Task Order 6 Task Area 3

Patient Safety Event Qualitative Thematic Analysis

Contract Number: 75R60223D00001
COR's Name: Aaron Fisher
Date of Submission: 9/25/2025

Table of Contents

List of Tables

Table of Figures

Executive Summary

Background:

- 2,856 patient safety events from 2019-2024 were received from the OPTN Contractor, 95% of which had qualitative free text information. A representative sample of 739 of the 2,856 events were manually coded.
- The goal was to examine the qualitative patient safety data using a standardized coding taxonomy to identify patterns and trends, highlight gaps in patient safety, and better characterize the nature of events occurring from 2019-2024.

Methodology

- A stratified sampling strategy ensured representative coverage across event types and year received.
- Manual coding was performed using a standardized Codebook across six root nodes:
    - **Primary Impacted Individual:** The primary individual who was directly impacted or had the potential to be directly impacted by the patient safety event.
    - **Primary Outcome:** The primary physical outcome or effect of the patient safety event on the primary impacted individuals.
    - **Primary Phase**: The primary phase in the transplant process when the error or incident occurred.
    - **Primary Type:** The implied or visible error.
    - **Primary Responsible Party:** The primary responsible party for the root cause of the event.
    - **Primary Cause:** The primary root factor that led to an incident.
- Interrater reliability and coder training were used to mitigate subjectivity.
- To ensure overall estimates reflected the full patient safety event population in the MCL, not just the manually coded subset, each event was assigned a weight using the population of the stratum divided by the number of events sampled from the stratum.
- Analysis was limited by a decline in data quality over time and variation in event description completeness.

Results:

Analysis of Manually Coded Patient Safety Events

Event Phase and Impact

- The highest number of safety events occurred in the **Procurement and Transplant phases**.
- The majority of events impacted transplant patients.
- **Living donor events** showed unique patterns – 21% of living donor events were attributed to human factors.

Responsible Parties

- **OPOs and Transplant Centers** were most frequently responsible for errors.
- **Other parties** (e.g., labs, couriers, etc.) were rarely cited but sometimes involved.

Causes of Events

- **Process or System Design Causes**
    - 94% of **Process or System Design** errors were attributed to **Policy/Protocol Deviations**.
- **Human Factors**

    - Human Factors caused 12% of all weighted events but caused 21% of weighted events in the living donor phase.
- **Communication Failures**
    - Generally low (0-10%), but **most prevalent during procurement**.
    - **May be undercounted** due to the coding structure treating Communication and Process or System Design as distinct primary classifications under the Primary Cause root node.
- **Environment and Equipment**
    - **Rare** (<4%) but contributed to some complex multi-actor events.

Event Classifications and Outcomes

- **Administrative and Documentation errors** were the most frequent Primary Types.
- **No Detectable Harm** was used when there was an error, and the error may have caused harm, but no harm was detectable at the time of documentation; this may undercount actual harm.
- **Unclassifiable** causes were more common when the phase was unclear or qualitative data was poor.
- **Living Donor Adverse Events** were always coded as errors due to OPTN reporting requirements, though true causality was often ambiguous.

Discussion

- Most patient safety events resulted from **deviations, deficiencies, and failures of existing policies**, not their absence.
- **Documentation and verification process failures** are common root issues.
- The **decline in data quality** over time undermined the ability to detect trends and assign causes accurately.

Recommendations

- **Standardize reporting requirements** to include mandatory root cause analyses and corrective action plans
- **Target policy compliance gaps** with retraining, signage, and refresher sessions on high-risk procedures.
- **Stress-test policies** under real world conditions, especially at high-risk transition points.

Recommended Next Steps

- **Code the full dataset (2019-2024)** to reduce sampling bias and improve confidence in findings.
- **Implement continuous coding** of patient safety events using a standardized protocol.
- **Develop a reporting dashboard** to promote transparency and highlight trends.
- **Refine existing NLP for continued use or transition to LLM** to support manual coding processes and reduce error

Background

Ensuring patient safety in organ transplantation is fundamental to protecting patients and donors, sustaining public trust, and improving outcomes across the transplant continuum. Systematic monitoring of patient safety events, including adverse events and near misses, enables a deeper understanding of what is happening (trends), why it happens (root causes), and where to intervene (policy and practice). These insights have the capacity to inform targeted education and training, strengthen member accountability, and support effective federal oversight of the OPTN. By analyzing both events that caused harm and those intercepted before harm occurred, HRSA can identify systemic vulnerabilities, and promote evidence-based policies that reduce risk and prevent recurrence. In short, a comprehensive patient safety lens is essential to preventing harm, improving quality, and fostering a culture of safety in transplantation.

The Master Case List (MCL) contains 2,856 patient safety complaints from 2019-2024, of which 2,720 (95%) include free text narratives. The MCL captures the Incident Type of each event with a Primary Classification (e.g., communication, testing, allocation) and Subclassification (e.g., inaccurate/insufficient donor OR information, HLA inaccurate results reported, bypasses) but fails to codify context beyond the type of each event. HRSA received the MCL with free text fields from the OPTN Contractor in March 2025. Several elements needed for rigorous analysis and actionable oversight were not present in the structured data in the MCL: these include standardized, context-rich information on the Primary Impacted Individual, Primary Outcome, Primary Phase, Primary Type, Primary Responsible Party, and Primary Cause, along with clear definitions, and examples for each information element. For the remainder of this report, we refer to these information elements as "root nodes."

This is the first comprehensive systematic analysis of the qualitative transplant patient safety data delivered to HRSA by the OPTN contractor. The 2019-2024 window offers a uniquely rich view of patterns and issues across phases and organizations, providing insights that can directly inform policy, OPTN guidance, monitoring, and targeted education, thereby strengthening the transplant ecosystem.

This analysis addresses the need for the standardized coding of each event to identify patterns in who was impacted, how they were impacted, what type of event occurred, where/when it occurred, and why it happened. A holistic understanding of these dimensions across 2019-2024 will allow HRSA to surface gaps in transplant patient safety, identify systemic risks along the care continuum, and translate findings into policy recommendations, targeted training, and continuous process improvement. Most importantly, with routine receipt and consistent coding of qualitative safety data, HRSA will obtain clear insight into emerging risks and corrective actions, enabling proactive, evidence-based oversight of the OPTN now and going forward.

Methodology

This section outlines steps taken to create a tailored taxonomy for coding qualitative patient safety event information; develop the Excel file known as the Patient Safety Codebook (the Codebook); manually code stratified random samples of the MCL data; assess manually coded samples for inter-

rater reliability; train, validate, and apply an NLP classifier to perform classifications on all MCL data not sampled; perform summarization and analysis on the results from the manually coded classifications.

Patient Safety Taxonomy

To transform qualitative free text fields into analyzable, decision-ready information, we used existing patient safety coding taxonomies to develop a tailored taxonomy for coding the qualitative transplant patient safety information. The development of this taxonomy was informed by a review of the MCL patient safety data, identifying patterns and themes, as well as a systematic literature review of existing coding practices. Key articles, guidelines, and taxonomies were identified from the literature review, and coding structures were compared across taxonomies.

The resulting final taxonomy is primarily based on the Joint Commission on Accreditation of Healthcare Organizations' Patient Safety Event Taxonomy with elements of the Primary Care Patient Safety (PISA) taxonomy developed at Cardiff University. The six root nodes are elements which describe a patient safety event:

- **Primary Impacted Individual:** The primary individual who was directly impacted or had the potential to be directly impacted by the patient safety event.
- **Primary Outcome:** The primary physical outcome or effect of the patient safety event on the primary impacted individuals.
- **Primary Phase**: The primary phase in the transplant process when the error or incident occurred.
- **Primary Type:** The implied or visible error.
- **Primary Responsible Party:** The primary responsible party for the root cause of the event.
- **Primary Cause:** The primary root factor that led to an incident.

Each node is comprised of at least a primary classification, which takes on one of several values. Additionally, Primary Impacted Individual, Primary Phase, Primary Type, Primary Responsible Party, and Primary Cause, also consist of a subclassification, which takes on one of several values determined by the node's classification.

Codebook Development

The tailored taxonomy is represented in the Codebook Excel file. The Codebook also contains clear definitions and transplant and/or procurement specific examples for each node, enabling standardized coding across the six root nodes and harmonization of Primary Classification and Subclassification terms. In addition to the Codebook, a PowerPoint deck was produced to show the structure of the coding taxonomy (the Taxonomy). The supplemental Codebook Excel file and the associated PowerPoint file provide comprehensive details on the Taxonomy and are hyperlinked in Appendix A of this report.

The Taxonomy was refined using three pilot coding exercises. A subset of 52 patient safety events from the MCL were each manually coded by four coders. The inter-rater reliability was determined, and the Taxonomy was evaluated for reliability, clarity, and comprehensiveness. This updated version of the Taxonomy was shared with HRSA for their review. HRSA provided feedback, notably to increase focus on procurement rather than transplant specific events, and this feedback was incorporated.

Sampling

A total of 739 patient safety events were sampled from the MCL to be coded manually by human coders in four iterations of 206, 233, 225, and 75 events. The sample for each iteration was chosen using stratified random sampling with balanced allocation to ensure an appropriate number of records were coded across all years and member types. Samples were stratified based on complaint receipt year (2019, 2020, 2021, 2022, 2023, 2024) and subject member type (OPO, Transplant Center, Other), for a total of 18 strata. All samples were drawn randomly without replacement, and events were not reused across samples. The goal in creating each sample was to draw an equal amount from each stratum, however for strata with smaller populations, this was not possible, so the entire stratum was used. Each sample was coded in its entirety by one or more trained coders.

One consideration for this sampling technique is that the sample of manually coded data is not representative of the target population of patient safety events in the MCL. There are different distributions of each stratum in the sampled data when compared to the population. Given that the goal is to perform calculations that provide information representing all patient safety events in the MCL, not just the manually coded data, weights were calculated for each stratum to be used during analysis to allow for unbiased overall estimation. Each weight is the population of the stratum divided by the number of events sampled from the stratum.

Manual Coding Protocol

Patient safety events were divided and manually coded by a group of three regular coders and three reviewing coders for sample set one and one regular coder and one reviewing coder for sample sets two through four, one sampling set at a time. Reviewing coders were drawn from the set of Codebook developers, while regular coders served as trainees. Training for regular coders involved studying the Codebook Taxonomy, double coding previously validated patient safety events, and trial periods of coding with progressively less frequent oversight from reviewing coders. Throughout the coding of the sampled patient safety events, coders and reviewing coders met frequently to establish standard procedures and discuss complex or ambiguous events. Coders established standard methodologies for consistent coding across events.

Coding a patient safety event involves applying the rules and definitions in the Codebook. First, a coder must determine if a patient safety event has any qualitative data. If not, every node for the event is coded as "Not Enough Qualitative Information to Code." Next, for each node, a coder selects the correct primary classification and, if applicable, subclassification for the event. There are two notable classifications patient safety events can fall into. First, a patient safety event may be reported which does not represent an error. Here an error represents wrongdoing by a party intentionally or unintentionally, not a policy or bylaw violation. If there is no error, then every node's classification and subclassification is coded as "No Error." Second, patient safety events are sometimes reported with insufficient information to determine the correct value for one or more nodes' classifications or subclassifications. This may be due to lack of information, unreliable information, or contradictory information. In this case, a node's primary classification or subclassification is "Unclassifiable." If more than 50% of nodes are unclassifiable, then the entire event is labeled as "Not Enough Qualitative Information to Code."

Reviewing coders double coded 63% of events to serve as validation and allow for calculation of inter-rater reliability. When coders were done with a sample, they compared a subset of their results to the results of a reviewing coder to make sure that the overlapping safety events were identical.

While double coding was performed across all sample sets with the reconciliation of any disagreements, inter-rater reliability was formally evaluated for sample set four. For sample set four, two coders independently labeled the same sampled narratives across all primary classifications and subclassifications. We calculated the percentage agreement between the two coders. Additionally, for each label, we calculated the percentage agreement across the entire sample set were also calculated. Following this review, the coder and reviewing coder discussed their results and revised discrepancies to ensure agreement on every event in sample set four.

NLP Methodology

Because manual coding is labor-intensive and can vary across reviewers, we supplemented the manual approach of coding four samples of data with natural language processing (NLP) to increase scale, speed, and consistency by applying a single model per root node to every narrative. All data cleaning, transformation, and modeling of patient safety text data were performed using Python.

Data cleaning and transformation included removing repetitive or immaterial header text from the input patient safety data along with irrelevant dates. Because there were multiple columns of free text data, once data pre-processing was complete, the five free text columns were combined. Additionally, as the free text contained multiple acronyms, some with multiple expansions, subject matter experts developed a list of acronyms and their potential expansions. These acronyms were then replaced within the text to improve the model's ability to understand the free text data. Following data pre-processing, we performed tokenization to allow for reasonable feeding of text chunks into the chosen text classification model, BERT.

Using the Codebook, we curated high-precision cue phrases as a weak supervision source and paired them with manually coded patient safety data from sample sets one, two, and three, iteratively, to fine-tune an existing transformer-based classifier. Snorkel LabelModel in Python was used to implement the weak supervision framework. After turning the high-precision phrases into label functions, we applied the label functions to an unlabeled pool of patient safety events to get a matrix of noisy votes. Then we trained the Snorkel LabelModel to de-noise these votes and carefully incorporate into our training data. The identified loss function was binary cross-entropy, the chosen optimizer was the Adaptive Gradient Algorithm due to its success with sparse data, and the scheduler was a linear warmup. We used cross-validation and hold-out data and iterated to improve performance. There were multiple iterations to retrain the model as new manually coded sample sets were ready to be incorporated as training data, refine necessary pre-processing, and develop confusion matrices to understand the mislabeling between modeled datasets and validation datasets.

Primary model evaluation metrics included accuracy, weighted average F1 score, and macro average F1 score. Weighted average scores take the average of per-class metrics weighted by class size and were

used to better reflect overall performance. Macro averages helped us understand how the model performs on rare categories. Accuracy was calculated by taking the number of events that were assigned the correct classification and dividing it by the total number of events. The F1 score is the harmonic mean of the precision and the recall, balancing how accurate predictions are and how many true cases were found. These evaluation metrics were calculated during cross-validation during model training.

We then applied each model, one per root node, to assign standardized categories determined by the developed Taxonomy, to sample set four of the patient safety data. To evaluate this automated classification performance, we compared the NLP models' predictions to the manually validated codes in sample set four. Each of the six models corresponded to one of the primary classification root nodes. This direct comparison provided a practical benchmark of model performance in the same context where human agreement was evaluated. The evaluated metrics for this validation were accuracy, average precision, and average recall. The accuracy is again calculated by taking the number of events that were assigned the correct classification and dividing it by the total number of events. The precision identifies the percentage of correct events out of all events predicted as each category label. High precision directly relates to a smaller percentage of false positives. The recall takes the set of all events that were true classifications and calculates the percentage that the model was able to identify. High recall directly relates to a smaller percentage of false negatives. To calculate average precision and average recall, we took the mean across all categories for each primary classification.

Lastly, we incorporated sample set four into the fine-tuning of the models and applied each model to assign standardized categories to the non-sampled patient safety data.

Data Analysis of Manually Coded Samples

Four manually coded sampling sets were used for all data analysis.

The results below first demonstrate the performance of manual coding processes as compared to the natural language processing model. Next, high level summaries were presented summarizing the completeness of data, proportions of data types, and trends over time. Following descriptive statistics for the overall dataset, analyses were provided for each root node. These analyses followed similar structures and initially stratified the results by categories and year before providing some general recommendations. The remaining analyses were based on stratifications of interest and were accompanied by interpretations, key takeaways, and recommendations as appropriate.

Results from each root node were tabulated and analyzed year-over-year. For overall population counts in each root node, we accounted for stratum by summing the calculated sampling weights corresponding to the stratum from which an event was sampled. Because there was not uniform sampling of events, we computed aggregations to scale up each event in significance based on the extent its stratum was under- or over-sampled. All figures in the patient safety analysis results are in terms of these weights – notably, this explains why counts of weighted patient safety events are not integers or why there may be rounding errors when counts of weighted safety events are rounded to integers.

Results

Inter-Rater Reliability of Sample Set Four

Overall, the two coders for sample set four achieved 76% agreement across all primary classifications and subclassifications, indicating that coders assigned the same classifications on roughly three out of four patient safety events. There was higher concordance observed in classifying the Primary Impacted Individual primary and subclassifications, as well as the Primary Responsible Party subclassification. There was lower concordance observed for Primary Cause and Primary Type primary and subclassifications. Note that this simple percent agreement does not adjust for agreement expected by chance. Table 1 presents the percent agreement across all primary and subclassifications.

Percentage agreement between coders on sample set four	
Node	**Percent Agreement**
Primary Impacted Individual Classification	84%
Primary Impacted Individual Subclassification	84%
Primary Outcome Classification	76%
Primary Phase Classification	79%
Primary Phase Subclassification	77%
Primary Type Classification	71%
Primary Type Subclassification	71%
Primary Responsible Party Classification	77%
Primary Responsible Party Subclassification	84%
Primary Cause Primary Classification	68%
Primary Cause Subclassification	68%

Table 1: Percent Agreement Between Coders on Sample Set Four

All event classifications were iteratively reconciled to create the final sampling sets for analysis.

NLP Classifier Evaluation

Just as the inter-rater reliability analysis revealed variation in agreement across primary classifications, the performance of the NLP models also differed depending on the root node. Across models, agreement with the manually coded benchmark varied, with some models achieving stronger performance than others. For example, overall accuracy across models ranged from 68% for the Primary Impacted Individual model to 25% for the Primary Type model. The average precision ranged from 65% to 15%, and the average recall ranged from 61% to 8%. This points to the models on average doing slightly better at reducing false positive classifications than reducing false negative classifications. Table 2 summarizes the evaluation metrics for each model on sample set four.

NLP classifier performance on sample set four (compared to manual coding)			
Primary Classification Model	**Accuracy (%)**	**Average Precision**	**Average Recall**
Primary Impacted Individual	68%	65%	61%
Primary Outcome	37%	34%	24%
Primary Phase	38%	30%	26%
Primary Type	25%	15%	15%
Primary Responsible Party	28%	15%	8%
Primary Cause	35%	28%	24%

Table 2: NLP classifier performance on sample set four

In addition to the inference results on sample set four, we also assessed classifier performance during the modeling stage using k-fold cross-validation. The cross-validation metrics were consistent with the inference findings with overall accuracy across models from 66% for Primary Impacted Individual to 16% for Primary Type and weighted-average F1 scores from 0.64 to 0.15. This suggests the models generalize reasonably well, while still reflecting variability in classification difficulty across event types. Table 3 summarizes the average evaluation metrics for each model from the cross-validation process.

NLP Model Training Cross-Validation Results				
Primary Classification Model	**Accuracy (%)**	**Weighted Average Precision (%)**	**Weighted Average Recall (%)**	**Weighted Average F1 Score**
Primary Impacted Individual	66%	63%	66%	0.64
Primary Outcome	27%	30%	27%	0.24
Primary Phase	41%	43%	41%	0.39
Primary Type	16%	24%	16%	0.15
Primary Responsible Party	32%	30%	32%	0.26
Primary Cause	34%	31%	34%	0.31

Table 3: NLP Model Training Cross-Validation Results

Given the low performance measure results across all models, the NLP coded non-sampled patient safety data was not analyzed further. However, the spreadsheet of all classified data is in TO6 TA3 NLP Metrics and Inference Excel document for review.

With additional improvements, such as an adjusted taxonomy with fewer classification categories, more labeled data, expanded phrase coverage, fine-tuning tweaks, ongoing QA, and potential augmentation or transition to a large language model (LLM) with human oversight, this same workflow can be applied for routine prospective use at HRSA.

Analysis of Manually Coded Patient Safety Events

Patient Safety Events Suitable for Analysis

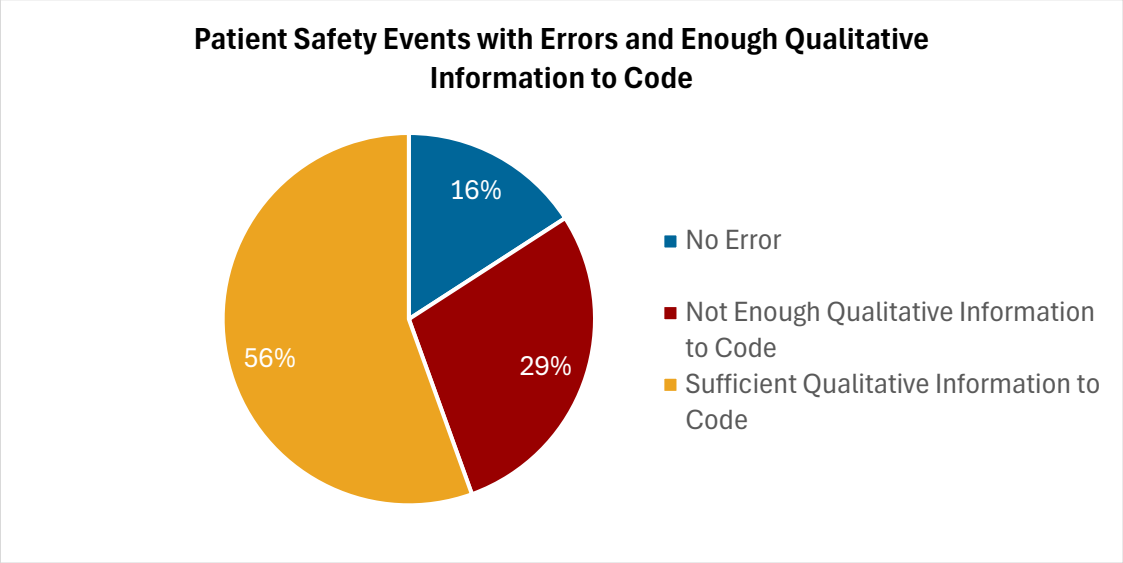

Figure 1: Patient Safety Events with Errors and Enough Qualitative Information to Code

The analysis in this report focuses on patient safety events that have sufficient qualitative information to code and which represent an error. This amounted to 56% of the weighted patient safety events represented by this dataset, about 1585 events. References to "weighted patient safety events" or "weighted events" in subsequent sections refer to this subset of patient safety events. A patient safety event had insufficient qualitative information to code if at least half of its primary nodes were Unclassifiable, that is, a specific value for the node was indeterminate. Of the weighted patient safety events represented by this report, 29% had insufficient qualitative information to code.

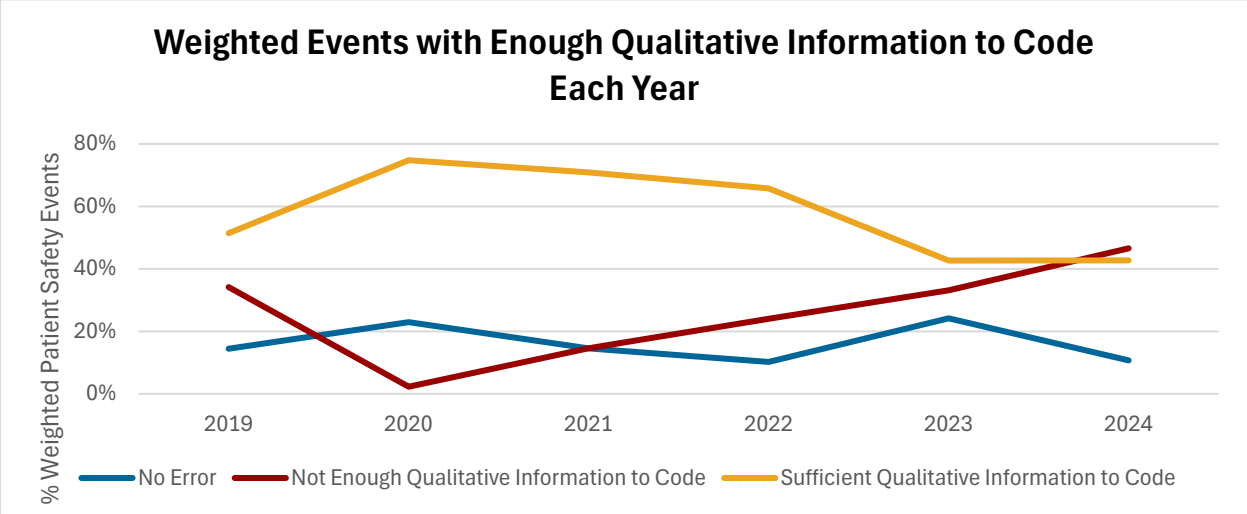

Figure 2: Events with Enough Qualitative Information to Code Each Year

The portion of weighted patient safety events with insufficient qualitative information to code increased between 2019 through 2024 from 34% to 47%. This resulted from shorter, less descriptive qualitative information being provided in later years. As will be shown in later sections, this also drove an increase in Unclassifiable categorizations within primary nodes.

Number of Codable Patient Safety Events Each Year

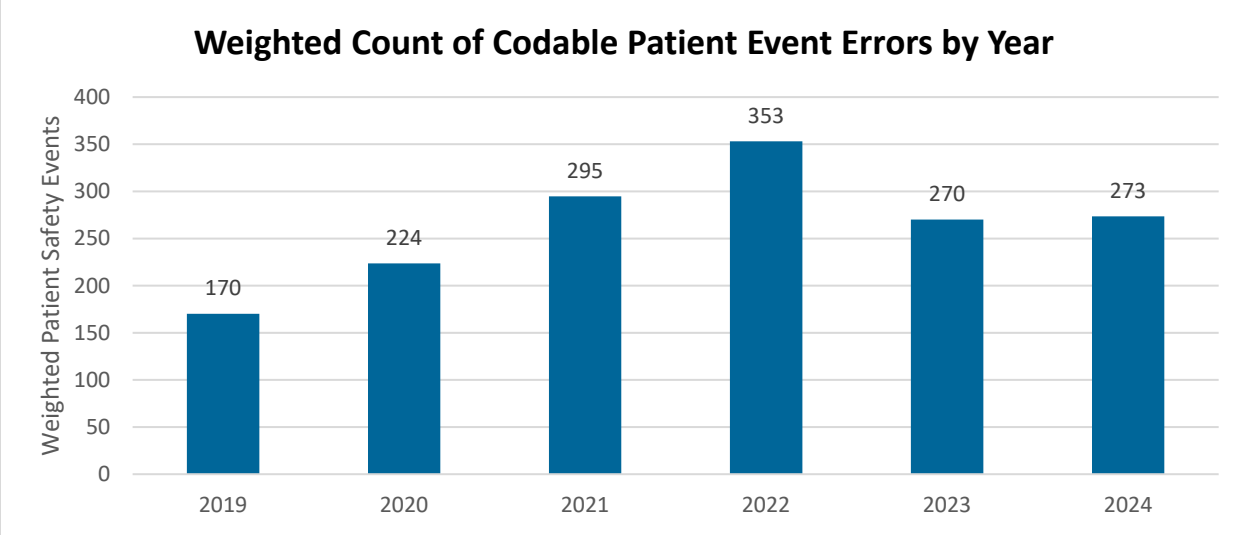

Figure 3: Patient Safety Events Each Year

The weighted count of codable patient safety events each year has increased since 2019, from 170 to 273. However, there has not been a year over year increase each year. The weighted count of events representing patient safety events that contained sufficient qualitative data to code and found to be errors increased from 2019 to 2022 before decreasing again and flattening off.

Root Node One: Primary Impacted Individual

Most Frequently Impacted Individuals

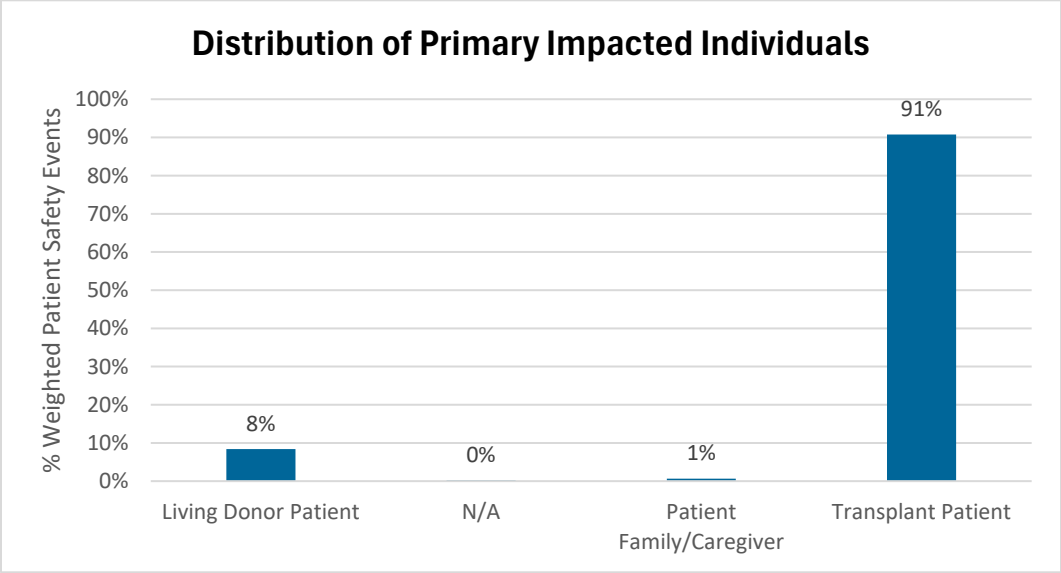

Figure 4: Distribution of Primary Impacted Individuals

Of the 1585 weighted events represented by the portion of the dataset under analysis, 91%, involved transplant recipients**.** This reflects the clinical centrality of transplant patients in the U.S. transplant ecosystem and the reality that most patient safety surveillance efforts are recipient focused.

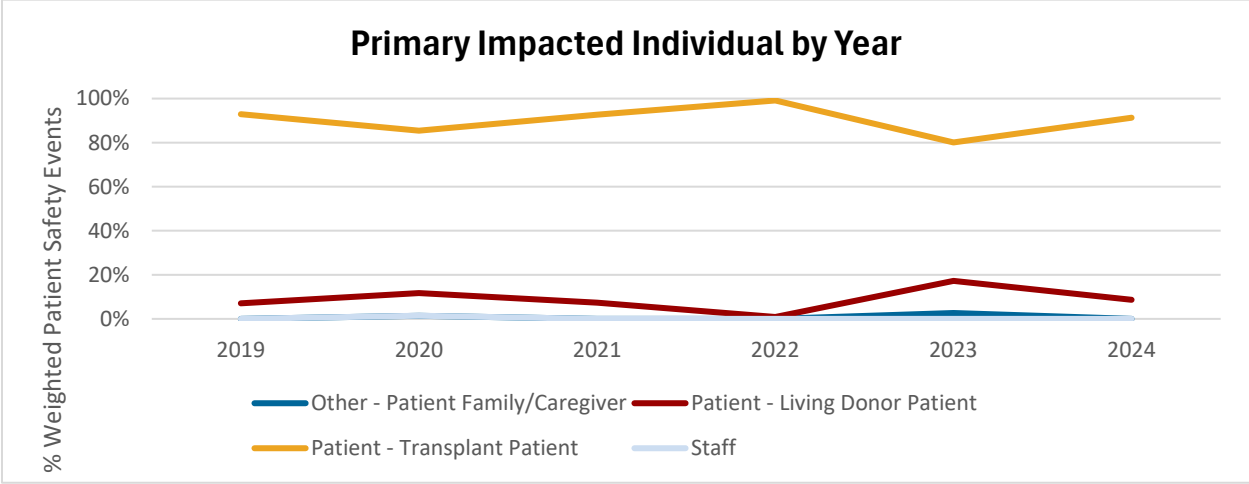

Figure 5: Primary Impacted Individual by Year

From 2019-2024, the majority of weighted patient safety events involved transplant recipients though their proportional representation varied from 80% to 99% annually. Even in years with increasing reporting from other categories (e.g., 2023), transplant patients remain the primary impacted group.

Groups Experiencing the Most Severe Outcomes

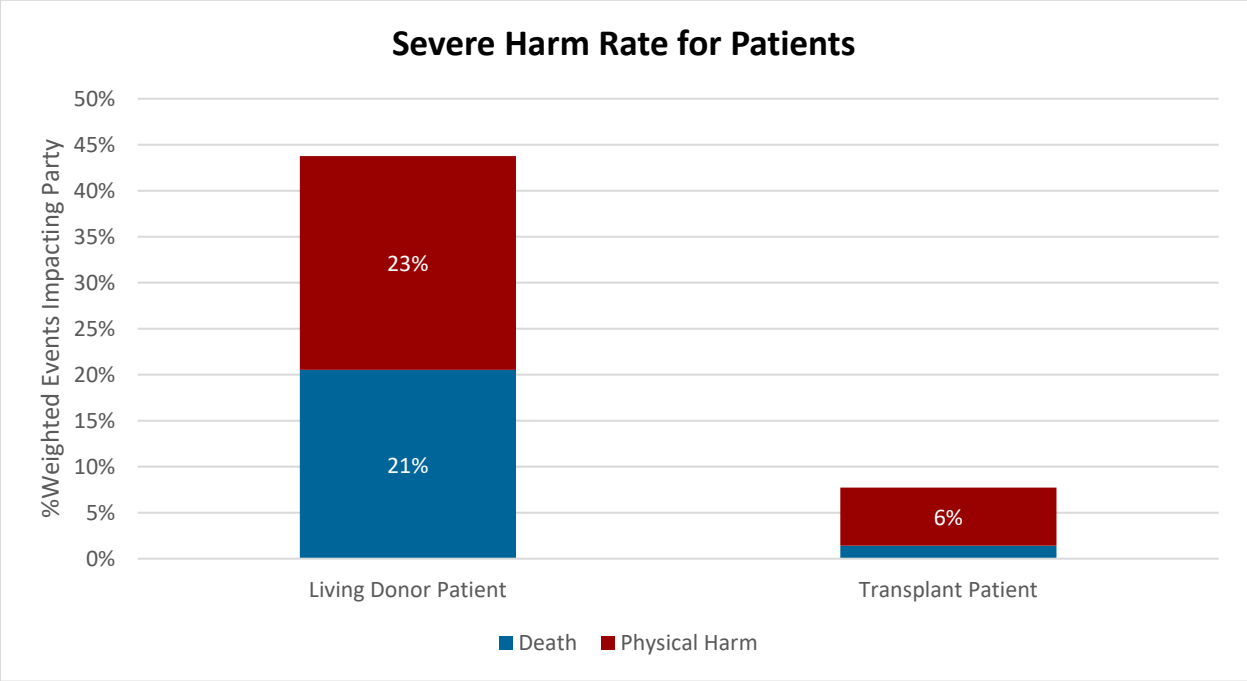

Figure 6: Severe Harm Rate for Patients

The vast majority of the 1585 weighted patient safety events (91%) involve Transplant Patients as the Primary Impacted Individual, with Living Donors accounting for only 9% of the total events. However, when considering the severity of outcomes, particularly those resulting in Death or Physical Harm, events with living donors appear more severe.

Among the 133 weighted Living Donor events, 21% or 27 weighted events resulted in Death and 23% or 31 weighted events resulted in Physical Harm, meaning that 44% of weighted reported Living Donor events involved the most serious outcomes. In comparison, 1% of weighted events involving Transplant Patients resulted in Death and 6% result in Physical Harm, totaling 7% of weighted Transplant Patient events with serious outcomes. Although Transplant Patients account for a greater number of weighted events with serious outcomes due to their much larger volume, the severity appears higher for Living Donors when considering the rarity of events and the nature of reporting. It is important to contextualize this trend within the OPTN Policy Framework, which mandates reporting of certain events specific to Living Donors, including events that may not necessarily reflect systemic failure. For example, a death within two years of donation (such as from a motor vehicle accident) is reportable regardless of its relationship to the donation itself. This nuance introduces complexity into interpreting the severity profile of weighted Living Donor events, as not all Deaths or Physical Harm reflect preventable safety breakdowns.

Outcomes for Living Donors and Transplant Patients

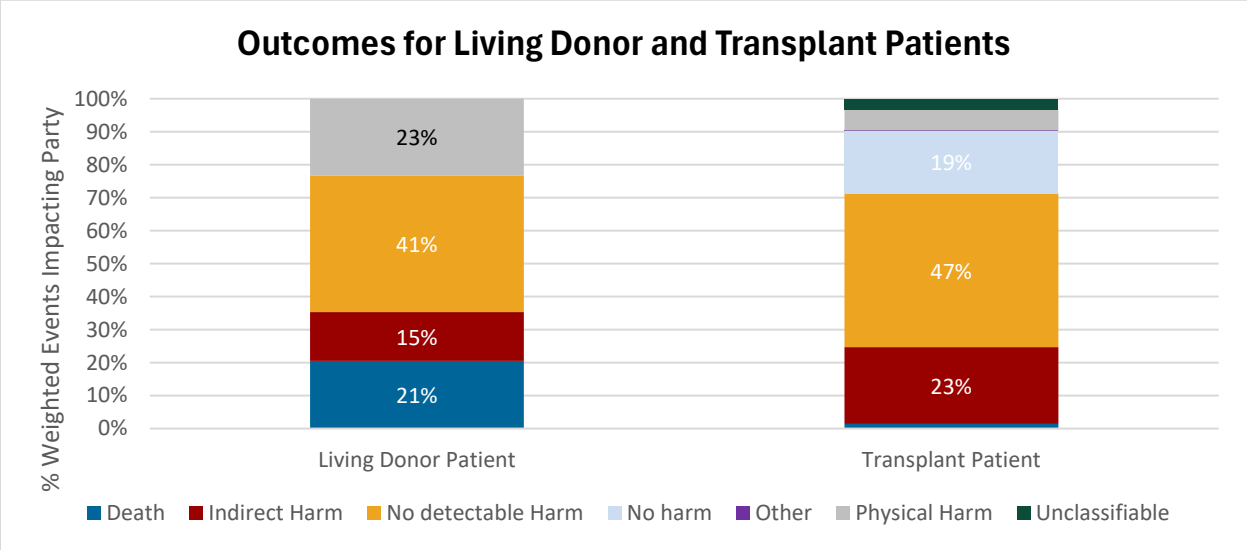

Figure 7: Outcome Distributions for Patients

In contrast, of the 1438 weighted events impacting Transplant Patients, fewer events result in Death and the overall harm burden is more distributed, with a notable portion of events categorized as Indirect Harm (23%), No Detectable Harm (47%) and No Harm (19%). This reflects the broad clinical exposure Transplant Patients have across complex care episodes, where communication failures, procedural complications, and coordination breakdowns can lead to a range of consequences short of Death.

Key Takeaways

- Living donor events appear more severe because OPTN policy requires reporting of sentinel outcomes like death, organ failure, and surgical complications.
- Transplant patient safety events show a more typical clinical distribution, with a majority resulting in No Detectable Harm and a smaller share leading to Physical Harm or Death.
- Discrepancies in harm distribution reflect underlying differences in reporting expectations, not necessarily differences in risk or system quality.

Root Node Two: Primary Outcome

Most Common Outcomes

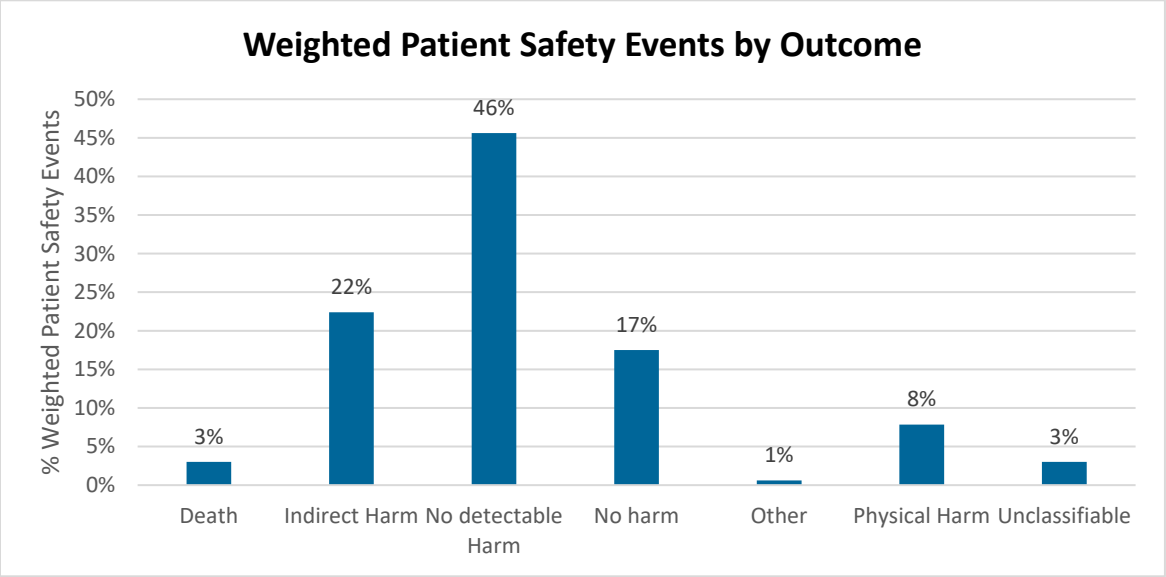

Figure 8: Patient Safety Event Outcome Distribution

Across all six years of the 1585 weighted patient safety events in the U.S. Organ Transplant System, the most common outcome is consistently No Detectable Harm, comprising more than 46% of all outcomes. No Detectable Harm is a classification that still represents an individual's exposure to error, even if harm could not be observed or quantified. These events suggest a need for improved follow-up and documentation to rule out latent effects. The next most frequent outcome is Indirect Harm, making up 22% of weighted errors. These events involve delays, confusion, procedural errors, or other issues that impact care quality but stop short of causing Physical Harm or Death. No Harm makes up 18% of weighted events. In these events, it is apparent that the patient experienced no harm.

Physical Harm and Death remain relatively infrequent outcomes. Physical Harm comprises approximately 8% of total errors, while Death appears in less than 5% of complaints.

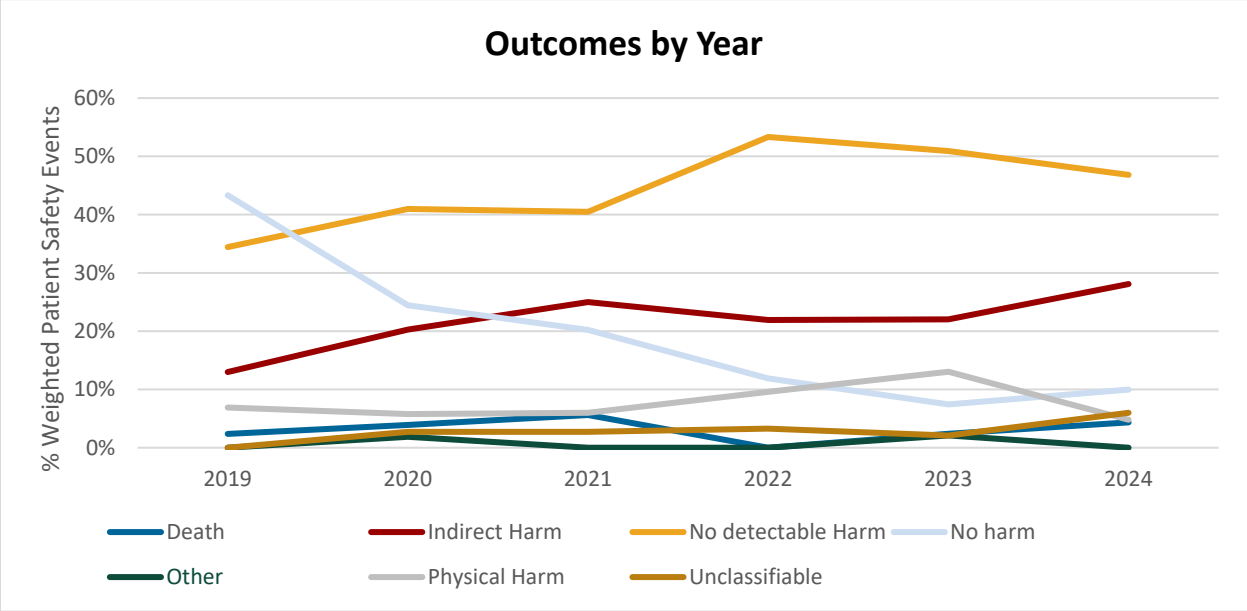

Figure 9: Outcomes by Year

Patient safety events resulting in No Detectable Harm and Indirect Harm have increased while weighted events resulting in No Harm have decreased sharply from 43% to 10%. In considering the rise in No Detectable Harm outcomes, it must be considered that this is the code used when no information about harm is reported or inferable. Indirect Harm is often inferable simply from type, for example an Allocation Out of Sequence event always causes Indirect Harm and so would not be affected by the exclusion of specific outcome information. However, No Harm is generally only selected when affirmative evidence of No Harm is found. This must be understood in the context of the observable increase in qualitative events without enough information to code, essentially as a limitation on the extent to which one can conclude No Harm events are actually decreasing. Similarly, the rise in Unclassifiable events in 2024 is not a reflection of novel outcomes but is instead attributable to declining report quality. This decline in qualitative detail comprises the HRSA's ability to code and analyze the dataset, and assess impact, and should be addressed with clearer reporting guidance and accountability mechanisms.

Causes of Different Outcomes

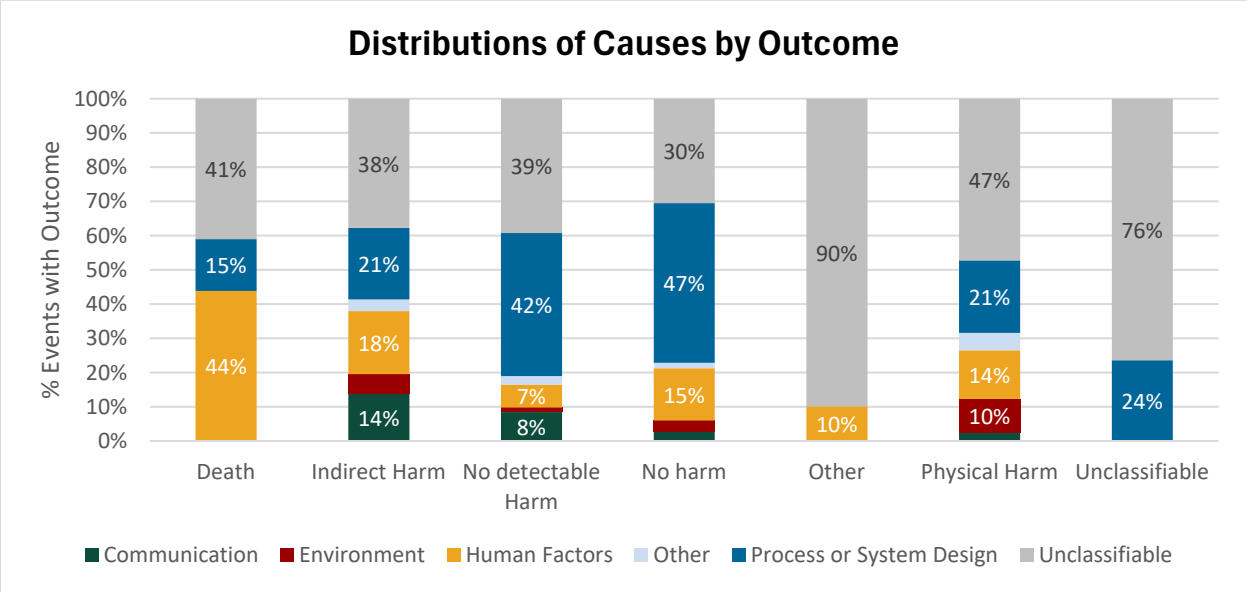

Figure 10: Distributions of Causes by Outcomes

Death is most commonly attributed to Human Factors (44%) and Unclassifiable causes (41%).

When Human Factors led to Death, in 21 weighted events, the Human Factor is most frequently a patient factor: in 68% of cases, death is caused by Human Factors. However, a substantial number of Human Factor causes are Unclassifiable at the subclassification level. While human factors are a major contributor to death outcomes in the weighted analysis, this is heavily influenced by OPTN policy requiring mandatory reporting of all living donor adverse events, which include donor deaths within two years of donation. In our sample analysis, living donor adverse events account for nearly half of all reported deaths. Importantly, many of the human factors linked to death in the living donor phase stem from causes not directly related to transplant processes (e.g., MVAs, drug overdose) but are nonetheless captured under current reporting requirements. As such, the high proportion of weighted living donor events, driven by regulatory reporting requirements, may amplify the apparent impact of human factors on death outcomes. This nuance is essential when interpreting the broader influence of human factors across all phases of transplant care.

While analysis at the subclassification level clarifies the role of Human Factors, the large proportion of Unclassifiable causes raises concern at the primary classification level. In particular, the fact that such a high percentage of Death events lack a clearly identifiable root cause reveals that event reports often omit critical contextual or causal information. This lack of specificity undermines the transplant system's ability to derive actionable insights and implement effective preventive measures. Improving the quality and completeness of narrative and structured data in patient safety event reporting should be a priority to ensure that the system learns from its most serious failures.

Indirect Harm weighted events show a broader distribution, with Unclassifiable (38%) and Process or System Design (21%) as leading causes. Notably, Communication issues account for 14% of these

weighted events, which is higher than for any other outcome category, indicating that breakdowns in how information is transmitted, interpreted, or responded to are a substantial contributor to severe, yet non-physical, consequences.

No Harm is similar in distribution to Indirect Harm, with Process or System Design (47%) and Unclassifiable (30%) as leading causes. Communication issues are also comparable, causing 15% of weighted events leading to No Harm.

Physical Harm weighted events are most frequently attributed to Unclassifiable causes (47%), a concerning pattern that mirrors the trend seen in Death outcomes. This again points to gaps in reporting quality, especially for higher severity outcomes.

Types of No Harm Events

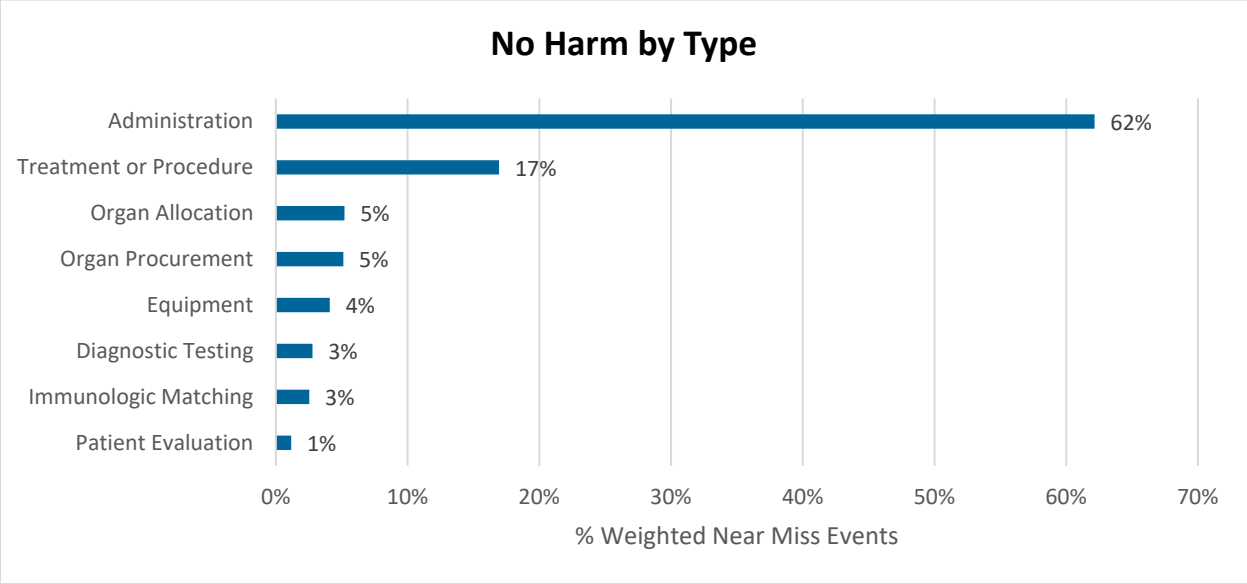

Figure 11: No Harm Type Distribution

The data shows that administrative issues are by far the leading cause of the 274 weighted No Harm events, accounting for 62% of the 274 events. This suggests that Errors in Documentation, Delays or Failures to Communicate or Report Information, while not resulting in direct harm, frequently reach the point of impact or potential impact.

Administration Event Subclassifications	
Subclassification	**% Weighted No Harm Events**
Delay in Communicating or Reporting Information	10%
Errors in Documentation	44%
Failure to Communicate or Report Information	1%
Other	6%

Table 4: Administration Event Type Subclassifications

Within this classification, consisting of 170 weighted events, the largest single contributor is Errors in Documentation, which alone accounts for 44% of weighted No Harm events. The high proportion highlights documentation as a major vulnerability in the transplant safety ecosystem, particularly given how critical accurate records are to patient care and organ handling.

Treatment or Procedure Event Subclassifications	
Subclassification	**% Weighted No Harm Events**
ABO Verification Protocol Error	9%
Inappropriate Vessel Storage	8%

Table 5: Treatment or Procedure Event Type Subclassifications

The second most common category is Treatment or Procedure related events (17% or 46 weighted events), with ABO Verification Protocol Errors (9%) and Inappropriate Vessel Storage (8%) identified as key areas. These are significant because they involve technical steps where mistakes could easily lead to more severe outcomes if not caught in time, reinforcing the need for strong process controls and redundancy.

Other classifications such as Organ Allocation (5%), Organ Procurement (5%), and Equipment Issues (4%) appear less frequently as causes of weighted No Harm events, but they still represent meaningful points of failure. Notably, miscommunication during organ offers and procurement injuries are among the listed contributing factors, signaling that operational handoffs in the transplant workflow, where communication is often fast paced and complex, remains a risk.

Key Takeaways

- Administrative events, especially documentation errors, dominate weighted near-miss incidents
- Procedural errors, including ABO verification errors and vessel storage, also represent a significant share of weighted incidents
- The diversity of No Harm sources spanning allocation, procurement, equipment and testing shows that vulnerabilities exist at multiple steps of the transplant workflow

Event Types Causing Indirect Harm

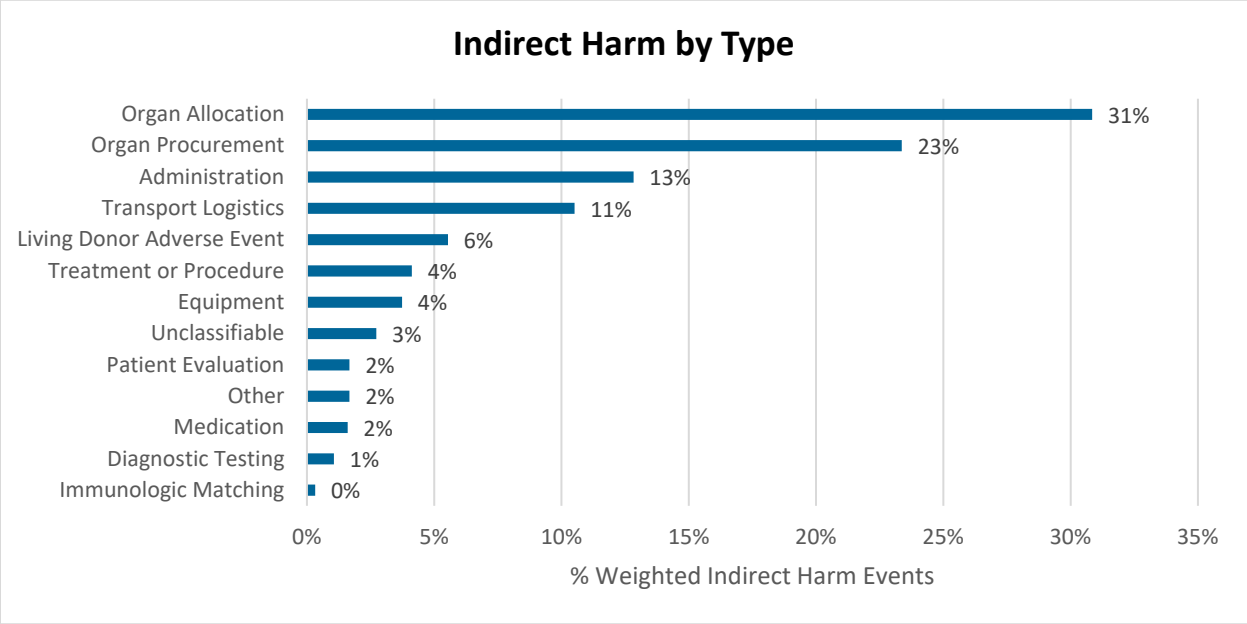

Figure 12: Indirect Harm Type Distribution

Indirect Harm events, which comprise 356 of the 1585 weighted events, are most commonly attributed to failures or breakdowns within the Organ Allocation (31%) and Organ Procurement (23%) pathways, together accounting for over 50% of all Indirect Harm reports. Organ Procurement and Organ Allocation incidents represent critical disruptions that affect the timeliness, quality, and success of transplantation efforts.

Within Organ Allocation, Allocation Out of Sequence (13%) and Late Declines of Organ Offers (8%) emerge as the most frequent subclassifications, suggesting persistent issues in coordination, timing, and decision making that prevent optimal organ utilization. Additionally, Organ Offer Miscommunication (2%) and other allocation-related errors (8%) further compromise successful procurement, allocation, and transplantation.

Within Organ Procurement, Organ Injury During Procurement or Packaging (20%) is the single most frequently cited subclassification across the entire Indirect Harm landscape. This points to surgical technique, preservation quality, or transport handling as high-risk moments for error, underscoring the need for increased standardization and vigilance.

Administration Errors (13%) including Failure to Communicate or Report Information (7%) and Errors in Documentation (5%) often trigger a cascade of downstream consequences, from misinformed clinical decisions to missed transplant opportunities. Transport Logistics (11%) also contribute significantly to Indirect Harm, particularly when organs are Lost in Transit or Delayed by Airlines (7% and 2% respectively). These delays not only risk organ viability but also introduce uncertainty into tightly timed transplant workflows.

Key Takeaways

- Organ Allocation (31%) is the top source of Indirect Harm, especially from Allocation Out of Sequence decisions and Late Declines
- Organ Procurement (23%) causes significant harm through Organ Injury During Recovery or Packaging

Root Node Three: Primary Phase

When Errors Tend to Occur

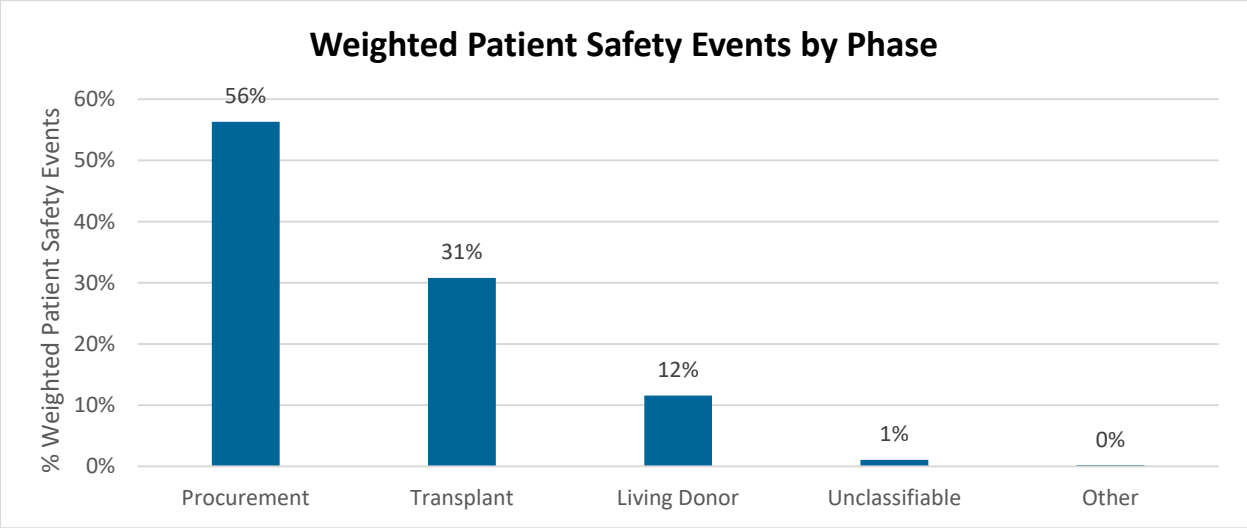

Figure 13: Patient Safety Event Phase Distribution

Of the 1585 weighted patient safety events, events tend to occur, with descending frequency, during the Procurement, Transplant and Living Donor Phase. Relatively few events occur in the Living Donor phase (12%). This reflects the relatively small number of living donors and is inflated by mandatory reporting of Living Donor Adverse Events. A small portion of weighted events happen during Other phases, phases which are identifiable but not defined in the Codebook. A majority of weighted events, 56%, happen during the Procurement phase. This reflects the characteristics of the phase: not every procurement leads to a transplant, many actors are involved, and there are many potential points of friction. Fewer weighted events occur during transplant, 31%, reflecting that the Transplant phase is relatively less complex than procurement from a systemic point of view. Few phases are unclassifiable, showing that the phase during which a patient safety event occurs is usually clear.

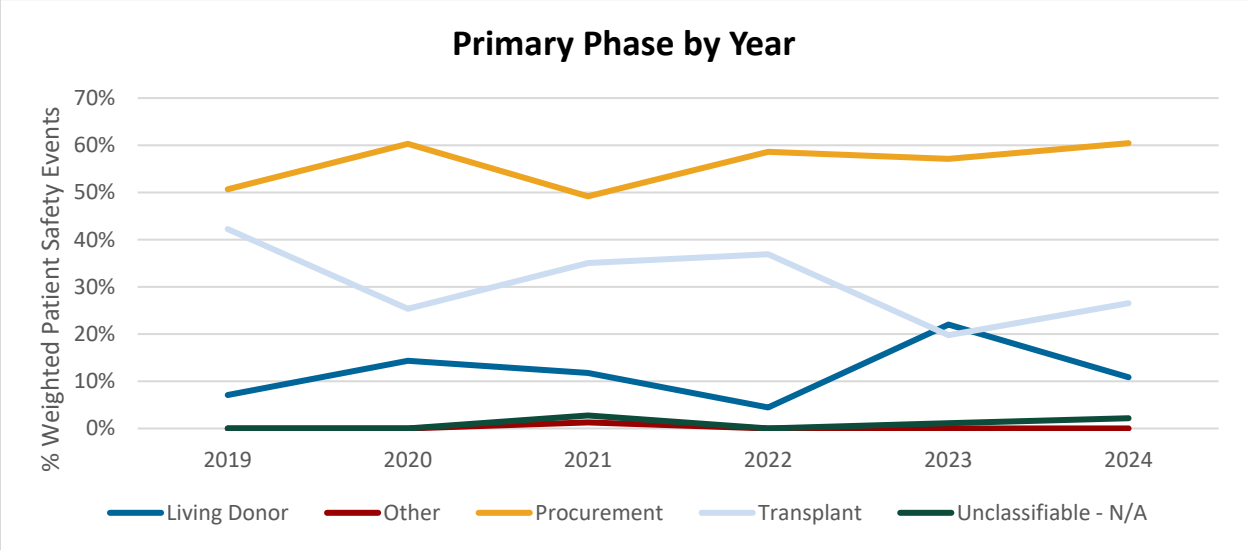

Figure 14: Primary Phase by Year

In each year between 2019 and 2024, except for 2021, Procurement phase weighted events represent at least 50% of weighted events. In 2021, weighted Procurement events, although not a majority, were still a large minority of weighted events, 49%. The portion of weighted events occurring during the Procurement phase increased over this time period, from 51% to 60%. The second most common phase is the Transplant phase, which decreased from 42% to 27%. Other phases, such as the Living Donor phase represent a smaller portion of weighted events.

When the phase an event occurs is apparent, it will likely either be a Transplant event or Procurement event due to the characteristics of the transplant ecosystem. The Procurement phase has more moving parts and actors involved, like OPOs, transplant centers, donor hospitals, labs, recovery surgeons, and couriers. These actors also interact, creating potential points of friction or errors during handoffs. Procurement also involves time pressure and encompasses many allocation errors, leading to a higher rate of errors.

Patient Safety Events Occurring During the Transplant Phase

Weighted Events During the Transplant Phases				
Pre-Transplant	**Transplant**	**Post-Transplant**	**Unclassifiable**	**Total**
203	115	137	32	488

Table 6: Weighted Events During the Transplant Phases

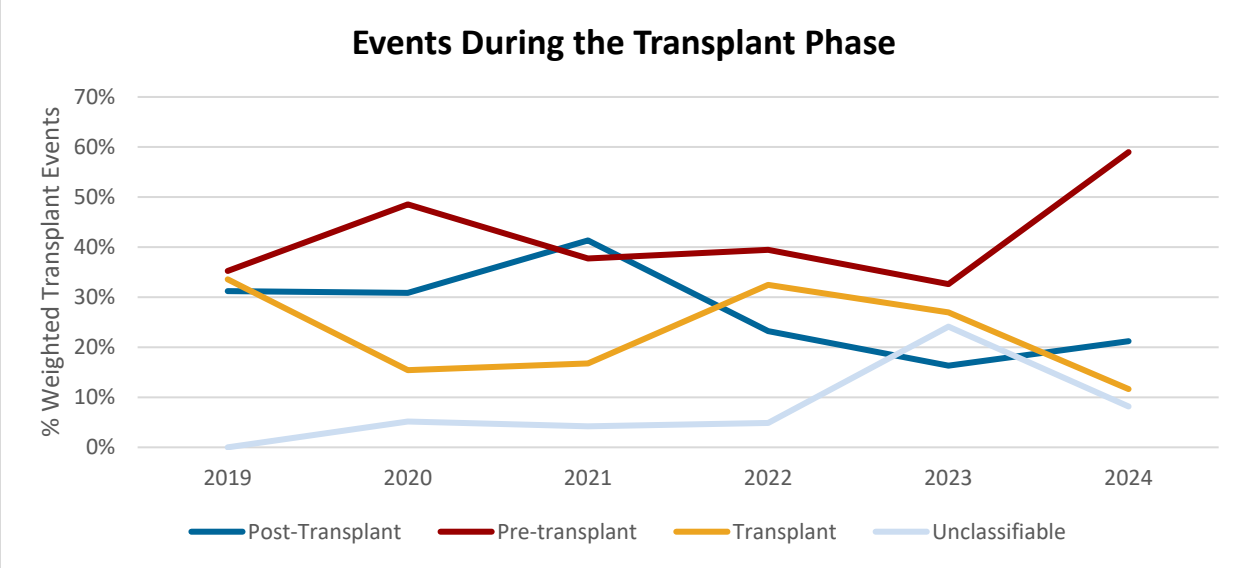

Figure 15: Transplant Phases by Year

Of the 488 weighted patient safety events occurring during the Transplant phase, Pre-Transplant events historically represented the majority and increased in frequency. While the Pre- and Post-Transplant phases both cover extensive periods of time, the Pre-Transplant phase involves more actors and therefore presents an inherently increased level of risk. Transplant and Post-Transplant events decreased in frequency.

Patient Safety Events During the Living Donor Phase

Weighted Events During Living Donor Phases			
Living Donor Evaluation	**Living Donor Surgery**	**Living Donor Recovery and Follow Up**	**Total**
28	121	35	184

Table 7: Weighted Events During Living Donor Phases

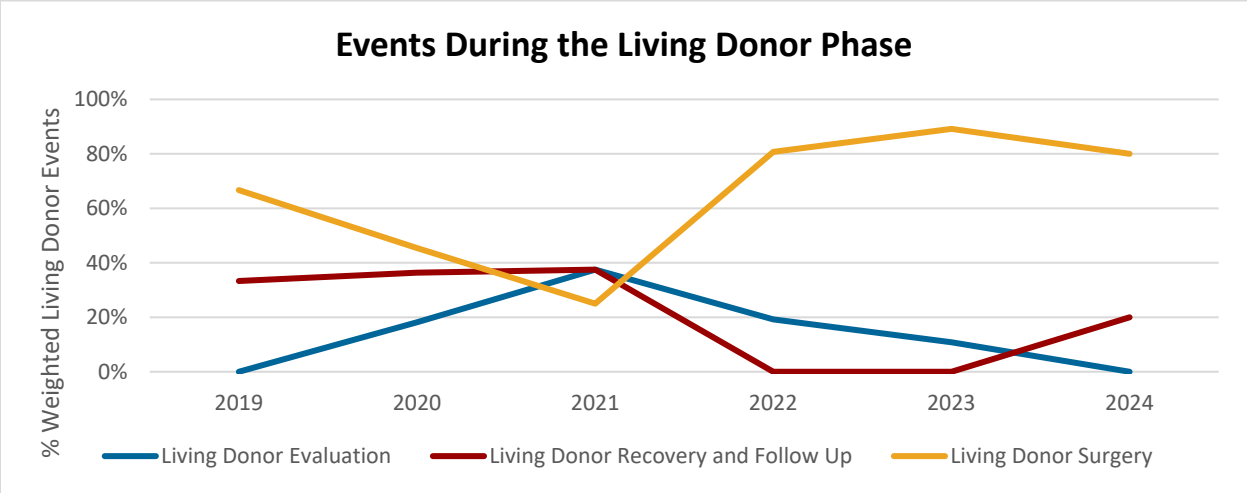

Figure 16: Living Donor Phases by Year

Of the 184 weighted patient safety events occurring during the Living Donor Phase, the Living Donor Surgery phase has historically represented most weighted events and is increasing. Further analyses identify this is driven by Aborted Recovery After Anesthesia Initiation Living Donor Adverse Events. Living Donor Evaluation phase weighted events decreased to zero since 2021 but increased from 2023 to 2024, while Living Donor Recovery and Follow Up weighted events exhibit less clear trends, decreasing from 2021 to 2022, then increasing from 2023 to 2024.

Key Takeaways

- Weighted errors tend to occur more frequently during the Procurement (56%) and Transplant (31%) phases rather than the Living Donor phases (12%).
- From 2019-2024, weighted errors during the Procurement Phase have increased by 9% while those during Transplant have decreased by 15%.
- During Transplant phases, weighted errors happened most frequently during the Pre-Transplant Phase (41%).
- During the Living Donor Phases, weighted errors recently happened most frequently during the Living Donor Surgery Phase (between 81% and 89% between 2022 and 2024).

Causes at Each Phase

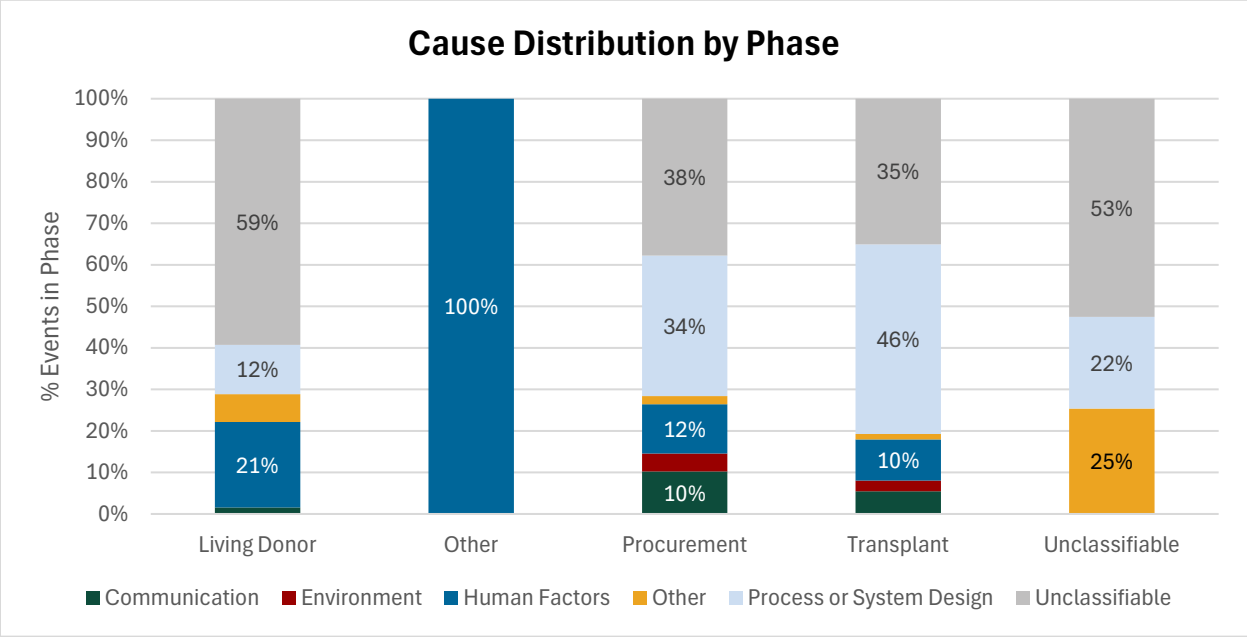

Figure 17: Cause Distributions by Phase

A substantial portion of weighted events, 40% of the 1585 weighted patient safety events, have unclassifiable causes. Of the 184 weighted Living Donor phase events, the majority of causes are Unclassifiable. For the 4 weighted Other events, cause is never Unclassifiable, but this must be conditioned on the finding that Other is a rare phase. Of the 893 weighted Procurement and 488 weighted Transplant phase events, a large portion of events, more than a third, have an Unclassifiable cause. When the phase during which an error occurred is Unclassifiable, the cause is also more likely than not Unclassifiable. Since phase is usually easy to determine, only Unclassifiable for 17 weighted events, when it is difficult to determine, the qualitative information is likely also lacking information about cause.

The issue of Unclassifiable causes speaks to the completeness and quality of the qualitative data. When discussing the Primary Cause root node, further analyses found that in later years, it became more difficult to determine cause because of the decreasing quality of information provided.

Of identifiable causes, Process or System Design causes contributed many patient safety events during most phases. Communication caused a small portion of weighted patient safety events across phases, between 0% and 10%. Procurement phase weighted patient safety events were the most likely to be caused by Communication issues, 10%, showing issues of communication between the many different actors involved during the Procurement phase. Environment, which included both cultural issues and equipment issues, caused very few weighted events overall, between 0% and 4%. Human Factors caused a relatively small portion of weighted errors overall. However, Human Factors caused a large portion of Living Donor Phase weighted events at 21% and all the Other phase weighted patient safety events.

Process or System Design Errors by Phase

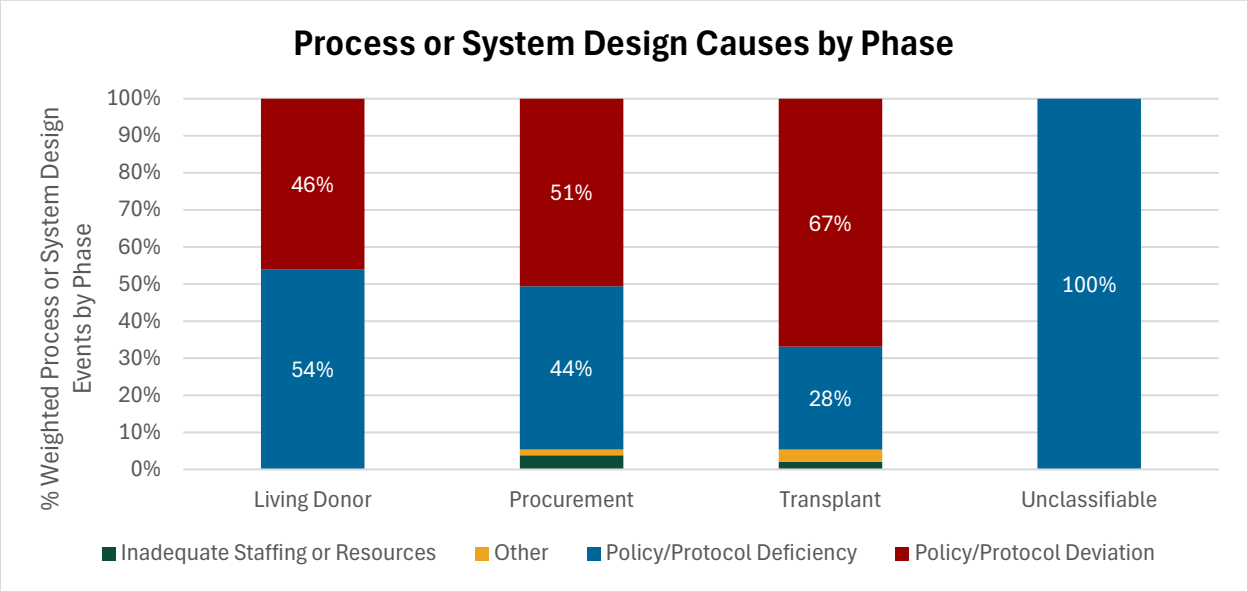

Figure 18: Process or System Design by Phase

Most weighted patient safety events due to Process or System Design causes across phases were caused by Policy/Protocol Deficiencies or Policy/Protocol Deviations. In the Living Donor and Procurement phases, the split between Deficiencies and Deviations is relatively even. However, during the Transplant phase, there were over twice as many Deviations as Deficiencies. This perhaps suggests the existence of more effective policies and protocols during the Transplant phase. Another notable finding is that patient safety events caused by Human Errors when the phase is Unclassifiable are always Policy/Protocol Deficiencies. Deficiencies are more serious than Deviations. Ideally, more serious events have more complete information.

Parties Responsible for Errors at Each Phase

Primary Responsible Parties at Each Phase						
Primary Responsible Party	**Living Donor**	**Other**	**Procurement**	**Transplant**	**Unclassifiable**	**% Weighted Events**
Clinical Lab	0%	100%	1%	1%	9%	2%
Donor Hospital	0%	0%	2%	0%	0%	1%
Histocompatibility Lab	0%	0%	6%	2%	0%	4%
OPO	0%	0%	50%	4%	22%	30%
Other	11%	0%	15%	0%	25%	10%
Patient	4%	0%	0%	0%	0%	0%
Transplant Center	45%	0%	8%	83%	44%	36%
Transportation	0%	0%	7%	1%	0%	4%
Unclassifiable	40%	0%	9%	9%	0%	12%
Total	100%	100%	100%	100%	100%	100%

Table 8: Primary Responsible Party by Phase

The Primary Responsible Party during each phase is largely dependent on the phase. During the Living Donor phase, transplant centers are responsible for a large minority of weighted errors, 45%, reflecting their prominent role in Living Donor transplants. Clinical labs are always the responsible party for events that occur during the Other phase but note this accounts for a small number of events.

During Procurement, the OPO is responsible for 50% of weighted events. This is reasonable as OPOs are central to procurement. However, the Procurement phase contains a substantial portion of weighted errors from other actors, like Recovery Surgeons, who make up a portion of the Other category.

The Transplant phase is dominated by patient safety events for which transplant centers are responsible in comparison to the Procurement phase, for which OPOs are primarily responsible. Transplant centers are responsible for 83% of weighted Transplant phase events. Other than a minority of processes like OPO disease reporting responsibilities, the Transplant phase is defined to include primarily events for which transplant centers are responsible.

Responsibility for Errors during Living Donor Phases

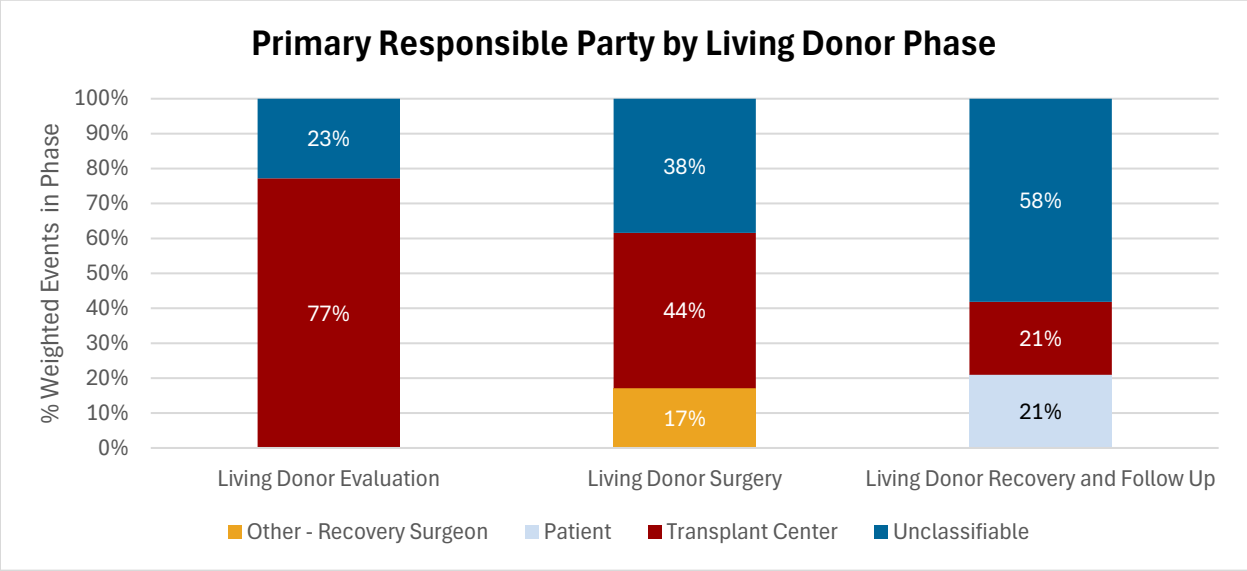

Figure 19: Primary Responsible Party by Living Donor Phase

Responsibility varies across the Living Donor phases in an expected fashion, based on the transplant lifecycle. During the Living Donor Evaluation phase, transplant centers are responsible for nearly 80% of the 28 weighted events. The transplant center is the core actor during this phase. During Living Donor Surgery (121 weighted events), the transplant center is responsible for fewer weighted events, with some events caused by recovery surgeons. This reflects that the weighted events in this phase subclassification are surgical in nature, including Aborted Recovery after Anesthesia Initiation and organ injuries. During the Living Donor Recovery and Follow Up phase (35 weighted events), there are weighted patient safety events caused by the patient or Unclassifiable causes. These are explained by Living Donor Adverse Events, specifically Living Donor Deaths within two Years Post-Donation.

Living Donor Adverse Events are patient safety events affecting living donors which require mandatory reporting. Notable common Living Donor Adverse Events are Aborted Recovery after Anesthesia Initiation and Living Donor Deaths Within Two Years Post-Donation. Especially in the case of Living Donor Deaths Within Two Years Post-Donation, causation is often either unclear or unlikely to be due to transplant related factors or due to patient behavior (ex. motor vehicle accidents).

Responsibility for Errors in Transplant Phases

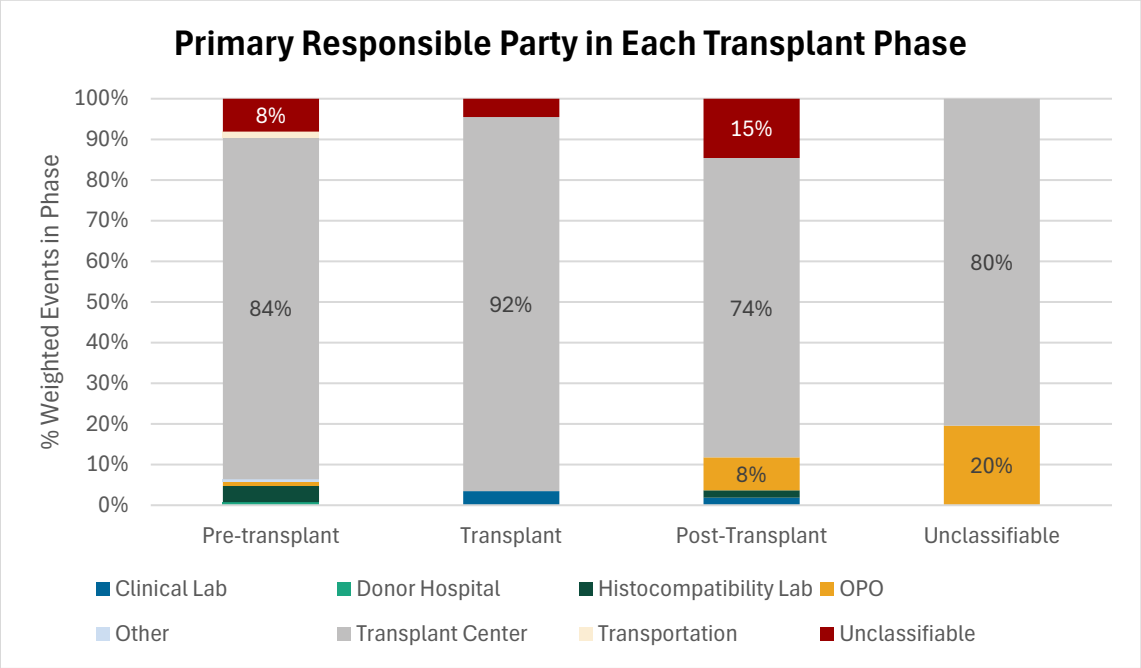

Figure 20: Responsible Party by Transplant Phase

Transplant Centers are responsible for most of the 488 weighted events across the transplant phases. Notably, OPOs are more likely to be Primary Responsible Parties after transplant (15% of 137 weighted events) than during (5% of 115 weighted events) or before (8% of 203 weighted events) transplant. After transplant, OPOs may fail to or be delayed in communicating information (e.g., potential disease transmission). The Unclassifiable phase subclassification here includes events that spanned across multiple phases of the transplant and procurement process. It is notable that these also include more events where OPOs are the primary responsible party.

Key Takeaways

- Transplant Centers are responsible for the largest proportion of weighted errors during the Living Donor (45%) and Transplant (83%) phases. During procurement, OPOs are responsible for the largest proportion of weighted errors (50%).
- Within Living Donor phases, Transplant Centers are responsible for a decreasing number of weighted events as phases proceed (77% during Evaluation but 21% during Follow Up). The opposite trend occurs for Unclassifiable Primary Responsible Parties, increasing from 23% during Evaluation to 58% during Follow Up.
- During Transplant, the Transplant Center is responsible for over 83% of weighted events during the Pre-Transplant, Transplant, and Post-Transplant phases.

Severe Harm Rate by Phase

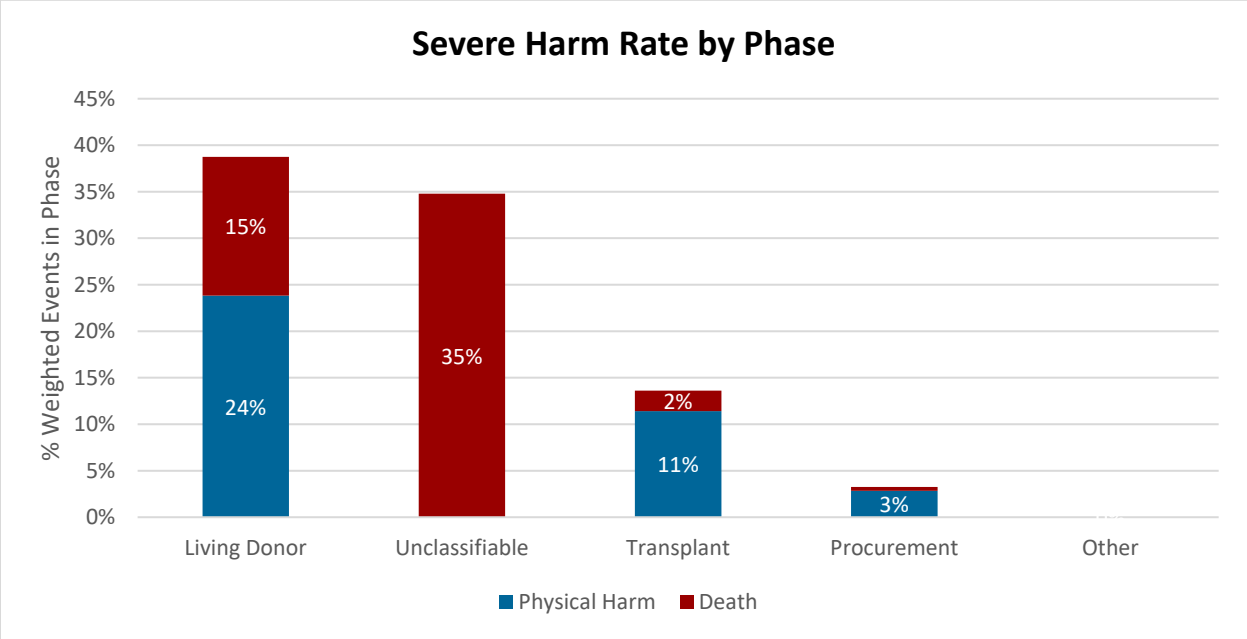

Figure 21: Severe Harm Rate by Phase

Weighted patient safety events during the Living Donor Phase (184 events) are the most likely to result in Physical Harm or Death. However, recall that weighted events during the Living Donor Phase are often Living Donor Adverse Events which must be reported but do not necessarily suggest wrongdoing by actors in the transplant ecosystem.

Procurement phase weighted events (893 weighted events) rarely result in severe harm, less than 5% of weighted outcomes, and a very low rate of death. This may suggest that either severe harm is effectively prevented or the types of errors that occur during the Procurement phase are unlikely to result in severe harm. Alternatively, Procurement phase events may rarely result in severe harm because they occur in the context of patients who are either being evaluated for eligibility or actively undergoing organ procurement. Unlike living patients, patients eligible for organ donation are unable to advocate for themselves when harm has been done. This unique status contributes to a structural blind spot in patient safety surveillance where adverse events related to patients undergoing organ procurement may be underreported or deprioritized, despite their potential to affect downstream clinical outcomes and systemic trust in the transplant process. Weighted Transplant phase events (488 events) are more likely to result in harm, at 13% of weighted events. However, this phase includes the transplant surgery itself and therefore presents an increased inherent risk of harm compared to Procurement.

Weighted events that occur during an unclassifiable phase were likely to be fatal, but this rate is a portion of only 17 weighted events. These represent a small portion of total deaths, but ideally events involving severe harm have complete information in their qualitative data.

Severe Outcomes in Living Donor Phases

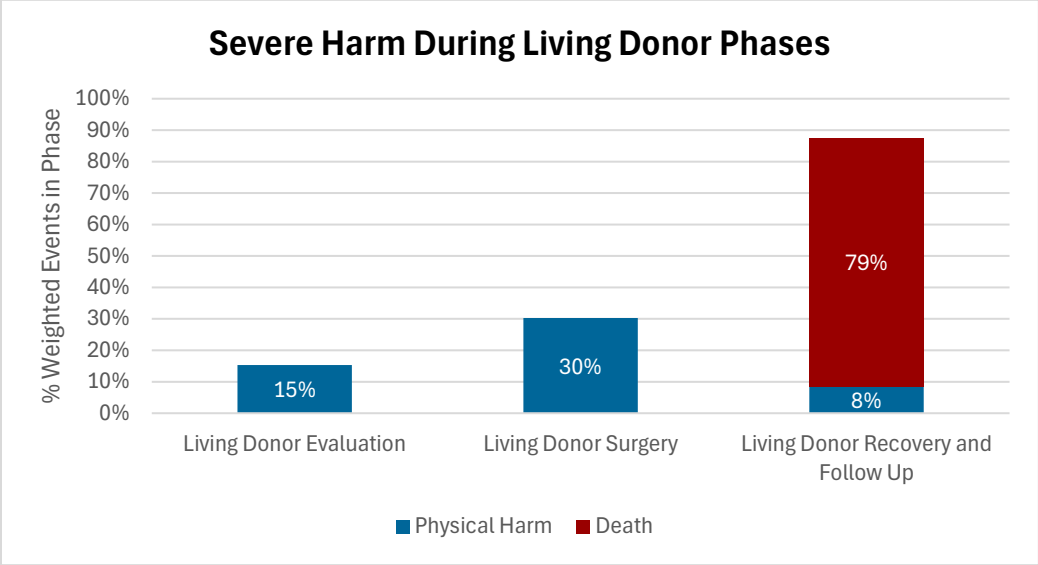

Figure 22: Severe Harm Rates During Living Donor Phases

Severe harm is defined as the Primary Outcomes of both Physical Harm and Death. The rate of severe harm is substantially higher during Living Donor Recovery and Follow Up (35 weighted events) than the preceding phases and Living Donor Recovery and Follow Up is the only Living Donor phase when deaths occurred. This shows the influence of the mandatorily reported Living Donor Death Within Two Years Post-Donation Living Donor Adverse Event. Similarly, the Living Donor Surgery phase (121 weighted events) includes Aborted Recovery Procedures After Anesthesia Initiation. This data reflects mandatory reporting requirements; it does not necessarily show serious issues with any Living Donor processes.

Severe Outcomes in Transplant Phases

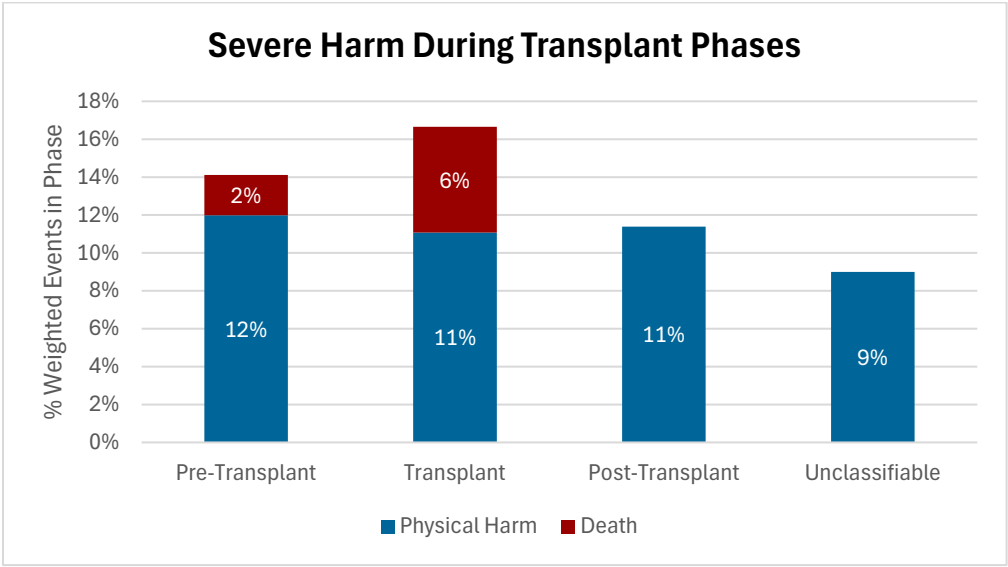

Figure 23: Severe Harm During Transplant Phases

When Transplant Phase (488 weighted events) subclassification is classifiable, severe harm, especially Physical Harm, happens at relatively comparable rates across phases. Rates of severe harm are highest during the Transplant phase at 17% of 115 weighted events, with 6% of patient safety events being fatal. However, this is a low number of weighted events, 13 resulting in physical harm and 6 in death respectively. Additionally, this phase includes the transplant surgery itself, contributing to a higher inherent risk during this phase.

Key Takeaways

- Living Donors have a high rate of physical harm and death (24%), but this is driven by Living Donor Adverse Event reporting requirements.
- Transplant weighted patient safety events are more likely to result in more physical harm and death than Procurement phase weighted events (13% vs. 3%).
- Transplant phase weighted events resulting in physical harm and death are spread evenly between Pre-Transplant (14%), Transplant (17%) and Post-Transplant (11%) phases.

Types of Errors during Transplant Phases

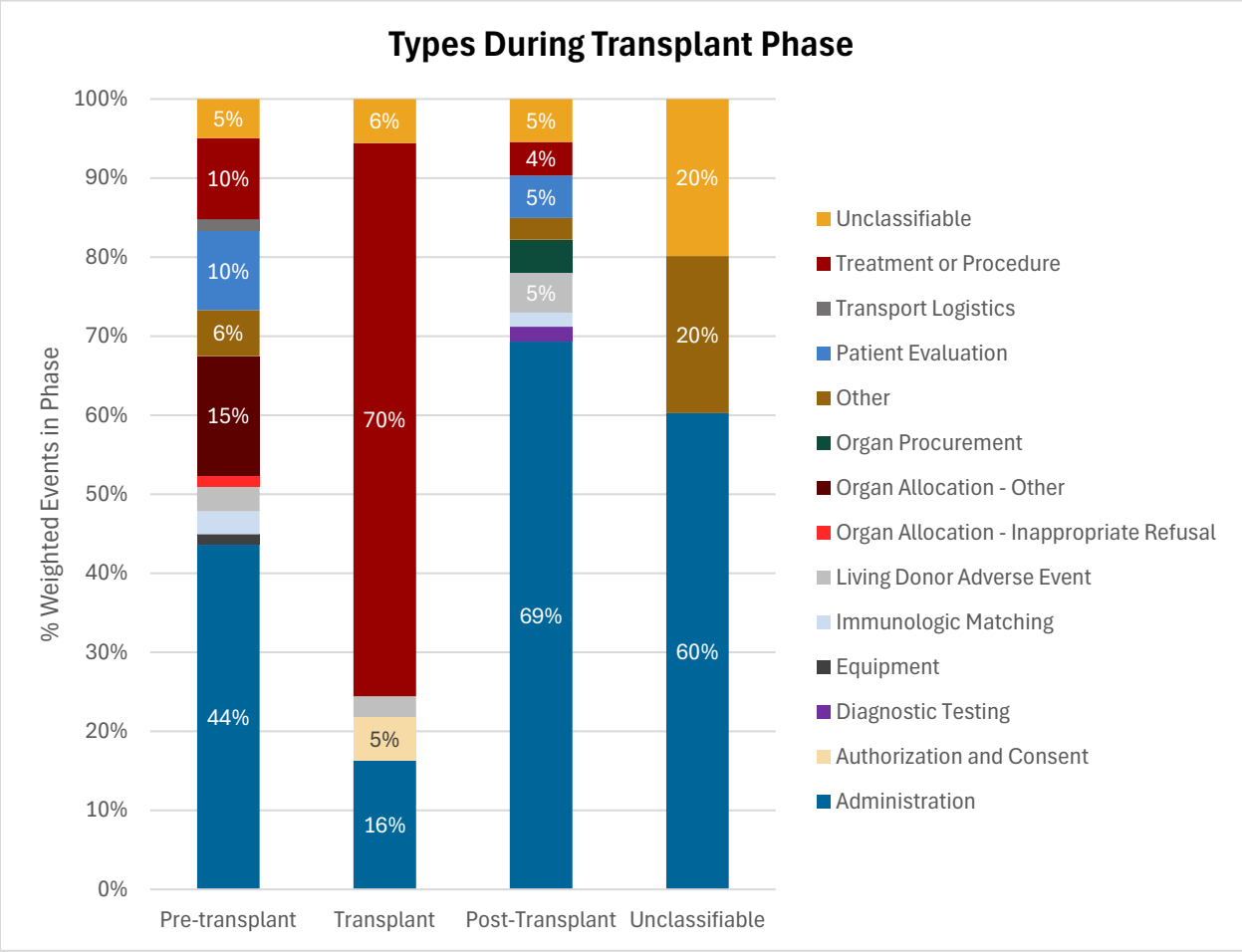

Figure 24: Type Distribution by Transplant Phase

Within the transplant phase, 203 weighted patient safety events occurred Pre-Transplant, 115 during Transplant, 137 Post-Transplant and 32 occurred during an Unclassifiable part of the Transplant phase. Pre-Transplant, the distribution of weighted patient safety event types is notably more varied than other Transplant phases, with the most common type, Administration, representing only 44% of weighted events and the remainder of weighted events being split across 11 different types. However, Organ Allocation weighted patient safety events represent 15%, and Treatment or Procedure represent 10% of events occurring in the Pre-Transplant phase. These latter two categories reflect organ allocation and ABO Verification Errors respectively. In transplant and post-transplant phases, a single type becomes more dominant. During the Transplant subclassification phase, Treatment or Procedure makes up 70% of weighted patient safety events where Inappropriate Vessel Storage accounts for the majority of events. During the Post-Transplant phase, Administration dominates, making up 69% of patient safety events. This reflects the characteristics of each phase. Pre-Transplant, there are more actors and varied responsibilities. The transplant center participates in allocation with OPOs, sends materials to and receives materials from labs, etc. Each additional actor introduces increased risk of error. On the contrary, the Transplant phase centers on surgery, and the follow up phase focuses on post-transplant patient treatment.

Treatment or Procedure Errors During the Transplant Subclassification Phase

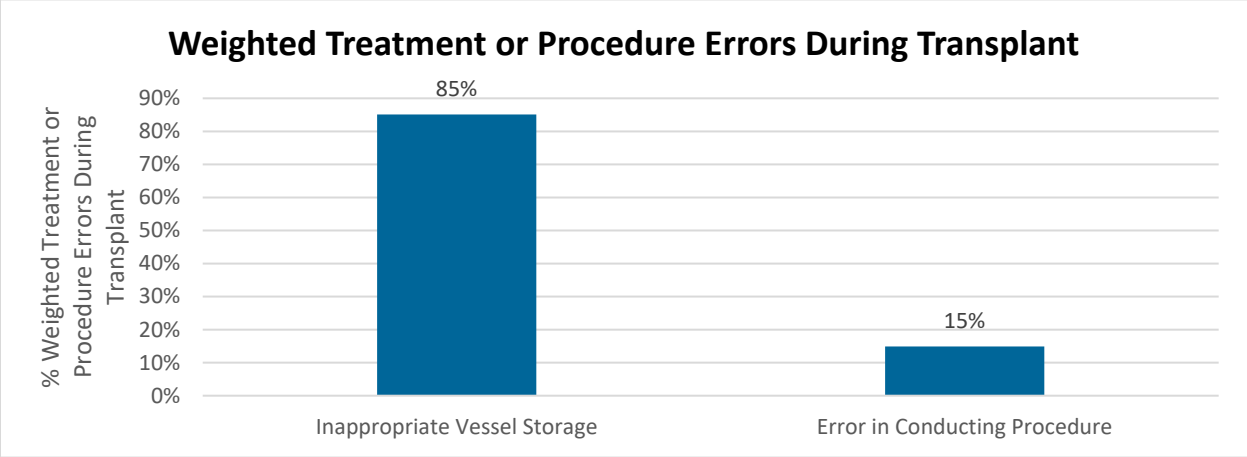

Figure 25: Treatment or Procedure Errors During Transplant

There were 81 weighted Treatment or Procedure events during the Transplant phase. Treatment or procedure errors tend to be associated with Inappropriate Vessel Storage events rather than Errors in Conducting Procedure. The positive interpretation is that issues like complications or mistakes made while performing surgeries are rare. However, there is a notable rate of failing to store donor vessels in accordance with OPTN Policy.

Post-Transplant Administration Errors

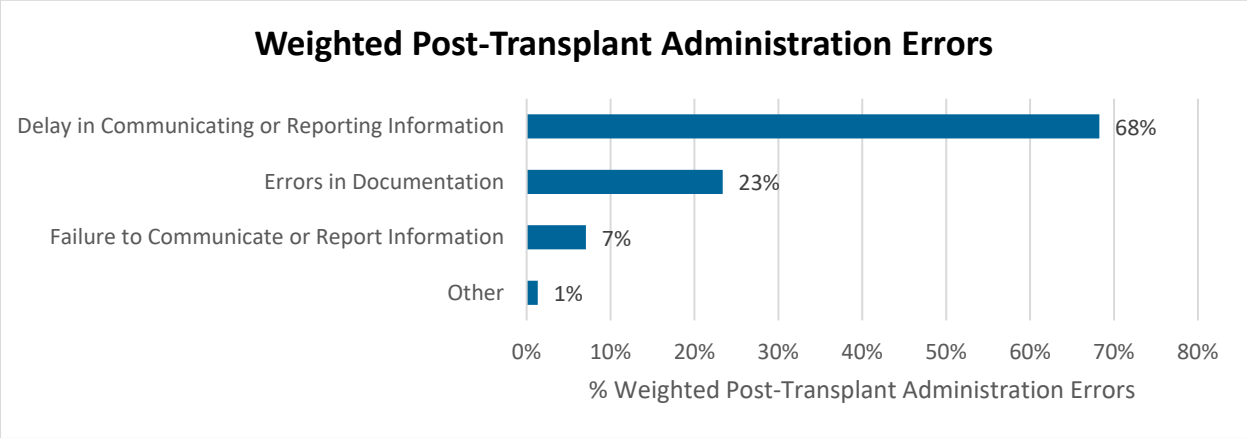

Figure 26: Post-Transplant Administration Errors

There were 94 weighted administration errors during the Post-Transplant phase. Post-Transplant errors are mostly Delays in Communicating or Reporting Information. These are often late reports of transmissible diseases to OPOs and the OPTN by transplant centers.

Key Takeaways

- Weighted Administration errors were most common during the Pre-Transplant (44%) and Post-Transplant phase (69%), while during Transplant, Treatment or Procedure errors were most common (70%). However, errors in Pre-Transplant varied in type.
- Weighted Treatment or Procedure errors during Transplant were overwhelmingly Inappropriate Vessel Storage errors, (85%).
- Weighted Post-Transplant Administration errors were largely Delays in Communicating or Reporting Information (68%).

Types of Errors During Living Donor Phases

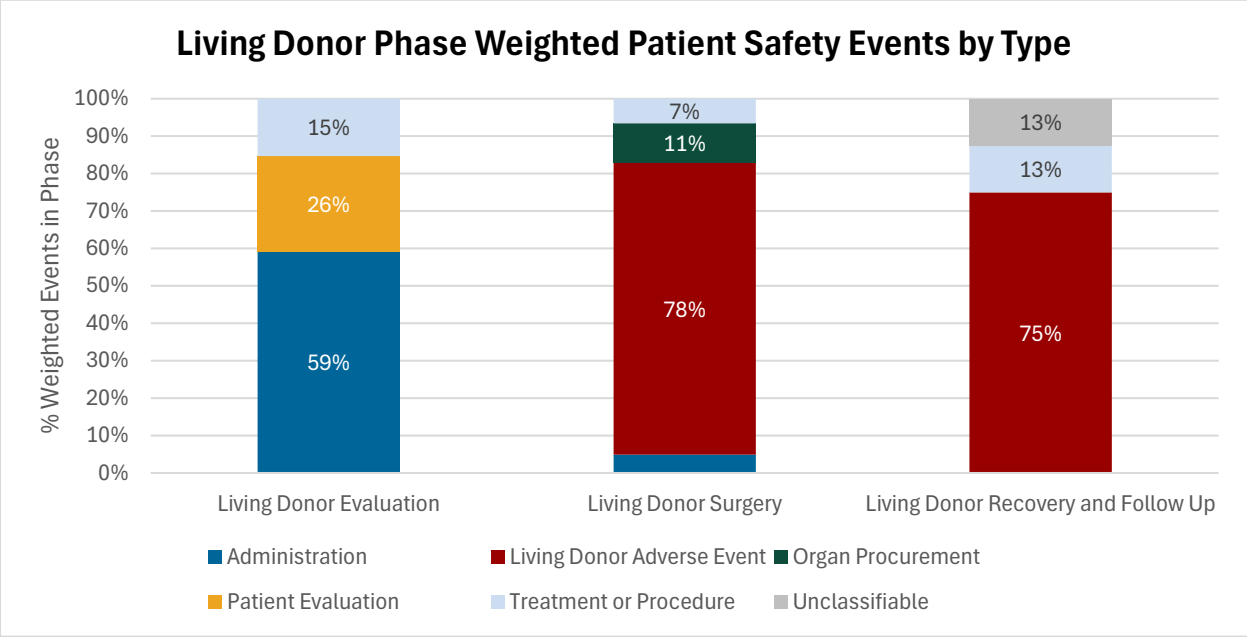

Figure 27: Type Distribution by Living Donor Phase

Within the Living Donor Phases, 28 weighted events occurred during Living Donor Evaluation, 121 occurred during Living Donor Surgery, and 35 occurred during Living Donor Recovery and Follow Up. Living Donor phase patient safety events consist of a small subset of types. During the Evaluation phase, there are Treatment or Procedure, Patient Evaluation, and Administration errors. However, during Living Donor Surgery and Living Donor Recovery and Follow Up, events are dominated by Living Donor Adverse Events. This reinforces prior discussion – Living Donor phase analysis is heavily influenced by mandatory reporting of Living Donor Adverse Events.

Living Donor Evaluation Events

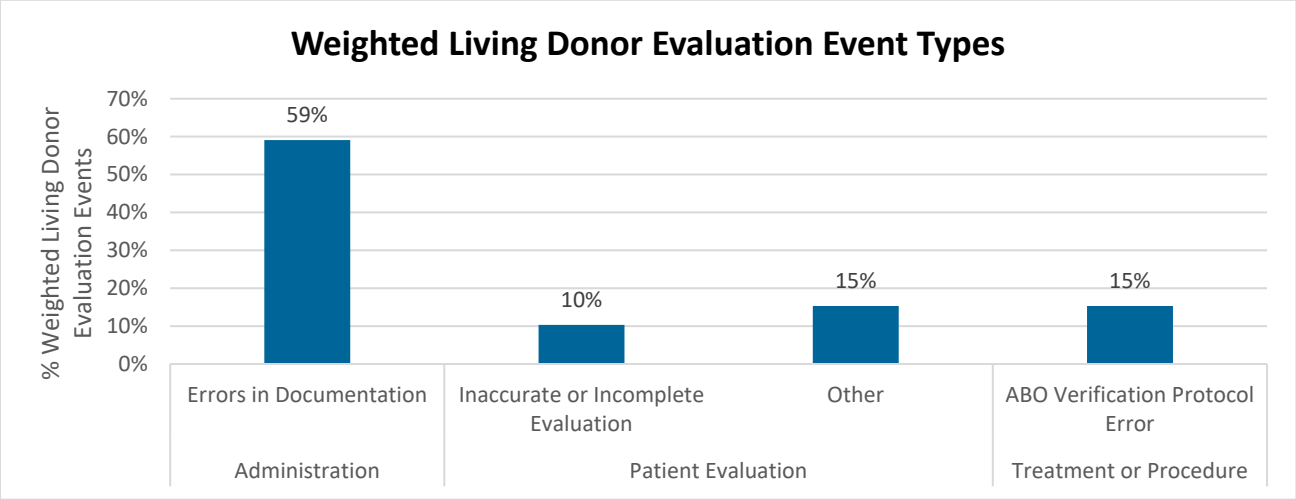

Figure 28: Living Donor Evaluation Event Types

Errors in Documentation make up a majority of the 28 weighted errors that occur in the Living Donor Evaluation phase, 59%. Errors in Documentation is a broad category, encompassing errors ranging from near miss typos to improperly recording crucial information in health records (e.g., failing to accurately record a donor's HLA typing). The unifying factor in many documentation error events is an ineffective verification process. There are also substantial numbers of more serious errors during the Living Donor Evaluation Phase, like ABO Verification Protocol Errors.

Living Donor Surgery

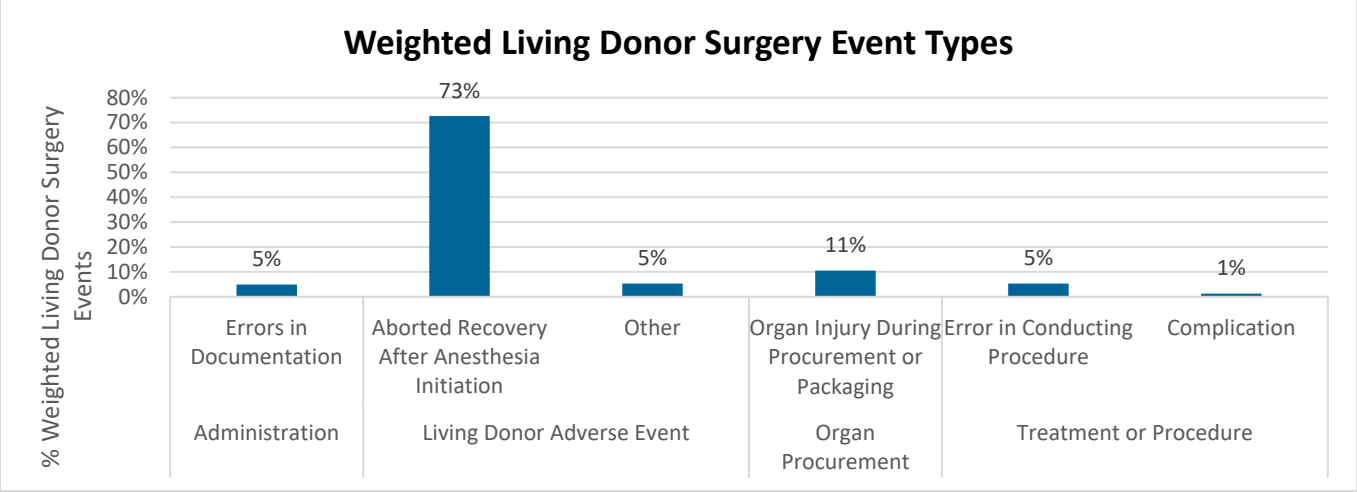

Figure 29: Living Donor Surgery Event Types

During living donor surgeries, the most common patient safety events of the 121 weighted events are Aborted Recoveries after Anesthesia Initiation. This reflects mandatory reporting requirements. However, other complications during Living Donor Surgery were rare, representing 1% of Living Donor Surgery weighted events.

Living Donor Recovery and Follow Up

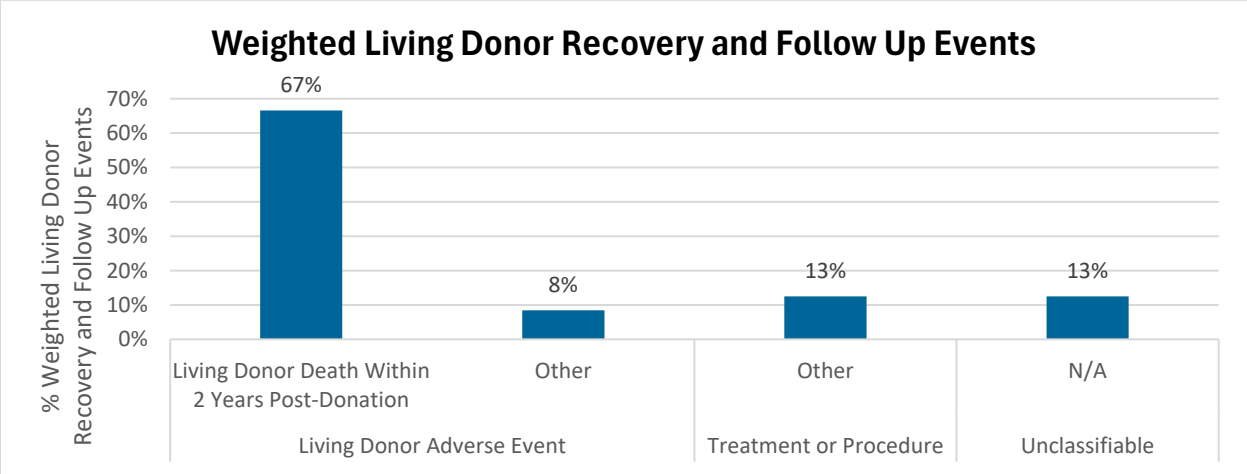

Figure 30: Living Donor Recovery and Follow Up Event Types

Living Donor Deaths within 2 Years Post-Donation compose most of the 35 weighted Living Donor Recovery and Follow Up phase events (6%). These were often due to patient behavior reasonably believed to be unrelated the transplant ecosystem, like motor vehicle accidents. However, in at least one case, a patient died of a drug overdose, and in another, a patient died of suicide.

Key Takeaways

- Weighted Administration errors were most common during the Living Donor Evaluation Phase (59%) while during Surgery and Recovery and Follow Up, Living Donor Adverse Events were most common (78% and 74%).
- During Living Donor Evaluation, most weighted Administration events were Errors in Documentation (59%)
- During Living Donor Surgery, the most common weighted event Primary Type was Aborted Recovery After Anesthesia Initiation, 73% of Living Donor Adverse Events.
- During Living Donor Recovery and Follow Up, Living Donor Deaths Within Two Years Post-Donation were the most common type, 67% of Living Donor Adverse Events.

Types of Errors During the Procurement Phase

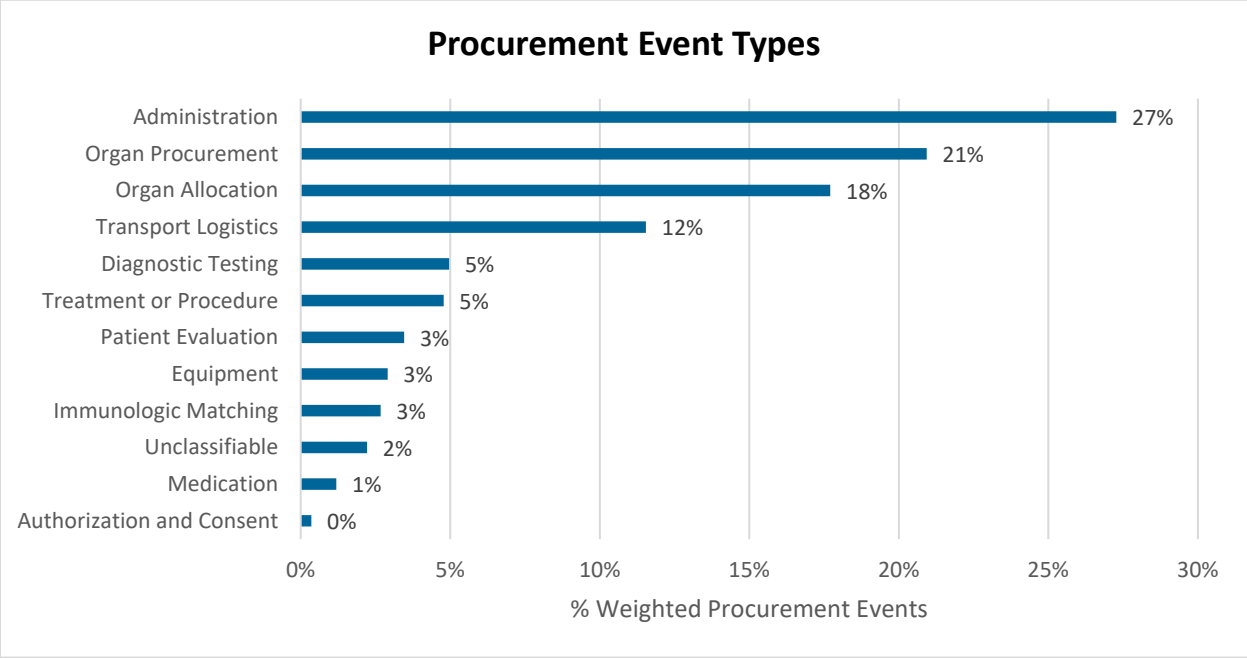

Figure 31: Procurement Event Type Distribution

The 863 weighted patient safety events during the Procurement phase are varied in type, reflecting the complexity of the Procurement phase. However, four types of patient safety events, a third of the event type categories, make up 75% of patient safety events during the procurement phase. These are Administration, 27%, Organ Procurement, 21%, Organ Allocation, 18%, and Transport Logistics, 12%.

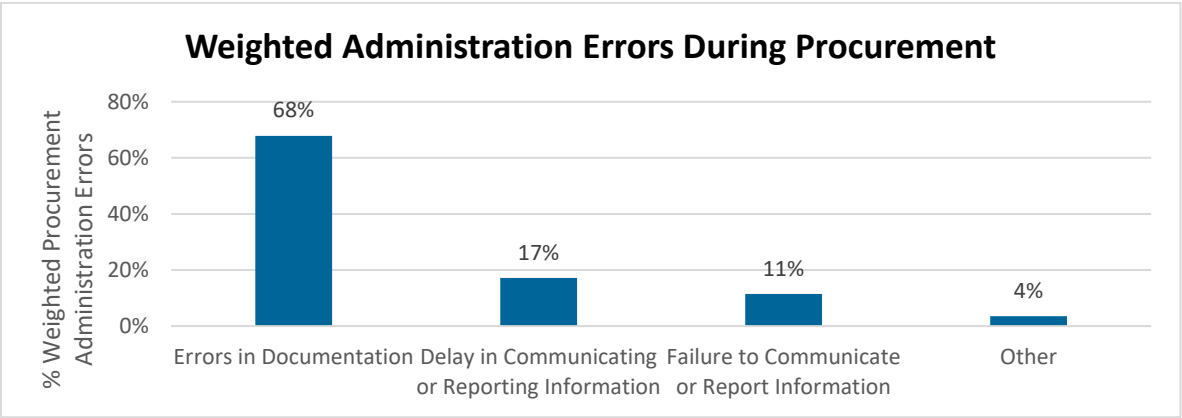

Figure 32: Procurement Phase Administration Event Types

Errors in Documentation make up a substantial majority of Procurement errors (68% or 243 weighted events). An Error in Documentation can represent the first in a chain of increasingly serious errors, like mislabeling a package containing an organ so it gets sent to the wrong transplant center. Errors in Documentation often represent failures to effectively verify information. Delays and Failures to

Communicate information make up a substantial minority of Procurement phase weighted patient safety events, nearly 30%. These show difficulties surrounding communication during the Procurement phase.

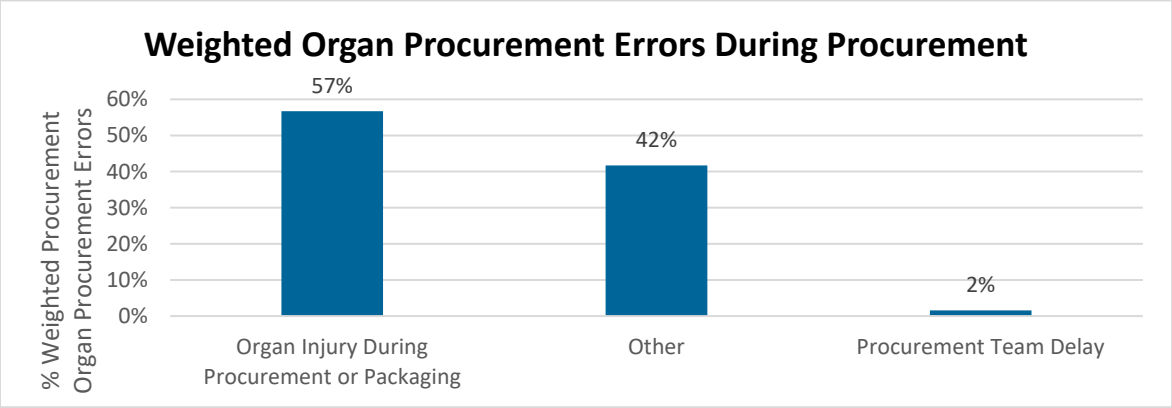

Figure 33: Procurement Phase Organ Procurement Errors

Most of the 187 weighted Organ Procurement errors during the Procurement Phase are organ injuries (57%). These are serious, as they often lead either to reallocation or to organ non-use, and they are often reported without a clear cause. Other organ procurement injuries make up a large minority (42%) of organ procurement injuries. Other Organ Procurement weighted events, 84 weighted events, frequently include kidney laterality swaps, sometimes resulting in transplant centers receiving the wrong kidney. Procurement Team Delays are rare, representing 2% of weighted Organ Procurement errors.

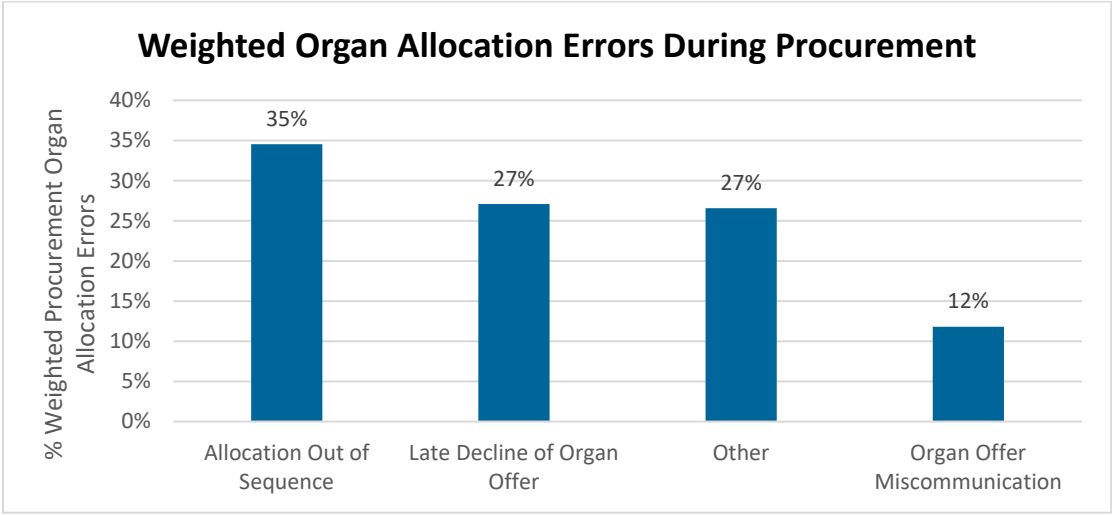

Figure 34: Procurement Phase Organ Allocation Errors

Allocation out of Sequence and Late Declines together make up a substantial proportion of the 158 weighted organ allocation events during procurement, 62%. This reinforces known policy concerns surrounding allocation out of sequence and late declines, notably concerns about OPOs gaming the

organ allocation system and transplant centers causing unnecessary organ discard, respectively. Other Organ Allocation patient safety events along with Organ Offer Miscommunication make up smaller but not insubstantial portions of Organ Allocation events during Procurement.

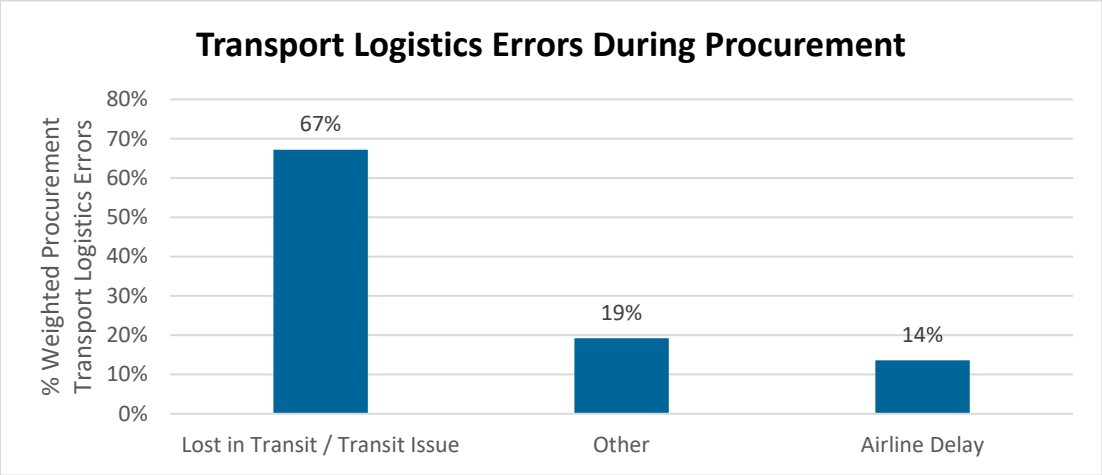

Figure 35: Procurement Phase Transport Logistics Issues

The 103 weighted Transport Logistics events during Procurement are mostly composed of organs being lost in transit or experiencing other issues unrelated to airline delays, like couriers getting lost and delivering organs to the wrong location. However, this may reflect the underlying distribution of the ways in which organs are transported.

Key Takeaways

- Four event types make up 75% of Patient Safety events during Procurement: Administration, (27%) Organ Procurement (21%), Organ Allocation (18%), and Transport Logistics (11%).
- Of the Administration errors during procurement, Errors in Documentation are most common, 68% of Administration events.
- Of the Organ Procurement errors during Procurement, Organ Injury During Procurement or Packaging Make up 56% of Organ Procurement Events. However, Other events, like kidney laterality swaps, make up 41% of Organ Procurement events.
- Organ Allocation Events during the Procurement Phase are made up largely of Allocations Out of Sequence (35%) and Late Declines (27%). However, Other events make up a substantial portion of Organ Allocation events during Procurement.
- Transport Logistic Events during Procurement are made up largely of Lost in Transit / Transit Issue events (67%).

Event Types Driving Discard during the Procurement Phase

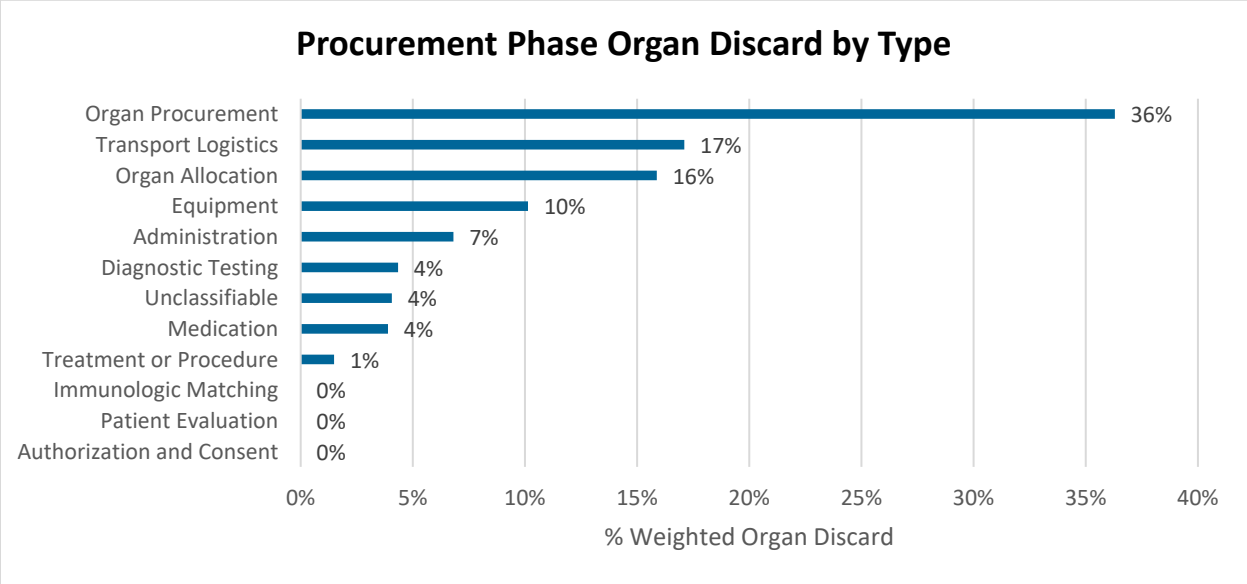

Figure 36: Procurement Phase Organ Discard by Type

During the Procurement phase, organs were discarded in 146 weighted patient safety events. There is variation in the errors that lead to organ discard. However, four patient safety event types, a third of the twelve observed, accounted for 79% of organ discard due to Procurement phase weighted patient safety events. These were Organ Procurement, 36%, Transport Logistics, 17%, Organ Allocation, 16%, and Equipment, 10%. Note that Organ Procurement led to more organ discard than the next two types of patient safety events combined.

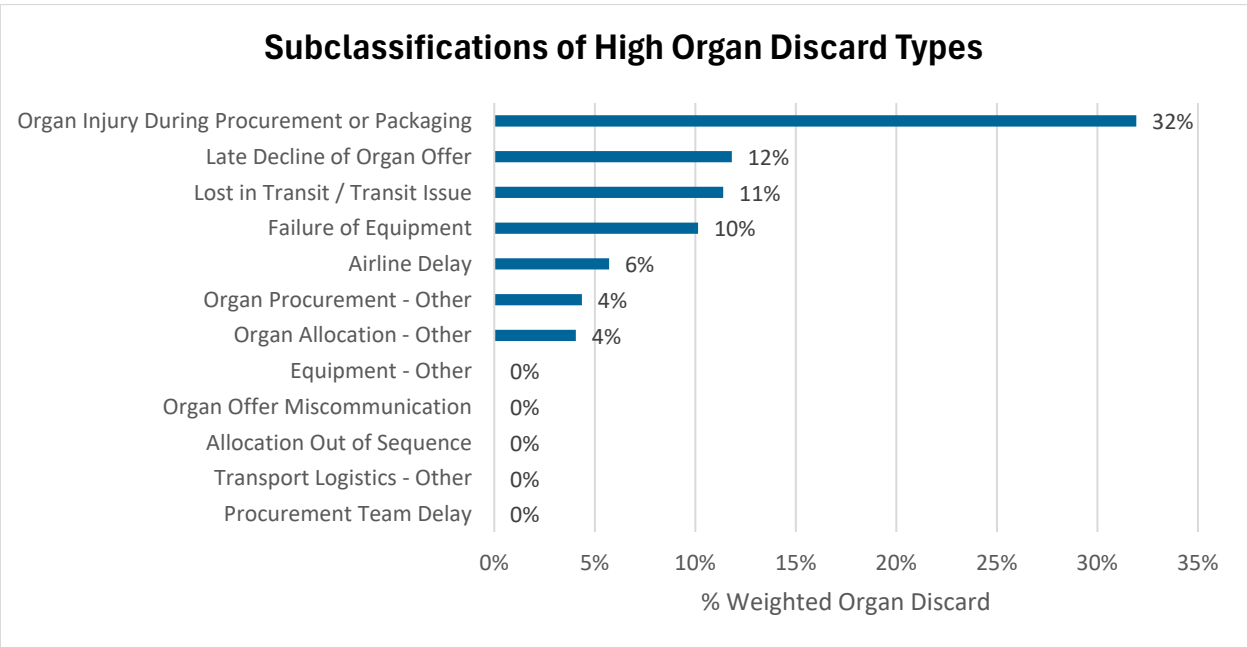

Figure 37: Subclassifications of High Organ Discard Types

Organ non-use was analyzed using the "Non-Utilization?" column included in the raw dataset provided by the OPTN contractor. Together organ injuries, late declines, and issues of organs lost in transit led to over 50% of organ non-use due to Procurement phase weighted patient safety events. Notably, Organ Injury during Procurement or Packaging accounted for 32% of organ discard due to weighted errors during the Procurement phase, nearly as much as the next three types combined. In practice, organ injury occurred more often during the recovery surgery rather than during the packaging process. Also notable is that five subclassifications, Procurement Team Delay, Transport Logistics – Other, Allocation Out of Sequence, Organ Offer Miscommunication, and Equipment – Other, did not contribute to Organ Discard due to errors during the Procurement phase at all.

Key Takeaways

- During Procurement, four weighted patient safety event Primary Types (33% of the 12 possible types) account for 79% of organ discard reported with weighted patient safety events: Organ Procurement (36%), Transport Logistics (17%), Organ Allocation (16%), and Equipment (10%).
- In terms of subclassifications, three subclassifications caused 50% of weighted discard: Organ Injury During Procurement or Packaging (32%), Late Decline of Organ Offer (12%), Lost in Transit / Transit Issue (11%).

Root Node Four: Primary Type

Common Types of Events

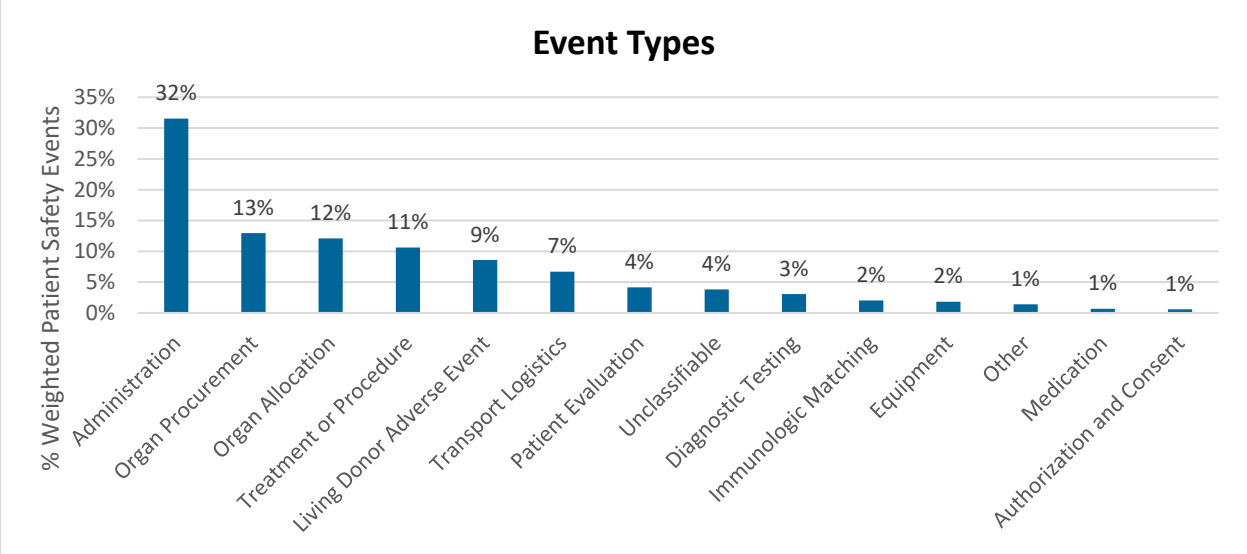

Figure 38: Patient Safety Event Type Distribution

About 75% of the 1585 weighted patient safety events are comprised of four Primary Type Primary Classifications, about 30% of the 14 possible Primary Classifications for the Primary Type root node: Administration, Organ Procurement, Organ Allocation, and Treatment or Procedure.

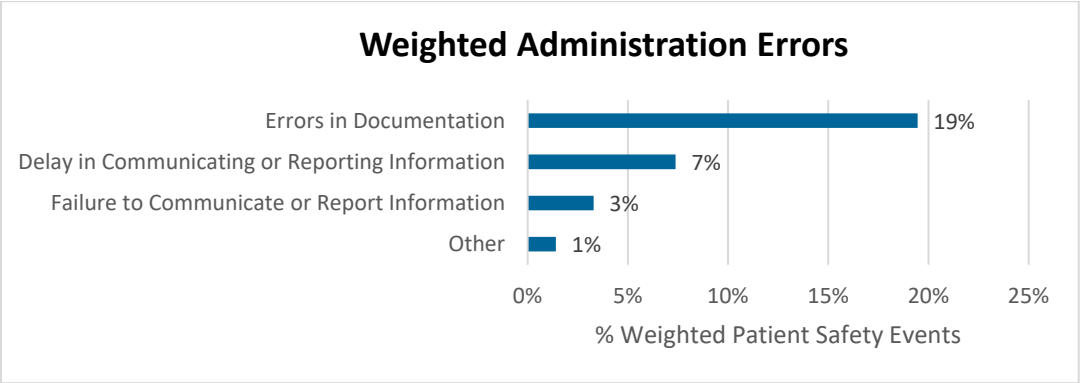

Figure 39: Administration Event Types

Administration errors make up 500 weighted safety events. Weighted patient safety events are frequently Errors in Documentation, 308 weighted patient safety events. This is a broad category spanning any error in documentation, from typos while listing patients to mislabeling organ pumps. However, Delays in and Failures to Communicate or Report Information together make up 10% of weighted Patient Safety events. These include a variety of communication issues, like failing to timely report infectious diseases and communication issues between OPOs and Transplant Centers during Pre-Transplant.

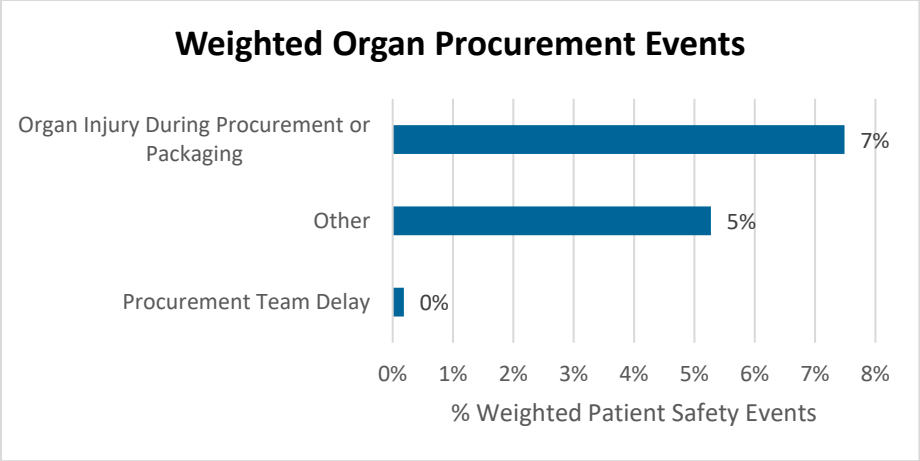

Figure 40: Organ Procurement Events

Organ Procurement events made up 204 weighted patient safety events. The majority of organ procurement events are organ injuries, 119 weighted events. However, a portion of organ procurement issues were Other organ procurement issues that were identifiable but did not align with the Codebook's coding system. These included events like kidney laterality swaps unexplained by some other upstream error.

Organ Allocation Events

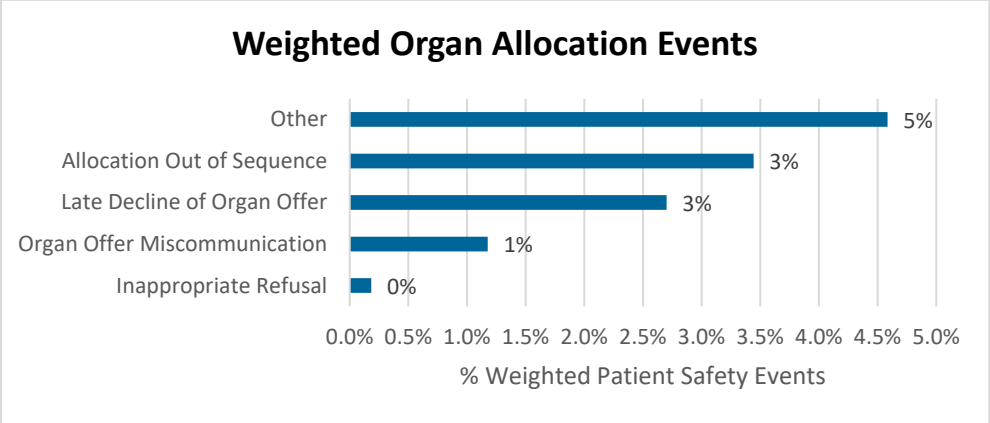

Figure 41: Organ Allocation Events

There were 192 weighted Organ Allocation events. Common Organ Allocation events include Allocations Out of Sequence and Late Declines. These are known issues in the transplant ecosystem.

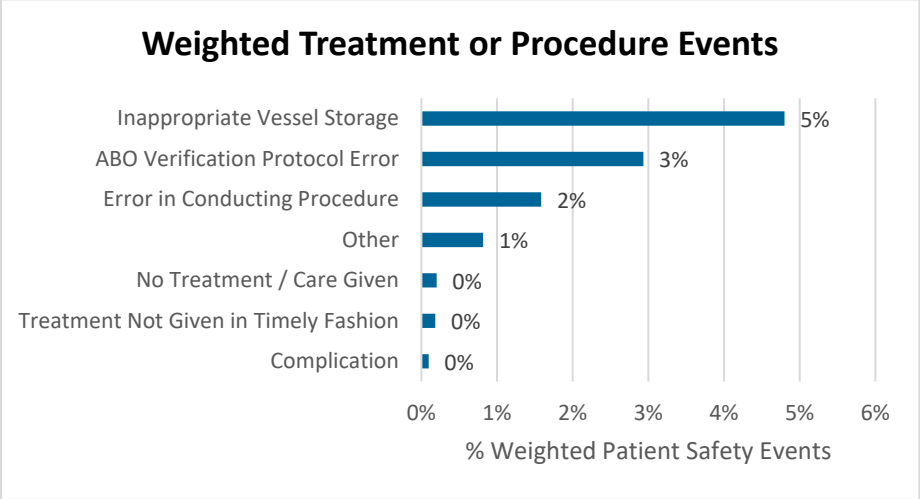

Figure 42: Treatment or Procedure Events

There were 168 Treatment or Procedure weighted events. Treatment or Procedure errors tend to be primarily Inappropriate Vessel Storage or ABO Verification Protocol errors, together making up nearly 75% of weighted Treatment or Procedure errors.

Patient Safety Event Types by Year

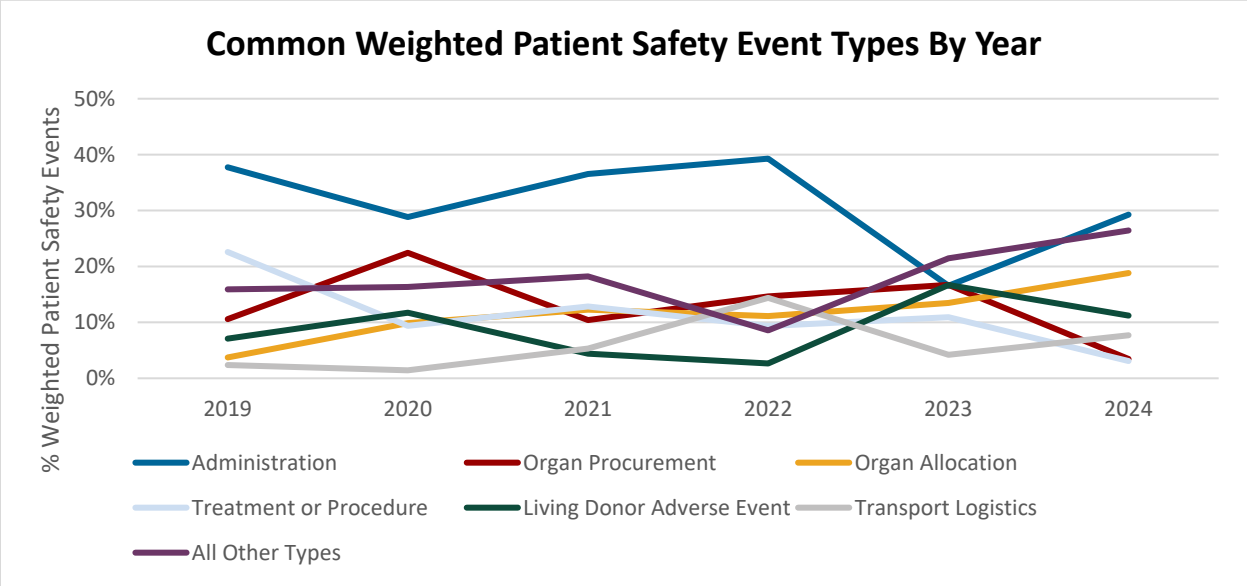

Figure 43: Common Patient Safety Event Types by Year

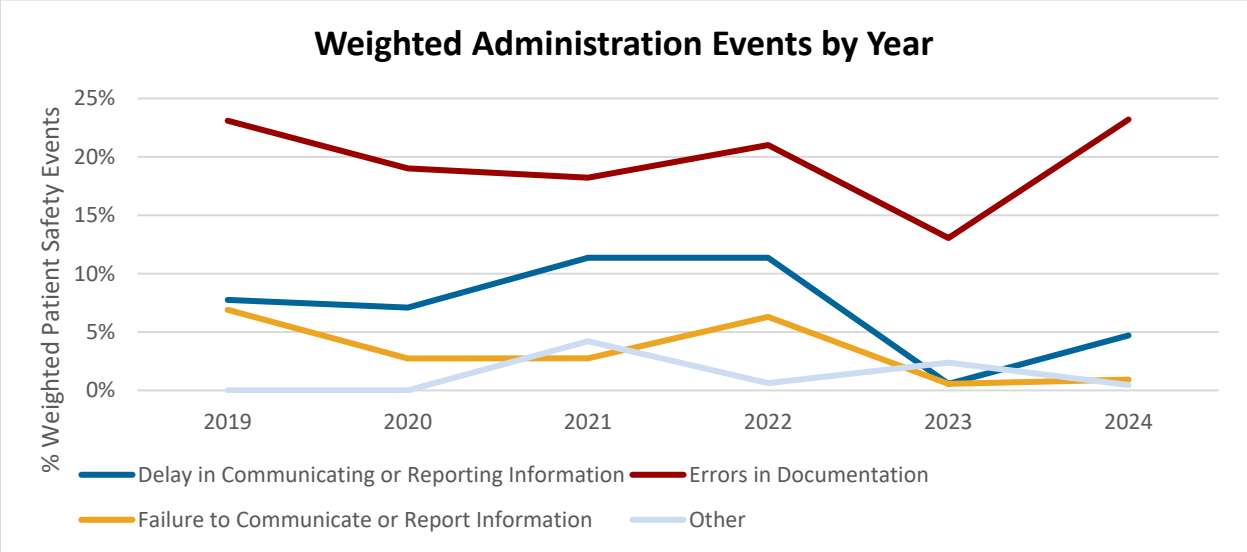

Figure 44: Administration Events by Year

From 2019 to 2024, Administration-related events consistently represented the largest share of weighted events, peaking at 39% in 2022 before dipping to a low of 17% in 2023 and rebounding to 29% in 2024. Within this category, Errors in Documentation were the most frequent subclassification, often comprising over half of Administration weighted events each year. These documentation errors commonly involved data entry mistakes during the registration of patients on the transplant waitlist or during the recording of HLA typing results in DonorNet. Such missteps at upstream junctures in the

transplant workflow introduce risk that can propagate through allocation matching and downstream clinical decision highlighting the need for robust verification protocols and training.

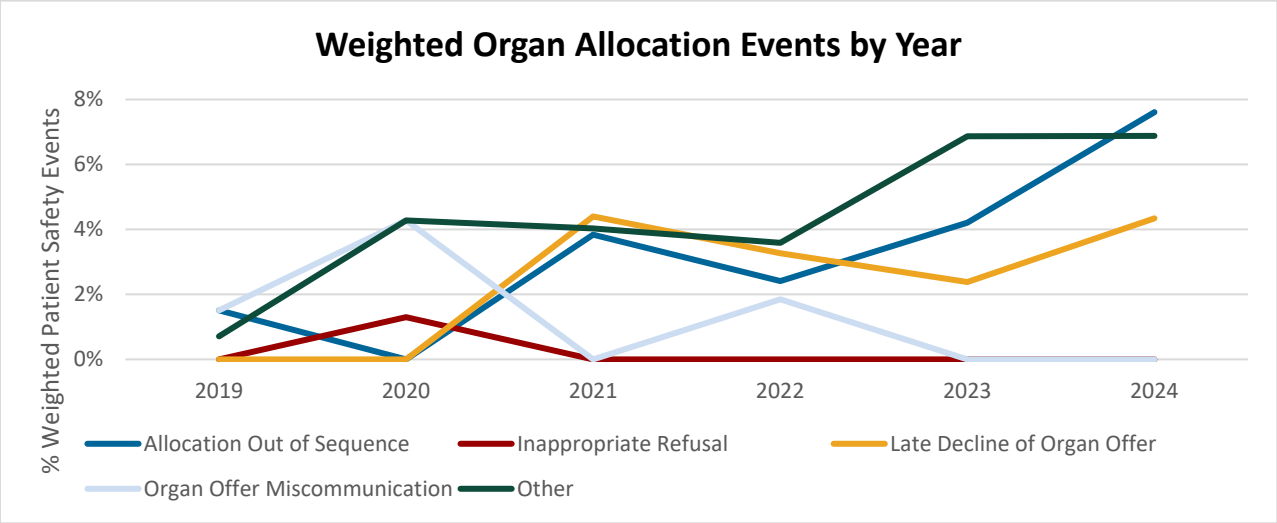

Figure 45: Organ Allocation Events by Year

A notable trend is this steady rise of Organ Allocation-related weighted events, increasing from just 4% in 2019 to 19% by 2024. This nearly fivefold growth indicates a growing vulnerability in the allocation and offer processes. The main driver of this category is Allocation Out of Sequence which surged in 2024, accounting for 8% of total events. Other allocation issues, such as Late Declines and Miscommunication of Organ Offers remain relatively stable but contribute to the overall burden. These trends may reflect increasing procedural complexity, which is especially germane given the implementation of new allocation policies during this timeframe, or gaps in automated guardrails within the overall allocation system.

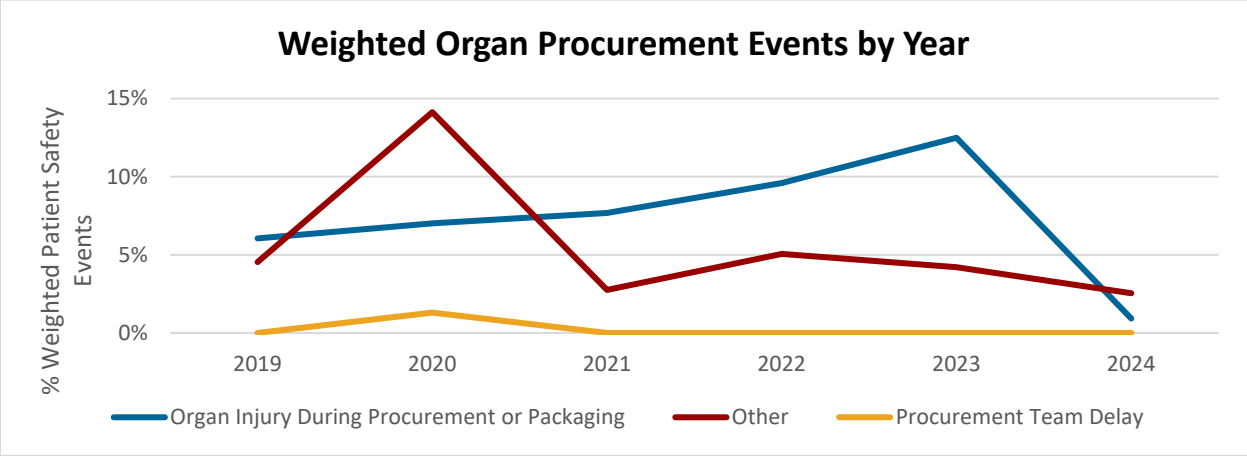

Figure 46: Organ Procurement Events by Year

Organ Procurement weighted events were prominent in 2020 (22%) but have since sharply declined to just 3% in 2024. Most Organ Procurement weighted events previously stemmed from Organ Injury During Procurement or Packaging, suggesting that targeted interventions, such as improved team coordination and packaging protocols, may have been effective in curbing this type of harm.

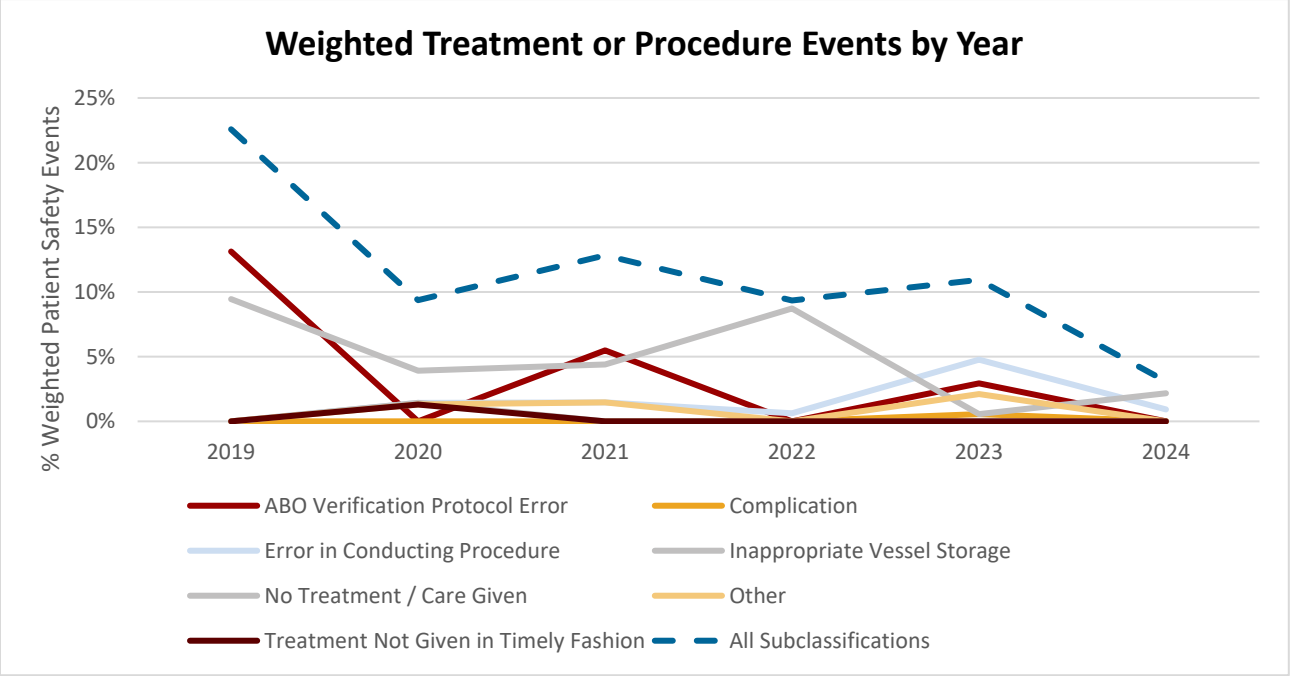

Figure 47: Treatment or Procedure Events by Year

The Treatment or Procedure category of weighted patient safety events has seen a reduction over time, dropping from 23% in 2019 to just 3% in 2024. This decline is largely due to near elimination of high frequency issues such as ABO Verification Protocol Errors and Inappropriate Vessel Storage. These improvements could suggest that there may have been successful implementation of safeguards and checklist-based procedures in clinical workflows.

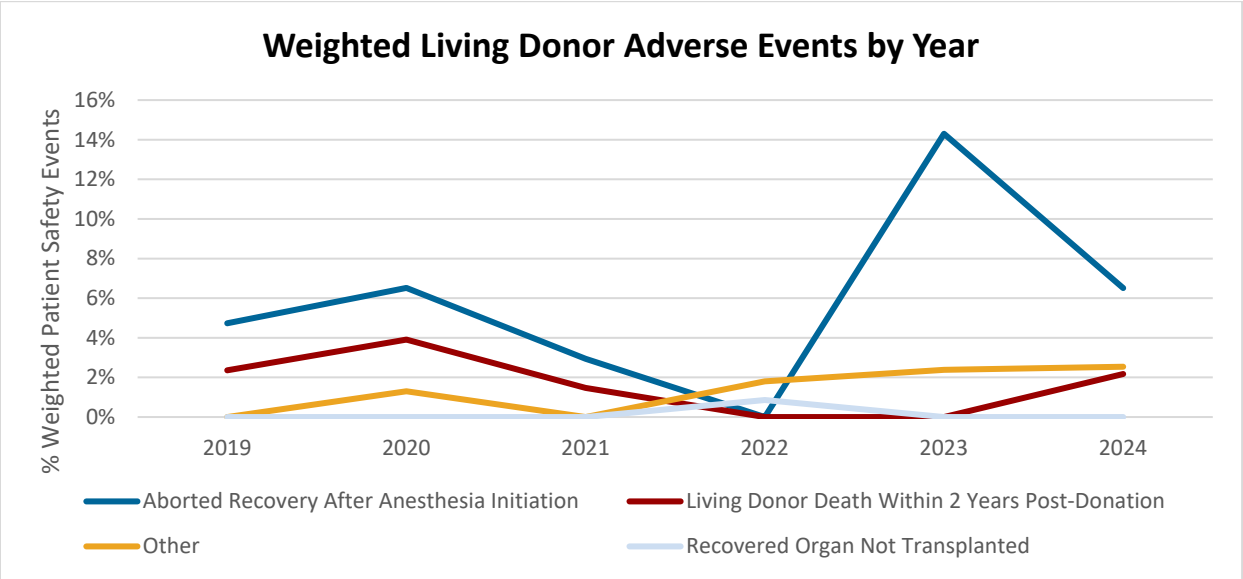

Figure 48: Living Donor Adverse Events by Year

Living Donor Adverse Events have fluctuated considerably across years, peaking at 17% of weighted events in 2023 and declining to 11% in 2024. The most common issue within this category is Aborted Recovery After Anesthesia Initiation, which spiked sharply in 2023. This volatility may reflect variation in surgical preparedness, living donor evaluations, or intraoperative decision-making protocols that vary by case or center.

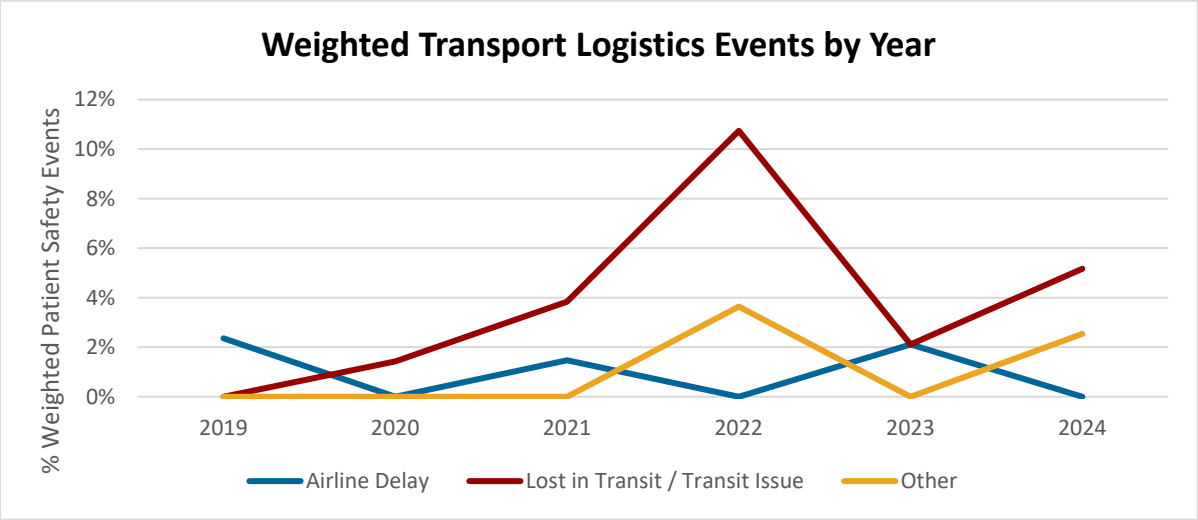

Figure 49: Transport Logistics Events by Year

Transport Logistic weighted patient safety events, including Lost in Transit, have increased in prominence, particularly in 2022 (14%) and 2024 (8%). One plausible driver of this trend is the implementation of new organ allocation policies that expand geographic sharing, resulting in organs traveling longer distances to get to their intended recipients. While organ allocation policy changes aim

to improve equity and access, they may unintentionally introduce new logistical vulnerabilities, such as missed connections, mislabeling issues, or longer exposure to risk during transit.

Key Takeaways

- Administrative errors, especially Errors in Documentation, are consistently the top source of weighted events.
- Organ Allocation issues are steadily increasing and now represent nearly one fifth of weighted events.
- Treatment or Procedure and Organ Procurement issues have sharply declined.
- Transport Logistics and Living Donor Adverse Events are volatile and require continued monitoring and more comprehensive root cause analysis.
- The rise in unclassifiable events signals the need for standardized and comprehensive patient safety documentation.

Causes of Different Types of Events

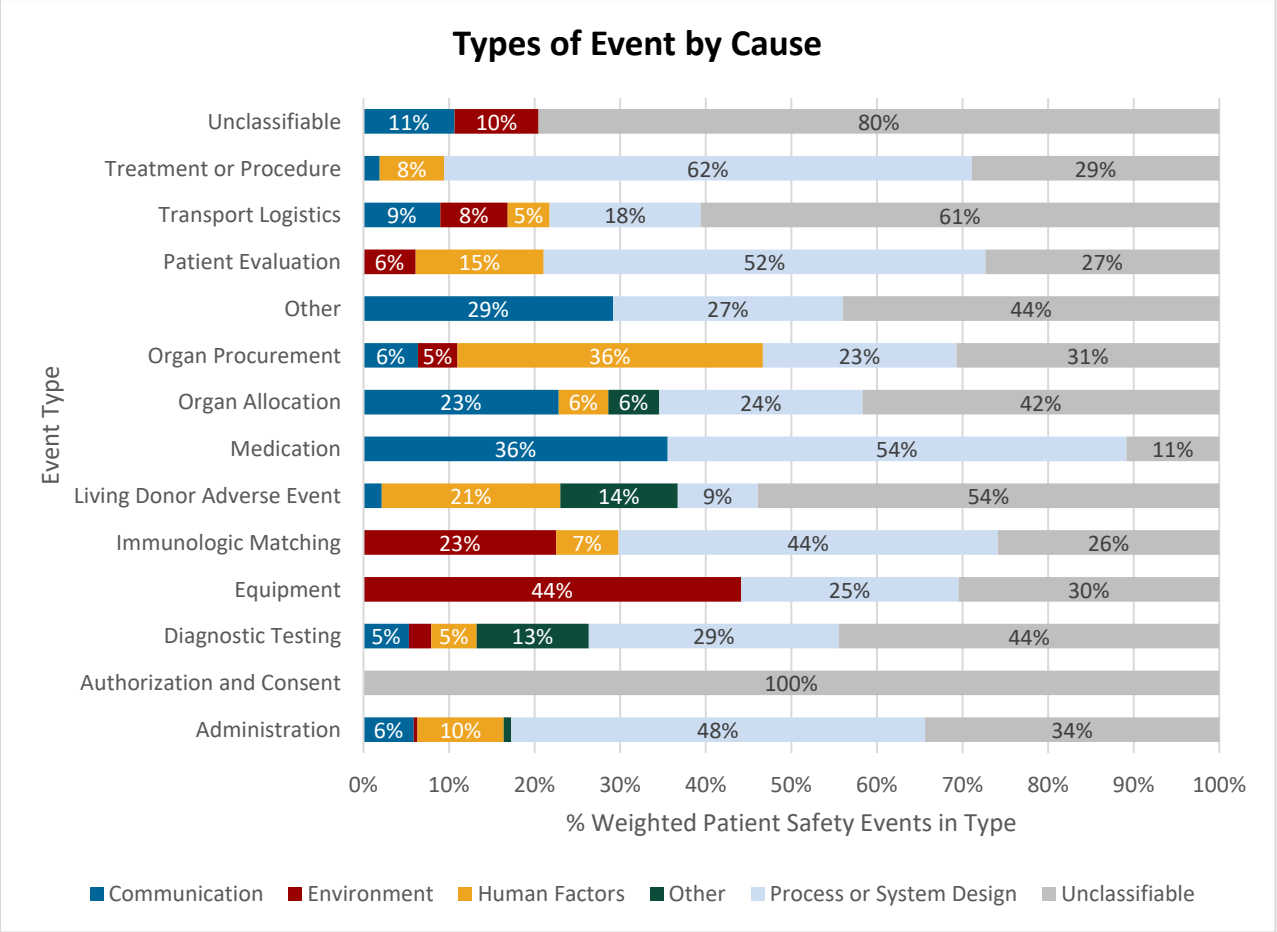

Figure 50: Cause Distribution of Event Types

Across nearly every Primary Type, Process or System Design flaws are the most frequently identified contributing factor to weighted patient safety events, revealing how complex or poorly structured

workflows contributed to errors in high-stakes contexts. For example, 18% of weighted Transport Logistics events are attributed to Process or System Design, underscoring how procedural weakness, such as coordination breakdowns or insufficient contingency planning, can lead to organ delays that can ultimately lead to organ non-use. Similarly, Administration issues are also largely tied to Process or System Design (48%), frequently involving challenges with entering data into systems like DonorNet or during patient registration on the waitlist.

Human Factors also play a significant role in certain domains, particularly where time constraints, manual handling, physical handling of organs, and coordination-intensive tasks are central. For example, in Organ Procurement, 36% of weighted errors stem from Human Factors, often arising during the retrieval, labeling, or packaging of organs, where coordination between surgical and transport teams under tight timelines is critical. These operations often occur in external or unfamiliar hospitals, operating rooms, or transit hubs, where teams must adapt quickly to variable environments, rely heavily on memory or verbal communication, and perform manual steps such as double-checking organ labels, securing containers, and ensuring accurate handoffs.

Communication failures are another prominent cause, particularly in Organ Allocation (23%) and Medication errors (36%). These likely represent scenarios where delays, miscommunications, or misinterpretations of information impacted the timeliness of decisions. In Organ Allocation, this may include Late Declines, Organ Offer Miscommunication, or Allocation Out of Sequence, all events that can have critical downstream consequences.

Recall that Administration, Organ Procurement, Organ Allocation, and Treatment or Procedure errors make up 75% of weighted patient safety events.

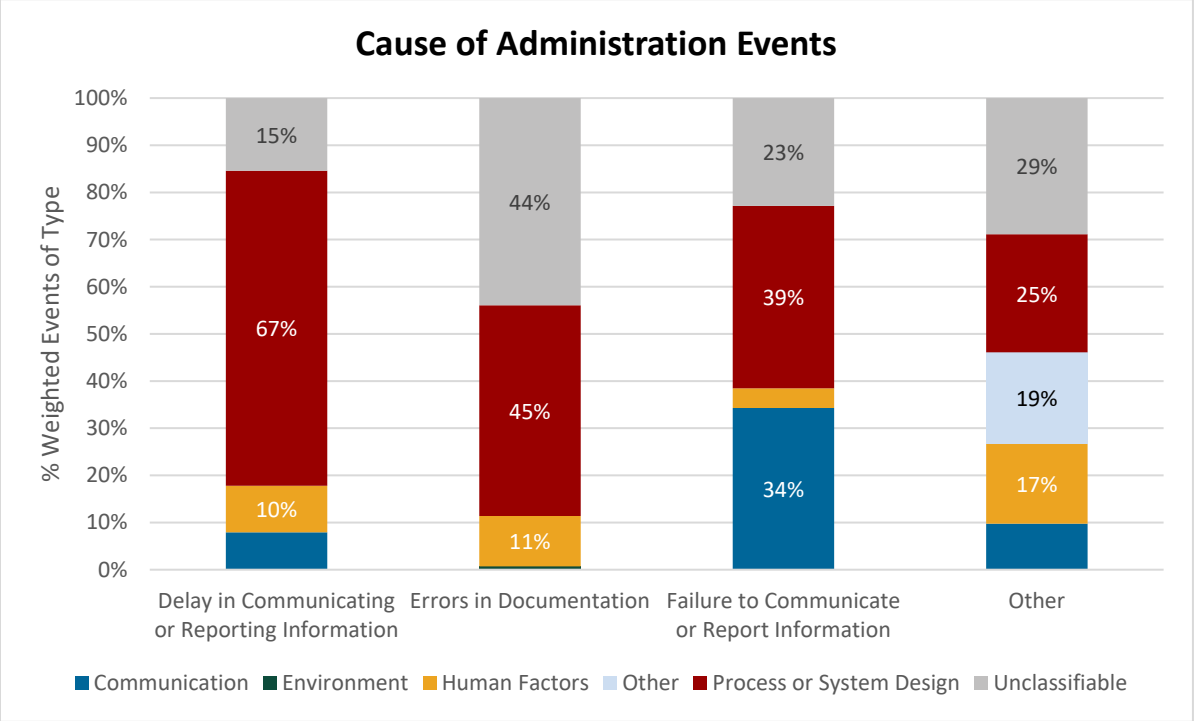

Figure 51: Causes of Administration Events

Process or System Design causes substantial portions of all types of the 500 weighted Administration Errors. Notably, Errors in Documentation (308 weighted events) and Delays in (117 weighted events) or Failures to (52 weighted events) Communicate or Report Information, Process or System Design issues are primarily associated with policy issues, whether a Deviation or a Deficiency rather than a staffing issue. These policy-caused errors make up larger portions of Delays in Communicating or Reporting Information and Errors in Documentation than in Failures to Communicate or Report Information. Delays frequently include late reports of infectious diseases, in which case there ought to be a policy in place. Errors in Documentation frequently occur as manifestations of errors in processes where there are verification policies. However, Failures to Communicate or Report Information tended to be communication issues not governed by policy or caused by policy deficiency.

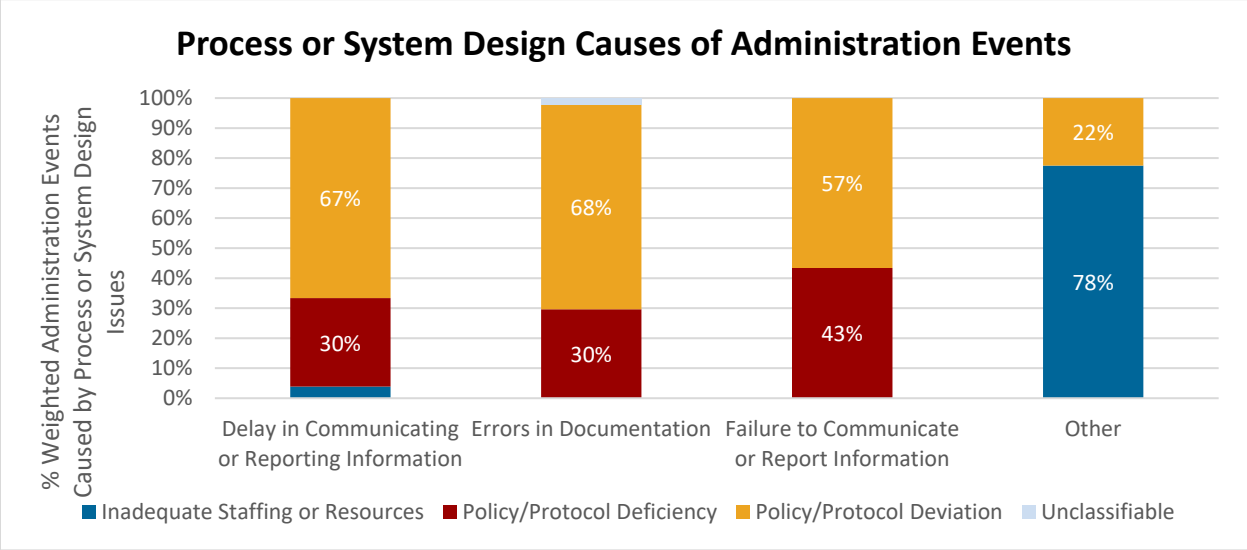

Figure 52: Process or System Design Causes of Administration Events

When an Administration event occurs, and the cause is a Process or System Design issue, 241 weighted events, the cause tends to be a Policy/Protocol Deviation rather than a Deficiency. This is a relatively positive result. If a patient safety event is due to some sort of failure regarding policy, it is preferable that there was an attempt to regulate the conduct in question.

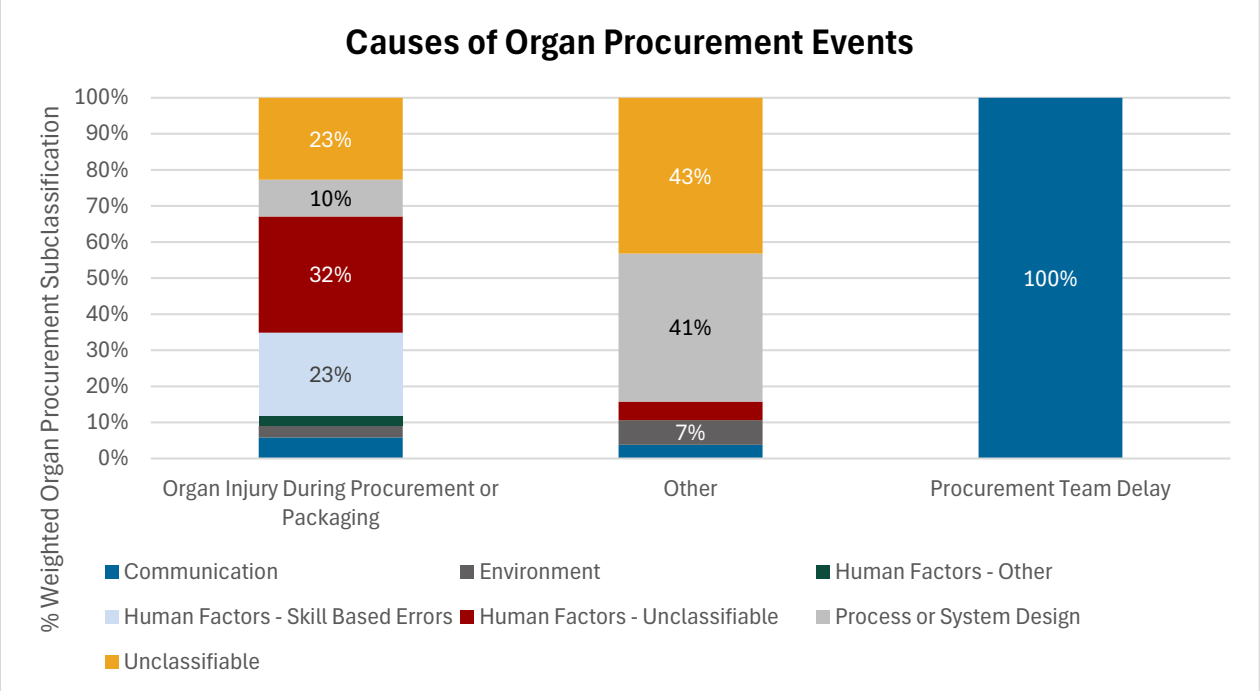

Figure 53: Causes of Organ Procurement Events

There were 205 weighted Organ Procurement events. Organ injuries (119 weighted events) are mostly caused by Human Factors. However, the specific type of Human Factor that causes an Organ Injury is more likely than not unclear from the qualitative information presented. Frequently, the type of organ injury that occurred is stated flatly with no attempt made to describe the cause of the error. Other Organ Procurement patient safety events (84 weighted events) tend to either be caused by Human Factors or Unclassifiable causes. This tracks with Other events frequently either being process issues, like swapping left and right kidneys without later verification or odd, one-off events that happen during procurement, like accidentally discarding organs for inexplicable reasons.

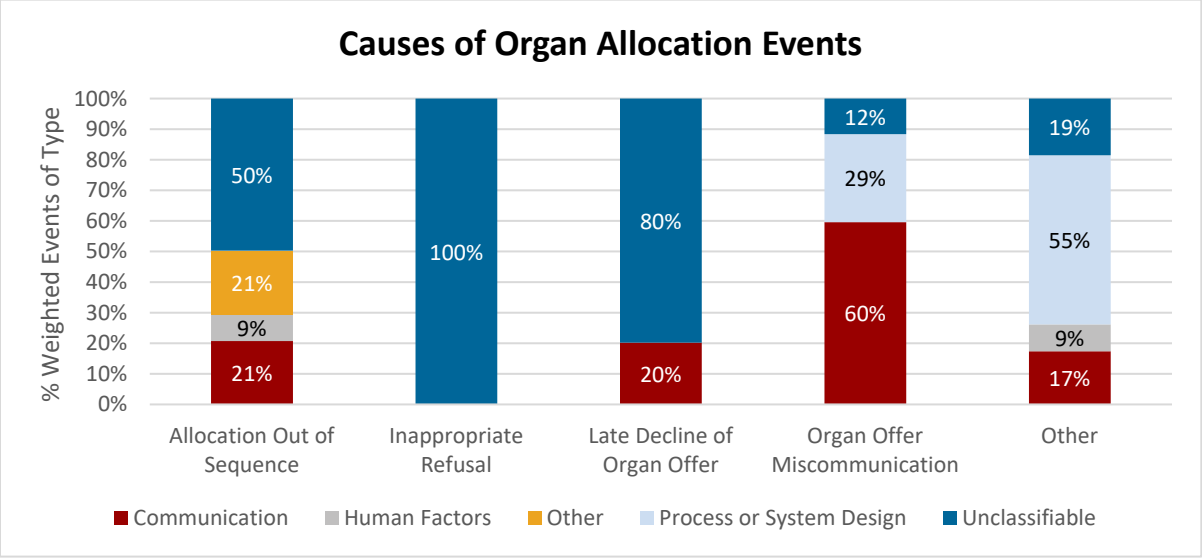

Figure 54: Causes of Organ Allocation Events

Each subclassification of the 192 weighted Organ Allocation event tends to have a very distinct distribution of Causes, associated with the fact that each Organ Allocation event type is qualitatively distinct from each other. Notable however is that the two most common Organ Allocation events, Allocations out of Sequence (55 weighted events) and Late Declines of Organ Offers (43 weighted events) also have very high portions of weighted patient safety events with Unclassifiable causes.

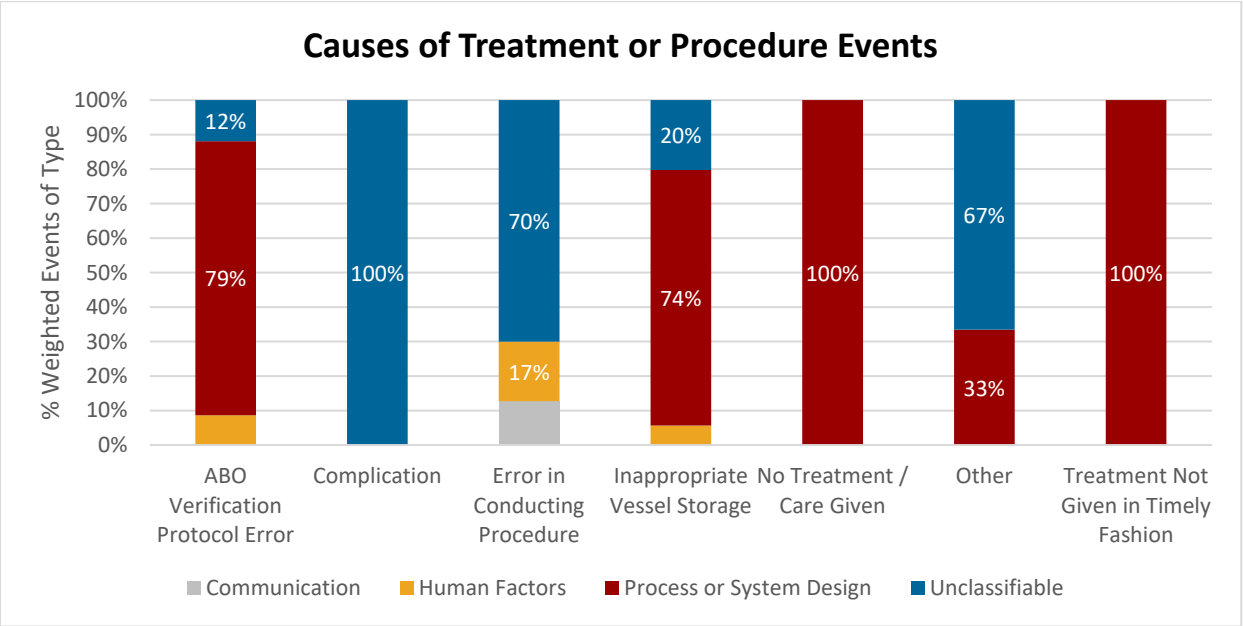

Figure 55: Causes of Treatment or Procedure Events

Each subclassification of the 168 Treatment or Procedure weighted patient safety events each have very distinct distributions of causes. However, the two most common types of Events, ABO Verification Protocol Errors and Inappropriate Vessel Storage errors tend to be caused by Policy/Protocol Deviations. This is a desirable result: OPTN policy is clear in these areas, so there ought to be adequate policies governing these types of events.

Key Takeaways

- Process or System Design is the top contributor to weighted errors across most types, including Administration, Transport Logistics, and Treatment or Procedure.
- Where weighted errors are caused by Process or System Design problems, those problems tend to overwhelmingly be Policy/Protocol issues, primarily Policy/Protocol Deviations and secondarily Policy/Protocol Deficiencies.
- Human Factors drive weighted errors in high pressure or settings where organs are being physically handled, like during Organ Procurement
- Communication is a leading cause in Organ Allocation and Medication events

Most Harmful Event Types

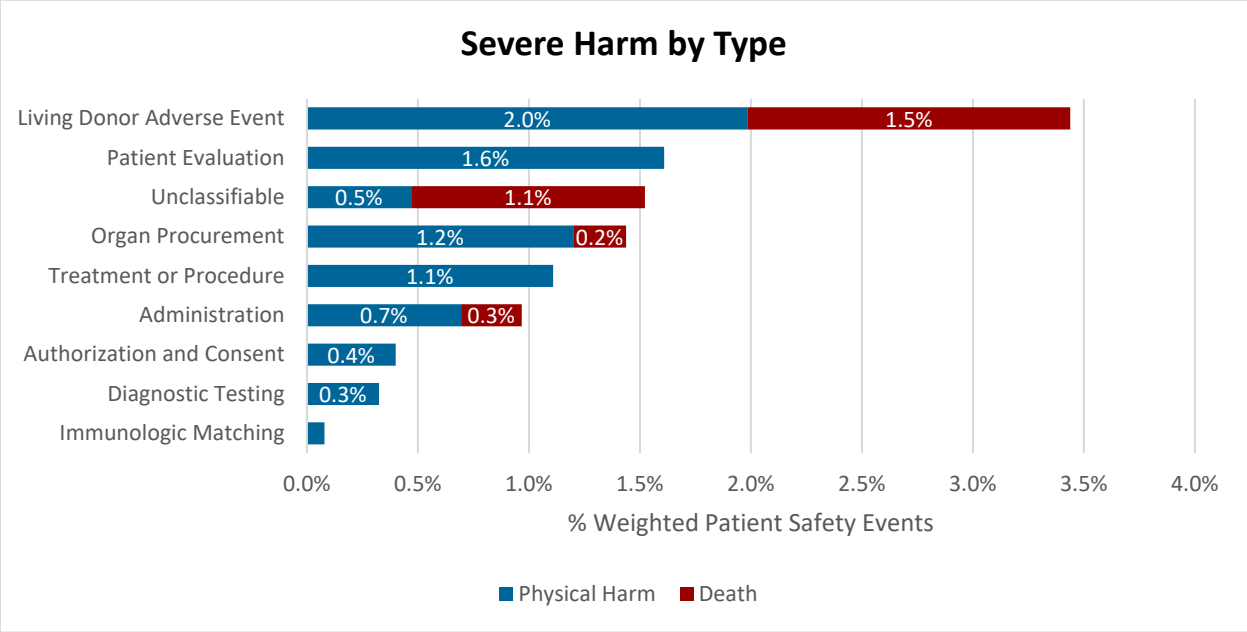

Figure 56: Severe Harm Rates of Different Event Types

The most severe outcomes, defined as events resulting in Physical Harm or Death, are most concentrated in two primary categories: Living Donor Adverse Events and Organ Procurement. Living Donor Adverse Events accounted for the highest percentage of overall harm, comprising 2.0% Physical Harm or 31 weighted events and 1.5% Deaths or 23 weighted events, for a total of 3.5% of total weighted patient safety events or 55 weighted events. While this rate reflects the inherent risks of surgically intervening on otherwise healthy individuals, it is important to note that OPTN policy requires reporting of specific Living Donor Safety Events. For example, OPTN requires the reporting of any living donor death within two years of donation, regardless of cause. While this reporting requirement ensures critical oversight and transparency, it also introduces some interpretive complexity: not all post-donation deaths necessarily reflect errors or failures in the transplant care system. For example, fatalities due to causes such as motor vehicle accidents are still counted within this category, despite being largely unrelated to clinical or procedural deficiencies. As such, while the high rates of reported harm warrant vigilance, they also partially reflect a broader reporting net rather than systemic shortcomings in living donor safety. Organ Procurement also contributes to severe outcomes, 1.2% Physical Harm. These outcomes stem from Organ Injury During Procurement or Packaging.

Beyond these two types, Patient Evaluation contributes 1.6% of the Physical Harm cases or 26 weighted events, underscoring the importance of thorough and accurate assessments prior to listing and transplant. Mistakes in this domain may not be immediately visible but can reverberate downstream during critical moments of care. Treatment or Procedure errors result in 1.1% of Physical Harm, often related to intraoperative complications or inadequate post-operative monitoring. Administration weighted events account for 1% of severe outcomes, which is a reminder that even documentation or clerical lapses, such as HLA typing transcription errors, can lead to significant downstream consequences.

Root Node Five: Primary Responsible Party

Responsibility for Patient Safety Events

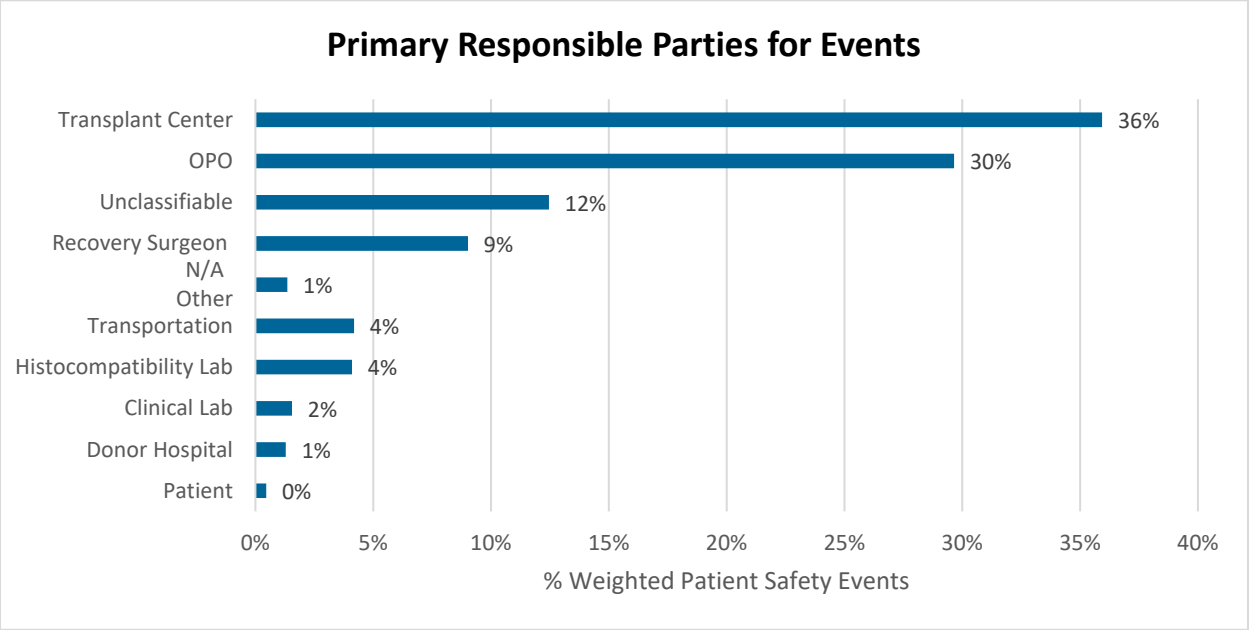

Figure 57: Primary Responsible Party Distribution

Transplant Centers and OPOs together are the primary responsible parties in 66% of the 1585 weighted patient safety events or 520 weighted events. This reflects central roles in the transplant ecosystem. Recovery surgeons are responsible for a notable 9% of weighted events or 143 weighted events, largely representing the number of organ injury patient safety events. Weighted patient safety events with unclassifiable primary responsible parties make up a fairly small number of weighted patient safety events, 198, but below, we see that the number of these events is increasing year over year.

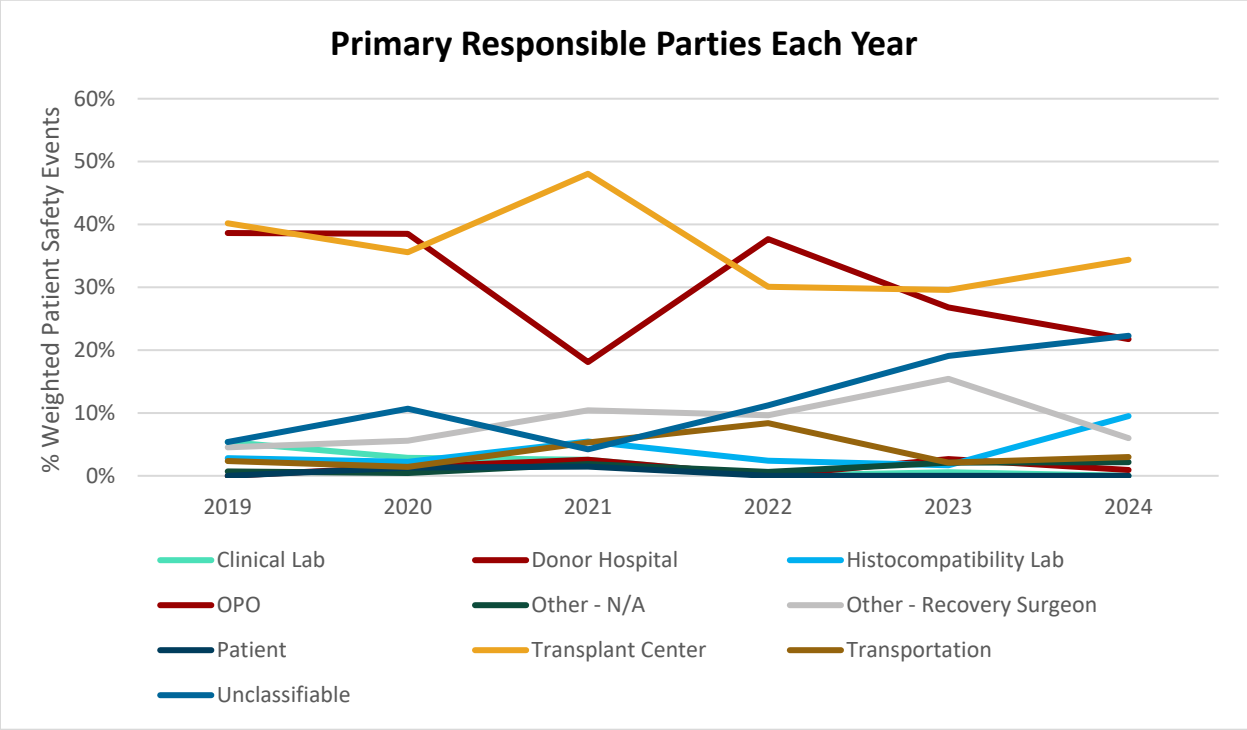

Figure 58: Primary Responsible Party by Year

Between 2019 and 2024, the share of weighted patient safety events for which OPOs and Transplant Centers were the Primary Responsible Party has fallen from 40% to 34% and from 39% to 22% respectively. However, the share of weighted patient safety events for which the Primary Responsible Party is Unclassifiable has risen from 5% to 22%. This shows a trend in which substantially less qualitative information is provided in more recent patient safety events for which qualitative information is provided. Other Primary Responsible Parties make up a smaller portion of patient safety events.

Reasons Different Parties Err

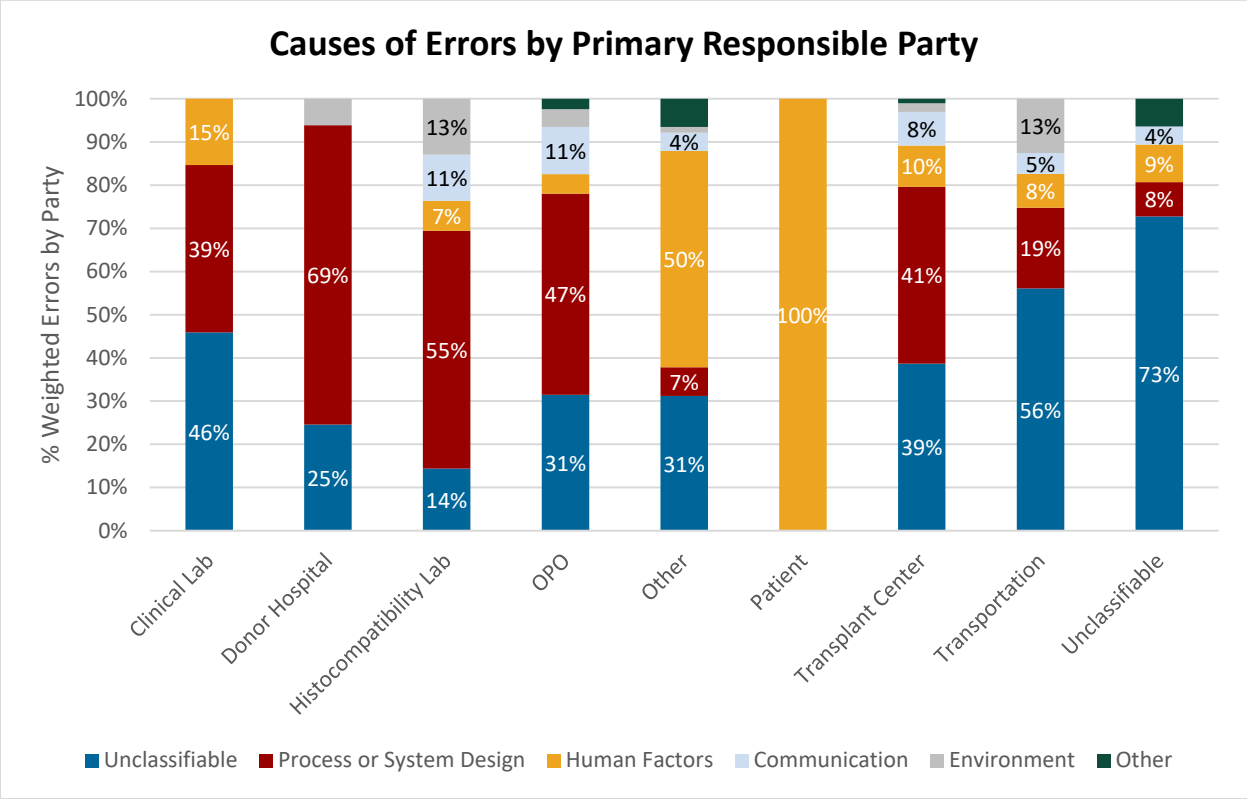

Figure 59: Cause Distribution by Primary Responsible Party

A substantial portion of the causes of weighted patient safety events were Unclassifiable for most of the actors, totaling 626 of the 1585 weighted patient safety events. Every Primary Responsible Party had patient safety events with Unclassifiable causes at least 10% of the time, and two-thirds had events with Unclassifiable causes at least 30% of the time.

Of events with classifiable causes, Process or System Design errors caused 550 of the 1585 weighted events. It was a particularly prominent cause for OPOs, Donor Hospitals, Clinical Labs, Histocompatibility Labs and Transplant Centers. This is because these entities are the most tightly regulated and ought to have internal policies and protocols to govern their behavior. Policy/Protocol Deviations in particular are fairly agnostic to the reason for deviation. So, where a Primary Responsible Party errs for reasons unrelated to Process or System Design, it suggests there was no on point policy or protocol governing their actions.

Other Primary Causes contributed to errors. When OPOs or Histocompatibility labs erred, a substantial portion of the time, Communication was the cause (11% of weighted errors for which OPOs and Histocompatibility labs were responsible). This reflects how OPOs sit in a central location in a network of hospitals and labs and how Histocompatibility labs must communicate both with OPOs and Transplant Centers. A portion of weighted errors for which Transplant Centers were responsible, 8%, were caused

by Communication. This reflects Transplant Centers' duties to report information about infectious diseases as well as their role during the organ allocation process.

Clinical Labs erred due to Human Factors, making up 15% of their weighted errors. Donor Hospitals erred due to Environment causes, 6% of the time. Histocompatibility Labs erred due to Communication and Environment issues, 11% and 13% of the time, respectively. Communication errors by Histocompatibility Labs reflect that they must communicate both with OPOs and Transplant Centers. Environment causes for Clinical and Histocompatibility labs include equipment failure, reflecting issues with lab equipment and software. Other actors often erred due to Human Factors. This is driven by recovery surgeons injuring organs, which were classified as due to human factors.

Where the actor was Unclassifiable, the cause was likely Unclassifiable too. This is reasonable; when the actor or scenario that caused an event is unclear, it is harder to determine the reason an event happened.

Weighted events caused by Process or System Design	
Primary Responsible Party	**Process or System Design**
Clinical Lab	10
Donor Hospital	14
Histocompatibility Lab	36
OPO	219
Other	11
Patient	
Transplant Center	233
Transportation	12
Unclassifiable	16
Total	550

Table 9: Weighted Events Caused by Process or System Design

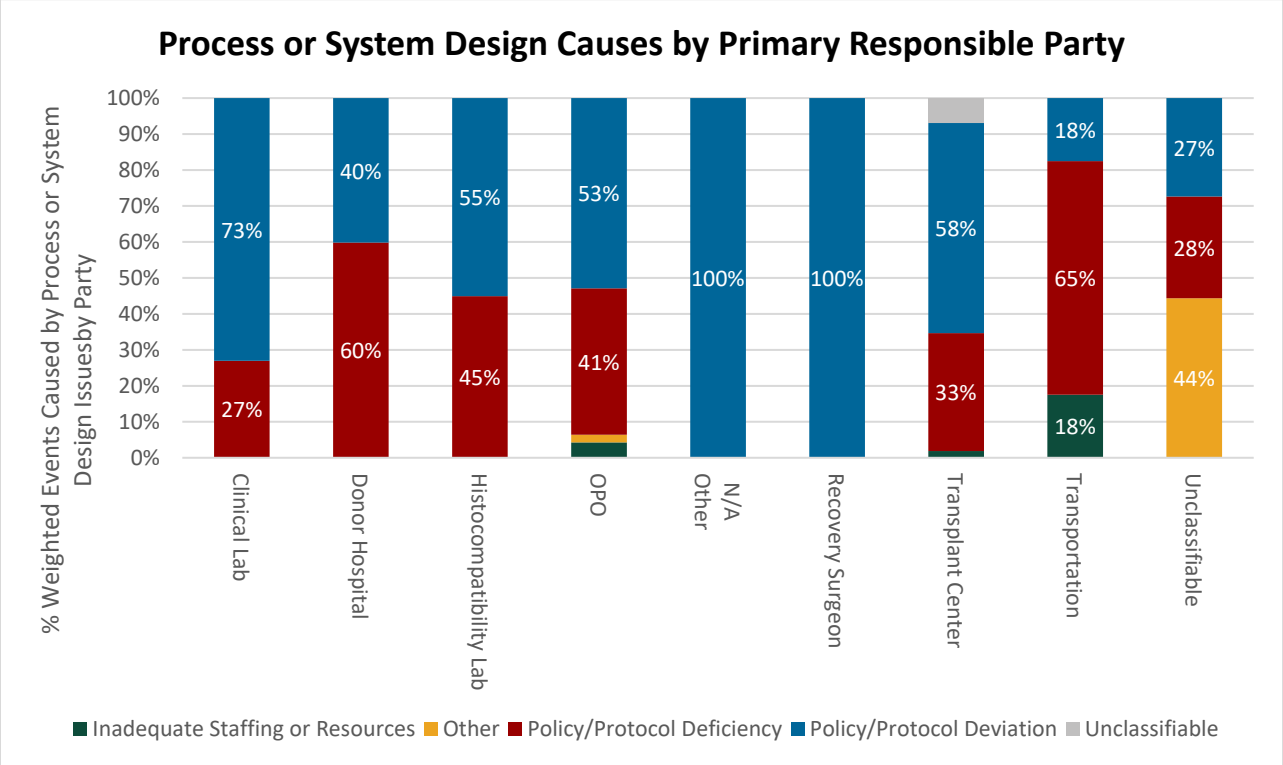

Figure 60: Process or System Design Causes by Primary Responsible Parties

When a weighted patient safety event is caused by a process or system design issue, and the Primary Responsible Party is not Unclassifiable, 535 of the 1585 patient safety events, the cause is almost always rooted in a policy or protocol issue, regardless of the Primary Responsible Party. However, there is variance in whether issues are Policy/Protocol Deviations or Deficiencies. Histocompatibility Labs, Clinical Labs, OPOs, and Transplant Centers have more Deviations than Deficiencies, speaking to the existence of more robust policies or processes that are less vulnerable to other forms of error. Donor Hospitals and Transportation actors have more Deficiencies than Deviations, perhaps suggesting blind spots in policy or internal protocols.

Weighted events caused by Human Factors	
Primary Responsible Party	**Weighted Events**
Clinical Lab	4
Donor Hospital	0
Histocompatibility Lab	5
OPO	21
Other	82
Patient	7
Transplant Center	54
Transportation	5
Unclassifiable	17
Total	196

Table 10: Weighted Events caused by Process or System Design

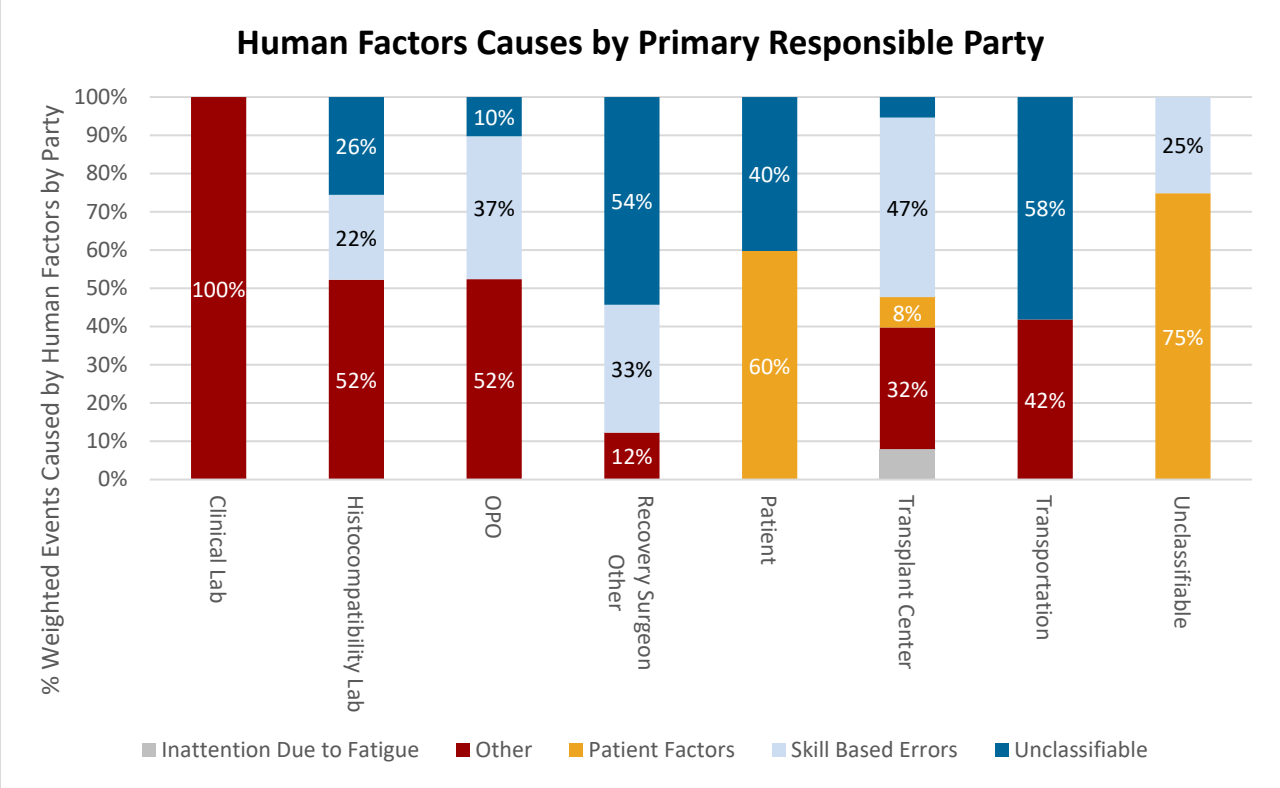

Figure 61: Human Factor Causes by Primary Responsible Party

Primary Responsible Parties err for a variety of Human Factors reasons. Clinical Labs, Histocompatibility Labs, OPOs, and Transportation services erred for Other reasons. This category was frequently used for basic but identifiable errors like inattention not due to fatigue, for example, a courier carelessly entering the wrong address into a GPS system. Histocompatibility labs, OPOs, Recovery Surgeons, and Transplant Centers saw substantial amounts of Skill Based Errors. In each of these Primary Responsible Parties,

patient safety events occur where required skill was not demonstrated, like incorrectly reviewing a CT scan. Human Factors causes were often Unclassifiable. In these instances, some kind of error best attributed to carelessness was demonstrated but no specific cause was given. A notable to the exception above are Patients. Patients were largely responsible for Living Donor Deaths within 2 Years Post-Donation.

Weighted events caused by Communication	
Primary Responsible Party	**Weighted Events**
Clinical Lab	0
Donor Hospital	0
Histocompatibility Lab	7
OPO	51
Other	7
Patient	
Transplant Center	44
Transportation	3
Unclassifiable	8
Total	121

Table 11: Weighted Events Caused by Human Factors

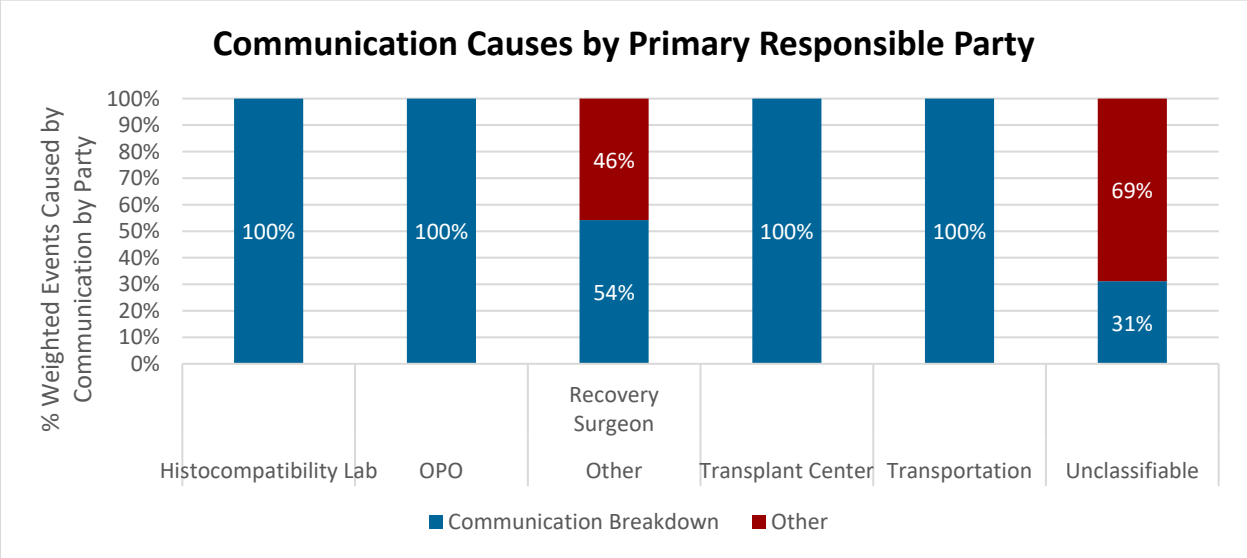

Figure 62: Communication Issues by Primary Responsible Parties

Where the Primary Responsible Party is identifiable and weighted patient safety events are due to Communication issues, Communication issues are often Communication Breakdowns. Other Communication causes are overrepresented for Recovery Surgeons and Unclassifiable Primary Responsible parties due to only having a sample size of two unweighted events for each Primary Responsible Party.

Responsible Parties with the Highest Rates of Severe Harm

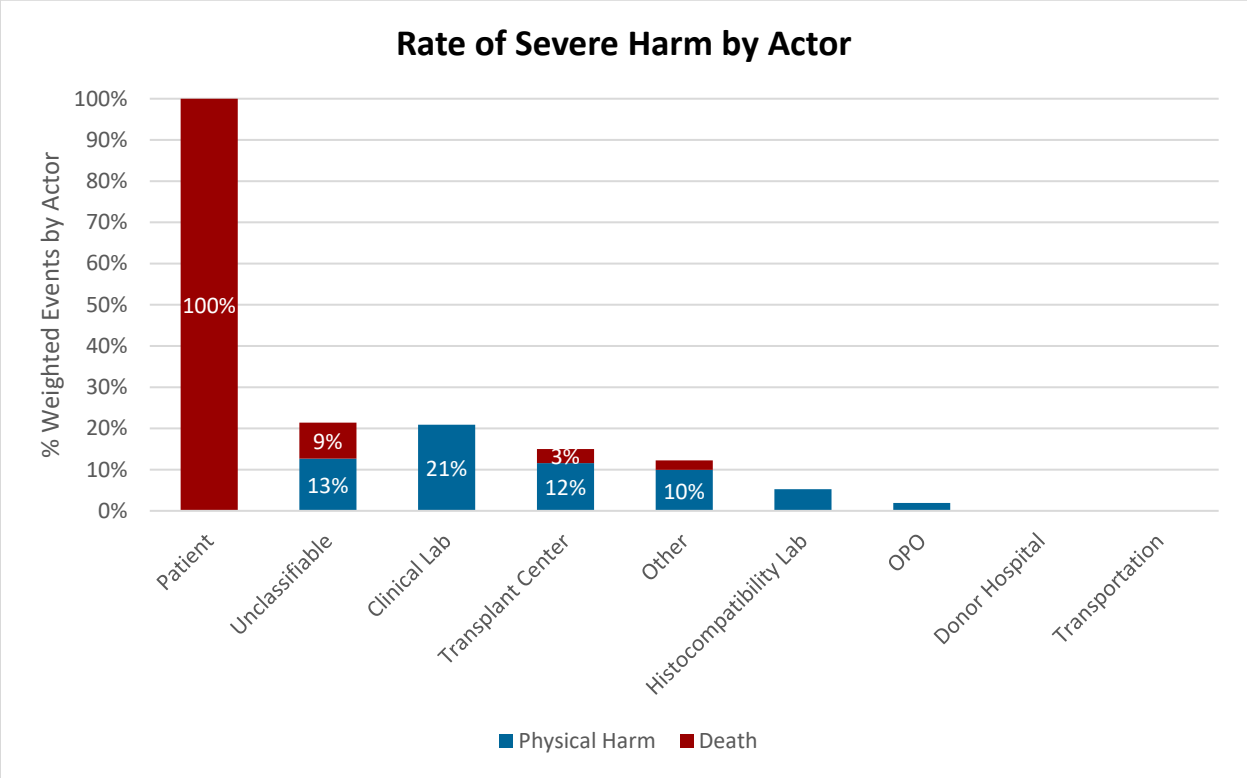

Figure 63: Rates of Severe Harm by Primary Responsible Party

Volume of Severe Harm by Primary Responsible Party			
Primary Responsible Party	**Physical Harm**	**Death**	**Total**
Transplant Center	66	20	86
Unclassifiable	25	17	42
Patient	0	7	7
Other	16	4	20
Histocompatibility Lab	3	0	3
Clinical Lab	5	0	5
OPO	9	0	9

Table 12: Weighted Patient Safety Events Leading to Severe Harm by Primary Responsible Party

When Patients are the Primary Responsible Party for weighted patient safety events, the result is always fatal. However, this represents only two events, both Living Donor Deaths within Two Years Post-Donation, which have mandatory reporting requirements.

When the primary responsible party is Unclassifiable, harm is severe 22% of the time. Harm is fatal 9% of the time. However, this is influenced upwards both by Living Donor Deaths within Two Years Post-

Donation and Living Donor Aborted Recoveries After Anesthesia Initiation. The implication in these cases is often that the Living Donor Adverse Event is due to factors outside of the control of the transplant ecosystem.

Clinical Lab weighted patient safety events result in physical harm 21% of the time. However, there is a low sample size of events where Clinical Labs are the Primary Responsible Party. Patient Safety Events caused by Transplant Centers are relatively likely to lead to physical harm or death (15%). A higher number here is not unexpected due to the risk of harm to patients due to the inherent risks of harm during the transplant process.

Key Takeaways

- The Primary Responsible Party for most weighted patient safety events was either an OPO (30%) or a Transplant Center (36%).
- The portion of weighted patient safety events for which the Primary Responsible Party is Unclassifiable is increasing (5% in 2019 to 22% in 2024).
- The cause of weighted patient safety events for which each Primary Responsible Party varies greatly, but for Primary Responsible Parties other than Patient and Other, Process or System Design and Unclassifiable Responsible Parties dominate.
- Where the Primary Responsible Party is not Unclassifiable, Process or System Design issues are either Policy/Protocol Deviations or Policy/Protocol Deficiencies at least 80% of the time.
- Human Factors subclassifications vary greatly by Primary Responsible party and the distributions defy short characterization. While Skill Based Errors are common where the cause is not Unclassifiable, Other Human Based Factors are not at all uncommon. Inattention Due to Fatigue is rare, making up 8% of Human Factors caused weighted patient safety events where the Primary Responsible Party was a Transplant center only.
- Communication events were almost always Communication Breakdowns regardless of Primary Responsible Party, except for the 46% of Communication – Other weighted events where Recovery Surgeons were primarily responsible and excluding patient safety events where the Primary Responsible party was unclassifiable.
- Transplant Centers had a higher rate of weighted patient safety events for which they were responsible leading to physical harm than death (13% vs 2%).
- When OPOs were the Primary Responsible party, the patient safety event never resulted in death in any of the coded events.
- When Transplant Centers were primarily responsible, the outcome was fatal 2% of the time.

Root Node Six: Primary Cause

Most Common Causes of Patient Safety Events

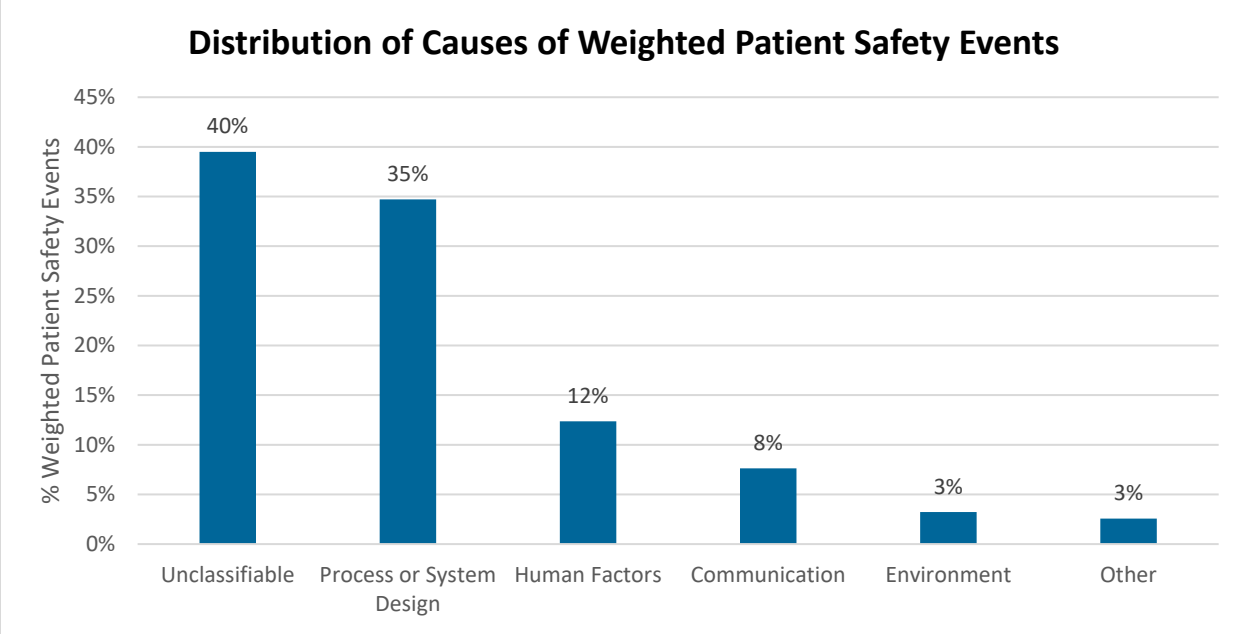

Figure 64: Causes of Weighted Patient Safety Events

Of the 1585 weighted patient safety events, a large minority are Unclassifiable, 40% or 626 weighted events. When not Unclassifiable, Events are largely caused by Process or System design, at 35% of weighted events or 550 weighted events.

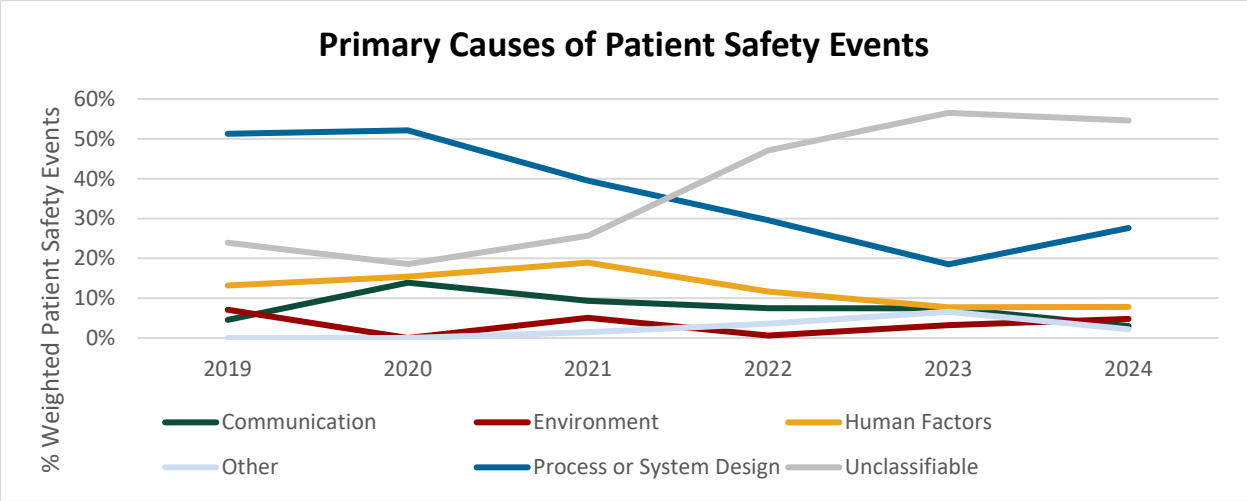

Figure 65: Primary Cause by Year

The key trend is that the portion of weighted events with Unclassifiable causes have increased sharply, from 24% to 55% of weighted patient events, while every other cause has either declined or remained fairly flat. The second most common cause, Process or System Design, has declined sharply, from 51% to 28%. The implication is that the portion of patient safety events with qualitative data from which causation is determinate has gone from a substantial majority to a large minority. This is due to a decline in the amount of information provided with qualitative information.

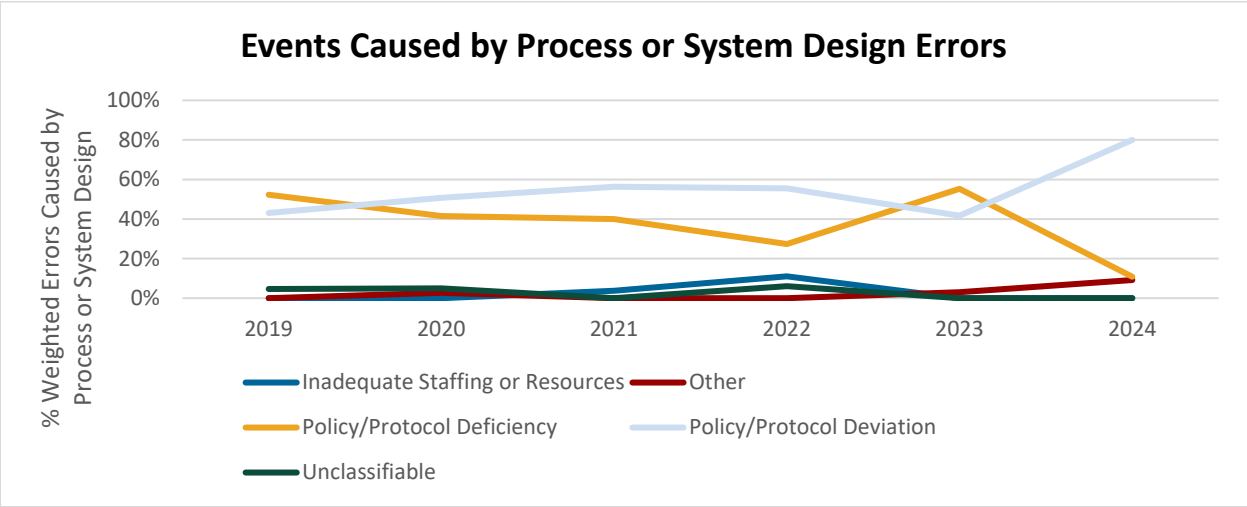

Figure 66: Process or System Design Causes by Year

Process and System Design causes are usually Policy/Protocol Deficiencies or Deviations, 507 of 550 weighted patient events. This reflects that processes within the transplant system are well documented

– in order to have a Policy/Protocol Deficiency or Deviation, an error must be in scope of some policy or protocol.

Causes Leading to Severe Harm

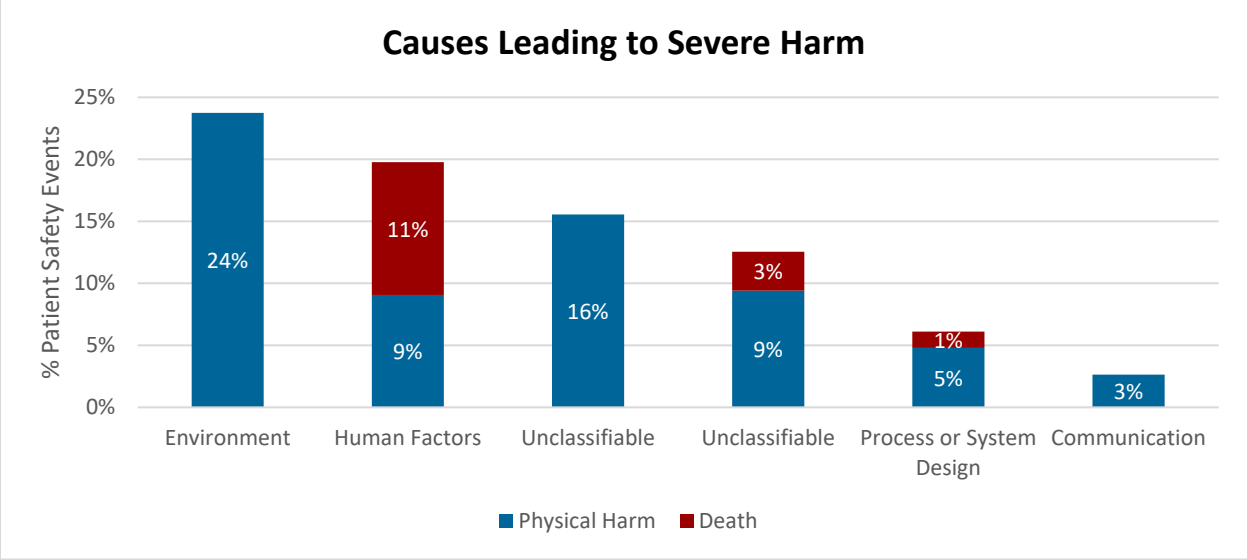

Figure 67: Severe Harm Rates of Causes

Volume of Severe Harm by Primary Cause			
Primary Cause	**Physical Harm**	**Death**	**Grand Total**
Communication	3	0	3
Environment	12	0	12
Human Factors	18	21	39
Other	6	0	6
Process or System Design	26	7	34
Unclassifiable	59	20	79
Total	125	48	173

Table 13: Weighted Events Leading to Severe Harm by Cause

Environment has a high rate of harm at 24%. Environment included both organizational culture issues but also issues caused by equipment such as organ pump failures. Human Factors also had a high rate of severe harm at 20%. The high rate of death caused by Human Factors is driven by the mandatory reporting of Living Donor Adverse Events. Other and Unclassifiable causes also have high rates of severe harm. Ideally, more severe outcomes would have easily codable qualitative information. Notably, the rate of harm from Process or System Design issues is relatively low, while the incidence of events caused by Process or System Design issues is high. This suggests the existence of effective guardrails preventing the worst forms of harm from issues caused by Process or System Design.

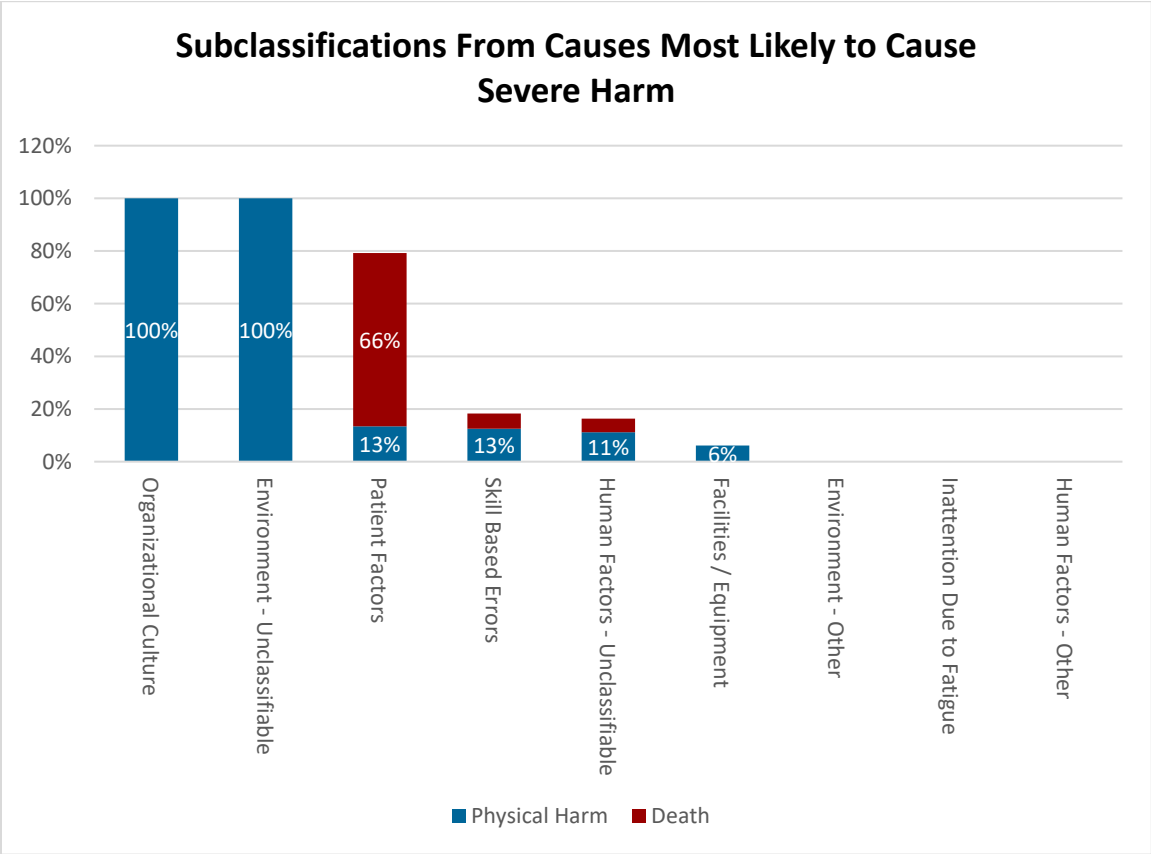

Figure 68: Severe Harm Rate of Cause Subclassifications

The above chart presents subclassifications from the Environment and Human Factors causes. Of the Environment-caused events, events where the cause is either Organizational Culture or the subclassification was Unclassifiable, the outcomes is much more likely to be severe harm than if there Is a facilities or equipment issue. Patient factors are quite likely to be deaths, but these are likely to be Living Donor Deaths within 2 Years Post-Transplant which are often not caused by patient safety breakdowns.

Key Takeaways

- The Primary Cause of weighted patient safety events is increasingly Unclassifiable (24% in 2019 vs 55% in 2024).
- The portion of weighted patient safety events caused by Process or System Design issues has been driven down (51% to 28%) by the rise in Unclassifiable causes.
- Process or System Design issues are consistently and largely dominated by Policy/Protocol Deviations and Deficiencies (ex. 95% of Process or System Design caused events in 2019 and 91% in 2024).
- Process or System Design causes lead to physical harm or death 6% of the time, but Unclassifiable causes lead to physical harm or death 12% of the time.

Discussion

Implications for Policy and Practice

The following recommendations are derived from an extrapolated analysis based off a weighted patient safety sample from a dataset of patient safety events received by the OPTN contractor for the years of 2019 through 2024. These recommendations aim to address two overarching challenges: The first is the need to reduce the proportion of events with indeterminate or poorly understood causes. The second is recognizing the need to focus on critical points in the system where small improvements can yield significant reductions in harm, such as administrative procedures, procurement practices, transport logistics, and policy adherence, that disproportionately contribute to patient harm. The goal is not only to reduce errors but also improve compliance with existing standards and ensure that safety interventions are informed by clear and actionable data.

1. Standardize and Enforce Root Cause Documentation

Accurate and complete root cause documentation is essential for understanding the underlying causes that result in patient safety events. A substantial proportion of reported events still lack root cause information or include qualitative descriptions that are difficult to code or analyze using our current taxonomy. A standardized, structured reporting system should require submission of root cause analysis and corrective action plans. Such a system would support better aggregation, pattern detection, and continuous improvement. It should include mandatory fields, drop down classifications, data validation processes and clear instructions to reduce ambiguity and increase data usability.

2. Target Policy Gaps Through Systematic Qualitative Review

While many errors reflect failure to follow existing policy, some do reveal true policy gaps – areas where existing guidance is ambiguous, missing, or not tailored to operational reality. A systematic review of qualitative reports coded as Policy/Protocol Deficiency can identify these gaps and inform policy revision.

3. Improve Policy Adherence Through Targeted Reinforcement

Many events classified as Policy/Protocol Deviation are the result of failing to follow existing policy rather than lack of policy altogether. To reduce these issues, organizations should implement targeted reinforcement strategies such as signage in clinical or procurement work areas, focused retraining for high-error roles or scenarios, and regular refresher models. Vessel storage missteps are a clear example of where greater policy reinforcement could yield substantial benefit.

4. Prioritize Verification Processes in Administrative Tasks

Administrative errors are frequent and often stem from failures in verification processes or deviations from existing policy, particularly during data entry into the Waitlist or DonorNet systems or documenting HLA information. Sending verification steps, such as dual sign offs or automated checks during critical data handling steps, can help prevent errors at the outset. These improvements should be accompanied by retraining initiatives to reinforce expectations around data accuracy and procedural adherence.

5. Strengthen Training at Critical Operational Interfaces

Events often occur at the junction between clinical and administrative domains, where information such as test results, organ data, or patient identifiers is transferred. Staff at these transition points – transplant coordinators, procurement teams, and hospital administrators – should receive focused training that emphasizes inter-team communication and fail-safe protocols. These interventions are particularly relevant for phases with frequent handoffs such as Organ Procurement and Organ Allocation.

6. Address Human Factors in Time-Sensitive, High-Pressure Settings

Human error tends to cluster in Primary Responsible Parties that involve time pressure, complex decision making, or field-based operations, such as transplant surgeons, organ transporters, or recovery surgeons/teams. Mitigating these risks requires not only policy adherence, but also investments in team-based coordination protocols, and tools that aid in real-time decision support. Greater effort should be made to stress test key workflows under realistic operational constraints and to provide simulation testing before policies and workflows are formalized.

7. Mitigate Process and System Design Vulnerabilities

Process or System Design issues account for the largest share of Primary Causes across multiple Primary Types. These vulnerabilities are often exposed when systems fail to account for real world complexity such as cross-team handoffs, coordination gaps, or reliance on manual workflows. Interventions should focus on developing stronger safeguards in process design (e.g., automated stopgaps, or system alerts that prevent or catch errors before they occur), and stress testing policies under different use case scenarios.

8. Improve Patient Safety Event Reporting and Investigation Guidance

A significant proportion of events across all root nodes remain unclassifiable, either due to minimal qualitative documentation, overly general narratives, or unclear cause. Improving clarity in reporting guidance and training reviewers on how to classify edge cases could reduce the unclassifiable bucket over time.

9. Enhancements to Coding Taxonomy Based on Coding and Analysis

The commonality of certain specific types of events that do not have specific types in the coding taxonomy only became clear after coding several hundred events or analyzing the data. An example would be events in which kidneys are laterally swapped, and the incorrect kidneys are sent to transplant centers. Only after coding more than two sample sets did it become clear that this event repeated to the extent it should have its own type. Such events are identifiable in type and were therefore placed into Other subclassifications. Creating specific types for these events would enhance the Codebook and allow for better tracking of these events.

Opportunities in AI Event Classification

Although the NLP classifiers did not achieve consistently strong performance across all primary classifications, this work represents an important first step in applying AI to the large-scale coding of patient safety events. While overall performance varied across the six models, there are encouraging signs that automated approaches can add value. In particular, the Primary Impacted Individual model achieved performance approaching a threshold that could be viable for certain retrospective analyses, even though it and the other models remain below the level needed for operational use. This finding

illustrates that some classification tasks are more tractable than others and may be ready for partial automation sooner than others, providing proof of concept for the broader approach.

At the same time, several challenges emerged that explain current performance levels. Notably, patterns emerged between the manual and automated coding: models tended to perform worse in the same primary classifications where human coders also had lower agreement. This suggests that the challenges are not solely due to model limitations but are also driven by the inherent ambiguity, overlap, and complexity of the Taxonomy itself. In other words, the same factors that make some events difficult for humans to code consistently also make them difficult for algorithms to classify reliably.

Several additional factors further constrained model performance. The limited size of the manually labeled dataset (739 training records) restricted the ability of the BERT models to learn complex distinctions. The free-text event descriptions often contained acronyms and shorthand, even after replacing commonly known acronyms with long form versions, which reduced clarity and likely contributed to misclassification. The limited input window of BERT also meant that some longer narratives were truncated, even after data cleaning efforts, resulting in the loss of potentially relevant context. Moreover, the models showed a tendency to over-predict broader categories such as Unclassifiable, Not Enough Qualitative Information to Code, and Other, while under-predicting more specific categories. This pattern is consistent with weak supervision, where models tend to gravitate toward broader or more generalizable categories when uncertainty is high.

These findings point to clear opportunities for improvement. Refining the Taxonomy to reduce overlap and ambiguity among categories would likely benefit both manual and automated coding. Expanding the volume and diversity of labeled training data would strengthen supervised learning, expanding phrase coverage for labeling functions, implementing fine-tuning adjustments, and establishing continuous quality assurance checks would also likely yield incremental gains. Exploring models with larger input windows or alternative architectures could address the issue of truncated narratives.

Beyond these refinements, a transition to or augmentation with large language models (LLMs) presents a promising future direction. LLMs can better capture contextual nuance, handle longer text spans, and flexibly adapt to diverse inputs. However, before such models could be deployed in practice, data privacy and PII safeguards must be addressed, particularly given the sensitivity of patient safety data. With appropriate safeguards and human oversight, LLMs could support not only retrospective analyses but also prospective applications, such as automatically coding incoming events in near-real time, enabling continuous monitoring, automated reporting, and integration into dashboards for decision support.

In sum, while the current weakly supervised BERT classifiers did not reach thresholds necessary for operational use, they provide a valuable foundation. They demonstrate both the feasibility of applying AI to patient safety event classification and the pathways forward to enhance performance through taxonomy refinement, data expansion, model innovation, and eventual integration into real-time pipelines.

Limitations of Manually Coded Analysis

Quality and Amount of Available Data

The quality and amount of qualitative data varied greatly across the dataset. As shown in the Analysis section, the amount of data with sufficient qualitative information to code decreased over time. However, within the events with sufficient qualitative information to code, the quality and amount of information also varied.

In the best cases, qualitative information had detailed investigation findings, clearly identified root cause analyses, and clearly identified corrective action plans. On the other extreme, especially in the later years of the dataset, very little information was provided, leading to multiple fields being coded as unclassifiable.

This was the single most significant limiting factor when coding but is influenced by many other downstream issues. A reduction in provided content leads to further confusion and human error when developing manual codes, especially when nuance and complexity are high. Limited qualitative information not only limited the volume of patient safety events that could be categorized across primary classifications and subclassifications but also increased the significance of the subjective dimension of coding.

Finally, there is a risk of underreporting from certain types of actors as well as a risk of underreporting for certain types of patient safety events.

Manual Coding Process

Subjectivity

The manual coding process was inherently subjective, as it required applying human judgment. This was mitigated first by a Codebook that effectively distinguished between classifications and subclassifications within nodes. Secondly, the influence of subjectivity was mitigated by interrater reliability checks. However, the impact of subjectivity increased when less qualitative information was provided.

Coding One Row Per Incident

Only one primary classification was coded for each node, but the qualitative information reported with a patient safety event may involve more than one event. Multiple events within a single record of qualitative information may appear in two ways. First, as a single causal chain of events. For example, kidney pumps are mislabeled, so the wrong kidneys are put on each pump and ultimately sent to the wrong transplant center. In this case, the primary error is the first error, although another reasonable interpretation is that the primary error is the most significant. Second, multiple events may arise out of the same cause but not as a single causal chain. For example, due to poor concern for patient wellbeing at a transplant center, multiple patients are improperly evaluated, one dies, and others experience no harm. In this case, primary means most significant: the most serious harm is recorded, the most common error type is recorded, and so on. The influence of subjectivity in either case is reduced when highly descriptive qualitative information is provided.

Sampled Data

The figures discussed represent estimates of the true quantities and percentages of patient safety events in the full population of data. This will introduce some degree of error. The original intent was to manually code a subset of data and use this subset to train an NLP model. To create a training dataset, the data was divided into strata and events were sampled from each stratum. In doing so, weights were created so events could be counted as if they came from a set of data where their source strata were weighed the same as the original population. This is a standard practice and superior to using unweighted counts. However, it does introduce a degree of error. See the Methodology section for a full discussion of the weighting process.

Codebook

No Formal Definition for Patient Safety Events

The absence of a clear definition of a patient safety event created challenges for coders in determining whether a safety event had occurred. The contractor's approach appeared to equate safety incidents with policy violations; however, a safety event can still occur in the absence of a policy violation. The aim was to code all incidences of patient harm that might reasonably have been prevented by the transplant system, not only those within the scope of OPTN policy. For example, when an investigator implied that an OPO's behavior seemed like a successful end run around OPTN organ allocation policy, this was coded as a patient safety event. However, there was no clear threshold here. This injects a level of subjectivity. However, this subjectivity was reduced by interrater reliability checks.

Limitations From Outcome Definitions

No Detectable Harm and No Recorded Harm

No detectable harm is used to code events where harm is not discussed, but at other nodes, unclassifiable would have been used. For most nodes, when it is not possible to determine the appropriate value, Unclassifiable is used. This includes when no information about a node is provided. However, No Detectable Harm is used in two situations:

- An attempt to detect harm was made, and no harm was detected, and
- Harm was not mentioned in the qualitative information.

Especially for events with little qualitative information, harm is often not mentioned at all, and it is not clear that harm would have been mentioned had it occurred. In these instances, it is possible that there was harm that was potentially detectable but simply not mentioned in the qualitative information.

No Error and Not a Patient Safety Event

No Error encompasses two different outcomes. First, it was used if a Primary Responsible Party did not err. For example, the transplant of a HIV positive organ was reported but the recipient properly consented. Because there was proper consent, there was no error. Second, No Error was used if a Primary Responsible Party did err, but it was not a patient safety event. For example, a hospital was found to be choosing funeral homes for deceased patients, which may have been unlawful and therefore a kind of error. However, there was no potential negative effect on patient safety, so it was not a patient safety event. No Error was used in both situations as it was the best fit, but it was not entirely accurate.

Cause

Human Factors and Process or System Design Issues

Where a primary responsible party deviated from policy or protocol or had a policy or protocol that was found deficient, the Cause of the patient safety event was coded as a Process or System Design – Policy/Protocol Deviation or Policy/Protocol Deficiency respectively; however, this may cause an undercounting of Human Factors causes. The specific actor responsible for an error would have erred for a certain reason or with a certain mental state. For example, a lab may be responsible for an error in documentation because a verification step failed. This would get coded as a Policy/Protocol Deviation. But the underlying reason may be that a laboratory technician willfully skipped the step, was improperly trained and never knew about the step, or simply made a mistake.

This coding aligns with the corrective action plans provided in the qualitative information, which generally prescribe improvements to policy where policy is deficient or re-trainings and other reminders about policy when actors deviate from policy. Further, Policy/Protocol Deviations specifically are very rarely intentional. They are overwhelmingly simple mistakes like typos.

Communication Breakdowns as Process or System Design Issues

Communication Breakdowns can be thought of as Process or System Design issues, as they often result from flaws in processes. For example, an OPO and transplant center may not be aligned on which communication channels to use during different stages of the organ allocation process. As a result, one fails to timely receive information from the other. However, the Codebook places Communication Breakdowns under the Communication Primary Cause primary node. Given the coders' and the qualitative information's focus on processes, this may lead to an undercounting of Communication Breakdowns.

Living Donor Adverse Events

There are mandatory reporting requirements for certain events experienced by Living Donors, called Living Donor Adverse Events in the Codebook. To reflect the seriousness with which policy treats these events, these are always treated as errors. Essentially, coders were much more risk averse in finding that the outcome of a Living Donor Adverse Event was No Error.

However, it is unclear whether many Living Donor Adverse Events resulted from errors. In the extreme case, the most reasonable inference is that no actor in the transplant center was responsible. For example, a Living Donor died less than two years after donation in a motor vehicle accident. This was possibly but likely not related to their living donor status. In other cases, notably recovery operations aborted after anesthesia initiation, evidence of causation is entirely omitted from the qualitative information. In these events, causes that arise during the recovery operation will be stated, but generally they are not traced back to any failure or lack thereof during the evaluation stage.

Conclusion

Many errors across the transplant ecosystem reduce to problems with the execution of existing processes or policies. However, there has also been a measurable decrease in the quality of qualitative information, making it more difficult to determine the cause of patient safety events. This analysis

highlighted the need to increase policy compliance, create policies and practices that better align with the realities of the transplant system in practice, and to increase the quality and availability of qualitative information.

Most patient safety events with qualitative information are caused by Policy/Protocol Deviations or Deficiencies and are Administration events or other events that can be fairly described as data entry or documentation errors. These events overwhelmingly affect transplant patients, they predominantly occur during procurement and throughout the transplant process, to include pre-transplant and post-transplant processes, and OPOs and transplant centers are responsible for most errors. The cause of events is increasingly unclear in recent years due to a decline in the quality of qualitative information.

The frequency of Policy/Protocol Deviations suggests the failure to adhere to existing policies. Qualitative information often prescribes effective corrective action plans in response to these Policy/Protocol Deviations, but the actions suggested ought to be applied more broadly. These include signage in work areas, focused re-trainings for high-risk roles, and more frequent refresher trainings. As Policy/Protocol Deviations are frequent, these could have a substantial effect on reducing the overall rate of patient safety events. Similarly, Administration errors often reduce to the failure of a verification process during data entry. These are frequent, so targeting the enforcement of verification policies and processes could have a substantial effect on reducing the overall patient safety event rate. To the extent policies are deficient, a review of events coded as caused by Policy/Protocol Deficiencies could help identify policy gaps.

Patient safety events occur less due to wrongdoing or negligence on the part of individuals or institutions and more because of the realities of the transplant system in practice. Where different actors interface, different domains meet or time pressure and stress are high, there may be an increased risk of Policy Deviation. Staff at transition points should receive training focused on inter-team communication. These and similar issues may not become clear until policy meets the real world. To respond to the effects of stress and time pressure, policy must be stress tested before it is implemented.

While the provided qualitative information is rich and informative in the best cases, it is often unclear and has become increasingly unclear in recent years. Reporting processes should be standardized to include a root cause analysis for each patient safety event.

In terms of the NLP modeling, although the weakly supervised BERT models did not yet reach operational thresholds, this evaluation underscores the opportunity for expanding NLP and LLM approaches to transform how patient safety events are classified and monitored. Building on the foundation established here, future efforts could move toward scalable AI pipelines that automatically code events, feed into dashboards, and support near real-time reporting. With continued refinement of taxonomies, expansion of labeled data, and safeguards for privacy and PII, AI can evolve from a retrospective research tool into a proactive system that augments manual review, reduces burden, and delivers more timely insights to strengthen patient safety.

There are several potential next steps given the findings and recommendations presented. Coding the entire population of events from 2019 to 2024 would reduce any uncertainty introduced from analyzing a subset of the patient safety events. Implementing a standardized coding process to track patient

safety data at a regular cadence would enhance the understanding of the patient safety data and increase the ability to respond to it. Generating a standardized report or dashboard to present patient safety data would promote transparency regarding trending patient safety events. Finally, identifying low hanging fruit recommendations based on this report to implement and track would improve patient safety.

Appendix A: Supporting Documents

1. TO6 TA3 Patient Safety Event Taxonomy- PowerPoint
2. TO6 TA3 Patient Safety Event Codebook – Excel
3. TO6 TA3 Patient Safety Event Qualitative Analysis
4. TO6 TA3 NLP Metrics and Inference