 fewer transplants than expected. *(Performed two living donor transplants.)*
- WL mortality rate O/E = 1.89 (p=0.03) 7 more deaths than expected.

Pre-transplant metrics for kidney program with the second highest CPM of 2.97:

- Acceptance rates: organ-based O/E = 0.00 (p<0.60), offer-based O/E=0.00 (p<0.15) *(Received just two offers.)*
- Transplant rate O/E = 0.00 (p<0.01) 32 fewer transplants than expected.
- WL mortality rate O/E = 1.55 (p=0.09) 9 more deaths than expected.

Pre-transplant metrics for kidney program with the third highest CPM of 2.04:

- Acceptance rates: organ-based[5] O/E = 0.38 (p<0.01), offer-based O/E=0.31 (p<0.01).
- Transplant rate O/E = 0.54 (p<0.01) 35 fewer transplants than expected.
- WL mortality rate O/E = 1.26 (p=0.17) 6 more deaths than expected.

[5] CPM is now using an offer-based acceptance rate model, but was originally computed using both organ and offer-based acceptance rate models, as in Wolfe RA, LaPorte FB, Rodgers AM, Roys EC, Fant G, Leichtman AB. Developing organ offer and acceptance measures: when 'good' organs are turned down. Am J Transplant 2007;7:1404-11.

Several of the programs identified by CPM had already involuntarily inactivated or withdrawn from the OPTN during this period.[6]

### A New, More Balanced Approach to Performance Monitoring

Monitoring graft and patient survival is vital to ensuring that transplant recipients continue to have good outcomes and that donated organs are used effectively. However, given the substantial net-benefit for most patients of organ transplantation compared to waiting on organ-replacement therapy (if applicable), transplant centers with excellent post-transplant outcomes may not be adequately serving their waitlisted patients if few patients are actually getting transplanted. The use of pre-transplant metrics in conjunction with post-transplant graft and patient survival metrics for performance monitoring (Figure 3) may ultimately be in the best interest of end-stage organ failure candidates on the waitlist.[7]

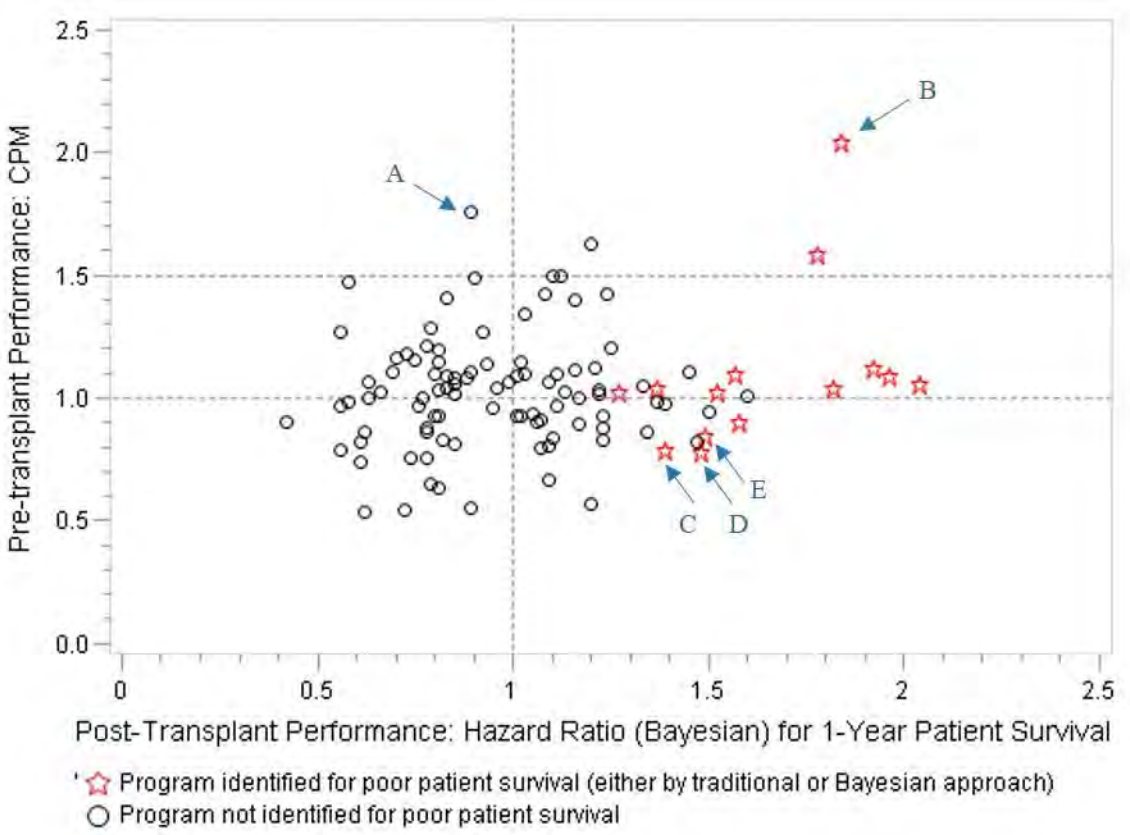


**Figure 3. CPM vs. 1-Year Patient Survival Hazard Ratio (Bayesian), for Adult* Liver Programs (n=108). Reference lines indicate the proposed MPSC review threshold of CPM=1.5, as well as the "average" or expected value of 1.0 for both CPM and patient survival hazard ratios. Results based on July 2012 SRTR PSR cohort: pre-transplant metrics derived on calendar year 2011 data; post-transplant O/E derived from recipients**

[6] Pre-transplant metrics for programs that closed during the evaluation period may be even more outlying due to abrupt inactivity associated with program closure and a residual waitlist that was not immediately transferred to other program(s).

[7] Axelrod, D. A. "Balancing accountable care with risk aversion: Transplantation as a model." *American Journal of Transplantation* 13.1 (2013): 7-8.

**transplanted between Jan, 2009 – Jun 30, 2011.** ***(* Programs having more than 50% pediatric patients on their waitlist during 2011 were excluded from this analysis.)***

While patient survival rates were better than expected (Bayesian HR=0.89) for Program A, this program had a very high CPM of 1.75, suggesting a potential need for review with respect to pre-transplant performance (Figure 3). In fact, this program was accepting liver offers at a rate only 58% of expected ($p<0.01$), was transplanting patients at a rate 63% of expected ($p<0.01$), and had a mortality rate 2.1 times greater than expected ($p<0.01$). All considering, this program may have room for improvement in the area of waitlist management/transplant activity in order to more effectively serve the patients on its waitlist. Of course, as further explained in the compliance monitoring section of this document, the CPM merely provides a trigger for further review; in and of itself, this metric does not provide a definitive indication that a systemic issue exists that requires improvement/corrective action.

Transplant program B (Figure 3) may have been in need of process improvements in both pre and post-transplant patient care. To go along with a patient survival hazard ratio of 1.84, this program was only accepting offers at a rate 25% of expected ($p<0.01$) and transplanting patients at a rate 27% of expected ($p<0.01$). Program B's waitlist mortality rate was also 1.26, or 26% higher than expected; however, this difference was not statistically significant due to a relatively small number of deaths.

Transplant programs identified for exceptionally poor graft or patient survival rates should, of course, assess whether process improvements are needed, irrespective of how quickly they are transplanting patients on their list. Review of transplant program processes to identify meaningful process improvement areas that improve patient outcomes is, after all, the overriding purpose of MPSC's review of survival rate data. However, some centers identified for review based on moderately poor graft or patient survival rates may, upon closer review, have no obvious need for process improvement. And some of these programs may have excellent pre-transplant metrics, in terms of transplanting patients on their waitlist and mitigating waitlist mortality. The establishment of a pre-transplant metric will provide the MPSC additional information regarding the program's service to its patients to consider when reviewing post-transplant outcomes in addition to its identifying programs that may need improvement in waitlist management.

For example, liver programs C, D, and E (highlighted in Figure 3) have moderately lower than expected patient survival rates (i.e., higher than average Bayesian hazard ratios, between 1.39 and 1.49) and would have been identified for review based on either the traditional identification method, the new Bayesian method, or both methods. However, in aggregate these programs may be serving their waitlist population quite well, given their exceptionally low CPM values. Statistical profiles of these programs reveal that each was accepting liver offers at a rate higher than expected, was transplanting patients at a rate more than 80% above expected, and had waitlist mortality rates lower than (and not statistically different from) expected.

- Liver program C
  - Post-transplant (Bayesian hazard ratios, 1-year survival): patient HR=1.39, graft HR=1.39
  - Pre-transplant: acceptance O/E=1.10 ($p=0.41$), transplant rate O/E=1.88 ($p<0.01$), waitlist mortality rate O/E=0.83 ($p=0.46$)

- Liver program D
  - Post-transplant (Bayesian hazard ratios, 1-year survival): patient HR=1.48, graft HR=1.37

- Pre-transplant: acceptance O/E=1.33 (p=0.03), transplant rate O/E=1.83 (p<0.01), waitlist mortality rate O/E=0.88 (p=0.62)

- Liver program E
  - Post-transplant (Bayesian hazard ratios, 1-year survival): patient HR=1.49, graft HR=1.38
  - Pre-transplant: acceptance O/E=1.06 (p=0.77), transplant rate O/E=1.92 (p<0.01), waitlist mortality rate O/E=0.97 (p=0.99)

Incorporating the CPM into the Bylaws will allow the MPSC to more formally take into account pre-transplant information when reviewing programs already identified based on post-transplant metrics, to better assess – from a perspective broader than just recipient outcomes – whether each program is effectively serving its patient population. In this way, the CPM has the potential to reduce the emphasis on post-transplant metrics, in particular in cases of borderline-high graft or patient survival hazard ratios, when such programs are reviewed by the MPSC. The CPM will provide the MPSC a tool to evaluate transplant program performance more holistically, including both post-transplant and pre-transplant outcomes.

**Calculating the CPM**

The SRTR's program-specific waitlist mortality rates, geography-adjusted transplant rates, and offer acceptance rates – adjusted for case-mix and in the form of observed-to-expected (O/E) ratios – are combined into a single composite indicator of pre-transplant performance, the CPM. To account for the statistical uncertainty in O/E's due to varying sample sizes among programs, an approximation to the empirical Bayes estimation method is used to "shrink" each O/E ratio toward the neutral value of 1.0 before combining them. The CPM is a weighted average of these "shrunken" O/E's, with weights determined by the CPM Work Group of the MPSC.

The CPM is calculated in 5 steps.

Step 1. Apply logarithmic transformation of O/E ratios (symmetry)
Step 2. Apply negative sign for transplant and acceptance rates (directional consistency)
Step 3. Account for statistical uncertainty due to finite sample sizes (Empirical Bayes method)
Step 4. Combine into a single metric by applying component weights (composite approach)
Step 5. Apply antilog function (return to familiar O/E scale)

**Table 1: CPM Component Weights**

| Program Type | Mortality Rate | Transplant Rate | Acceptance Rates |
|---|---|---|---|
| Liver | 0.50 | 0.25 | 0.25 |
| Kidney | 0.00 | 0.50 | 0.50 |

An example CPM calculation as well as additional information about the Empirical Bayes method is provided in the appendix.

**The CPM "Safety Net"**

One benefit of using a composite metric approach for performance evaluation is that the CPM may identify a potential waitlist management issue at a program with *borderline-low* performance in all three metrics, when no *single-metric* threshold would have triggered a review. However, a risk associated with using a composite approach is that extremely poor performance in one particular metric – which may in and of itself be cause for concern – may be offset by good or average performance in the other metric(s). Though it is unlikely that either transplant or acceptance rates could be extremely low while the other was high due to their high correlation, their correlations with waitlist mortality rates are much lower. It is possible for a program to have an extremely high (and statistically higher than average) waitlist mortality rate that is offset by good or average transplant and/or acceptance rates. Though the CPM Work Group agreed that identification of programs based on pre-transplant performance should be primarily driven by the CPM, programs with an extremely high waitlist mortality rate should not be ignored, even if the CPM does not reach the 1.5 threshold. Improvement in patient care or a reevaluation of listing practices may be needed, despite transplant and acceptance rates that conform to national expectations.

Consequently, in addition to program identification using the CPM > 1.5 trigger, the following "safety net" is also part of this proposal.

CPM Safety Net: Waitlist mortality O/E > 2.0 and p-value (one-sided) < 0.05

As shown in Figure 4, only rarely do mortality rates exceed twice the expected value in programs with more than a few patients on their waitlist.

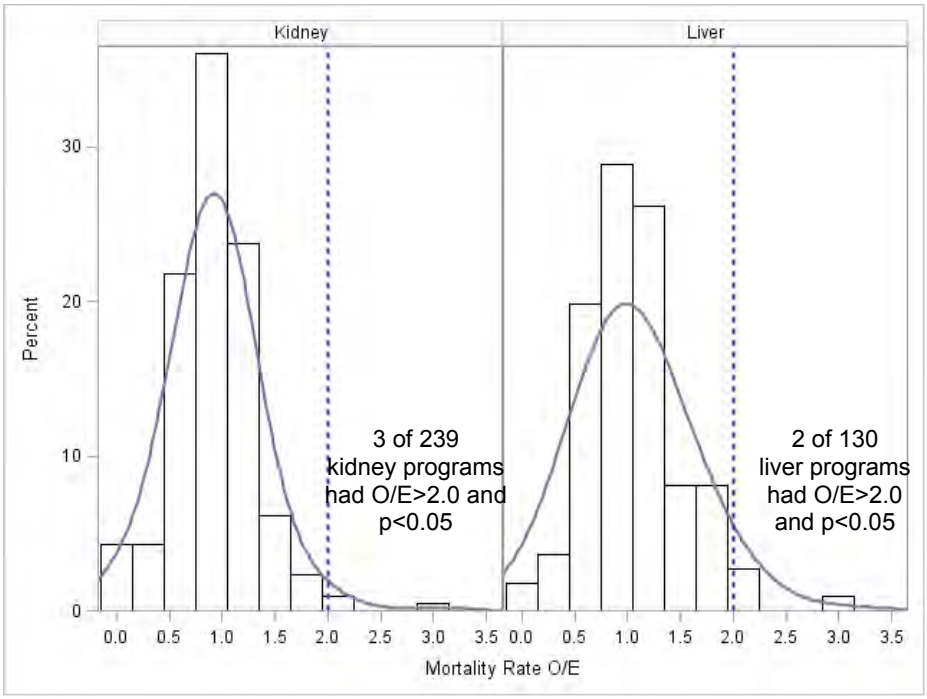


**Figure 4. Waitlist Mortality Rate O/E distribution for kidney and liver programs. Programs with less than 20 person-years on the waitlist in 2011 were excluded to avoid including**

**outlier O/E ratios that are driven by small sample sizes and are generally not statistically different from 1.0. Results are based on July 2012 SRTR PSR cohort (calendar year 2011).**

The safety net threshold of 2.0 was chosen because it is extreme and represents a 100% increase in waitlist deaths relative to expectations; the intent is to identify very few programs. Even though the waitlist mortality rate was excluded from the CPM for kidney programs due to the reasons explained below, the CPM Work Group concluded that in those rare instances when more than twice the expected number of deaths occurred, such programs should be reviewed.

**The Anticipated Number of Programs to be Identified by the CPM and Safety Net**

Based on the July 2012 PSR cohort, the following number of programs would have been identified with the CPM methodology:

19 of 239 kidney programs (7.9%):
- 16 with CPM > 1.5
- 3 with mortality O/E>2.0 and p-value (one-sided) <0.05 ("safety net")

8 of 130 liver programs (6.1%):
- 6 with CPM > 1.5
- 2 with mortality O/E>2.0 and p-value (one-sided) <0.05 ("safety net")

Overall, 27 of 369 programs (7.3%) would have been identified for pre-transplant performance review. However, 13 of these programs were already under review by the MPSC for poor post-transplant outcomes or functional inactivity, or had previously inactivated or withdrawn.

Thus, a total of 14 programs (approximately 4% of all kidney and liver programs) would have been *newly identified* for MPSC review by the CPM methodology based on the July 2012 cohort.

Though this number will vary from one review cycle to the next, previous analyses have shown the number of programs exceeding the CPM and mortality rate thresholds has not changed greatly in recent years.

**The Underlying Risk-Adjusted Models behind the CPM**

*Waitlist mortality rates*

Waitlist mortality rates measure the number of deaths among waitlisted candidates at a program relative to the number of patient-years after listing during a specific one year cohort period (e.g., July 1, 2012 to June 30, 2013). Patient deaths are identified as waitlist removals for reason of patient death, as well as by supplementary data sources including the Social Security Death Master file and CMS data. Some patient deaths *after* removal from the waitlist, provided they occurred within the specific one-year cohort period, are counted. Deaths after removal for transplant are not counted. Deaths after removal for patient recovery or transfer to another center are also not counted, unless the death occurred within 60 days of removal. Since some deaths *after removal* are counted, this metric is more accurately described as "mortality after listing" as opposed to "waitlist mortality."

There are two mortality rate models, one for kidney and one for liver candidates. These rates are adjusted for candidate factors such as age, gender, blood type, diagnosis, and lab MELD score (liver), which are associated with the likelihood of candidate mortality. By adjusting for these

factors, centers with a disproportionate number of patients having a higher likelihood of death; for example, programs with older liver patients with high MELD scores, are not disadvantaged by the metric. Rather, each program's observed number of patient deaths during the one year period of time is evaluated relative to the expected number of deaths, which is based on the number of person-years and case-mix of patients on their waitlist.

The models, in particular for liver patients, have strong predictive power with respect to waitlist mortality. This predictive ability is measured by the c-statistic, which ranges from 0.50 (no ability to discriminate) to 1.0 (perfect ability). The c-statistics for these models are 0.66 and 0.87 for kidney and liver candidates, respectively.

The following factors are currently included in the waitlist mortality models:

| Liver | Kidney |
|---|---|
| candidate age | candidate age |
| candidate blood type | candidate blood type |
| candidate diagnosis | candidate diagnosis |
| candidate status (laboratory MELD) | candidate gender |
| candidate gender | candidate race/ethnicity |
| candidate race/ethnicity | candidate waiting time |
| candidate waiting time | |

Additional documentation for these models can be found on the SRTR's website (http://www.srtr.org/csr/current/Tech_notes.aspx). These models may be periodically updated based on more recent data. These updates may result in changes to the factors included in the models as well as the model coefficients.

*Geography-adjusted transplant rates*

Transplant rates measure a program's frequency of transplanting patients using either living or deceased donor organs, relative to the number of patient-years on their waitlist. There are two transplant rate models, one for kidney and one for liver candidates. These rates are adjusted for candidate factors such as age, blood type, CPRA (kidney), and MELD score (liver) that are associated with the likelihood of transplantation. By adjusting for these factors, centers with a disproportionate number of patients having a lower likelihood of receiving a transplant; for example, programs with high CPRA blood group B kidney candidates, are not disadvantaged by the metric. Rather, each program's observed number of transplants during the one year period of time is evaluated relative to the expected number of transplants, given the number of person-years and case-mix of patients on their waitlist.

The models have good or excellent predictive power in distinguishing candidates that are likely to be transplanted from those that are not. This predictive ability is measured by the c-statistic, which ranges from 0.50 (no ability to discriminate) to 1.0 (perfect ability). The c-statistics for these models are 0.63 and 0.84 for kidney and liver candidates, respectively.

An important addition to the transplant rate models (based on member feedback) was the *DSA supply-to-demand* component. The supply-to-demand ratio in a DSA is calculated as the number of deceased liver (or kidney) donors recovered during the year divided by the number of waitlisted liver (or kidney) candidates in the DSA at the start of the period. This risk-adjustment factor was

added to address concerns about transplant rates not being a fair or reliable way to measure the performance of programs in DSAs with relatively few viable organ donors relative to the number of candidates. Without this adjustment factor, transplant programs in geographic areas with relatively fewer donors tended to have lower transplant rates and higher CPMs. However, after incorporating the supply-to-demand adjustment factor into the models, this bias was mitigated (Additional information available upon request).

The following factors are currently included in the geography-adjusted transplant rate models:

| Liver | Kidney |
|---|---|
| candidate age | candidate age |
| candidate blood type | candidate blood type |
| candidate previous transplant | candidate previous transplant |
| candidate status (match MELD) | candidate CPRA |
| candidate waiting time | candidate CPRA x previous transplant (interaction) |
| DSA supply-to-demand ratio | candidate waiting time |
| | DSA supply-to-demand ratio |

(Because the addition of the supply-to-demand ratio as an adjustment factor was solely for the CPM, the *geography-adjusted* transplant rate models differ slightly from the transplant rate models published on www.SRTR.org. Additional information available upon request) These models may be periodically updated based on more recent data. These updates may result in changes to the factors included in the models as well as the model coefficients.

*Offer acceptance rates*

The offer acceptance rate models predict the likelihood of a deceased donor kidney (or liver) offer being accepted, based on characteristics of both the donor as well as the candidate to which the offer is being made. There are two acceptance rate models, one for kidneys and one for livers. These models are used to determine the number of expected acceptances for each transplant program, for comparison with their observed number of acceptances during a one year time period. In this way, an acceptance rate O/E ratio is computed for each kidney and liver program.

Only organs that were ultimately accepted and transplanted are included. Candidates that were bypassed by the OPO and thus did not actually receive an offer are also excluded. Offers for candidates that could not have accepted due to already having been transplanted, a positive crossmatch, or requiring a multi-organ transplant where the other organ was not available, are excluded as well. Offers received after (higher allocation sequence number) the candidate that ultimately accepted the offer are also excluded.

An acceptance rate model's ability to distinguish offers that are likely to be accepted from those that are not (i.e., predictive power) is measured by the c-statistic, which ranges from 0.50 (no ability to discriminate) to 1.0 (perfect ability). The c-statistic for the liver acceptance rate model is 0.91. The kidney acceptance rate model, which is being redeveloped, has a c-statistic of 0.70.

The models adjust for factors that affect the likelihood of organ acceptance, based on national offer acceptance and refusal data. For example, donor factors such as age, cause of death, hypertensive history, and serological status can affect the quality, expected longevity, and desirability of the organ and are included in the risk-adjustment models, along with other donor

factors. Because these factors are included, a program's expected number of acceptances among kidney offers from hypertensive age 60+ donors, for example, will be lower than if the same number of offers were received from non-hypertensive kidney donors age 18-39. In this way, programs that receive more "marginal" kidney offers will not be disadvantaged, as their observed number of acceptances will be compared against the national expected number of acceptances for the same types of organ offers.

Similarly, candidate factors associated with likelihood of offer acceptance are included in the models. For example, candidate age, CPRA (kidney), and MELD score (liver) have been shown to relate to the odds of an offer being accepted. Programs with a disproportionate number of candidates having clinical and demographic characteristics associated with increased offer selectivity – for example, unsensitized pediatric kidney patients – will also not be disadvantaged, since their observed number of acceptances will be compared against the national expected number for the same types of organs and candidates.

The following factors are currently included in the offer acceptance rate models:

Liver

donor age
donor blood type
donor DCD
donor history of cancer
donor BUN
donor circumstances of death
donor HTLV
donor administered insulin
donor administered antihypertensives
donor EBV (nuclear antigen)
donor liver biopsy performed
donor liver biopsy (% macro vesicular fat)
donor PHS increased risk
donor protein in urine
donor SGPT/ALT
donor number of transfusions
candidate laboratory MELD score
candidate match MELD and status (1A, 1B)
candidate difference between match & lab MELD
candidate serum sodium
candidate albumin
candidate dialysis in prior week
candidate received HCC exception points
candidate previous malignancy
candidate height
candidate Willing to Accept ABO Incompatible
candidate Willing to Accept HBV Core antibody Positive

Kidney

donor age
donor cause of death
donor gender
donor race/ethnicity
donor height
donor blood type
donor serum creatinine
donor hypertension
donor hepatitis (B or C) status
donor location (local, regional, national)
candidate age
candidate gender
candidate race/ethnicity
candidate height
candidate diagnosis
candidate CPRA
HLA mismatch (A-locus) offer
HLA mismatch (B-locus) offer
HLA mismatch (DR-locus) offer
adult donor / pediatric recipient offer
size of program's waitlist

| Liver | Kidney |
|---|---|
| candidate Willing to Accept HCV Antibody Positive | |
| candidate Willing to Accept Liver Segment | |
| candidate max distance willing to accept | |
| candidate minimum age willing to accept | |
| candidate time on list | |
| candidate-donor gender match | |
| candidate-donor ABO compatibility | |
| offer sequence number | |
| donor location (local, regional, national) | |
| estimated travel time | |

These models are currently being redeveloped and or refined and may continue to be improved periodically based on more recent data. These updates may result in changes to the factors included in the models as well as the model coefficients.

**The Evolution of the CPM Methodology:**

The CPM concept was first presented to the MPSC in July of 2009. Due to positive feedback, the CPM Working Group was formed to further explore the utility of this metric. The CPM Work Group first met (by phone) in October 2009 and has had 11 subsequent meetings, including an in-person meeting in Chicago in April 2010.

The CPM concept was presented at the American Transplant Congress in 2011 as well as the Transplant Management Forum in both 2011 and 2012 (Additional information available upon request). In addition, in December 2011 a 52-question survey focusing on pre-transplant processes and requesting feedback on the use of pre-transplant metrics for performance monitoring was sent to 47 kidney and 30 liver programs. In 2012, CPM and other pre-transplant metrics were discussed at the 2012 PSR consensus conference.[8]

Based on feedback from these various venues, CPM work group deliberations, and discussions with HRSA and the current and previous SRTR contractor, the following key decisions were made as the methodology was developed and refined over the past six years.

*Rationale for developing a "composite" metric*

A composite metric approach was pursued for the following reasons:

1. Incorporate acceptance rates but temper their impact

   A previous joint Board-MPSC work group and the GAO emphasized the importance of using acceptance rates for program monitoring, but the work group recommended they not be used as a stand-alone measure for identifying programs to review.

[8] Kasiske, B. L., et al. "Report of a consensus conference on transplant program quality and surveillance." *American Journal of Transplantation* 12.8 (2012): 1988-1996.

2. Identify programs in need of process improvement that would not be identified using any single metric alone

   One benefit of using a composite metric approach for performance evaluation is that the CPM may identify a potential waitlist management issue at a program with borderline-low performance in all three metrics, in cases where no single-metric threshold would have been breached. The use of multiple variables together, as in a multivariable regression model, is a common way to increase predictive power.

3. Provide a convenient summary statistic to help prioritize MPSC resources.

   The CPM approach combines three dimensions of pre-transplant information, plus the statistical uncertainty associated with each, into a single value. In this way, the CPM provides a high-level assessment of "aggregate" pre-transplant performance that can be used to identify programs for review. If it is found that too many (or few) programs are being identified, the threshold can be raised (or lowered) from the initial proposed threshold of 1.5.

4. No single metric perfectly reflects the true state of a program with respect to pre-transplant performance.

   Each pre-transplant metric has strengths and weaknesses in terms of its ability to reliably characterize the pre-transplant performance of a transplant program. All rely on accurate reporting of data and are subject to limitations of our ability to adequately adjust for case-mix and other mitigating factors. And some metrics may be more vulnerable to potential manipulation than others. For these reasons, the CPM Work Group felt more comfortable relying on a composite metric as the primary pre-transplant trigger, as opposed to putting all their "eggs in one basket" for identifying programs to review.

5. Mitigate the effect of geography (local organ supply relative to demand)

   Since transplant rates are highly influenced by geography, use of a composite metric would help mitigate the impact of local supply-to-demand dynamics outside the control of transplant programs, since it includes other measures less influenced by geography. This particular rationale for using a composite metric approach became less relevant after the CPM Work Group recommended that the risk-adjusted transplant rates used in the CPM be explicitly adjusted for the supply-to-demand ratio of each DSA.

*Rationale for the chosen CPM component weights*

The initial weights proposed for the CPM were the same for both kidney and liver programs: 50% for waitlist mortality rates, 25% for transplant rates, and 25% for acceptance rates. An analysis performed for the committee showed that this choice of weights differed very little – in terms of the programs identified as having outlying CPM values – from use of a simple average (33%, 33%, 33%).

At its in-person meeting in April, 2010, the CPM Work Group discussed the inclusion of waitlist mortality rates as a component of the CPM. It was concluded that this factor should be removed from the CPM for kidney programs, for the following two reasons:

- There are insufficient data to adjust for cardiovascular risk factor(s) of kidney candidates

- Waitlist mortality is less under the control of transplant programs, since waitlisted patients are often cared for by nephrologists or primary care physicians

The CPM component weights for kidney programs were subsequently modified to 0%, 50%, 50% for waitlist mortality, transplant, and acceptance rates, respectively.

Only a very small number of transplant programs are known to have had severe, structural problems (e.g., gross staffing shortage) in managing their waitlist to such an extent that caused patients to be at risk. One reason for this may be simply that such egregious cases are very rare; however, another contributing factor may be that programs have not yet been routinely reviewed based on pre-transplant performance, which highlights the impetus for this proposal. With such a small sample size, in terms of the number of historically known problem programs, use of mathematical optimization (e.g., regression analysis) to determine the appropriate weights was not possible. Consequently, the weights were judgmentally derived and agreed upon by the CPM Work Group and MPSC. The use of expert opinion in developing composite metrics is not without precedent.[9]

These weights were *not* chosen by analyzing data for the two high profile cases of waitlist mismanagement in the mid-2000's. The CPM Work Group gained confidence in this new metric when it was demonstrated that both programs would have stood out as outliers with respect to CPM during their crisis periods, had the metric been available at the time. This "validation" of the CPM is described further in the *Additional Evidence Supporting this Proposal* section of this document.

Finally, though the waitlist mortality rate is given the highest weight (50%), it is important to recognize that transplant and acceptance rates both actually have a much larger influence on the liver program CPM. Counterintuitively, despite the smaller component weights of 25%, these factors contribute more to the program-to-program variability in CPM. This is because there is far more program-to-program variability in risk-adjusted transplant and acceptance rates compared to risk-adjusted mortality rates, where differences tend to be smaller and/or statistically insignificant.

***Decision to adjust the transplant rate for geography inequities (supply-to-demand)***

Transplant rates can be highly influenced by a transplant program's geographic location. Due to substantial differences in the available supply of donor organs relative to demand (size of the local waitlist) across DSA's, transplant rates vary significantly by geography. Some transplant programs responded to a CPM/pre-transplant process metrics survey with concerns that the use of transplant rates wouldn't be fair in light of geographic differences in access to organs, in particular for geographically isolated programs in areas of high waitlist demand (Additional information available upon request).

[9] Saisana, Michaela, and Stefano Tarantola. *State-of-the-art report on current methodologies and practices for composite indicator development*. European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, Technological and Economic Risk Management Unit, 2002.

In response, the CPM Work Group agreed that a supply-to-demand adjustment be added to the transplant rate model, mitigating the bias of transplant rates with respect to the local supply-to-demand ratio (Additional information available upon request). The Work Group believes that this important revision to the transplant rate model is a significant step in responding to member feedback and creating a fairer metric that is better suited for evaluating transplant program performance.

The Work Group also discussed whether "supply" should include all deceased (kidney or liver) donors recovered in the DSA or if ECD, DCD, and/or high KDPI donors should be excluded. After data review and deliberation, the group determined that *all* recovered (kidney or liver) donors should be used to define DSA supply-to-demand ratios (Additional information available upon request).

The "demand" is defined as the total number of waitlisted patients on the liver (or kidney) waitlist in the DSA. Both active and inactive patients are included.

***Decision to include both deceased and living donor transplants in transplant rate model***

The SRTR contractor has been producing two types of risk-adjusted transplant rates: one that includes only deceased donor transplants, and the other that includes both deceased and living donor transplants. For kidney transplant programs, as well as a few liver programs, the difference in results (O/E) can be significant depending on which approach is used.

The CPM Work Group debated this issue extensively and reviewed data analyses (Additional information available upon request). They reached a consensus and agreed that transplant rates used in the CPM should include both living and deceased donor transplants, for the following reasons:

1. Philosophically, programs should be evaluated on whether they are effectively serving their waitlisted patients using whatever sources of transplantable organs are available, whether from living or deceased donors.

2. Kidney programs that perform a high percentage of living donor transplants tend to be more selective in accepting deceased donor offers. Excluding living donor transplants would have a disproportionate and unfair effect on such programs.

3. It is impossible from currently available data to distinguish these four categories of waitlisted patients:

    a. Waiting strictly for a deceased donor offer
    b. Waiting primarily for a deceased donor offer; may pursue living donation
    c. Primary intent is living donation (e.g., KPD); considering deceased donor offers
    d. Sole intent is to receive a living donor transplant

While living donor transplants can be excluded from the numerator of the transplant rate calculation, without the ability to identify these groups, candidates waiting solely or primarily for a living donor transplant cannot be excluded from the denominator (patient-

years). Consequently, it is not currently possible to develop a pure, "deceased donor activity only" transplant rate model.

One weakness of using the "all donors" version of the transplant rate is that it might unfairly and artificially inflate transplant rates for programs who add to the waitlist candidates that are solely intent on pursuing a living donor transplant, with no intention of accepting a deceased donor organ offer. And, as of September 1, 2014, all candidates for transplantation – even those that are only interested in a living donor transplant – are now required to be added to the waitlist. However, programs are now able to indicate in UNet$^{sm}$ that patients are being added strictly for living donation, offering the possibility of a more refined transplant rate calculation in the future.

***Decision to include inactive patient-years in the transplant and mortality rate denominators***

Both active and inactive patients are included in the denominators ("patient-years") of the transplant and mortality rate models. These are included because of philosophical ("intent-to-treat") as well as practical considerations ("avoiding gameability"). When a candidate is added to the waitlist, the transplant program is considered to have formally communicated both to the patient and the OPTN an intent-to-treat through the modality of transplantation. The Work Group agreed that transplants and deaths should be evaluated relative to all waitlisted patients, even those who are temporarily in inactive status.

Only including active patients renders the possibility of manipulating the transplant rate metric by setting groups of patients to inactive status during a time a program is having difficulty managing its waitlist. Furthermore, another component of the CPM – the acceptance rate – already has the potential to be affected by setting patients to inactive status, since such patients will not receive offers. The Work Group did not want all of the metrics to be able to be influenced by setting patients to inactive status.

Analyses have shown little or no relationship between programs' CPM and the percent of their list that is in inactive status (Additional information available upon request). Programs were found to have low, moderate, and high CPM all along the inactivity spectrum from those having very few inactive patients to programs with upwards of 80% of their list inactive.

***Choice of CPM > 1.5 threshold for identifying programs***

The CPM was developed such that higher values were associated with higher waitlist mortality rates, lower transplant rates, and/or lower acceptance rates (Step 2 in *Calculating the CPM*, above). This choice of scaling ("lower is better") was arbitrary but was selected to parallel post-transplant outcomes (graft failure and patient death O/E's).

The CPM threshold of 1.5 was also initially considered because it mirrors the O/E > 1.5 threshold traditionally used in evaluating post-transplant outcomes. As shown in Figures 1 & 2, the value of 1.5 occurs in the tail of the distribution, identifying only a relatively small number of programs that appear to be outliers in terms of pre-transplant performance. The CPM Work Group reviewed statistical profiles of programs with CPM > 1.5 and agreed that based on this underlying data, such programs should be reviewed for pre-transplant performance. Finally, both of the high profile programs with severe waitlist mismanagement issues would have had a CPM exceeding 1.5 during their crisis periods.

### *CPM and pediatric patients*

The CPM Work Group had extensive deliberations about the implications of CPM on pediatric patients. Currently, pre-transplant metrics are not produced separately for pediatric and adult patients. Rather, these metrics (and, in turn, CPM) include performance on both pediatric and adult patients together. Though most programs tend to predominantly serve either pediatric patients or adult patients, some programs serve a significant mix of both types of patients.

**Table 2**: CPM for Pediatric vs. Adult Programs
Based on January 2010 PSR Cohort

| Organ | Predominant Program Type* | Number of Programs | CPM Stats | | CPM Percentiles | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Mean | Stdev | P5 | P25 | P50 | P75 | P95 |
| **Kidney** | **Adults** | 209 | 0.99 | 0.21 | 0.74 | 0.85 | 0.94 | 1.08 | 1.34 |
| **Kidney** | **Peds** | 34 | 0.99 | 0.17 | 0.76 | 0.90 | 0.97 | 1.04 | 1.36 |
| **Liver** | **Adults** | 104 | 1.04 | 0.28 | 0.68 | 0.83 | 0.96 | 1.24 | 1.59 |
| **Liver** | **Peds** | 22 | 1.03 | 0.25 | 0.80 | 0.91 | 0.99 | 1.06 | 1.46 |

** Programs with more than 50% of pediatric patients considered predominantly pediatric programs.*

In an analysis shown to the CPM Working Group in April 2010 (Table 2), the CPM distribution was shown to be very similar for predominantly adult vs. pediatric programs, suggesting CPM is not biased with respect to pediatric programs.

Creating separate pre-transplant metric O/E's, and CPM values, for pediatric patients separately from adult patients has been discussed and could be considered for a future revision to the CPM methodology.

### *Decision to switch to an offer-based model using all organs*

The CPM was initially developed using both an offer-based and an organ-based model. In the offer-based model, an organ refused for multiple candidates on a program's waitlist would be counted as multiple refusals, whereas the organ-based model would just consider it as one refused organ. Furthermore, these models only included "good" organs, excluding ECD's, DCD's, and other "marginal" or *difficult to place* organs from the calculations.[10]

The current SRTR contractor has recently made improvements to the liver acceptance rate model, including moving to a single, offer-based model that includes all organs that were ultimately accepted and transplanted. Based on a sensitivity analysis that showed very little changes in the CPM when switching to the new modeling approach, the CPM Work Group agreed to adopt the new approach as part of the CPM (Additional information available upon request).

[10] Wolfe RA, LaPorte FB, Rodgers AM, Roys EC, Fant G, Leichtman AB. Developing organ offer and acceptance measures: when 'good' organs are turned down. Am J Transplant 2007;7:1404-11.

This model is described in more detail in the *The Underlying Risk-Adjusted Models behind the CPM* section of this document.

***Decision to remove transplant rates from the "safety net" component of the CPM approach***

A final change adopted in 2014 by the CPM Work Group and MPSC was to remove the transplant rate component of the CPM safety net. Previously, the safety net element of the CPM approach would identify programs with mortality rate O/E > 2.0 (one-sided $p<0.05$) *or* transplant rate O/E < 0.25 (one-sided $p<0.05$). Due to concerns about using the transplant rate as a stand-alone metric, the Work Group decided to remove this element of the safety net but leave the mortality rate component.

Further evolution of the CPM may be needed after the committee has had a chance to review some programs based on pre-transplant performance. These reviews may reveal false positives, and false negatives may also be discovered. Modifications to the CPM methodology, the review threshold, and/or component models may be needed in the future.

**Alternative Approaches Considered:**

During the course of developing and refining the CPM methodology, engaging the transplant community, and in committee deliberations, the following alternative approaches for pre-transplant performance monitoring were considered but ultimately not endorsed by the committee.

1. Using acceptance rates alone

2. Using transplant rates alone

3. Using a transplant rate threshold, a mortality rate threshold, and an acceptance rate threshold independently

4. Using a metric such as Life Years from Listing (LYFL) that combines both pre and post-transplant performance

5. Using a metric that measures how well hospitals are serving patients with end-stage organ disease in their geographic area, not just those that have been waitlisted

6. Using a statistical process control (SPC) technique such as CUSUM in lieu of developing a cohort-based pre-transplant metric for identifying programs

Further explanation of these alternatives and the rationale for proposing the CPM approach can be found in the appendix.

**Intended Effects of this Proposal:**

It is intended that this proposal will identify a small number of transplant programs with extreme, outlying pre-transplant performance indicators. The MPSC will inquire about such programs in an attempt to determine if process improvements are necessary. If so, it is expected that these programs' pre-transplant performance will eventually "normalize" to some degree, reducing the overall variability among programs in pre-transplant metrics.

The MPSC may also become aware of programs with potentially exemplary pre-transplant performance based on these new metrics. Understanding the practices of these programs may provide insight into ways other programs may be able to improve in terms of effectively serving their waitlists.

It is also expected that some transplant centers may become less risk averse due to the increased emphasis on pre-transplant metrics. This may lead to an increase in the number of liver and/or kidney transplants performed nationally, as well as a decrease in liver and/or kidney discard rates. This may also lead to increases in the number and characteristics of donors recovered for the purpose of transplantation.

**Potential Unintended Consequences:**

*Increased aversion to adding higher-risk patients to the waitlist*

It is possible that some centers may decide to add fewer patients to the waitlist, in particular patients considered to be harder to transplant and/or with higher likelihood of waitlist mortality, in response this proposal. Research suggests that this behavior change – listing fewer high-risk patients – has already been taking place due to concerns with post-transplant monitoring[11]. It is unknown whether this aversion to listing patients will increase beyond the current level.

It is important for transplant programs to be fully aware of the effectiveness of the statistical adjustment for various risk factors before deciding to be more selective in listing patients. For example, both the liver transplant rate model and waitlist mortality model adjust for each patient's MELD score at listing, since patients with higher MELD scores are likely to be transplanted more quickly, but also have a higher likelihood of death after listing. Some programs have expressed concern about listing low-MELD patients in light of pre-transplant performance evaluation; however, analyses have shown little to no discernible relationship between liver programs' CPM and the percent of their waitlist with MELD of 18 or higher. Programs were found to have low, moderate, and high CPM regardless of whether they had just 10% or upwards of 50% of their patients with MELD of 18+. (Additional information available upon request).

These pre-transplant statistical models have c-statistics ranging from 0.63 to 0.91, suggesting good or excellent ability to predict likelihood of organ acceptance, transplantation, and mortality after listing. *Programs performing better than national averages with respect to certain patient subpopulations may actually worsen their pre-transplant O/E's by changing listing practices.* On the other hand, for some programs, tightening patient selection criteria may be warranted, not to manipulate the metrics, but because process improvements are needed in the area of pre-transplant patient care and reducing waitlist mortality.

*Use of CPM by payers*

The CPM was developed to be used within the confines of the OPTN for the purpose of identifying process improvement opportunities. It was not developed for use in determining transplant hospital reimbursement or for public consumption, and the OPTN does not intend to have CPM published on any publicly available website.

[11] Schold JD, Arrington CJ, Levine G. Significant alterations in reported clinical practice associated with increased oversight of organ transplant center performance. Prog Transplant 2010;20:279-87.

The CPM (and CPM-specific subcomponent model results) will be provided to the MPSC's PAIS for review of transplant programs. Transplant programs will also be able to access their own pre-transplant results.

**Additional Supporting Evidence:**

In addition to the previously referenced analyses, the following information is presented in support of this proposal.

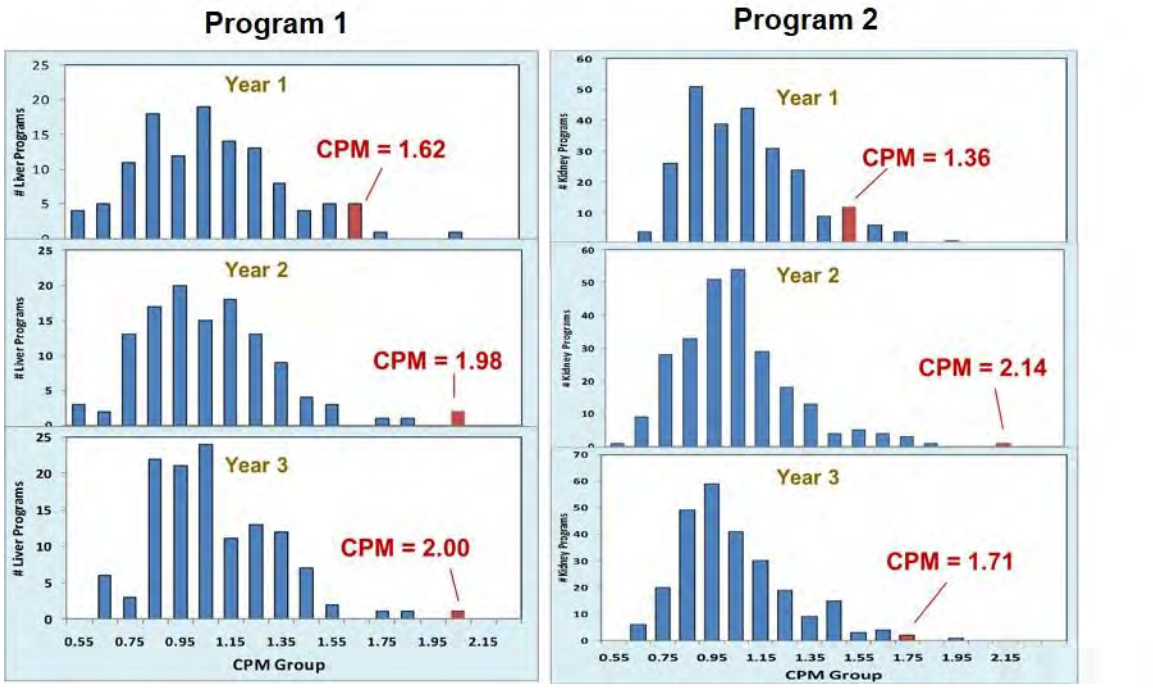


**Figure 5. Retrospectively calculated CPM for two programs grossly unable to manage their waitlist. These programs had outlying CPM values during their crisis periods.**

Though motivated (in part) by the two programs found to have severe waitlist mismanagement in the mid-2000's, the CPM was not designed by analyzing data from these two high profile cases. However, as shown in Figure 5, both programs would have stood out as outliers with respect to CPM during their crisis periods, had the metric been available at the time. Though this analysis does not represent a complete "validation" of the CPM, it provided the working group and MPSC with increased confidence that the metric would, at minimum, achieve the goal of identifying egregious waitlist management issues.

In addition to analyses supporting the analytical decisions made in developing the CPM, clinical research suggests that centers' *median waiting time to transplant* is a more important center-level factor than recipient outcomes for predicting survival of waitlisted patients.[12] This research highlights the importance of including both pre-transplant waitlist management and post-transplant patient and graft survival in reviews of overall program performance.

[12] Schold, Jesse D., et al. "The pivotal impact of center characteristics on survival of candidates listed for deceased donor kidney transplantation." *Medical care* 47.2 (2009): 146-153.

**Expected Impact on Living Donors or Living Donation:**

Since the transplant rate being used in the CPM counts both deceased and living donor transplants, it is conceivable that some programs may consider ways to develop or expand their living donor transplant services to further meet the needs of their waitlisted patients.

**Expected Impact on Specific Patient Populations:**

The Bylaw revision has no known impact for specific patient populations.

**Expected Impact on OPTN Strategic Plan, and Adherence to OPTN Final Rule:**

This proposal is consistent with the OPTN Final Rule, which stresses the importance of reviewing inter-transplant program variability in waitlist mortality. In addition, the proposal addresses the OPTN key goal of increasing access to transplants. Specifically, the proposal helps the OPTN meet the objective of promoting the best use of donated organs. The proposal provides a meaningful metric to identify potential issues with pre-transplant performance and waiting list practices, in addition to identifying best practices. The strategic plan includes many areas for focus including sharing best practices with the transplant community as well as facilitating patient access. This metric will be used to evaluate member programs that may not be effectively managing the waiting list and identify opportunities to increase access to transplantation.

**Plan for Evaluating the Proposal:**

The MPSC will monitor whether the new methodology is identifying those transplant programs that are truly underperforming in meeting the needs of their waitlisted patients. The distribution of CPM will also be closely tracked to see if changes to the methodology and/or review threshold are needed.

**Additional Data Collection:**

This proposal does not require additional data collection.

**Expected Implementation Plan:**

If successful, this proposal will be considered by the Board of Directors in June 2015. If approved by the Board, implementation of this proposal will be guided by the SRTR's schedule for producing pre-transplant program performance metrics, as well as the development of a process for routine calculation of the CPM and dissemination of this information to the MPSC and members. It is expected that the review of pre-transplant performance will initially be implemented for liver and kidney programs only since all of the models used by the CPM analysis have not yet been developed for other organs.

This proposal will not require programming in UNet℠. The CPM (and CPM-specific subcomponent model results) will be provided to the MPSC's PAIS subcommittee for blinded review of transplant programs, but the OPTN does not intend to publish the CPM on any publicly available website. Transplant programs will be able to access their own CPM and pre-transplant results.

**Communication and Education Plan:**

The proposal addresses new methodology for assessing member performance and process improvement. Communication and education efforts will address awareness of the new system, the factors that are used to assess pre-transplant outcomes, and how the metrics will be used in member monitoring.

Information about the new model would be included in ongoing efforts to inform members about monitoring of member performance, including educational presentations such as webinars or e-learning modules. The OPTN Evaluation Plan would also be updated with information about the metrics and their applicability to monitoring of member performance.

In addition, notification of the amended bylaw requirements would be included in the following routine communication vehicles:

- Policy notice
- System notice
- Article on OPTN website and member e-newsletter
- Notification to a listserv group for transplant administrators

**Compliance Monitoring:**

This proposal introduces a new tool for performance monitoring of transplant programs. Currently, the MPSC's review of performance monitoring is limited to post-transplant patient and graft survival and functional inactivity defined by the lack of transplant activity for a specified period. The addition of monitoring of pre-transplant waiting list management will provide the MPSC with a more balanced view of a program's overall performance. The MPSC anticipates that the review of pre-transplant metrics by the Performance Analysis and Improvement Subcommittee (PAIS) will be similar to our current process for review of post-transplant survival and functional inactivity. The MPSC will evaluate the effectiveness of this process and consider revisions to the process as well as the CPM methodology and review threshold, if deemed appropriate.

Falling below either the CPM or mortality rate safety net threshold contained in the proposal will trigger an inquiry by the PAIS. The inquiry will request information relevant to the transplant program's waiting list management process and any unique clinical aspects (i.e., potential mitigating factors) that may influence its ability to meet the thresholds. In its review, the PAIS will consider other available metrics as well as information submitted by the member to determine if the program is truly underperforming and in need of assistance to improve. The PAIS will have the same options as those available for post-transplant and functional inactivity reviews, including:

> **Release from reporting**: the PAIS may recommend releasing a program from review if satisfied that the issues that led to review have been addressed by the program and/or the program's pre-transplant performance has improved. Releasing a program from reporting does not mean that the program is no longer subject to performance reviews conducted by the PAIS. Rather, the program is released from actively reporting to the PAIS at that time. A program can be introduced back into the PAIS performance reviews if, in subsequent cohorts, it does not meet the performance thresholds established by the PAIS.

**Continue to report**: In its simplest form, a recommendation for continued monitoring by the PAIS is a recommendation for continued reporting for the next meeting cycle. The subcommittee will request the submission of additional information to further assess factors contributing to a program's lower than expected performance and the program's improvement efforts.

**Informal Discussion**: Programs may be offered the opportunity to meet with the PAIS informally, through a teleconference. An informal discussion provides the members of the PAIS the ability to ask questions of program personnel in real time, and allows the program personnel to address issues that are sometimes hard to summarize in the paper submissions. Programs can be invited to participate in an informal discussion with the PAIS if the program has not been able to identify steps to improve patient outcomes, there has been an apparent lack of progress in implementing the site visit recommendations, or if the PAIS simply wishes to discuss particular issues with the program. An informal discussion does not constitute an adverse action.

**Peer Visit**: Some programs may be recommended to undergo a peer review site visit. Typically, programs must be under review for at least two MPSC cycles before the PAIS makes this recommendation, and the program has not been able to identify steps to improve patient outcomes and/or there has been an apparent lack of progress in implementing improvements. The peer visit team would generally include a transplant surgeon, physician, and administrator, and is supported by a UNOS staff member. Typically, the panel would be on-site for two days to conduct interviews of all key personnel to the program, including ancillary support, as well as an in-depth review of the relevant patient charts. At the conclusion of the site visit, the panel would provide the center with a preliminary (verbal) summary of its findings. A formal report would be submitted to the PAIS for issuance to the program.

Once a program has undergone a peer visit and received the report, the PAIS would request a plan for quality improvement be submitted in response to the recommendations contained within the report. The Committee would continue to monitor the program's progress in implementing the site visit recommendations.

**Voluntary Inactivation**: In those rare instances where the review of a program raises concerns for patient safety, the MPSC may recommend that a member inactivate a program or a component of a program, or withdraw its designated transplant program status. Programs that do not voluntarily inactivate or withdraw membership status may be recommended for other action, such as probation or member not in good standing under Bylaws, Appendix L. 15. OPTN Determinations and Actions.

**Bylaw Proposal:**

Proposed new language is underlined (example) and language that is proposed for removal is struck through (~~example~~).

## D.10 ~~Additional Transplant Program Requirements~~ Transplant Program Performance Reviews

The MPSC will conduct reviews of transplant program performance to identify underperforming transplant programs and require the implementation of quality assessment and performance improvement measures.

Transplant program performance reviews will be used to determine if the lower than expected performance can be explained by patient mix or some other unique clinical aspect of the transplant program. If a program's performance cannot be explained by patient mix or some other unique clinical aspect of the transplant program, the member, in cooperation with the MPSC, will adopt and promptly implement a plan for quality improvement. The member's failure to adopt and promptly implement a plan for quality improvement will constitute a violation of OPTN obligations.

As part of this process, the MPSC may conduct a peer visit to the program at member expense. The MPSC may also require, at its discretion, that the member participate in an informal discussion. The informal discussion may be with the MPSC, a subcommittee, or a work group, as determined by the MPSC. The informal discussion will be conducted according to the principles of confidential medical peer review, as described in *Appendix L* of these Bylaws. The informal discussion is not an adverse action or an element of due process. A member who participates in an informal discussion with the MPSC is entitled to receive a summary of the discussion.

The MPSC may recommend that a member inactivate a program or a component of a program or withdraw its designated transplant program status based on patient safety concerns arising from review of the program's graft and patient survival. If the program fails to inactivate or withdraw its designated transplant program status when the MPSC recommends it do so, the MPSC may recommend that the Board of Directors take appropriate action as defined in *Appendix L: Reviews, Actions, and Due Process* of these Bylaws.

### A. Pre-Transplant Performance Reviews

MPSC review of transplant program performance can be triggered through a review of pre-transplant metrics including waiting list mortality rate, transplant rate, and offer acceptance rates.

The MPSC will review a transplant program based on pre-transplant performance if the program meets *either* of the following criteria over a 1-year period:

- The composite pre-transplant metric (CPM) is greater than 1.5
- The waiting list mortality rate observed to expected ratio is greater than

2.0 and the one-sided p-value is less than 0.05

## B. Post-Transplant Performance Reviews

MPSC review of transplant program performance can be triggered through a review of the one-year graft and patient survival rates. The MPSC will review a transplant program if it has a low survival rate compared to the expected survival rate for that transplant program. The MPSC utilizes performance metrics produced by the Scientific Registry of Transplant Recipients (SRTR) as the principal tool to identify transplant programs that have lower than expected outcomes.

For programs performing 10 or more transplants in a 2.5 year period, the MPSC will review a transplant program if it has a higher hazard ratio of mortality or graft failure than would be expected for that transplant program. The criteria used to identify programs with a hazard ratio that is higher than expected will include *either* of the following:

1. The probability is greater than 75% that the hazard ratio is greater than 1.2.

2. The probability is greater than 10% that the hazard ratio is greater than 2.5.

For programs performing 9 or fewer transplants in a 2.5 year period, the MPSC will review a transplant program if the program has one or more events in a 2.5 year cohort.

## D.~~10~~ 11 Additional Transplant Program Requirements

## ~~A. Transplant Program Survival Rates~~

~~The MPSC will conduct reviews of transplant program performance to identify underperforming transplant programs and require the implementation of quality assessment and performance improvement measures. One measure of transplant program performance is triggered through a review of the one-year graft and patient survival rates. The MPSC utilizes performance metrics produced by the Scientific Registry of Transplant Recipients (SRTR) as the principal tool to identify transplant programs that have lower than expected outcomes.~~

~~For programs performing 10 or more transplants in a 2.5 year period, the MPSC will review a transplant program if it has a higher hazard ratio of mortality or graft failure than would be expected for that transplant program. The criteria used to identify programs with a hazard ratio that is higher than expected will include *either* of the following:~~

~~1. The probability is greater than 75% that the hazard ratio is greater than 1.2.~~

~~2. The probability is greater than 10% that the hazard ratio is greater than 2.5.~~

~~For programs performing 9 or fewer transplants in a 2.5 year period, the MPSC will review a transplant program if the program has one or more events in a 2.5 year cohort.~~

~~The MPSC review will be to determine if the higher hazard ratio or events can be explained by patient mix or some other unique clinical aspect of the transplant program. If a program's performance cannot be explained by patient mix or some other unique clinical aspect of the transplant program, the program, in cooperation with the MPSC, will adopt and promptly implement a plan for quality improvement. The member's failure to adopt and promptly implement a plan for quality improvement will constitute a violation of OPTN obligations.~~

~~As part of this process, the MPSC may conduct a peer visit to the program at member expense. The MPSC may also require, at its discretion, that the member participate in an informal discussion. The informal discussion may be with the MPSC, a subcommittee, or a work group, as determined by the MPSC. The informal discussion will be conducted according to the principles of confidential medical peer review, as described in *Appendix L* of these Bylaws. The informal discussion is not an adverse action or an element of due process. A member who participates in an informal discussion with the MPSC is entitled to receive a summary of the discussion.~~

~~The MPSC may recommend that a member inactivate a program or a component of a program or withdraw its designated transplant program status based on patient safety concerns arising from review of the program's graft and patient survival. If the program fails to inactivate or withdraw its designated transplant program status when the MPSC recommends it do so, the MPSC may recommend that the Board of Directors take appropriate action as defined in *Appendix L: Reviews, Actions, and Due Process* of these Bylaws~~.

### ~~B~~A. Patient Notification Requirements for Waiting List Inactivation

[*Subsequent headings affected by the re-numbering of this policy will also be changed as necessary*.]

## Appendix M. Definitions

**Composite Pre-Transplant Metric (CPM)**

The composite pre-transplant metric (CPM) is an aggregate, pre-transplant observed to expected ratio that combines observed to expected ratios of waiting list mortality rate, transplant rate including deceased and living donor recipients, and offer acceptance rates into one number. The CPM for kidney programs does not include an observed to expected ratio for waiting list mortality rate.

## Appendix A

## Alternative Approaches Considered

1. Using acceptance rates alone

Instead of using a composite metric approach, a simpler alternative would be to use offer acceptance rates alone to identifying programs for MPSC review. Of the three metrics – transplant, mortality, and acceptance rates – it is the acceptance rate over which transplant programs may have the most direct influence. However, the following reasons support use of the composite approach as opposed to relying on acceptance rates alone for performance monitoring.

- A prior Joint Board-MPSC Work Group reviewed acceptance rate data, and though they found the acceptance rate to represent a potentially useful and powerful metric for identifying anomalous transplant program performance, the group recommended that acceptance rates *not* be used as a stand-alone metric for identifying programs.
- Though acceptance rates are correlated with transplant rates, not explicitly including transplant rates in pre-transplant performance monitoring could result in future programs with systemic waitlist management issues not being identified for corrective action. Two high profile cases in the mid-2000's of transplant programs being grossly unable to meet the needs of their waitlisted patients was one of the key motivating factors for the CPM. An analysis showed that while acceptance rates were generally lower than expected for these two programs, it was their transplant rates that most stood out as being aberrant compared to the rest of the country (Additional information available upon request).
- During the lengthy process of vetting and refining the CPM over the course of five years, the CPM Work Group on several occasions reviewed data "profiles" of programs that would be identified for further review by CPM. Using acceptance rates alone would identify a different set of transplant programs with noticeably different pre-transplant profiles compared to using CPM. For example, based on data from year 2011:
    - One liver program had a slightly above average acceptance rate O/E of 1.01 but a waitlist mortality rate of over 3 times expected (p=0.03). Such a program would be identified by the mortality rate "safety net" component of the CPM approach but not by an acceptance rate only approach.
    - Conversely, several liver programs had low acceptance rates but much lower than expected waitlist mortality rates and would thus not be identified by the CPM approach.
- It is possible for a center that is not adequately meeting the needs of its waitlisted patients to set a significant percentage of its patients to inactive status in order to avoid receiving offers that they have no capacity to accept. This type of manipulation of data would artificially increase their acceptance rates due to a reduced denominator. On the other hand, since the transplant and mortality rate models include both active and inactive patients, these metrics cannot be manipulated in this way. The CPM's use of multiple metrics helps to mitigate against the potential for data manipulation associated with any one metric.
- An analysis completed for the CPM Work Group showed that kidney programs that perform more living donor transplants (as a percentage of their total kidney transplants) tend to be more selective in accepting organ offers, as evidenced by lower (case-mix adjusted) acceptance rates (O/Es) (Additional information available

upon request). In this way, use of acceptance rates alone, which exclusively focus on deceased donor transplantation, might result in unfair pre-transplant performance assessments for programs that rely on a mix of living and deceased donor organs to meet the needs of their waitlisted patients. The CPM, on the other hand, incorporates transplant rates that include both living and deceased donor transplants, as well as acceptance rates.

2. Using transplant rates alone

   During the several year process of developing, scrutinizing, and refining the CPM approach, one alternative suggestion was to use transplant rates alone for evaluating programs' pre-transplant performance. The two high profile cases of waitlist mismanagement, after all, had very low transplant rates. However, this was ultimately not proposed for the following reasons:

   - The 2008 GAO Report[3] emphasized the use of acceptance rates for program monitoring.
   - The OPTN MPSC Process Improvement Working Group's recommendation to use acceptance rates but not as a stand-alone trigger. CPM accomplishes this goal by utilizing acceptance rates in conjunction with mortality and transplant rates to prioritize programs for review.
   - Though CPM would have identified the two high profile programs as "standing out," it was not designed by strictly focusing on those two outliers. The CPM does not rely on the assumption that all future, severe waitlist management problems will manifest the same way as the previous two, where low transplant rates were the leading indicator. Acceptance rates were also extremely low for these programs and may be a leading indicator in some cases.
   - Of the three metrics, acceptance rates are arguably under the most direct influence of transplant programs.
   - The CPM Work Group believed that waitlist mortality rates should play an important role in pre-transplant monitoring, especially for liver programs.

   After the incorporation of a geographic (supply-to-demand) adjustment in the transplant rate model, the use of transplant rates alone for identification of programs was again considered. However, when presented with the option to use only supply-to-demand adjusted transplant rates in April 2013, the CPM Work Group concluded that the composite metric approach was still preferred.

3. Using a transplant rate threshold, a mortality rate threshold, and an acceptance rate threshold independently

   The CPM Work Group was presented with the option of identifying programs based on all of the metrics individually, but instead endorsed the composite metric approach, out of concern that too many programs would be identified using this alternate method. It turns out that setting independent thresholds to identify programs performing 50% "worse" in any of the three metrics (mortality O/E > 1.5, acceptance O/E < 0.67, transplant rate O/E < 0.67, and statistically significant) would have identified an order of magnitude more programs than the CPM approach. Based on an analysis using the July 2012 cohort, 64 kidney and 59 liver programs would have been identified using this alternate approach, a total of 123 programs.

Though these thresholds could be calibrated to identify approximately the same number of programs as CPM, there are a vast number of possible ways to accomplish this. And the CPM Work Group preferred the composite approach since it simplifies the multi-dimensional suite of pre-transplant metrics into a single dimension for monitoring.

4. Using a metric such as Life Years from Listing (LYFL) that combines both pre and post-transplant performance[13]

   Programs may be serving their waitlisted patients equally well in different ways: by performing slightly fewer transplants but maintaining excelling post-transplant outcomes; or by having slightly inferior post-transplant outcomes but performing more transplants. The expected (risk-adjusted) survival of patients after being added to the waitlist may be the same at both programs. LYFL would capture how well transplant programs are serving the patients on their waitlist by measuring patient survival after listing, including both pre and post-transplant survival time; this metric could also be quality-of-life adjusted, for example by discounting survival time after graft failure (e.g., on dialysis).

   Though such a metric may be a very useful and perhaps consumer/patient-friendly way to quantify transplant programs' overall performance, it may not be ideally suited for identifying areas for process improvement. A program with extremely poor post-transplant metrics could have an unextraordinary LYFL if they had very high transplant rates. Conversely, a program with poor pre-transplant performance might be offset in the LYFL by very good post-transplant patient and graft survival metrics. In these ways, the use of a *pre and post*-transplant composite metric such as the LYFL may obscure the need for process improvement in either pre or post-transplant processes.

   The LYFL is also partially redundant with current post-transplant outcomes metrics, so it may not make sense for the MPSC to use both the recently revised Bayesian outcome metrics and LYFL to monitor programs. Still, the LYFL is a metric that deserves further study to determine how best to define it and what purpose it most effectively serves.

5. Using a metric that measures how well hospitals are serving patients with end-stage organ disease in their geographic area, not just those that have been waitlisted.

   The CPM Work Group discussed on several occasions the desirability of measuring performance relative to those patients truly in need, not just those that transplant hospitals choose to add to the waitlist. Expanding the denominator in this way could help alleviate the potential unintended consequence of altered listing practices due to monitoring pre-transplant performance.

   However, the OPTN does not collect data on transplant patients prior to registration on the waitlist. Historically, transplant program performance on patients prior to registration has been outside the purview of the OPTN. Furthermore, even if such data were available and able to be used by the OPTN, it may be difficult to attribute performance to specific transplant programs that serve the same or overlapping populations of end-stage organ failure patients with respect to geographic location (e.g., in the same "health service areas").

[13] Kasiske, B. L., et al. "Report of a consensus conference on transplant program quality and surveillance." *American Journal of Transplantation* 12.8 (2012): 1988-1996.

6. Applying a statistical process control (SPC) technique such as CUSUM in lieu of developing a cohort-based pre-transplant metric for identifying programs.

   An SPC technique such as CUSUM (Cumulative Sum) may actually be more adept than CPM at rapidly detecting sudden disruptions in a transplant program's ability to serve its waitlisted patients. The SRTR contractor has already developed post-transplant CUSUM charts to help programs identify potentially concerning trends in patient and graft outcomes. HRSA and both the OPTN and SRTR contractors have discussed the possibility of applying the CUSUM methodology to pre-transplant metrics as well.

   However, the CPM Work Group considers CUSUM and CPM to be complementary tools, as opposed to one being a potential replacement for the other, just as post-transplant CUSUM charts[14,15] are complementary to post-transplant O/E's. Post-transplant CUSUM charts are being used strictly for transplant programs' self-improvement (QAPI) efforts, not for OPTN/MPSC oversight purposes. Pre-transplant CUSUM charts could potentially be used in the same way. They would not alleviate the need for a cohort-based pre-transplant metric like CPM, but rather serve in a complementary role to help alert programs to potentially concerning trends in pre-transplant performance. Additional work is needed in this area.

**Method for 'Shrinking' O/E Ratios: Approximation to Empirical Bayes/BLUP**

*Rationale for using Empirical Bayes methodology*

During the initial attempts to combine O/E ratios, it became clear that accounting for the variability due to small sample sizes would be critical. Failing to address statistical uncertainty would lead to "small sample false positives," where the programs having the most aberrant composite metric values were those with the smallest sample sizes, and thus those with the least empirical evidence of a potential problem. These false positives would obscure the metric's ability to identify true positives, programs with sufficient evidence to suggest a pre-transplant performance issue exists and who actually need to implement improvement measures.

Use of p-values to address sampling variability was considered. For example, all O/E ratios not statistically different from 1.0 could be set to 1.0. However, this approach would be highly reliant on the arbitrary significance level threshold of 0.05 (or other chosen alpha level). For example, a program with a transplant rate O/E of 2.50 (p=0.04) would have 2.5 used in the composite metric, whereas a program with transplant rate O/E of 2.5 (p=0.06) would have a 1.0. Furthermore, this approach forces an unsatisfactory binary decision – either use the nominal O/E (if $p<0.05$) or use 1.0 (if $p>=0.05$) – when in reality a better estimate of a program's true underlying performance may lie somewhere in the middle. For developing this composite metric, effective *estimation* was paramount and far more relevant than *hypothesis testing*.

Consequently, the Empirical Bayes estimation methodology – a.k.a., best linear unbiased prediction (BLUP) – was a natural choice for accounting for sampling variation in the O/E ratios

[14] Biswas, Pinaki, and John D. Kalbfleisch. "A risk-adjusted CUSUM in continuous time based on the Cox model." *Statistics in medicine* 27.17 (2008): 3382-3406.

15 Axelrod, D. A., et al. "Transplant Center Quality Assessment Using a Continuously Updatable, Risk-Adjusted Technique (CUSUM)." American journal of transplantation 6.2 (2006): 313-323.

before combining them. This approach is derived from a "random effects" (hierarchical) model, with patient-level (or organ offer-level) data contained within transplant hospitals. By estimating the random effect for each program, the result is a weighted average of the program's observed performance in a particular metric and the overall national average performance.

The approximation to the Empirical Bayes method results in a weighted average of the program's observed O/E ratio and the overall national average of 1.0, essentially "shrinking" the program's O/E toward 1.0. Conceptually, this approach is "Bayesian" because it starts with the proposition that a transplant program is no different from the rest of the nation (O/E=1.0), and then modifies that proposition based on the strength of evidence underlying the program's O/E ratio. It can also be thought of as an example of the "regression to the mean" phenomenon[16], where future values tend to be more similar to the long-run average than to very recent results. Programs with large sample sizes (small amount of statistical uncertainty in the O/E) will have a resulting value very close to their nominal O/E value (minimal "shrinkage"). Conversely, O/E's for programs with small sample sizes will be pulled closer to 1.0, reflecting the uncertainty in their observed performance.

It is highly unlikely that a program with a waitlist mortality rate O/E of 8.5, based on one patient death, is actually 8.5 times worse than the national average in terms of mitigating death on the waitlist. A better estimate of this program's true performance is somewhere closer to 1.0, perhaps just slightly above 1.0. The Empirical Bayes (BLUP) methodology recognizes that the 8.5 is not a realistic estimate, even if it happens to be statistically significantly different from 1.0. This methodology has been shown to perform better at estimating true underlying performance than using the nominal center effect estimates and has been used extensively in estimating institutional performance in healthcare.[17,18,19,20] Using shrinkage estimation for more reliable estimation of institution performance was also a recommendation from the Committee of Presidents of Statistical Societies.[21]

The CPM Work Group reviewed data on CPM by size of transplant program, as measured by number of waitlist candidates. Results showed that the percentage of programs with CPM > 1.5 was statistically no different for programs with 10-49 patients, 50-99, 100-249, 250-499, and 500+, suggesting that the CPM methodology is not biased toward small, medium, or large programs. However, extremely small programs (<10 candidates) tend to have CPMs close to 1.0; none of these programs had CPM > 1.5 due to the absence of strong evidence suggesting true deviance in pre-transplant performance.

*Addition of "limited translation rules"*

One weakness of the Empirical Bayes (BLUP) methodology is the potential for "overshrinkage." This phenomenon can occur when a program with truly aberrant performance is assumed, as in Empirical Bayes methodology, to be part of the same bell-shaped distribution as all other

16 Stigler, Stephen M. "Regression towards the mean, historically considered." Statistical methods in medical research 6.2 (1997): 103-114.

17 Robinson, G. K. That BLUP is a Good Thing: The Estimation of Random Effects. Statistical Science 6 (1991), no. 1, 15--32. doi:10.1214/ss/1177011926. http://projecteuclid.org/euclid.ss/1177011926.

18 Thomas, N., Longford, N. T. and Rolph, J. E. (1994), Empirical Bayes methods for estimating hospital-specific mortality rates. Statist. Med., 13: 889–903. doi: 10.1002/sim.4780130902

19 Christiansen CL, Morris CN. Improving the Statistical Approach to Health Care Provider Profiling. Ann Intern Med. 1997;127:764-768. doi:10.7326/0003-4819-127-8_Part_2-199710151-00065

20 Efron, Bradley, and Carl N. Morris. *Stein's paradox in statistics*. WH Freeman, 1977.

21 "Statistical Issues in Assessing Hospital Performance," Commissioned by the Committee of Presidents of Statistical Societies (COPSS), 2012.

programs. Since a key purpose of the CPM is to identify outlying performance, the possibility of overshrinkage was a valid concern.

To mitigate against overshrinkage, "limited translation rules"[22] are applied. The limited translation rule implemented in the CPM methodology is that shrinkage is constrained to extend no further than the 95% confidence limits for the O/E ratio. If shrinkage would shift an O/E ratio closer to 1.0 than either the upper or lower confidence limit, the modified O/E value is set equal to the confidence limit instead.

*Approximation to empirical Bayes/BLUP method for 'shrinking' O/E ratios*

The "shrunken" O/E's are derived by the following formula, which can be thought of as a weighted average between the program's O/E ratio and the national average, or 1.0, on a natural logarithm scale:

$$(1) \quad Shrunken\ (O/E) = \frac{\sigma_D^2}{(\sigma_D^2 + \sigma_r^2)} * \ln(O/E) + \frac{\sigma_r^2}{(\sigma_D^2 + \sigma_r^2)} * \ln(1.0)$$

The term $\frac{\sigma_D^2}{(\sigma_D^2+\sigma_r^2)}$ is the weight associated with the program's observed O/E ratio, while $\frac{\sigma_r^2}{(\sigma_D^2 + \sigma_r^2)}$ is the weight associated with the central value of 1.0. Together, these two weights sum to 1, or 100%. Since ln (1.0) is zero, the second half of the formula collapses and what remains is simply

$$(2) \quad Shrunken\ (O/E) = \frac{\sigma_D^2}{(\sigma_D^2 + \sigma_r^2)} * \ln(O/E)$$

$\sigma^2_D$ represents an estimate of the variance among programs (or, "program-to-program" variance) in the metric of interest, either transplant rate, mortality rate, or acceptance rate. $\sigma^2_r$ represents an estimate of the variance associated with the particular program's metric of interest. The higher the variance in the program's metric, the lower the weight and hence the greater the shrinkage. The greater the program-to-program variability, the higher the weight and hence less shrinkage.

This formula is actually derived as the best linear unbiased predictor (BLUP) from a random effects model where the response variable is normally distributed. It is also commonly referred to as the empirical Bayes estimator. This formula was adapted to accommodate the binomial (acceptance rates) and poisson-distributed (mortality and transplant rates) metrics that comprise the CPM, as follows:

Binomial case (acceptance rates):

- ✓ Program-specific variance: $\sigma^2_r$ calculated as p*(1-p)/N, where
  - p is the overall, national average acceptance rate across all programs
  - N is the number of offers received by this program
- ✓ Program-to-program variance: $\sigma^2_d$ is calculated as follows
  - First, the estimated acceptance rate for each program *assuming they all had the same, average case-mix of offers and candidates* is computed.

[22] Efron, Bradley, and Carl Morris. "Limiting the risk of Bayes and empirical Bayes estimators—Part II: The empirical Bayes case." *Journal of the American Statistical Association* 67.337 (1972): 130-139.

- This is done by starting with the overall odds of acceptance across all programs and multiplying by each program's acceptance rate O/E ratio.
- These odds of acceptance for each program are then converted back to acceptance probabilities as follows: probability=odds/(1 + odds).
- This results in a distribution of acceptance rates among programs with case mix differences removed, to isolate program-to-program differences.
- The sample-weighted variance of these acceptance rates is then calculated, with the weights equal to the number of offers associated with each rate, to obtain $\sigma^2_d$.

✓ The acceptance rate shrinkage weight for each program is calculated as a function of $\sigma^2_r$ and $\sigma^2_d$, as shown in (2) above.

✓ If shrinkage is found to exceed either the upper or lower 95% confidence limit, the value is set to the respective limit, to avoid potentially overshrinking. See "Addition of Limited Translation Rules" above.

✓ Shrunken acceptance rate O/Es are expressed in the probability ratio (or relative risk) scale – *not the odds ratio scale* – for inclusion in the CPM, to parallel mortality and transplant rate O/Es, which are expressed on the hazard ratio scale.

Poisson case (mortality and transplant rates):

✓ Program-specific variance: $\sigma^2_r$ calculated as $(1/K^2)*\lambda$, where
- K is the number of person-years for the specific program
- λ is the overall, national average mortality (or transplant) rate across all programs per person-year, multiplied by K
- $\sigma^2_r$ simplifies to the overall national mortality (or transplant) rate per person year divided by the number of person-years for the specific program.
- This is derived from Poisson model conditional on the (fixed) number of person-years observed for each center.

✓ Program-to-program variance: $\sigma^2_d$ is calculated as follows
- First, the estimated mortality (or transplant) rate for each program *assuming they all had the same, average case-mix of candidates* is computed.
- This is done by starting with the overall mortality (or transplant) rate across all programs and multiplying by each program's mortality (or transplant) rate O/E ratio.
- This results in a distribution of mortality (or transplant) rates among programs with case mix differences removed, to isolate program-to-program differences.
- The sample-weighted variance of these mortality (or transplant) rates is then calculated, with the weights equal to the number of person-years associated with each rate, to obtain $\sigma^2_d$.

✓ The mortality (or transplant) rate shrinkage weight for each program is calculated as a function of $\sigma^2_r$ and $\sigma^2_d$, as shown in (2) above.

✓ If shrinkage is found to exceed either the upper or lower 95% confidence limit, the value is set to the respective limit, to avoid potentially overshrinking. See "Addition of Limited Translation Rules" above.

Instead of adapting the normal-theory BLUP formula to accommodate acceptance rates (binomial) and mortality and transplant rates (Poisson), an alternative would be to simply estimate the program-specific odds ratios or hazard ratios from a random effects estimation in a mixed-effects binary or Cox regression model. This approach, however, was outside the purview and charge given to the OPTN, which was to identify ways to use the already available risk-adjusted

metrics produced by the SRTR contractor to develop a new MPSC tool for pre-transplant performance monitoring. A future enhancement to the CPM could include adoption of this alternate approach, where the shrunken O/E ratios (center effects) come directly as output from a multivariable model estimation procedure.

Another possible future enhancement is to adopt a more explicitly Bayesian approach using either a Gamma or Beta prior. The distribution parameters of the prior would be either selected to achieve the desired shrinkage (as in the recently adopted OPTN Bayesian methods for post-transplant outcomes[23]), or empirically derived in order to maintain the spirit of the approach currently used in the CPM.

**Example CPM Calculation**

Below is an example CPM calculation for a medium-sized liver program WXYZ-TX1.

Pre-transplant metrics for WXYZ-TX1, which had a waitlist size of 100-200 liver patients:

- Waitlist mortality rate O/E = 1.11
- Geography-adjusted transplant rate O/E = 0.64
- Acceptance rate O/E = 0.80

Step 1. Apply logarithmic transformation of O/E ratios (for symmetry)

Mortality rates: ln(1.11) = 0.104
Transplant rates: ln(0.64) = -0.446
Acceptance rates: ln(0.80) = -0.223

Step 2. Apply negative sign for transplant and acceptance rates (directional consistency)

Mortality rates: 0.104 (no change)
Transplant rates: (-1) • (-0.446) = 0.446
Acceptance rates: (-1) • (-0.223) = 0.223

Step 3. Account for statistical uncertainty due to finite sample sizes (Empirical Bayes method)

"Shrunken" mortality rate $O/E_{ln}$ = (0.104) • (mortality shrinkage weight) = (0.104) • (0.796) = 0.083
"Shrunken" transplant rate $O/E_{ln}$ = (0.446) • (transplant shrinkage weight) = (0.446) • (0.920) = 0.410
"Shrunken" acceptance. rate $O/E_{ln}$ = (0.223) • (acc. shrinkage weight) = (0.223) • (0.857) = 0.191

*Shrinkage weights are program-specific and depend on the sample size (number of person-years on the waitlist or number of offers received) as well as the overall program-to-program variability in the specific metric. See "Approximation to Empirical Bayes Method for 'Shrinking' O/E Ratios" section of the appendix for more detail on this methodology.*

*In Step 3, shrinkage is constrained (per "limited translation rules") so as not to extend beyond the 95% confidence limits for the original O/E ratio. If (on the O/E scale) the shrinkage extends*

[23] Salkowski, N., et al. "Bayesian methods for assessing transplant program performance." *American Journal of Transplantation* 14.6 (2014): 1271-1276.

*beyond either the lower or upper limit, then the value is set to the natural logarithm of the respective limit.*

Step 4. Combine into a single metric by applying component weights (composite approach)

$CPM_{log\ scale}$ = (0.50) • (0.083) + (0.25) • (0.410) + (0.25) • (0.191) = 0.192

*For kidney programs, note that the component weights would be 0.00, 0.50, and 0.50 for mortality, transplant, and acceptance rates, respectively.*

Step 5. Apply antilog function (return to O/E scale)

CPM = exp($CPM_{log\ scale}$) = **1.21**

*This program has an aggregate, pre-transplant performance metric 21% higher than expected. CPM values greater than one are generally associated with lower acceptance rates, lower transplant rates, and higher waitlist mortality rates.*

**CPM Calculation: Exclusions and Special Cases**

*Programs excluded from CPM calculation*

Programs with zero person-years for both mortality rate and transplant rate calculations are excluded from CPM calculations. Also excluded are programs that have already withdrawn (either voluntarily or involuntarily) before the start of the one-year cohort period.

*Zero deaths, acceptances, or transplants*

In cases where a program has zero deaths, zero acceptances, or zero transplants, formula (1) above cannot be computed since the logarithm of zero is undefined. In these circumstances, formula (1) is adapted by applying the shrinkage weights on the linear scale instead of the log scale. This leads to a weighted average between 0 and 1, which reduces to (1.0)*($\frac{\sigma_r^2}{(\sigma_D^2+\sigma_r^2)}$) as the estimated shrunken O/E ratio.

*Zero offers received*

If zero offers were received, the acceptance rate O/E is set to 1.0.