Accepted: 14 June 2016

DOI: 10.1111/ctr.12789

ORIGINAL ARTICLE

WILEY Clinical TRANSPLANTATION
The Journal of Clinical and Translational Research

The clinical relevance of Organ Procurement and Transplantation Network screening criteria for program performance review in the United States

Nicholas Salkowski[1] | Andrew Wey[1] | Jon J. Snyder[1,2] | Jeffrey P. Orlowski[3] | Ajay K. Israni[1,2,4] | Bertram L. Kasiske[1,4]

[1]Scientific Registry of Transplant Recipients, Minneapolis Medical Research Foundation, Minneapolis, MN, USA

[2]Department of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

[3]LifeShare Transplant Donor Services of Oklahoma, Oklahoma City, OK, USA

[4]Department of Medicine, Hennepin County Medical Center, University of Minnesota, Minneapolis, MN, USA

Correspondence
Bertram L. Kasiske, MD, Scientific Registry of Transplant Recipients, Minneapolis Medical Research Foundation, Minneapolis, MN, USA.
Email: bkasiske@cdrg.org

Abstract

The Organ Procurement and Transplantation Network is charged with overseeing the quality of transplant programs in the United States. However, there has been controversy over whether too many programs are being identified as underperforming. It has also been suggested that dramatic improvements in outcomes throughout the United States have made the thresholds for determining which deceased donor transplant programs are underperforming no longer clinically relevant. The Scientific Registry of Transplant Recipients compared actual and expected 1-y graft survival for transplant programs identified as underperforming in the most recent cohort (transplants from July 1, 2012 to December 31, 2014). For most organs, actual 1-y graft survival was substantially lower for programs identified as underperforming than for programs identified as performing as expected. Differences were smallest for kidney programs: median 1-y graft survival 89.2% vs 95.4% in large-volume programs identified and not identified for Membership and Professional Standards Committee review, respectively. Median expected graft survival was only slightly lower (94.8% vs 95.1%, respectively), suggesting that identified and not identified programs tend to have similar risk tolerances. An excess of 143 grafts were lost from kidney programs identified as underperforming. Transplant programs identified as underperforming generally have reduced 1-y graft survival that stakeholders may consider clinically relevant.

KEYWORDS

graft survival, mortality, Organ Procurement and Transplantation Network, organ transplantation, outcome assessment

1 | INTRODUCTION

Under the Organ Procurement and Transplantation Network (OPTN) Final Rule, OPTN is charged with the oversight of member transplant programs.[1] The OPTN Membership and Professional Standards Committee (MPSC) is charged with reviewing program performance and compliance. In part, the MPSC uses outcome measures developed by the Scientific Registry of Transplant Recipients (SRTR) as initial screening criteria to identify programs that may be underperforming and may require further MPSC review.

The criteria the MPSC uses to identify transplant programs that may be underperforming have been subject to controversy. There is a perception that too many programs are being flagged as underperforming and that this is causing programs to avoid performing high-risk transplants, thereby limiting access to transplant.[2–6] There is also a perception that improvements in outcomes throughout the US

have resulted in a situation in which differences in outcomes among deceased donor transplant programs identified for review are no longer clinically relevant. On December 1, 2015, the OPTN Board of Directors passed a resolution stating that "the MPSC is tasked over a period of 6 mo to provide the Board with a proposal for an improved program-specific reporting system that identifies substantive clinical differences in patient and graft outcomes." As a result of this resolution, SRTR examined differences in actual (unadjusted) and expected (adjusted) outcomes for kidney, liver, heart, and lung programs identified by the MPSC as underperforming, so that stakeholders could better judge whether the differences in actual outcomes among programs identified as underperforming and performing as expected are clinically relevant.

2 | PATIENTS AND METHODS

2.1 | Data

This study used data from the SRTR. The SRTR data system includes data on all donors, wait-listed candidates, and transplant recipients in the United States, submitted by the members of the OPTN, and has been described elsewhere.[7] The Health Resources and Services Administration, US Department of Health and Human Services, provides oversight of the activities of the OPTN and SRTR contractors.

2.2 | Patient population

All analyses were performed twice with two consecutive SRTR program-specific report (PSR) cohorts used to calculate outcomes after organ transplant in the United States, for reports released to the general public in June 2015 and December 2015. The June 2015 PSR cohort included all adult recipients of kidney, liver, heart, or lung transplants that occurred between January 1, 2012, and June 30, 2014; the December 2015 PSR cohort included recipients of transplants performed between July 1, 2012, and December 31, 2014. The 1-y time lag is necessary to calculate 1-y graft and patient survival. Of note, the two cohorts overlap, but the December 2015 cohort does not include transplants performed between January 1, 2012, and June 30, 2012, and adds transplants performed between July 1, 2014, and December 31, 2014, that were not included in the June 2015 cohort. Pancreas and intestine transplant programs were not included because SRTR does not currently evaluate graft failure for these programs.

2.3 | Analysis

Actual 1-y graft survival and patient survival were determined using the Kaplan—Meier method. Expected 1-y graft survival and patient survival were calculated for all transplants in the United States using the same models used to create the SRTR PSRs, as recently detailed.[8] The current risk adjustment models, which include variables at the time of transplant, are available online (http://www.srtr.org/csr/current/modtabs.aspx).

The MPSC uses flagging thresholds based on hazard ratios and probability distributions determined by a Bayesian methodology.[9,10] The MPSC uses two thresholds to identify underperforming programs for further scrutiny. If a program's hazard ratio exceeds either of these two thresholds, the MPSC will examine the program to determine whether further review or action is needed:

1. The probability is greater than 75% that the hazard ratio is greater than 1.2 or
2. The probability is greater than 10% that the hazard ratio is greater than 2.5.

We used these current MPSC flagging criteria to determine which programs were identified for review. Results for very small transplant programs, as defined by the MPSC (less than 10 transplants during the 2.5-y evaluation period), are shown separately.

All analyses were performed using R version 3.2.2 (https://www.r-project.org/).

3 | RESULTS

3.1 | Graft survival

Differences in actual graft survival between programs identified as underperforming by MPSC criteria ("flagged") and programs not identified as underperforming ("not flagged") varied by organ type and transplant volume (Table 1). Expected 1-y graft survival was slightly lower for kidney, liver, and lung programs, but higher for heart programs (Table 2). Differences between actual and expected graft survival were substantially more (a greater reduction for actual vs expected) for flagged programs than for not flagged programs, and these differences were greater for small-volume programs than for large-volume programs (Table 3). Overall, results were similar for both of the two cohorts examined, and we focus on results for the most recent cohort used to calculate the SRTR PSRs, that is, transplants performed between June 2012 and December 2014.

For large-volume (≥10 transplants) kidney transplant programs, the median actual graft survival was 6.2% lower among 17 flagged programs compared with 184 not flagged programs, 89.2% vs 95.4%, respectively (Table 1). Differences in graft survival between flagged and not flagged large-volume kidney transplant programs were greater for less active programs than for more active programs (Fig. 1). The expected graft survival was virtually identical between flagged and not flagged programs, suggesting similar proportions of high-risk transplants (Table 2). For small-volume (<10 transplants) kidney transplant programs, differences in actual graft survival between four flagged and 25 not flagged programs were even greater than differences for large-volume programs (Table 1). At the four flagged programs, the maximum graft survival was only 75.0%. Flagged kidney programs experienced 143 excess graft losses beyond what was expected from the risk adjustment models.

For large-volume (≥10 transplants) liver transplant programs, the median actual graft survival was 11.3% lower among eight flagged programs compared with 102 not flagged programs, 77.9% vs 89.1% (Table 1). Similar to what was observed for kidney programs, the

TABLE 1 Actual (Kaplan—Meier) 1-year graft survival in two consecutive program-specific report overlapping cohorts

Program	Performance	January 2012–June 2014 30-month cohort			June 2012–December 2014 30-month cohort		
Organ/ volume[a]	Lower performance[b]	N (%)[c]	Median 1-year graft survival (Min.–Max.), %	Difference (Yes—No), %	N (%)[c]	Median 1-year graft survival (Min.–Max.), %	Difference (Yes—No) %
Kidney ≥10	No	185 (91.1)	95.5 (90.0–100.0)	−5.7	184 (91.5)	95.4 (87.5–100.0)	−6.2
	Yes	18 (8.9)	89.8 (77.0–93.9)		17 (8.5)	89.2 (80.0–92.5)	
Liver ≥10	No	101 (91.0)	89.5 (72.9–100.0)	−8.1	102 (92.7)	89.1 (80.0–100.0)	−11.3
	Yes	10 (9.0)	81.4 (70.0–84.5)		8 (7.3)	77.9 (75.0–85.9)	
Heart ≥10	No	91 (91.9)	90.5 (71.4–100.0)	−12.5	91 (91.9)	91.2 (77.3–100.0)	−11.6
	Yes	8 (8.1)	78.0 (66.2–83.5)		8 (8.1)	79.7 (66.0–83.8)	
Lung ≥10	No	49 (90.7)	88.2 (80.9–100.0)	−12.8	49 (90.7)	87.5 (71.3–95.5)	−10.2
	Yes	5 (9.3)	75.4 (53.6–76.6)		5 (9.3)	77.3 (50.0–82.0)	
Kidney <10	No	23 (88.5)	100.0 (83.3–100.0)	−50.0	25 (86.2)	100.0 (75.0–100.0)	−41.7
	Yes	3 (11.5)	50.0 (0.0–50.0)		4 (13.8)	58.3 (0.0–75.0)	
Liver <10	No	12 (100.0)	100.0 (87.5–100.0)	–	15 (93.8)	100.0 (75.0–100.0)	−50.0
	Yes	–	–		1 (6.2)	50.0 (50.0–50.0)	
Heart <10	No	17 (89.5)	100.0 (50.0–100.0)	−75.0	23 (95.8)	100.0 (50.0–100.0)	−33.3
	Yes	2 (10.5)	25.0 (0.0–50.0)		1 (4.2)	66.7 (66.7–66.7)	
Lung <10	No	8 (72.7)	100.0 (83.3–100.0)	−100.0	8 (66.7)	100.0 (66.7–100.0)	−45.0
	Yes	3 (27.3)	0.0 (0.0–55.6)		4 (33.3)	55.0 (0.0–77.8)	

[a]Volume ≥ or <10 transplants over 30 months.
[b]Determined by standard Bayesian Scientific Registry of Transplant Recipients methods and Organ Procurement and Transplantation Network criteria.
[c]Number of programs (percentage of programs).

TABLE 2 Expected 1-year graft survival in two consecutive program-specific report overlapping cohorts

Program	Performance	January 2012–June 2014 30-month cohort			June 2012–December 2014 30-month cohort		
Organ/ volume[a]	Lower performance[b]	N (%)[c]	Median 1-year graft survival (Min.–Max.), %	Difference (Yes—No), %	N (%)[c]	Median 1-year graft survival (Min.–Max.), %	Difference (Yes—No), %
Kidney ≥10	No	185 (91.1)	95.0 (91.9–97.6)	−0.3	184 (91.5)	95.1 (92.7–97.9)	−0.3
	Yes	18 (8.9)	94.7 (92.4–96.2)		17 (8.5)	94.8 (92.4–96.2)	
Liver ≥10	No	101 (91.0)	88.9 (82.7–93.1)	−0.1	102 (92.7)	89.1 (84.3–97.0)	−0.4
	Yes	10 (9.0)	88.8 (85.3–90.5)		8 (7.3)	88.7 (82.5–92.4)	
Heart ≥10	No	91 (91.9)	90.3 (85.4–93.3)	1.7	91 (91.9)	90.6 (86.2–93.1)	0.9
	Yes	8 (8.1)	92.0 (90.3–92.9)		8 (8.1)	91.5 (88.7–93.2)	
Lung ≥10	No	49 (90.7)	87.2 (81.7–90.2)	−1.0	49 (90.7)	87.1 (80.3–90.5)	−0.4
	Yes	5 (9.3)	86.2 (81.2–86.6)		5 (9.3)	86.7 (84.2–87.8)	
Kidney < 10	No	23 (88.5)	97.1 (96.2–98.2)	−1.8	25 (86.2)	97.3 (93.7–98.5)	−1.4
	Yes	3 (11.5)	95.3 (94.3–96.1)		4 (13.8)	95.9 (93.5–96.8)	
Liver <10	No	12 (100.0)	90.8 (86.3–97.9)	–	15 (93.8)	91.7 (84.5–97.9)	−2.4
	Yes	–	–		1 (6.2)	89.3 (89.3–89.3)	
Heart <10	No	17 (89.5)	87.0 (80.2–95.1)	−1.8	23 (95.8)	89.0 (77.4–95.2)	1.2
	Yes	2 (10.5)	85.2 (82.7–87.6)		1 (4.2)	90.2 (90.2–90.2)	
Lung <10	No	8 (72.7)	89.0 (80.1–96.5)	−0.3	8 (66.7)	89.5 (74.6–95.8)	−1.6
	Yes	3 (27.3)	88.7 (87.9–90.9)		4 (33.3)	88.0 (76.8–93.9)	

[a]Volume ≥ or <10 transplants over 30 months.
[b]Determined by standard Bayesian Scientific Registry of Transplant Recipients methods and Organ Procurement and Transplantation Network criteria.
[c]Number of programs (percentage of programs).

differences in graft survival between flagged and not flagged liver transplant programs were greater for less active programs than for more active programs (Fig. 2). Expected graft survival was similar among flagged and not flagged programs (Table 2). The only small-volume (<10 transplants) liver transplant program that was flagged had 50% actual 1-y graft survival, compared with median actual 1-y graft survival of 100% for 15 not flagged programs (Table 1). Flagged liver programs experienced 41 excess graft losses beyond what was expected from the risk adjustment models.

For large-volume (≥10 transplants) heart transplant programs, the median actual graft survival was 11.6% lower among eight flagged programs compared with 91 not flagged programs, 79.7% vs 91.2%, respectively (Table 1, Fig. 3). The expected graft survival for the eight flagged programs was similar to that for the 91 not flagged programs (Table 2). The only small-volume (<10 transplants) heart transplant program that was flagged had 66.7% actual 1-y graft survival, compared with median actual 1-y graft survival of 100% for 23 not flagged programs (Table 1). Flagged heart programs experienced 28 excess

TABLE 3 One-year graft survival differences (actual–expected) in two consecutive program-specific report overlapping cohorts

Program	Performance	January 2012–June 2014 30-month cohort			June 2012–December 2014 30-month cohort		
Organ/ volume[a]	Lower performance[b]	N (%)[c]	Median 1-year graft survival (Min.–Max.), %	Difference (Yes—No), %	N (%)[c]	Median 1-year graft survival (Min.–Max.), %	Difference (Yes—No), %
Kidney ≥10	No	185 (91.1)	0.6 (−4.3 to 5.2)	−5.6	184 (91.5)	0.3 (−7.4 to 5.2)	−5.8
	Yes	18 (8.9)	−5.0 (−17.4 to 2.1)		17 (8.5)	−5.4 (−15.2 to 2.5)	
Liver ≥10	No	101 (91.0)	0.7 (−14.7 to 10.2)	−8.0	102 (92.7)	0.0 (−10.7 to 9.3)	−8.6
	Yes	10 (9.0)	−7.3 (−19.9 to 4.3)		8 (7.3)	−8.7 (−16.3 to 5.0)	
Heart ≥10	No	91 (91.9)	0.7 (−15.9 to 10.6)	−14.8	91 (91.9)	0.8 (−10.7 to 11.6)	−13.3
	Yes	8 (8.1)	−14.1 (−25.9 to 7.3)		8 (8.1)	−12.6 (−25.1 to 6.1)	
Lung ≥10	No	49 (90.7)	0.1 (−5.6 to 12.6)	−10.2	49 (90.7)	0.8 (−11.5 to 10.2)	−11.1
	Yes	5 (9.3)	−10.0 (−32.9 to 6.2)		5 (9.3)	−10.3 (−36.7 to 5.9)	
Kidney <10	No	23 (88.5)	2.8 (−13.8 to 3.8)	−48.9	25 (86.2)	2.6 (−21.3 to 6.3)	−40.8
	Yes	3 (11.5)	−46.1 (−95.3 to 44.3)		4 (13.8)	−38.2 (−93.5 to 20.5)	
Liver <10	No	12 (100.0)	8.3 (−0.7 to 13.7)	–	15 (93.8)	7.7 (−12.8 to 15.5)	−47.0
	Yes	–	–		1 (6.2)	−39.3 (−39.3 to 39.3)	
Heart <10	No	17 (89.5)	4.9 (−34.4 to 19.9)	−65.0	23 (95.8)	6.3 (−37.0 to 18.5)	−29.8
	Yes	2 (10.5)	−60.2 (−82.7 to 37.6)		1 (4.2)	−23.5 (−23.5 to 23.5)	
Lung <10	No	8 (72.7)	7.9 (−4.3 to 19.9)	−95.8	8 (66.7)	7.6 (−8.0 to 19.4)	−33.4
	Yes	3 (27.3)	−87.9 (−90.9 to 33.1)		4 (33.3)	−25.8 (−93.9 to 13.2)	

[a]Volume ≥ or <10 transplants over 30 months.
[b]Determined by standard Bayesian Scientific Registry of Transplant Recipients methods and Organ Procurement and Transplantation Network criteria.
[c]Number of programs (percentage of programs).

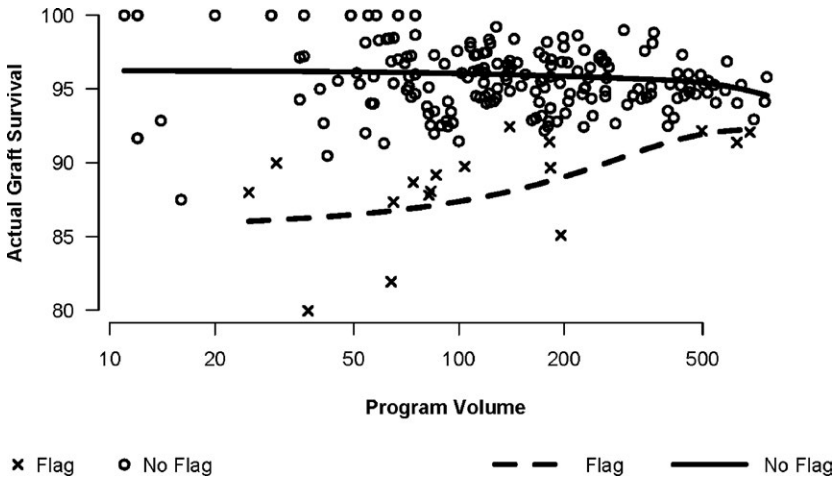

FIGURE 1 Comparison of actual 1-year graft survival (*y*-axis) for kidney transplant programs identified as performing as expected (circles) or underperforming (x's) by program volume, that is, the number of transplants in the 30-month cohort July 2012–December 2014 (*x*-axis)

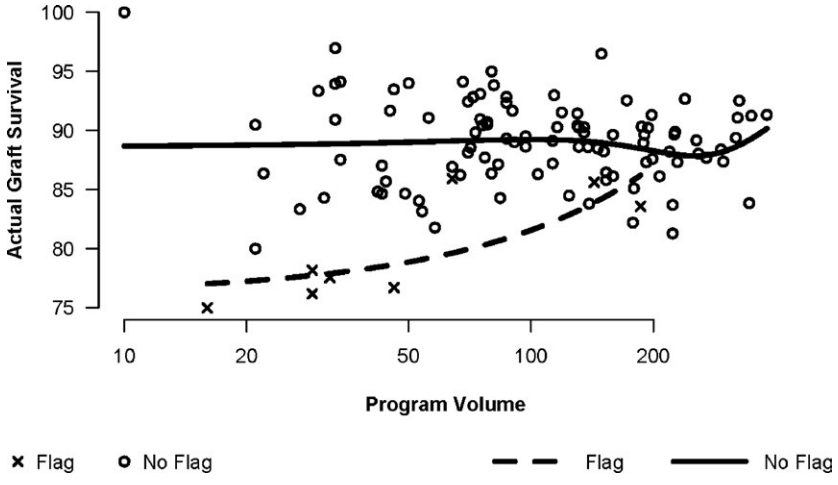

FIGURE 2 Comparison of actual 1-year graft survival (*y*-axis) for liver transplant programs identified as performing as expected (circles) or underperforming (x's) by program volume, that is, the number of transplants in the 30-month cohort July 2012–December 2014 (*x*-axis)

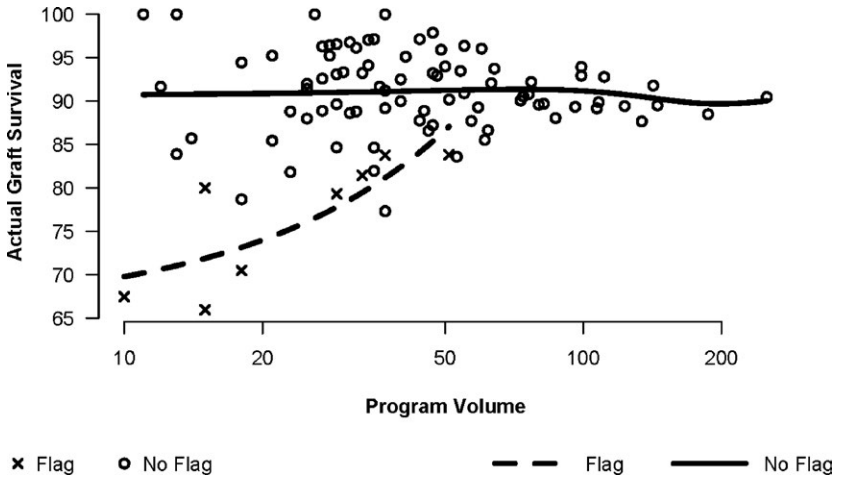

FIGURE 3 Comparison of actual 1-year graft survival (*y*-axis) for heart transplant programs identified as performing as expected (circles) or underperforming (x's) by program volume, that is, the number of transplants in the 30-month cohort July 2012–December 2014 (*x*-axis)

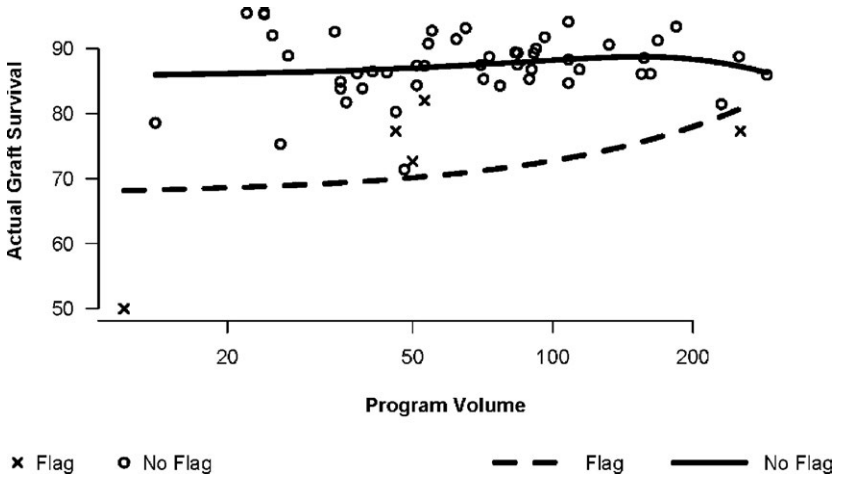

FIGURE 4 Comparison of actual 1-year graft survival (*y*-axis) for lung transplant programs identified as performing as expected (circles) or underperforming (x's) by program volume, that is, the number of transplants in the 30-month cohort July 2012–December 2014 (*x*-axis)

graft losses beyond what was expected from the risk adjustment models.

For large-volume (≥10 transplants) lung transplant programs, the median actual graft survival was 10.2% lower among five flagged programs compared with 49 not flagged programs, 77.3% vs 87.5%, respectively (Table 1, Fig. 4). The expected graft survival of the five flagged programs was similar to that of the 49 not flagged programs (Table 2). For small-volume (<10 transplants) lung transplant programs, maximum 1-y graft survival among the four flagged programs was only 77.8%, substantially lower than graft survival among the eight not flagged programs (Table 1). Flagged lung programs experienced 43 excess graft losses beyond what was expected from the risk adjustment models.

3.2 | Patient survival

We also examined patient survival at 1 y post-transplant. Not surprisingly, as most graft failure results in death, patient survival after heart and lung transplant was virtually identical to graft survival (Tables 4, 5, and 6). For large-volume (≥10 transplants) kidney transplant programs, the median actual patient survival was 3.3% lower for 19 flagged programs compared with 180 not flagged programs, 94.4% vs 97.7%, respectively (Table 4). Expected patient survival for flagged and not flagged kidney transplant programs was virtually identical (Table 5). For large-volume (≥10 transplants) liver transplant programs, the median actual patient survival was 6.5% lower among 11 flagged programs compared with 99 not flagged programs, 84.7% vs 91.2% (Table 4). Expected patient survival was only slightly higher for flagged than for not flagged liver programs (Table 5).

4 | DISCUSSION

It has recently been suggested that the threshold criteria used by the OPTN MPSC to identify underperforming transplant programs for further scrutiny may be identifying programs without meaningful or clinically relevant differences in graft or patient survival. These and other concerns led to an OPTN Board of Directors resolution calling for improving the program-specific reporting by identifying substantive clinical differences in outcomes. We examined the differences in outcomes for programs that were and were not identified as underperforming by the current system. We used the two most recent

TABLE 4 Actual (Kaplan—Meier) 1-year patient survival in two consecutive program-specific report overlapping cohorts

Program	Performance	January 2012–June 2014 30-month cohort			June 2012–December 2014 30-month cohort		
Organ/ volume[a]	Lower performance[b]	N (%)[c]	Median 1-year patient survival (Min.–Max.), %	Difference (Yes—No), %	N (%)[c]	Median 1-year patient survival (Min.–Max.), %	Difference (Yes—No), %
Kidney ≥10	No	179 (89.5)	97.8 (93.5–100.0)	−3.4	180 (90.5)	97.7 (93.5–100.0)	−3.3
	Yes	21 (10.5)	94.4 (88.9–96.1)		19 (9.5)	94.4 (88.2–95.7)	
Liver ≥10	No	102 (92.7)	91.1 (80.5–100.0)	−8.3	99 (90.0)	91.2 (81.0–100.0)	−6.5
	Yes	8 (7.3)	82.8 (77.0–88.6)		11 (10.0)	84.7 (75.0–90.8)	
Heart ≥10	No	90 (90.9)	90.5 (79.3–100.0)	−14.6	91 (92.9)	91.8 (78.7–100.0)	−12.5
	Yes	9 (9.1)	76.0 (66.2–83.8)		7 (7.1)	79.3 (66.0–85.7)	
Lung ≥10	No	51 (96.2)	87.6 (77.7–100.0)	−23.9	47 (87.0)	88.5 (74.3–96.2)	−10.3
	Yes	2 (3.8)	63.7 (53.6–73.8)		7 (13.0)	78.2 (50.0–84.3)	
Kidney <10	No	27 (96.4)	100.0 (100.0–100.0)	−50.0	30 (96.8)	100.0 (100.0–100.0)	−100.0
	Yes	1 (3.6)	50.0 (50.0–50.0)		1 (3.2)	0.0 (0.0–0.0)	
Liver <10	No	12 (100.0)	100.0 (77.8–100.0)	–	15 (100.0)	100.0 (66.7–100.0)	–
	Yes	–	–		–	–	
Heart <10	No	16 (88.9)	100.0 (50.0–100.0)	−75.0	19 (82.6)	100.0 (50.0–100.0)	−43.3
	Yes	2 (11.1)	25.0 (0.0–50.0)		4 (17.4)	56.8 (0.0–66.7)	
Lung <10	No	8 (72.7)	100.0 (83.3–100.0)	−100.0	8 (66.7)	100.0 (66.7–100.0)	−70.0
	Yes	3 (27.3)	0.0 (0.0–55.6)		4 (33.3)	30.0 (0.0–77.8)	

[a]Volume ≥ or <10 transplants over 30 months.
[b]Determined by standard Bayesian Scientific Registry of Transplant Recipients methods and Organ Procurement and Transplantation Network criteria.
[c]Number of programs (percentage of programs).

TABLE 5 Expected 1-year patient survival in two consecutive program-specific report overlapping cohorts

Program	Performance	January 2012–June 2014 30-month cohort			June 2012–December 2014 30-month cohort		
Organ/ volume[a]	Lower performance[b]	N (%)[c]	Median 1-year patient survival (Min.–Max.), %	Difference (Yes—No), %	N (%)[c]	Median 1-year patient survival (Min.–Max.), %	Difference (Yes—No), %
Kidney ≥10	No	179 (89.5)	97.4 (96.0–98.5)	0.1	180 (90.5)	97.5 (95.9–99.3)	0.0
	Yes	21 (10.5)	97.5 (96.3–98.0)		19 (9.5)	97.4 (96.3–98.2)	
Liver ≥10	No	102 (92.7)	91.0 (85.8–94.8)	0.5	99 (90.0)	91.1 (87.0–98.0)	1.1
	Yes	8 (7.3)	91.5 (88.0–92.9)		11 (10.0)	92.2 (84.8–94.1)	
Heart ≥10	No	90 (90.9)	90.6 (84.7–93.1)	0.8	91 (92.9)	90.9 (86.5–93.1)	0.5
	Yes	9 (9.1)	91.4 (87.4–93.4)		7 (7.1)	91.3 (89.0–93.5)	
Lung ≥10	No	51 (96.2)	87.7 (82.7–90.9)	−2.1	47 (87.0)	87.9 (81.8–91.2)	
	Yes	2 (3.8)	85.6 (84.1–87.2)		7 (13.0)	87.5 (86.0–89.6)	−0.4
Kidney <10	No	27 (96.4)	99.5 (99.0–99.7)	−3.4	30 (96.8)	99.5 (93.6–99.6)	−3.9
	Yes	1 (3.6)	96.1 (96.1–96.1)		1 (3.2)	95.5 (95.5–95.5)	
Liver <10	No	12 (100.0)	95.4 (87.9–98.9)	–	15 (100.0)	95.5 (87.0–98.5)	–
	Yes	–	–		–	–	
Heart <10	No	16 (88.9)	89.1 (79.9–93.9)	−4.2	19 (82.6)	89.0 (80.8–95.3)	0.9
	Yes	2 (11.1)	84.9 (80.8–89.0)		4 (17.4)	89.9 (69.1–90.1)	
Lung <10	No	8 (72.7)	90.2 (87.3–97.2)	1.3	8 (66.7)	89.5 (82.4–96.1)	1.5
	Yes	3 (27.3)	91.5 (89.2–93.9)		4 (33.3)	91.0 (85.3–92.3)	

[a]Volume ≥ or <10 transplants over 30 months.
[b]Determined by standard Bayesian Scientific Registry of Transplant Recipients methods and Organ Procurement and Transplantation Network criteria.
[c]Number of programs (percentage of programs).

cohorts and the same SRTR (Bayesian) methodologies used to provide information to the MPSC.

Not surprisingly, the smallest differences between programs that were identified as underperforming and those identified as performing as expected were for kidney programs, where expected 1-y graft survival is currently 95.4%. Even at this high level of overall graft survival, programs that were identified as underperforming had median graft survival of only 89.2%, that is, 6.2% less than programs not identified as underperforming. Although 89.2% is arguably a good clinical outcome, this 6% difference was associated with 143 excess graft losses. Differences for liver, heart, and lung transplant programs were greater, with median graft survival over 10% lower among flagged programs.

The differences in actual 1-y graft survival between programs identified as underperforming and those identified as performing as expected are proportionally less for programs with very large volume than for programs performing fewer transplants (Figs. 1–4). However,

TABLE 6 One-year patient survival differences (actual–expected) in two consecutive program-specific report overlapping cohorts

Program	Performance	January 2012–June 2014 30-month cohort			June 2012–December 2014 30-month cohort		
Organ/ volume[a]	Lower performance[b]	N (%)[c]	Median 1-year patient survival (Min.–Max.), %	Difference (Yes—No), %	N (%)[c]	Median 1-year patient survival (Min.–Max.), %	Difference (Yes—No), %
Kidney ≥10	No	179 (89.5)	0.4 (−3.5 to 3.4)	−3.5	180 (90.5)	0.5 (−3.7 to 4.0)	
	Yes	21 (10.5)	−3.0 (−8.5 to 1.4)		19 (9.5)	−3.1 (−8.2 to 1.5)	−3.6
Liver ≥10	No	102 (92.7)	−0.1 (−8.1 to 8.0)	−7.6	99 (90.0)	0.3 (−9.9 to 7.3)	−6.9
	Yes	8 (7.3)	−7.7 (−14.7 to 4.2)		11 (10.0)	−6.6 (−18.0 to 2.4)	
Heart ≥10	No	90 (90.9)	0.6 (−11.2 to 10.6)	−17.0	91 (92.9)	1.0 (−10.0 to 11.6)	−14.1
	Yes	9 (9.1)	−16.4 (−25.2 to 7.6)		7 (7.1)	−13.0 (−25.3 to 6.5)	
Lung ≥10	No	51 (96.2)	−0.1 (−9.5 to 11.8)	−21.9	47 (87.0)	0.6 (−10.1 to 10.5)	−11.4
	Yes	2 (3.8)	−21.9 (−33.6 to 10.3)		7 (13.0)	−10.8 (−37.0 to 4.4)	
Kidney <10	No	27 (96.4)	0.5 (0.3 to 1.0)	−46.6	30 (96.8)	0.5 (0.4 to 6.5)	−96.1
	Yes	1 (3.6)	−46.1 (−46.1 to 46.1)		1 (3.2)	−95.5 (−95.5 to 95.5)	
Liver <10	No	12 (100.0)	2.1 (−12.9 to 12.1)	–	15 (100.0)	2.6 (−23.6 to 3.0)	–
	Yes	–	–		–	–	
Heart <10	No	16 (88.9)	6.2 (−34.8 to 20.1)	−66.0	19 (82.6)	5.3 (−38.3–19.3)	−38.4
	Yes	2 (11.1)	−59.9 (−80.8 to 39.0)		4 (17.4)	−33.2 (−69.1 to 23.3)	
Lung <10	No	8 (72.7)	3.8 (−3.9 to 12.7)	−95.3	8 (66.7)	6.5 (−19.4 to 17.6)	−64.8
	Yes	3 (27.3)	−91.5 (−93.9 to 33.7)		4 (33.3)	−58.3 (−92.3 to 13.0)	

[a]Volume ≥ or <10 transplants over 30 months.
[b]Determined by standard Bayesian Scientific Registry of Transplant Recipients methods and Organ Procurement and Transplantation Network criteria.
[c]Number of programs (percentage of programs).

this does not necessarily mean that the clinical relevance of the smaller differences is also less; a 4% difference affecting twice as many patients may be equally as important as an 8% difference affecting half as many patients. Moreover, it has recently been argued that the remarkable improvement in graft survival after solid organ transplant makes it difficult to identify underperforming programs. However, all outcomes are relative, and the fact that excellent graft survival is expected does not preclude underperformance and the potential benefit of a rigorous surveillance program to improve performance and outcomes.

We examined differences between actual and expected 1-y graft survival to assess the clinical relevance of the current MPSC system for identifying underperforming programs. However, absolute differences between actual and expected graft survival depend on the risk level of grafts and recipients at each program. Programs that perform fewer "high-risk" transplants are more likely to have smaller absolute differences between actual and expected graft survival than programs that perform more "high-risk" transplants, even if the risk-adjusted hazard ratios used by the MPSC for assessing program performance are similar. For example, a 1% difference in actual and expected graft survival is more likely to occur at a program that only performs deceased donor transplants than at a program that only performs living donor transplants. Thus, an evaluation system based on differences between actual and expected patient and graft survival could encourage greater risk aversion.

We also compared expected outcomes in flagged and not flagged programs. The extent to which expected outcomes in flagged programs were greater than those in not flagged programs would suggest that transplants in flagged programs were less risky than transplants in not flagged programs. Similarly, the extent to which expected outcomes in flagged programs were less than those in not flagged programs would suggest that transplants in flagged program were more risky than transplants in not flagged programs. For kidney, liver, heart, and lung programs, expected 1-y graft and patient survival were very similar for flagged and not flagged programs, suggesting that flagged programs were performing transplants in patients and with organs of similar risk to not flagged programs (Tables 2 and 4). Thus, the general perception that performing higher risk transplants leads to a greater risk of being identified as underperforming is not substantiated by these results, at least to the extent that "risk" is defined based on measured risk factors accounted for by the risk adjustment models.

Differences between programs identified as underperforming in the system currently used by the MPSC depend on both outcomes (1-y patient and graft survival) and number of transplants performed. Statistical certainty in the outcomes is greater at programs performing more transplants than at programs performing fewer. As a result, a smaller absolute difference between observed and expected outcomes at a large-volume program may be identified as underperforming, while a larger absolute difference at a small-volume program is not identified as underperforming.

Although any definition of "clinical relevance" is arguably subjective, the results of this analysis demonstrate the differences in observed and expected survival for flagged vs not flagged programs. In the context of continuous quality improvement, identifying programs experiencing worse outcomes than those achieved at other programs is a worthy goal, even in the context of generally high graft and patient survival rates. Even so, the system for assessing program performance can and should be continuously evaluated and improved. For example, the risk prediction models SRTR uses might be improved by collecting additional clinical variables that predict risk. The addition of such variables might help allay any aversion to risk that may interfere with transplant programs' willingness to accept high-risk donors and

high-risk recipients, thereby improving access to transplant. OPTN's Data Advisory Committee has been tasked with such efforts to assess and improve the OPTN data. Similarly, efforts to educate transplant programs regarding the effectiveness of SRTR models in adjusting for risk could help reassure programs that they will not be penalized by performing high-risk transplants that are adjusted for in the SRTR models. Finally, if the MPSC determines that the current thresholds for defining underperformance are not meeting the current goals of the monitoring system, then SRTR could modify the thresholds to meet the stated goals. However one assesses the concept of clinical relevance or meaningful differences in clinical outcomes, any monitoring system should continue to make use of risk-adjusted analyses such that perceived disincentives to accepting higher risk donors or recipients can be minimized.

ACKNOWLEDGEMENTS

The authors thank SRTR colleague Nan Booth, MSW, MPH, ELS, for manuscript editing. This work was conducted under the support of the Minneapolis Medical Research Foundation, contractor for the Scientific Registry of Transplant Recipients, as a deliverable under contract no. HHSH250201500009C (US Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau, Division of Transplantation). As a US Government-sponsored work, there are no restrictions on its use. The views expressed herein are those of the authors and not necessarily those of the US Government. AI was partially supported by R01 HS 24527.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

Nicholas Salkowski, Andrew Wey, Jon J. Snyder, Ajay K. Israni, and Bertram L. Kasiske participated in study design, data acquisition, data analysis, interpretation, and writing of the manuscript. Jeffrey P. Orlowski participated in interpretation and writing of the manuscript.

REFERENCES

1. Department of Health and Human Services 42 CFR Part 12. Organ Procurement and Transplantation Network; Final Rule. http://www.gao.gov/special.pubs/organ/appendd.pdf. Accessed April 18, 2016.
2. Abecassis MM, Burke R, Klintmalm GB, et al. American Society of Transplant Surgeons transplant center outcomes requirements–a threat to innovation. *Am J Transplant*. 2009;9:1279.
3. Howard RJ, Cornell DL, Schold JD. CMS oversight, OPOs and transplant centers and the law of unintended consequences. *Clin Transplant*. 2009;23:778.
4. 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.
5. 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.
6. Kasiske BL, McBride MA, Cornell DL, et al. Report of a consensus conference on transplant program quality and surveillance. *Am J Transplant*. 2012;12:1988.
7. Leppke S, Leighton T, Zaun D, et al. Scientific Registry of Transplant Recipients: collecting, analyzing, and reporting data on transplantation in the United States. *Transplant Rev (Orlando)*. 2013;27:50.
8. Snyder JJ, Salkowski N, Kim SJ, et al. Developing statistical models to assess transplant outcomes using national registries: the process in the United States. *Transplantation*. 2016;100:288.
9. Salkowski N, Snyder JJ, Zaun DA, Leighton T, Israni AK, Kasiske BL. Bayesian methods for assessing transplant program performance. *Am J Transplant*. 2014;14:1271.
10. Salkowski N, Snyder JJ, Zaun DA, et al. A scientific registry of transplant recipients Bayesian method for identifying underperforming transplant programs. *Am J Transplant*. 2014;14:1310.

How to cite this article: Salkowski, N., Wey, A., Snyder, J.J., Orlowski, J.P., Israni, A.K., Kasiske, B.L. (2016), The clinical relevance of Organ Procurement and Transplantation Network screening criteria for program performance review in the United States. Clinical Transplantation, 30: 1066–1073. doi: 10.1111/ctr.12789