American Journal of Transplantation xxx (xxxx) xxx

Contents lists available at ScienceDirect

American Journal of Transplantation

journal homepage: www.amjtransplant.org

Brief Communication

Association of organ procurement organization volume with Centers for Medicare and Medicaid Services performance evaluations

Rocio Lopez [1,2], Sumit Mohan [3,4,5], James R. Rodrigue [6,7,8,9], Susana Arrigain [1,2], Deena Brosi [1,2], Ryan Lavanchy [1,2], Bruce Kaplan [1,2], Elizabeth A. Pomfret [1,2], Jesse D. Schold [1,2,10,*]

[1] Division of Transplant Surgery, Department of Surgery, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA
[2] Colorado Center for Transplantation Care (CCTCARE), Research and Education, Division of Transplant Surgery, Department of Surgery, Anschutz Medical Campus, University of Colorado, Aurora, Colorado, USA
[3] Division of Nephrology, Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, USA
[4] Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA
[5] Columbia University Renal Epidemiology Group, New York, New York, USA
[6] Transplant Institute, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[7] Department of Surgery, Harvard Medical School, Boston, Massachusetts, USA
[8] Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA
[9] New England Donor Services Inc, Waltham, Massachusetts, USA
[10] Department of Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA

ARTICLE INFO

Keywords:
transplantation
Organ Procurement Organizations
policy
quality metrics
Centers for Medicare and Medicaid Services

ABSTRACT

Under 2020 Centers for Medicare and Medicaid Services (CMS) conditions of coverage, Organ Procurement Organizations (OPOs) will be decertified if their 95% upper confidence limit for donation or transplant rate falls below the previous year's median (tier 3) and must recompete if either is below the 75th percentile (tier 2). This study aimed to examine the associations of CMS metrics with OPO volume and evaluate an alternate observed-to-expected tiering system using simulation analysis and CMS's OPO public report. In 2021, CMS tier 3 and 2 OPOs had significantly larger volumes than tier 1 OPOs (median = 2042 vs 2124 vs 1003; $P = .028$). In a simulation scenario in which OPOs should be CMS tier 2, large OPOs had 95% probability of needing to recompete vs 26% for the smallest OPOs. The observed-to-expected method misclassified OPOs as underperforming ~5% of simulated cases independent of volume. CMS methodology assigned a worse tier than

Abbreviations: CALC, cause, age, location consistent; CI, confidence interval; CMS, Centers for Medicare and Medicaid Services; DDP, deceased donor potential; DSA, donation service area; OPO, Organ Procurement Organization; OPTN, Organ Procurement and Transplantation Network; O-to-E, observed-to-expected; RR, rate ratio; SRTR, Scientific Registry of Transplant Recipients; UCL, upper confidence limit.

* Corresponding author. 1635 Aurora Court, Floor 7, Room 7085, C-318, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA.
E-mail address: jesse.schold@cuanschutz.edu (J.D. Schold).

https://doi.org/10.1016/j.ajt.2024.11.024
Received 6 August 2024; Received in revised form 21 November 2024; Accepted 23 November 2024
Available online xxxx

observed-to-expected to 24%-54% of OPOs across years. Results indicate that the current CMS methodology systematically identifies larger OPOs as underperforming and independent of quality, suggesting alternative statistical evaluations are needed to assess OPO performance accurately and improve donation processes of care and transplant rates.

1. Introduction

In December 2020, the Centers for Medicare and Medicaid Services (CMS) issued a final rule updating the conditions of coverage for Organ Procurement Organizations (OPOs).[1] This rule, effective from August 1, 2022, measures OPO performance and will significantly impact the viability of OPOs. Under this final rule, OPOs will be assessed yearly and undergo certification or decertification every 4 years based on performance data from a single year. As stated in policy, OPOs will be decertified if the upper limit of the 1-sided 95% confidence limit (UCL) for their donation or transplant rate is below the previous year's median (tier 3). OPOs with either of the 95% UCLs below the previous year's 75th percentile must compete to keep their donation service area (DSA) (tier 2); they can also apply and compete for any other open DSA. OPOs with both 95% UCLs above the previous year's 75th percentile will be recertified, have exclusive rights to their DSA, and be able to compete for any other open DSA (tier 1). The most recently released report from CMS indicates 24(42%) OPOs were classified as tier 3, 18 (32%) were tier 2, and 15 (26%) were tier 1.[2] If this performance under CMS metrics remains stable, up to 74% of OPOs will either be decertified or could lose their DSA through competition. This would vastly change the landscape of organ donation services in the United States.

As of July 2024, 56 OPOs of various sizes (measured by population served rather than geographic size) operate in the United States.[3] During the public comment period, a concern was expressed to CMS regarding the potential problems associated with using a 95% UCL, given potential bias against larger OPOs. In response, they stated as follows:

> "The purpose of the confidence interval (CI) was to ensure that the use of the threshold rate does not bias against small OPOs who may be prone to greater variability of rates due to smaller volumes. We do not concur with the commenters' assertion that our methodology is biased against large OPOs; they have a CI generated, but because they have more data, their CIs are proportionally smaller."[1]

A 1-sided 95% UCL provides an upper boundary below which the true population parameter would be with 95% certainty.[4,5] This limit is calculated based on the sample data and considers the inherent variability in the sample. The width of any CI is influenced by sample size (in this case, number of potential donors). Smaller sample sizes result in wider intervals because they provide less information about the population, leading to greater uncertainty and variability. In contrast, larger sample sizes yield narrower intervals and more precise estimates. Thus, with the current OPO performance metrics using fixed thresholds to define tiers, it is plausible that larger OPOs are less likely to fall into a specific tier, independent of actual performance. Therefore, we aimed to examine the association of CMS metrics with OPO volume using simulation models and empirical evidence over recent years and evaluate an alternate tiering system that may be less influenced by volume.

2. Methods

2.1. Data description

We used data from the CMS' 2023 OPO Interim Annual Public Aggregated Report, which provides donation and transplant rates and tier rankings for OPOs from 2019 to 2021.[2] This does not provide the cause, age, and location consistent (CALC) deceased donor potential (DDP) for each OPO nor the number of donors or organs transplanted. CALC DDP represents OPO volume and is the denominator for calculating donation and transplant rates.[2,6]

This study used data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donors, waitlisted candidates, and transplant recipients in the United States, submitted by the members of the Organ Procurement and Transplantation Network (OPTN). The Health Resources and Services Administration, US Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors. We used standard analysis files from March 2024 to calculate each OPO's yearly donor and organ count and calculated DDP based on reported CMS donation rate. We used OPO volume and DDP interchangeably to indicate each OPO's CALC DDP, referring specifically to the estimated number of deaths qualifying as potential donors, rather than geographic area or population size.

2.2. Statistical analysis

We conducted simulations to evaluate CMS performance under various scenarios, using donation rate as a primary example. Six scenarios were modeled to reflect varying underlying rates across OPOs, ranging from across-the-board improvements to declines and variations around key percentile thresholds. An in-depth explanation of the simulation process, including specific rate assumptions, is found in Supplementary Methods. We followed CMS methodology to determine tier rankings.[6] The User Guide for the OPO Annual Public Aggregated Performance Report outlines definitions, rate formulas, 1-sided 95% UCLs, and tier assignment.[6] We plotted the percentage of simulations falling into each tier by DDP and overall.

We evaluated an alternate tiering system using the observed-to-expected (O-to-E) rate ratios (RRs) approach from SRTR's transplant program assessments.[7] Using a Bayesian framework,

we estimated RR as the posterior mean as follows: (observed donors + 2)/(expected donors + 2). The RR's posterior distribution is a γ distribution with a shape parameter equal to the observed number of donors plus 2 and a rate parameter equal to the expected number of donors plus 2. OPOs were assigned a performance tier based on the probability that the RR is below critical thresholds (tier 3 if probability [RR $<$ threshold$_3$] $>$ probability$_3$); tier 2 if probability [RR $<$ threshold$_2$] $>$ probability$_2$); otherwise tier 1). We evaluated threshold values from 0.5 to 1.1 and probability values from 0.25 to 0.95 in 0.05 increments and restricted to rules with threshold$_2$ of $\geq$0.9 and threshold$_3$ of $<$0.9, resulting in a grid search of over 9000 possible O-to-E tiering systems. Using donation rate as an example, we used 2021 DDPs and simulated 100 000 evaluations under 3 scenarios (RR = 1, RR = 0.9, and RR = 0.8) and applied each of the 9000 potential rules. We narrowed the rules to those with a type I error rate of $<$0.03 across DDPs and selected the rule with equal probability for both tiers. We used 100 000 simulated evaluations, as detailed in Supplementary Methods, to compare the optimal O-to-E system with CMS's system by graphing the percentage in each tier by DDP and overall.

We applied the optimal O-to-E system to OPO donation and age-adjusted organ transplant RRs from 2019 to 2021 and determined an annual O-to-E tier based on these. We graphed the yearly DDP distribution across OPOs to display the range of OPO volumes. We summarized annual CMS and O-to-E tier rankings and reported the concordance rate between the 2 methods. We used Wilcoxon rank sum tests to compare the DDP for OPOs with discordant tiers between the 2 methods. We plotted DDP by CMS annual tier ranking and used Kruskal-Wallis tests to compare between tiers. Analyses were performed at a 0.05 significance level using SAS 9.4 (SAS Institute, Cary, NC).

3. Results

3.1. Evaluating CMS performance tiers based on OPO volume

As expected under simulation scenarios 1-3, where all OPOs have the same underlying rate, smaller DDPs show a wider 95% UCL spread, signifying less certainty in the rate estimate (Supplementary Fig. S1). In scenario 1, all OPOs should be tier 1, but 5% were misclassified as tier 2 or 3, consistent with the function of a 95% CI to preserve a type 1 error rate of 5% (Supplementary Fig. 2). This misclassification persisted across DDPs (Fig. 1A). In scenario 2, where all OPOs should be tier 2, 5% were incorrectly classified as tier 3 across DDPs (Supplementary Fig. 2). However, smaller OPOs in scenario 2 were less likely to be placed in tier 2, with 26% probability of having to recompete compared to 95% for the largest OPOs (Fig. 1B). In scenario 2, 17% of OPOs across DDPs were placed in tier 1, with smaller OPOs having 69% probability of being recertified compared with $<$1% of larger OPOs. In scenario 3, where all OPOs should be tier 3, 31% were classified as tier 1 or 2 across DDPs (Supplementary Fig. 2). However, in scenario 3, 33% of the smallest OPOs would be automatically recertified, and 46% would have to recompete. In comparison, 0% of the largest OPOs would be automatically recertified, and 6% would have to recompete (Fig. 1C).

When rates varied around the 75th percentile and median (scenarios 4 and 5), larger OPOs were more likely to be tier 2 or 3 than smaller OPOs despite having the same distribution of underlying donation rates (Fig. 1D, E). In scenario 6, where rates varied around the 25th percentile, smaller OPOs were more likely to be tier 1 or 2 (Fig. 1F). Overall, larger OPOs were more likely to be classified as tier 2 or 3 than smaller OPOs using the same assumptions.

3.2. Observed-to-expected tiering system

The optimal O-to-E tiering system placed an OPO in tier 3 if the donation rate was at least 15% worse than expected with greater than 95% probability (probability [RR $<$ 0.85] $>$95%), tier 2 if the donation rate was lower than expected with greater than 95% probability (probability [RR $<$ 1] $>$95%), and tier 1 otherwise.

In scenario 7, where OPOs should be CMS tier 2 and O-to-E tier 1, CMS methodology incorrectly placed 31% of OPOs in tier 1 and 4% in tier 3, while O-to-E methodology incorrectly placed 4% of OPOs in tier 2 and $<$0.05% in tier 3 across DDPs(Supplementary Fig. S3A, B). CMS methodology inaccurately categorized the smallest OPOs 71% of the time as tier 1 and rarely categorized the largest OPOs as tier 1 ($<$1%) (Supplementary Fig. S4A). O-to-E methodology misclassified OPOs as underperforming approximately 5% of the time, regardless of OPO volume (Supplementary Fig. S4B). In scenarios 8 and 9, when an OPO was truly underperforming, CMS and O-to-E methodologies were better at identifying underperformance in larger OPOs due to higher statistical power. The error rates were similar between the 2 methods in scenario 8 (Supplementary Fig. S4C, D). However, in scenario 9, O-to-E methodology incorrectly placed larger OPOs 57% of the time into tier 2, compared with only 40% by CMS methodology (Supplementary Figs. S3F and S4E). Overall, the O-to-E tiering system misclassified OPOs less frequently than current CMS methodology and significantly attenuated systematic biases for misclassification by volume.

3.3. Real-world data—OPO volume and performance tiers

The number of OPOs changed from 58 in 2019-2020 to 57 in 2021, although OPO volume increased with median DDP rising from 1514 (P25-P75: 902-2139) in 2019 to 1815 (P25-P75: 974-2467) in 2021 (Supplementary Fig. S5). The percentage of OPOs in CMS tier 3 increased from 28% in 2019 to 38% in 2020 and 42% in 2021 (Table). In contrast, 16%, 17%, and 12% of OPOs would be classified as O-to-E tier 3 in 2019, 2020, and 2021, respectively. DDP did not show a significant association with CMS tier ranking in 2019. However, in 2020 and 2021, CMS tier 3 and 2 OPOs were larger than CMS tier 1 OPOs (median: 1831 vs 2005 vs 952; $P = .097$, and 2042 vs 2124 vs 1003, $P = .028$, respectively) (Supplementary Fig. 6). DDP was not significantly associated with O-to-E tier ranking.

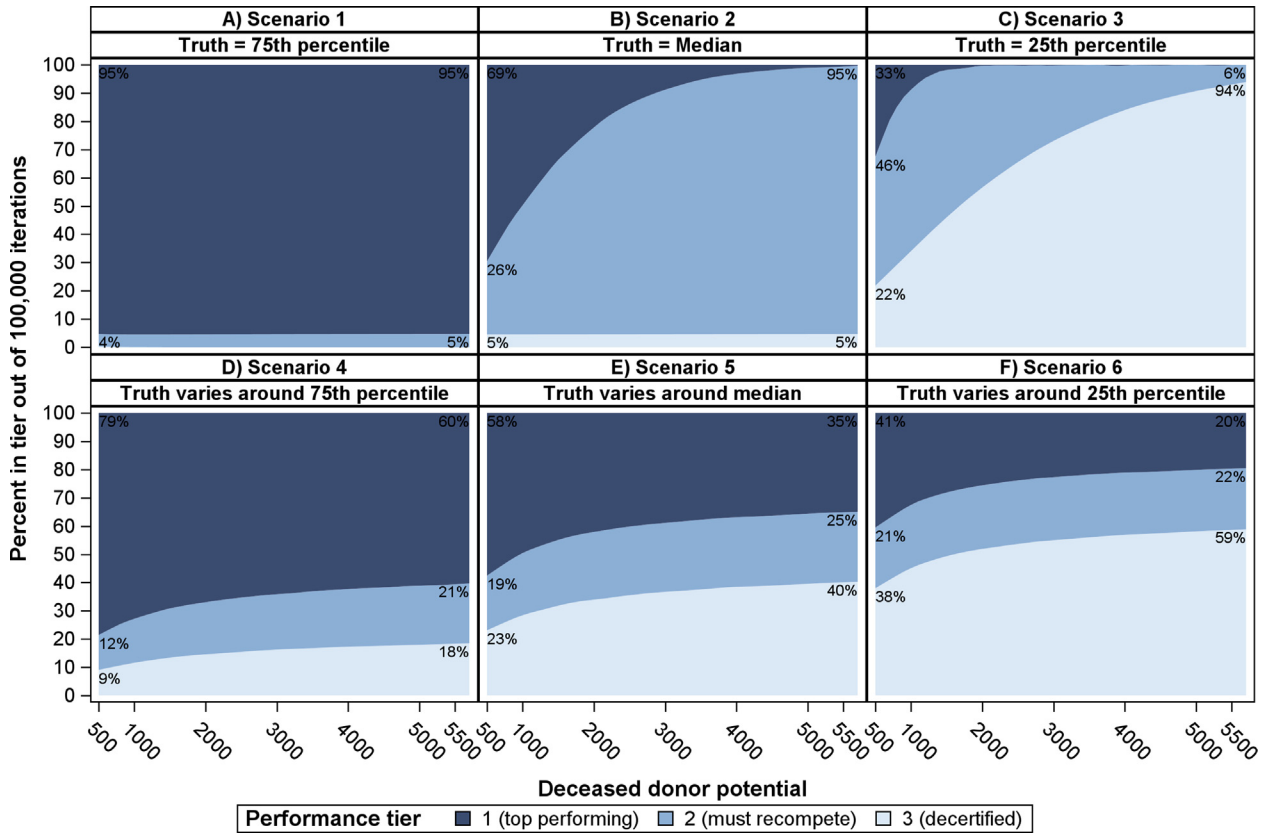

Figure 1. Percentage of simulations placed in each CMS performance tier by deceased donor potential for 6 simulation scenarios. For each deceased donor potential value, we simulated 100 000 CMS evaluations under each scenario and display the percentage of the simulations that fall into CMS tiers 1, 2, and 3. Scenario 1 (Figure 1a)—all OPOs have the same underlying rate: the previous year's 75th percentile (across-the-board improvement); scenario 2 (Figure 1b)—all OPOs have the same underlying rate: the previous year's median; scenario 3 (Figure 1c)—all OPOs have the same underlying rate: the previous year's 25th percentile (across-the-board decline); scenario 4 (Figure 1d)—all OPOs have different underlying rates that vary around the previous year's 75th percentile; scenario 5 (Figure 1e)—all OPOs have different underlying rates that vary around the previous year's median; and scenario 6 (Figure 1f)—all OPOs have different underlying rates that vary around the previous year's 25th percentile. CMS, Centers for Medicare and Medicaid Services; OPO, Organ Procurement Organization.

CMS methodology assigned a higher (worse) tier than O-to-E methodology to 24% of OPOs in 2019, 43% in 2020, and 54% in 2021 (Table). Two OPOs in 2019 and 1 in 2020-2021 had higher O-to-E tier ranking. Tier assignments by both methods for each OPO can be seen in Figure 2. In 2020-2021, DDP was significantly higher for OPOs with different tier assignments than those with the same tier under CMS and O-to-E tiering methods (Fig. 3).

4. Discussion

This study offers objective methodologic and statistical insights into current CMS OPO performance metrics. The results demonstrated that larger OPOs are systematically and disproportionally vulnerable to decertification independent of actual performance based on current CMS metrics. Consistent with this finding, the most recent CMS report shows that lower-performance tier (2 and 3) OPOs are significantly larger than highest-performance OPOs. As used in other transplant performance evaluations, an alternative O-to-E evaluation approach demonstrated significant discordance with current CMS rankings and reduced the association between OPO volume and performance tiers.

Simulation analyses demonstrated that large OPOs have an equal chance of being in the highest-performance CMS tier (tier 1) only when all perform at a tier 1 level, with underlying rates equivalent to previous year's 75th percentile. In other scenarios where OPOs perform uniformly, smaller OPOs were significantly more likely to be automatically recertified (tier 1) or able to compete for contract renewal (tier 2) based on statistical power. Smaller OPOs have wider CIs, increasing the likelihood of overlapping with a fixed threshold, although larger OPOs with narrower CIs are less likely to reach a given threshold. This relationship leads to tier assignments influenced by the estimates' precision instead of identifying outlier performance.

Based on CMS data, highest-performance CMS tier OPOs are significantly smaller than lower-performance tier (2 and 3) OPOs. This aligns with a 2020 study indicating smaller OPOs were more likely to meet fixed CMS thresholds.[8] The simulation analysis shows that current metrics are statistically biased against larger OPOs, independent of performance, but the study cannot determine whether larger OPOs perform worse or face a

Table
CMS and O-to-E tier rankings by year.

Factor	2019 (n = 58)	2020 (n = 58)	2021 (n = 57)
Deceased donor potential (OPO volume), median (P25-P75)	1514 (436-2139)	1589 (891-2269)	1815 (974-2467)
CMS tier, n (%)			
1 (top performer)	27 (46.6)	20 (34.5)	15 (26.3)
2 (must recompete)	15 (25.9)	16 (27.6)	18 (31.6)
3 (decertified)	16 (27.6)	22 (37.9)	24 (42.1)
O-to-E tier, n (%)			
1 (top performer)	33 (56.9)	37 (63.8)	33 (57.9)
2 (must recompete)	16 (27.6)	11 (19.0)	17 (29.8)
3 (decertified)	9 (15.5)	10 (17.2)	7 (12.3)
Tier consistency, n (%)			
CMS tier $\leq$ O-to-E tier[a]	44 (75.9)	33 (56.9)	26 (45.6)
CMS tier $>$ O-to-E tier	14 (24.1)	25 (43.1)	31 (54.4)
CMS tier 2 and O-to-E tier 1, n (%)	6 (10.3)	12 (20.7)	13 (22.8)
CMS tier 3 to O-to-E tier 1, n (%)	1 (1.7)	5 (8.6)	5 (8.8)
CMS tier 3 and O-to-E tier 2, n (%)	7 (12.1)	8 (13.8)	13 (22.8)

CMS, Centers for Medicare & Medicaid Services; OPO, Organ Procurement Organization; O-to-E, observed-to-expected; P75, 75th percentile; P25, 25th percentile.
[a] Only 2 OPOs in 2019 and 1 in 2020 and 2021 had a higher O-to-E tier ranking.

mathematical disadvantage related to the inverse relationship between UCL and sample size (potential donors).

Clear inefficiencies and inequities have been identified in the transplant system that may require policy intervention.[9] Improving the transplant system, including OPO performance, is essential for efficiently procuring and transplanting deceased donor organs while honoring donors and their families. Given evidence of significant variation in practice and outcomes of OPOs, appropriate performance assessment is needed to identify areas for improvement. However, the findings show that current CMS policy may unjustly identify high-performing large OPOs although failing to address smaller, poorly-performing ones, risking disruptions in the transplantation system without enhancing donation services.

This study also illustrated an alternative O-to-E ratio methodology that provides consistent error rates across OPO volumes, not only reducing bias against larger OPOs but also identifying underperformers effectively. This tiering system has an inverse relationship between the percentage of O-to-E tier 2 OPOs misclassified as tier 3 and the percentage of tier 3 OPOs misclassified as tier 2. Although larger underperforming OPOs may be misclassified as O-to-E tier 2, they would still need to recompete, allowing closer scrutiny. This approach may minimize disruption in the transplant system by preventing the decertification of high-performing OPOs classified as low-performing under current CMS methods. This method was used to highlight the potential for a more equitable evaluation system.

The number of OPOs in the lowest-performance CMS tier (tier 3) has increased.[2] Analyzing the real-world data using O-to-E methodology, shows many OPOs meet the national expectations for the given year but lack the year-over-year improvement needed to meet CMS thresholds. This impacts larger OPOs more owing to their narrower CIs.

Aside from methodologic issues highlighted in this study, CMS metrics overlook the underlying differences in DSA populations, only adjusting the transplant rate for decedent age.[1,10] Recent studies indicate that variations in DSA population characteristics can significantly impact performance measured by CMS metrics.[11,12] Notably, O-to-E methodology is neutral to risk adjustment, but this can be directly incorporated into expected rates. Accounting for statistical confounding when estimating expected events, like risk adjustment in transplant program evaluations, could also improve OPO performance assessments.

Some limitations should be considered when interpreting the results. Accuracy of simulation outcomes depends on the validity of the assumptions; if inaccurate, results can be misleading. However, assumptions are based on CMS data, providing a reliable foundation. Critical thresholds in evaluation systems should align with regulatory goals and maintain acceptable error rates. Stakeholders should discuss what constitutes acceptable errors while considering their consequences. The national average for the current year was used to estimate expected events in the O-to-E tiering system, but a 3- to 5-year rolling average could be appropriate, given less vulnerability to secular trends. CMS determines DDP using restricted-use National Center for Health Statistics Multiple Cause of Death data and SRTR data for donors and organ counts.[6] We lacked access to National Center for Health Statistics data and used SRTR donor

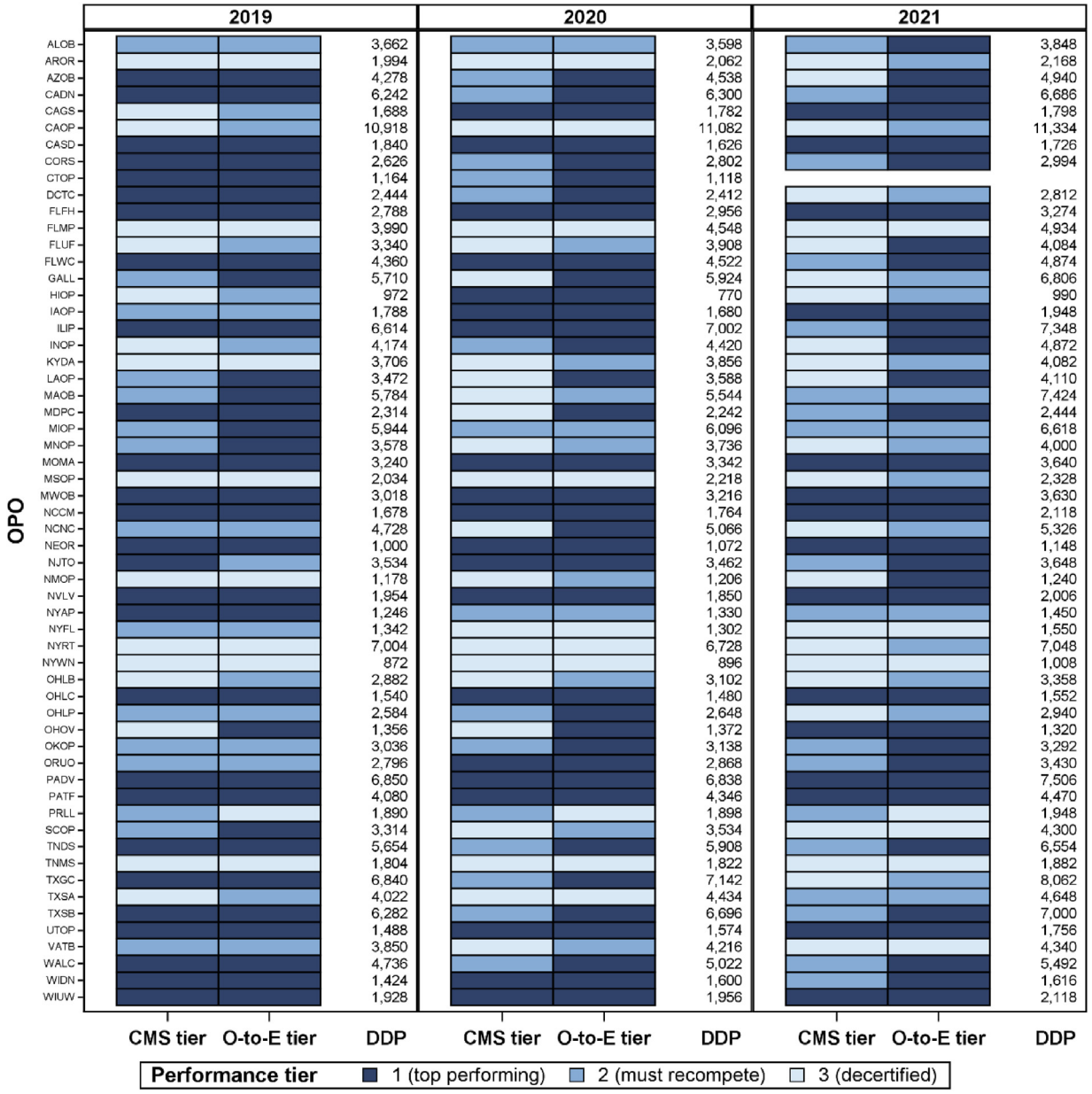

Figure 2. Performance tier assignments under CMS and O-to-E methodology by OPO. Effective from January 1, 2021, CTOP and MAOB merged into a single OPO, retaining the name of MAOB. CMS, Centers for Medicare & Medicaid Services; DDP, deceased donor potential; OPO, organ procurement organization; O-to-E, observed-to-expected. OPO abbreviations—ALOB, Legacy of Hope; AROR, Arkansas Regional Organ Recovery Agency; AZOB, Donor Network of Arizona; CADN, Donor Network West; CAGS, Sierra Donor Services; CAOP, OneLegacy; CASD, Lifesharing - A Donate Life Organization; CORS, Donor Alliance; CTOP, LifeChoice Donor Services; DCTC, Washington Regional Transplant Community; FLFH, OurLegacy; FLMP, Life Alliance Organ Recovery Agency; FLUF, LifeQuest Organ Recovery Services; FLWC, LifeLink of Florida; GALL, LifeLink of Georgia; HIOP, Organ Donor Center of Hawaii; IAOP, Iowa Donor Network; ILIP, Gift of Hope Organ & Tissue Donor Network; INOP, Indiana Donor Network; KYDA, Kentucky Organ Donor Affiliates; LAOP, Louisiana Organ Procurement Agency; MAOB, New England Organ Bank; MDPC, The Living Legacy Foundation of Maryland; MIOP, Gift of Life Michigan; MNOP, LifeSource Upper Midwest Organ Procurement Organization; MOMA, Mid-America Transplant Services; MSOP, Mississippi Organ Recovery Agency; MWOB, Midwest Transplant Network; NCCM, Lifeshare of the Carolinas; NCNC, HonorBridge; NEOR, Live On Nebraska; NJTO, New Jersey Organ and Tissue Sharing Network OPO; NMOP, New Mexico Donor Services; NVLV, Nevada Donor Network; NYAP, Center for Donation and Transplant; NYFL, Finger Lakes Donor Recovery Network; NYRT, LiveOnNY; NYWN, Upstate New York Transplant Services Inc; OHLB, LifeBanc; OHLC, Life Connection of Ohio; OHLP, Lifeline of Ohio; OHOV, LifeCenter Organ Donor Network; OKOP, LifeShare Transplant Donor Services of Oklahoma; ORUO, Pacific Northwest Transplant Bank; PADV, Gift of Life Donor Program; PATF, Center for Organ Recovery and Education; PRLL, LifeLink of Puerto Rico; SCOP, We Are Sharing Hope SC; TNDS, Tennessee Donor Services; TNMS, Mid-South Transplant Foundation; TXGC, LifeGift Organ Donation Center; TXSA, Texas Organ Sharing Alliance; TXSB, Southwest Transplant Alliance; UTOP, DonorConnect; VATB, LifeNet Health; WALC, LifeCenter Northwest Organ Donation Network; WIDN, Versiti Wisconsin, Inc; WIUW, UW Health Organ and Tissue Donation.

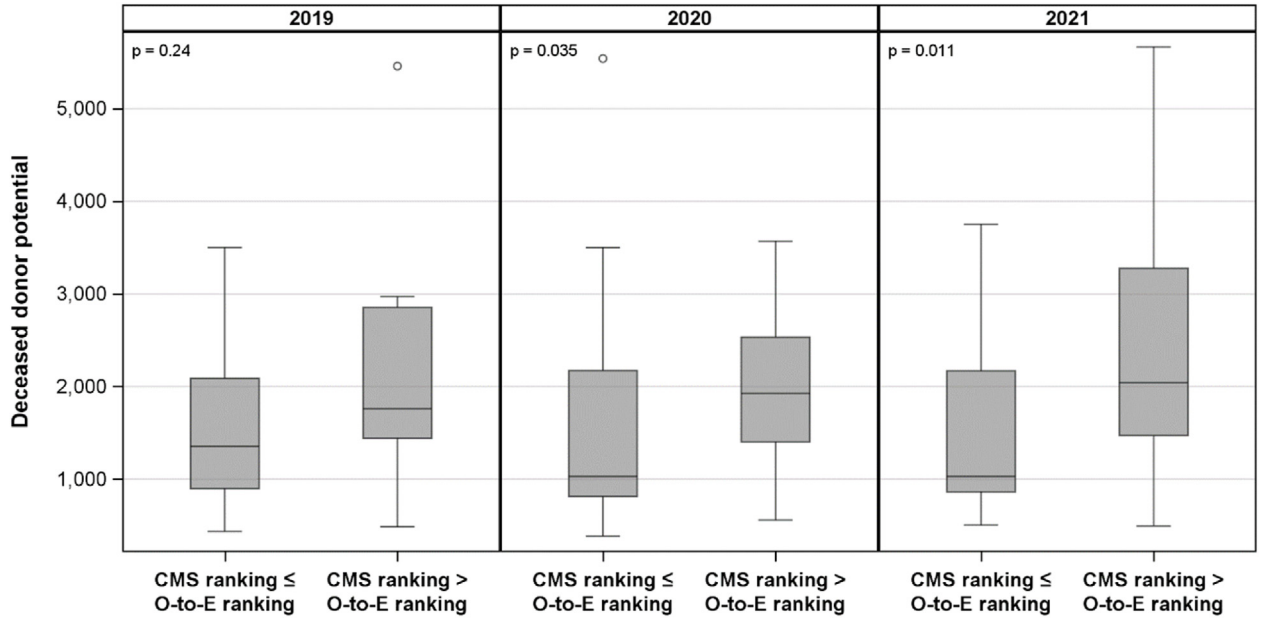

Figure 3. Comparison of deceased donor potential by performance tier consistency across CMS and O-to-E methods. Only 2 OPOs in 2019 and 1 in 2020 and 2021 had a higher O-to-E tier ranking. CMS, Centers for Medicare and Medicaid Services; OPO, Organ Procurement Organization; O-to-E: observed-to-expected.

data and CMS rates to estimate DDP. Although this method should yield accurate figures, minor variations may occur. CMS metrics include pancreata recovered or submitted for research in rate calculations.[1,6] Since this definition's release, the number of pancreata procured for research has increased.[13-15] CMS clarified pancreata are used for research if accepted for legitimate islet cell research.[16,17] However, as of September 2024, OPTN codes lack clarity on this issue, and new codes are being introduced.[16,17] Such changes to the numerator of the metrics could alter performance rates and tier assignments but would not affect the issue of narrower CIs for larger DDPs.

In conclusion, findings illustrate that current CMS evaluation system appears to be systematically biased against larger OPOs, with smaller OPOs more likely to be automatically recertified or able to compete for contract renewal even when actual performance is equivalent. Given the need to ensure accurate and fair regulation of OPO performance and the limitations of current metrics, it is imperative to establish reliable and validated measures that accurately reflect performance. Regulators should consider adopting unbiased metrics to assess OPO performance to ensure a fair and effective process, ultimately enhancing the overall quality and dependability of the transplantation system for patients.

Acknowledgments

The authors thank Drs Grace Lyden, Jonathan Miller, Ryutaro Hirose, and Jon Snyder from the Scientific Registry of Transplant Recipients (SRTR) at Hennepin Healthcare Research Institute in Minneapolis, MN, for their vital contributions to this study; Drs Snyder and Lyden for conceptualizing the initial simulation analysis for this study; Dr Lyden for conducting the initial simulation analysis for this study; and Drs Miller and Hirose for their contributions to the initial simulation analysis. Although they were unable to present these findings at the American Transplant Congress, their support and collaboration were essential in successfully carrying out the simulation analyses. The data reported in this study have been supplied by the Hennepin Healthcare Research Institute as the contractor for the SRTR. The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the SRTR or the US government.

Funding

No funding was used for the study.

Declaration of competing interest

The authors of this manuscript have conflicts of interest to disclose as described by the American Journal of Transplantation. R.L. reports research funding from Gift of Life Foundation and One Legacy Foundation, outside of the submitted work. S.M. receives grant funding from the Kidney Transplant Collaborative and the National Institutes of Health; and personal fees from Sanofi, Kidney International Reports, Specialist Direct and Health Services Advisory Group, outside of the submitted work; is a Board Member for the Alliance for Organ Donation and Transplantation; and recently completed a term as Chair of the OPTN data advisory committee and as an exofficio member of the Scientific Registry of Transplant Recipients Review Committee. E.A.P. is a member of the Board of Directors for Donor Alliance but does not receive any financial compensation for this

role. J.D.S. reports consultancy for eGenesis, Novartis, Veloxis, the Organ Donation Advocacy Group, and Sanofi Corporation; research funding from Department of Defense, Gift of Life Foundation, Kidney Transplant Collaborative, National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health/National Heart, Lung and Blood Institute, and One Legacy Foundation, outside of the submitted work; honoraria from eGenesis and Sanofi; and advisory or leadership roles as Data Safety Monitoring Board Member for Bristol Myers Squibb; is a member of the United Network for Organ Sharing Policy Oversight Committee; Chair of United Network for Organ Sharing Data Advisory Committee; and exofficio member of the Scientific Registry of Transplant Recipients Review Committee and speakers bureau for Sanofi; and reports being a statistical editor for the *American Journal of Transplantation*, *Kidney International*, and the *Journal of the American Society of Nephrology*. The other authors have no conflicts of interest to disclose as described by the American Journal of Transplantation.

Data availability

This study used data from Centers for Medicare and Medicaid Services' 2023 Organ Procurement Organization Interim Annual Public Aggregated Report, which are publicly available at https://qcor.cms.gov, and from the Scientific Registry of Transplant Recipients, available from Scientific Registry of Transplant Recipients with an approved data user agreement.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ajt.2024.11.024.

ORCiD

Rocio Lopez https://orcid.org/0000-0002-4319-420X
Sumit Mohan https://orcid.org/0000-0002-5305-9685
James R. Rodrigue https://orcid.org/0000-0001-5589-4919
Susana Arrigain https://orcid.org/0000-0001-7533-3932
Deena Brosi https://orcid.org/0000-0002-5294-1414
Ryan Lavanchy https://orcid.org/0009-0003-8950-9988
Bruce Kaplan https://orcid.org/0000-0003-3400-642X
Elizabeth A. Pomfret https://orcid.org/0000-0001-7970-1325
Jesse D. Schold https://orcid.org/0000-0002-5341-7286

References

1. Requirements for certification and designation and conditions for coverage: organ procurement organizations. Centers for Medicare & Medicaid Services (CMS). https://www.ecfr.gov/current/title-42/part-486/subpart-G. Accessed 28 October 2024.
2. OPO public performance report—2023 OPO Interim Annual Public Aggregated Report. CMS Quality Certification and Oversight Reports (QCOR). Accessed June 17, 2024. https://qcor.cms.gov/.
3. Organ procurement organization statistics. Scientific Registry of Transplant Recipients. Accessed June 17, 2024. https://www.srtr.org/reports/opo-specific-reports/interactive-report.
4. Sedgwick P. Understanding confidence intervals. *BMJ*. 2014;349:g6051. https://doi.org/10.1136/bmj.g6051.
5. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016;31(4):337–350. https://doi.org/10.1007/s10654-016-0149-3.
6. OPO annual public aggregated performance report—user guide: 2026 certification period. CMS Quality Certification and Oversight Reports (QCOR). Accessed June 17, 2024. https://qcor.cms.gov.
7. Scientific Registry of transplant recipients. Technical Methods for the Program-Specific Reports. https://www.srtr.org/about-the-data/technical-methods-for-the-program-specific-reports/. Accessed June 17, 2024.
8. Snyder JJ, Musgrove D, Zaun D, et al. The Centers for Medicare and Medicaid Services' proposed metrics for recertification of organ procurement organizations: evaluation by the Scientific Registry of Transplant Recipients. *Am J Transplant*. 2020;20(9):2466–2480. https://doi.org/10.1111/ajt.15842.
9. Evaluation of the NQF trial period for risk adjustment for social risk factors. National Quality Forum. https://www.qualityforum.org/Projects/s-z/SES_Trial_Period/Final_Report.aspx. Accessed 26 October 2024
10. Goldberg D, Doby B, Siminoff L, Shah M, Lynch R. Rejecting bias: the case against race adjustment for OPO performance in communities of color. *Am J Transplant*. 2020;20(9):2337–2342. https://doi.org/10.1111/ajt.15865.
11. Miller JM, Zaun D, Wood NL, et al. Adjusting for race in metrics of organ procurement organization performance. *Am J Transplant*. 2024;24(8):1440–1444. https://doi.org/10.1016/j.ajt.2024.01.032.
12. Lopez R, Mohan S, Schold JD. Population characteristics and organ procurement organization performance metrics. *JAMA Netw Open*. 2023;6(10):e2336749. https://doi.org/10.1001/jamanetworkopen.2023.36749.
13. Miller J, Wood NL, Lyden GR, et al. Impact of Centers for Medicare and Medicaid Services final rule on organ procurement organization metrics and procedural trends in the procurement of pancreata for research. *Pancreas*. 2024;53(2):e176–e179. https://doi.org/10.1097/MPA.0000000000002284.
14. Booker SE, Martinez C, Cartwright LA, et al. Pancreata recovered for research: has the increase impacted pancreas transplantation? *Am J Transplant*. 2024;24(12):2304–2306. https://doi.org/10.1016/j.ajt.2024.09.015.
15. Goldberg D, Chyou D, Wulf R, Wadsworth M. Temporal changes in procurement of pancreata for research. *Am J Transplant*. 2023;23(9):1465–1467. https://doi.org/10.1016/j.ajt.2023.05.002.
16. Organ Procurement Organization (OPO) conditions for coverage—definition clarification. Centers for Medicare & Medicaid Services (CMS). Accessed October 21, 2024. https://www.cms.gov/files/document/qso-24-04-opo.pdf.
17. Organ Procurement Organization (OPO) conditions for coverage—reporting data related to pancreata procured for research. Centers for Medicare & Medicaid Services (CMS). Accessed October 21, 2024. https://www.cms.gov/files/document/qso-24-19-opo.pdf.