Investigating Disparities, Clinical Variables, and Predictive Modeling in Organ Procurement

Benzon Carlitos Salazar

May 03, 2025

Contents

1 Introduction

Organ transplantation is a critical component of modern healthcare, offering life-saving interventions for patients with end-stage organ failure. However, the process of organ procurement, where we have donor referral to successful retrieval, is influenced by a complex interplay of medical, logistical, and systemic factors. Despite ongoing efforts to improve equity and efficiency in organ donation, disparities persist across racial, socioeconomic, and geographic lines, raising concerns about fairness and resource allocation.

My full code for this project can be found here: https://github.com/carrliitos/orchid_investigation

This project focuses on two interrelated themes in organ transplantation: disparities in organ procurement, clinical predictors of organ viability. The analysis will draw on the **Organ Retrieval and Collection of Health Information for Donation (ORCHID) dataset**, a publicly available resource from PhysioNet comprising over **133,101 deceased donor referrals** across six **Organ Procurement Organizations (OPOs)** in the United States.

Specifically, I will attempt to address the following questions:

1. **Disparities in Organ Procurement Outcomes** – To what extent do race, socioeconomic status, and geography influence procurement rates?

2. **Clinical Predictors of Procurement Success** – Which medical and laboratory parameters are most strongly associated with successful organ procurement?

These questions are critical to improving the equity and efficiency of the organ transplantation system. By identifying disparities and enhancing predictive capabilities, this work aims to support more informed and equitable decision-making in organ allocation.

1.1 Data Collection Process

1.1.1 Methods and Tools

I will be relying on publicly available data from the **Organ Retrieval and Collection of Health Information for Donation (ORCHID) dataset**, hosted on **PhysioNet**. The data collection process will involve the following steps:

1. **Data Acquisition:**

- The dataset will be downloaded directly from PhysioNet using their web interface.

2. **Data Structure and Organization:**

- The dataset consists of multiple CSV files, including **demographic, clinical, and procedural data** related to organ procurement.
- The tables will be linked using the unique **PatientID** field.

3. **Data Storage and Management:**

- The dataset will be stored in a **secure local repository** or **cloud storage** for easy access.
- R will be used for the data exploration, preprocessing, and analysis.

4. **Data Cleaning and Preprocessing:**

- Handling missing values by applying imputation or exclusion methods where appropriate.
- Standardizing variable formats for consistency in analysis.
- Filtering data based on relevant **study criteria**, such as time period, OPOs, or specific donor characteristics.

5. **Tools for Data Processing and Analysis:**

- **R (tidyverse, dplyr, ggplot2, caret, glmnet)** for data manipulation, visualization, and regression modeling.
- **R Markdown** for exploratory data analysis and documentation.

1.2 Data and Variables

The ORCHID dataset contains **demographic, clinical, and procedure information on deceased donor referrals**. Specifically, the dataset includes:

- **Demographic Data**: Race, age, sex, geographic region, socioeconomic proxies (hospital identifier, OPO region).
- **Clinical Variables**: Blood chemistry, hematology, arterial blood gas levels, infection status, cause of death, comorbidities.
- **Referral and Procurement Details**: Referral source, organ type, authorization status, OPO performance metrics.
- **Outcome Variable**: Whether the organ was successfully procured.

This dataset can be obtained from PhysioNet: https://physionet.org/content/orchid/2.0.0/

2 Data Extraction and Clean up

All code chunks will not be evaluated since clean up was already performed prior to knitting this document.

The data cleanup involved reading raw CSV files from the ORCHID dataset, transforming them into a structured format, and saving the cleaned data as Parquet files. The key steps included:

2.0.1 Reading Raw Data:

- CSV files for various event categories (e.g., ABG, CBC, Chemistry, Culture, Fluid Balance, Hemodynamics, etc.) were read into R.

2.0.2 Data Cleaning and Transformation:

- Unnecessary columns (`RowID`, unnamed columns) were dropped.
- Data was grouped by event type (e.g., `abg_name`, `cbc_name`, `chem_name`) and converted to a wide format using `pivot_wider()`.
- Numeric values were coerced into appropriate data types (`double`).
- An `opo_group` identifier was extracted from `PatientID` to categorize patients.
- `tidyr::fill()` was used to propagate missing values up and down within each group.
- Data was grouped by `PatientID`, `time_event`, and other relevant identifiers, then deduplicated using `slice_head()`.

2.0.3 Data Output & Cleanup:

- The cleaned datasets were saved in Parquet format for efficient storage and processing.
- Raw CSV files were removed after ensuring successful cleanup.
- Memory was cleared using `rm(list = ls(all.names = TRUE))` and `gc()` to optimize performance.

This process ensured that data was well-structured, deduplicated, and efficiently stored while maintaining data integrity.

Below is an example of what the whole clean up process looked like.

```
library(magrittr)

orchid_folder <- here::here("data-raw/physionet.org/files/orchid/2.0.0")
```

```
abg_events <-
  readr::read_csv(here::here(orchid_folder, "ABGEvents.csv"))

abg_events_proc <-
  abg_events %>%
  dplyr::select(-c("...1", "RowID")) %>%
  dplyr::group_by(abg_name) %>%
  dplyr::mutate(row_id = dplyr::row_number()) %>%
  tidyr::pivot_wider(names_from = abg_name, values_from = value) %>%
  dplyr::ungroup() %>%
  dplyr::select(-row_id) %>%
  dplyr::mutate(opo_group = as.factor(substring(PatientID, 0, 4)),
                abg_ventilator_mode = as.character(abg_ventilator_mode),
                PH = as.double(PH),
                PCO2 = as.double(PCO2),
                PO2 = as.double(PO2),
                HCO3 = as.double(HCO3),
                BE = as.double(BE),
                O2SAT = as.double(O2SAT),
                FIO2 = as.double(FIO2),
                Rate = as.double(Rate),
                TV = as.double(TV),
                PEEP = as.double(PEEP),
                PIP = as.double(PIP))

abg_events_final <-
  abg_events_proc %>%
  dplyr::group_by(PatientID, time_event, abg_ventilator_mode, opo_group) %>%
  tidyr::fill(dplyr::everything(), .direction = "updown") %>%
  dplyr::slice_head() %>%
  dplyr::ungroup()

abg_events_final %>% nanoparquet::write_parquet(here::here("data", "abg_events.parquet"))
rm(abg_events, abg_events_proc, abg_events_final)
```

```
# List of files to remove
files_to_remove <- c(
  "ABGEvents.csv",
  "calc_deaths.csv",
```

```
  "CBCEvents.csv",
  "ChemistryEvents.csv",
  "CultureEvents.csv",
  "FluidBalanceEvents.csv",
  "HemoEvents.csv",
  "referrals.csv",
  "SerologyEvents.csv"
)

# Remove files only if they exist
purrr::walk(files_to_remove, ~ {
  file_path <- here::here(orchid_folder, .x)
  if (file.exists(file_path)) {
    file.remove(file_path)
  }
})

rm(list = ls(all.names = TRUE)) # clear all objects including hidden objects
invisible(gc()) # free up memory
```

2.1 Purpose of Data Cleanup

The raw ORCHID dataset contains a wide variety of clinical event data stored across multiple CSV files. These files are structured in a long format with multiple redundant columns, inconsistent naming conventions, and varying data types. The primary objective of the cleanup was to create tidy, analysis-ready data by:

- Converting event-level long-form data into patient-level wide format.
- Ensuring consistency across numeric, categorical, and time-based variables.
- Removing duplicate rows, irrelevant columns, and noise.
- Storing processed outputs in a space-efficient, analysis-friendly format (`.parquet`).

This enables downstream applications such as predictive modeling, visualization, and exploratory data analysis to proceed without further data wrangling.

2.2 Overview of Files Processed

Raw File Name	Description	Output Filename
`ABGEvents.csv`	Arterial blood gas measurements	`abg_events.parquet`
`CBCEvents.csv`	Complete blood count lab events	`cbc_events.parquet`
`ChemistryEvents.csv`	Chemistry lab test results	`chem_events.parquet`
`CultureEvents.csv`	Microbial culture results	`culture_events.parquet`
`FluidBalanceEvents.csv`	Fluid input/output documentation	`fluid_balance_events.parquet`
`HemoEvents.csv`	Hemodynamic monitoring values	`hemo_events.parquet`
`SerologyEvents.csv`	Serological test outcomes	`serology_events.parquet`
`calc_deaths.csv`	Calculated patient mortality data	`calc_deaths.parquet`
`referrals.csv`	Referral information (specialty etc.)	`referrals.parquet`

2.3 Notes on Data Quality and Consistency

- **Column Removal**: Columns such as `RowID` and unnamed index columns (e.g., `...1`) were dropped to reduce noise.
- **Wide Format Transformation**: Measurements grouped by name (e.g., `abg_name`, `cbc_name`) were converted into wide format using `pivot_wider()` for one-row-per-timepoint structure.
- **Type Coercion**: Key clinical variables were cast into `double` to ensure numeric operations are consistent.
- **Group-level Missingness Imputation**: `tidyr::fill()` was used within each group (e.g., per `PatientID`) to forward- and backward-fill missing values where reasonable.
- **Deduplication Strategy**: Within each patient-time-event grouping, only the first row was retained using `slice_head()`, assuming earlier rows are most relevant.
- **Group Identifier**: An `opo_group` variable was extracted using the first four characters of `PatientID` to allow stratified subgroup analyses.

2.4 Folder Structure and Output Format

- Raw data is located in: `data-raw/physionet.org/files/orchid/2.0.0/`
- Cleaned files are saved to: `data/`
- All output files are written in **Parquet** format using the `nanoparquet::write_parquet()` function for fast I/O and minimal disk usage.

3 Analysis and Results

3.1 Objective 1: Analyze Disparities in Organ Procurement Outcomes

3.1.0.1 Goal To assess whether organ procurement rates differ across demographic and geographic groups—particularly by race/ethnicity, socioeconomic proxies, and OPO region—and quantify the magnitude and significance of these disparities.

- **Outcome Variable**
  - `organ_procured` (binary): 1 if organ was successfully procured, 0 otherwise.
- **Primary Exposure Variables**
  - **Age**
  - **Gender**
  - **Race**
  - **Organ Procurement Organizations (OPOs)**
- **Covariates (for adjustment)**
  - UNOS defined cause of death
  - UNOS defined mechanism of death
  - UNOS defined circumstances of death
  - Time of referral from hospital

3.1.1 Organ Procurement Outcomes by Demographic, Geographic, and Clinical Characteristics

Table 2: Organ Procurement vs. Outcomes

Characteristic	Overall (N = 132968)	Not Procured (N = 123466)	Procured (N = 9502)
Mean Patient Age (IQR)	58 (48, 71)	59 (50, 72)	40 (28, 54)
Age Category - no. (%)			
Less than 18 y.o.	4,535/132,968 (3.4%)	3,819/4,535 (84%)	716/4,535 (16%)
18-19 y.o.	1,112/132,968 (0.8%)	789/1,112 (71%)	323/1,112 (29%)
20-29 y.o.	6,729/132,968 (5.1%)	5,079/6,729 (75%)	1,650/6,729 (25%)
30-39 y.o.	9,607/132,968 (7.2%)	7,781/9,607 (81%)	1,826/9,607 (19%)
40-49 y.o.	13,926/132,968 (10%)	12,119/13,926 (87%)	1,807/13,926 (13%)
50-59 y.o.	25,730/132,968 (19%)	23,795/25,730 (92%)	1,935/25,730 (7.5%)
60-69 y.o.	33,976/132,968 (26%)	32,993/33,976 (97%)	983/33,976 (2.9%)
70+ y.o.	37,353/132,968 (28%)	37,091/37,353 (99%)	262/37,353 (0.7%)
Gender - no. (%)			
M	78,216/132,968 (59%)	72,374/78,216 (93%)	5,842/78,216 (7.5%)
F	54,752/132,968 (41%)	51,092/54,752 (93%)	3,660/54,752 (6.7%)
Race - no. (%)			
White / Caucasian	79,627/132,968 (60%)	73,843/79,627 (93%)	5,784/79,627 (7.3%)
Black / African American	25,155/132,968 (19%)	23,615/25,155 (94%)	1,540/25,155 (6.1%)
Hispanic	20,702/132,968 (16%)	18,942/20,702 (91%)	1,760/20,702 (8.5%)
Other / Unknown	7,484/132,968 (5.6%)	7,066/7,484 (94%)	418/7,484 (5.6%)
Organ Procurement Organization - no. (%)			
OPO4	33,616/132,968 (25%)	31,217/33,616 (93%)	2,399/33,616 (7.1%)
OPO1	32,079/132,968 (24%)	29,382/32,079 (92%)	2,697/32,079 (8.4%)
OPO6	22,905/132,968 (17%)	21,631/22,905 (94%)	1,274/22,905 (5.6%)
OPO2	16,142/132,968 (12%)	15,264/16,142 (95%)	878/16,142 (5.4%)
OPO5	15,725/132,968 (12%)	14,128/15,725 (90%)	1,597/15,725 (10%)
OPO3	12,501/132,968 (9.4%)	11,844/12,501 (95%)	657/12,501 (5.3%)
UNOS defined cause of death - no. (%)			
Anoxia	40,340/132,968 (30%)	36,458/40,340 (90%)	3,882/40,340 (9.6%)
Unknown Cause of Death	29,761/132,968 (22%)	29,761/29,761 (100%)	0/29,761 (0%)
Other	24,248/132,968 (18%)	24,007/24,248 (99%)	241/24,248 (1.0%)
CVA/Stroke	17,577/132,968 (13%)	15,314/17,577 (87%)	2,263/17,577 (13%)
Head Trauma	10,798/132,968 (8.1%)	8,159/10,798 (76%)	2,639/10,798 (24%)
Other Cause of Death	10,244/132,968 (7.7%)	9,767/10,244 (95%)	477/10,244 (4.7%)
UNOS defined mechanism of death - no. (%)			
Unknown Mechanism of Death	34,500/132,968 (26%)	34,500/34,500 (100%)	0/34,500 (0%)
Cardiovascular	27,259/132,968 (21%)	25,599/27,259 (94%)	1,660/27,259 (6.1%)
Natural Causes	25,871/132,968 (19%)	25,555/25,871 (99%)	316/25,871 (1.2%)
ICH/Stroke	16,472/132,968 (12%)	14,184/16,472 (86%)	2,288/16,472 (14%)
Other Mechanism of Death	11,824/132,968 (8.9%)	8,688/11,824 (73%)	3,136/11,824 (27%)
None of the Above	8,745/132,968 (6.6%)	8,482/8,745 (97%)	263/8,745 (3.0%)
Blunt Injury	8,297/132,968 (6.2%)	6,458/8,297 (78%)	1,839/8,297 (22%)

Table 2: Organ Procurement vs. Outcomes *(continued)*

Characteristic	**Overall** (N = 132968)	**Not Procured** (N = 123466)	**Procured** (N = 9502)
UNOS defined circumstances of death - no. (%)			
Death from Natural Causes	64,153/132,968 (48%)	60,438/64,153 (94%)	3,715/64,153 (5.8%)
Unknown Circumstances of Death	34,444/132,968 (26%)	34,444/34,444 (100%)	0/34,444 (0%)
Other Circumstances of Death	28,860/132,968 (22%)	24,424/28,860 (85%)	4,436/28,860 (15%)
Motor Vehicle Accident	5,511/132,968 (4.1%)	4,160/5,511 (75%)	1,351/5,511 (25%)

Table 2 summarizes the baseline characteristics of the study cohort (N = 132,968), stratified by whether an organ was successfully procured. Among all decedents, 9,502 (7.1%) had at least one organ procured, while 123,466 (92.9%) did not.

Age was strongly associated with organ procurement. The median age among those whose organs were not procured was 59 years (IQR: 50–72), compared to 40 years (IQR: 28–54) among those who underwent procurement. Stratified age group analysis revealed that younger patients had higher procurement rates; for instance, decedents aged 20–29 had a procurement rate of 25%, compared to only 16% among those under 18. Procurement likelihood declined progressively with increasing age, suggesting a potential age-related disparity that may reflect clinical ineligibility, comorbidity burden, or differences in referral or donor evaluation practices.

Gender differences in procurement were modest. Males comprised 59% of the cohort and had a procurement rate of 7.5%, compared to 6.7% among females. Although the absolute difference was small, this slight imbalance may stem from anatomical compatibility, clinical profiles, or unmeasured systemic factors. However, without adjustment for clinical variables, it is unclear whether this difference reflects a true disparity or confounding.

Race and ethnicity exhibited more pronounced disparities. White/Caucasian decedents accounted for 60% of the cohort, with a procurement rate of 7.3%. In contrast, Black/African American decedents (19% of the cohort) had a lower procurement rate of 6.1%, while Hispanic decedents (16%) had a notably higher rate of 8.5%. The lower procurement among Black patients may reflect structural inequities in referral, consent, or evaluation processes. Conversely, the relatively high procurement among Hispanic decedents may be influenced by geographic or demographic differences in donor identification and suitability. These trends highlight the importance of adjusting for clinical and regional context in disparity analyses.

Organ Procurement Organization (OPO) performance varied substantially across regions. Procurement rates ranged from 5.3% in OPO3 to 10.2% in OPO5. While differences in case mix may explain some variation, the consistently higher yield in OPO5 suggests operational or procedural differences in donor management and recovery efficiency. Conversely, lower rates in OPOs 2, 3, and 6 may reflect logistical challenges, under-referral, or systemic inefficiencies. These inter-regional disparities underscore the need for standardized performance evaluation and the potential for best-practice dissemination from higher-performing OPOs.

UNOS-defined cause of death was reported for all decedents. The most common causes were **anoxia** (30%) and **unknown cause of death** (22%). **Head trauma** had one of the highest procurement rates at **24%**, followed by **CVA/stroke** at **13%**, suggesting these conditions often preserve organ viability. In contrast, **other** and **unknown** causes of death had substantially lower procurement rates, **1.0%** and **0%**, respectively. The complete lack of procurement among decedents with an unknown cause of death highlights the potential impact of documentation gaps on organ recovery. Collectively, these patterns underscore the importance of both clinical suitability and accurate case classification in maximizing donation opportunities.

UNOS-defined mechanism of death was available for all decedents. The most frequent mechanisms were **unknown** (26%), **cardiovascular** (21%), and **natural causes** (19%). Procurement rates were highest among decedents with **other mechanisms of death** (27%) and **blunt injury** (22%), both of which are consistent with external trauma and greater clinical suitability for organ donation. In contrast, **natural causes** and **cardiovascular deaths** were associated with low procurement rates, **1.2%** and **6.1%**, respectively—likely reflecting comorbidities or rapid physiological deterioration. Notably, **no organs were procured** from the 34,500 patients (26%) with an **unknown mechanism of death**, underscoring how incomplete documentation may contribute to missed donation opportunities.

UNOS-defined circumstances of death further highlight these patterns. **Death from natural causes** was the most common circumstance (48%) but had a relatively low procurement rate of **5.8%**. In contrast, **motor vehicle accidents**, while accounting for only 4.1% of the cohort, were associated with a substantially higher procurement rate of **25%**, indicating the clinical viability of donors who experience sudden traumatic deaths. **Other circumstances of death**, a heterogeneous category comprising 22% of cases, had a procurement rate of **15%**, suggesting mixed donor suitability. As with other variables, **no procurement occurred among the 34,444 patients (26%) with unknown circumstances of death**, emphasizing the critical need for complete and accurate documentation to improve identification of potential donors.

Figure 1: Proportion of organ procurement by Race, Gender, Age Category, and OPO. Bars show within-group percentages; procurement is more common among younger, male, Hispanic, and OPO5 patients.

Figure 1 summarizes disparities in organ procurement rates across key demographic and regional groups. Notable patterns include higher procurement rates among younger individuals, males, Hispanic patients, and those referred within OPO5. These descriptive trends motivated subsequent hypothesis testing and regression analysis.

3.1.2 Hypothesis testing

We hypothesize **whether organ procurement is associated with specific demographic, geographic, or clinical characteristics.**

More specifically, we are tesing the following:

- H_{0} (null): There is no difference in organ procurement rates across the levels of the variable.
- H_{1} (alternative): There is a difference in organ procurement rates across the levels of the variable.

While my initial hypothesis testing plan included both demographic and clinical variables, I am excluding `Cause_of_Death_UNOS`, `Mechanism_of_Death`, and `Circumstances_of_Death` due to evidence of documentation bias. These variables are disproportionately complete among patients whose organs were procured, but often missing or marked as "unknown" for those who were not. Because this documentation typically occurs during or after donor evaluation, the completeness of these fields may be contingent on the procurement process itself. As a result, statistical associations involving these variables may reflect process-related bias rather than true underlying clinical differences.

I also excluded `age_category`, as it is a derived variable from `Age`, which was already included in the analysis using a Welch two-sample t-test. Including both would introduce redundancy and increase the risk of multiple comparisons without providing additional insight.

My final hypothesis testing focused on variables that are consistently recorded and not conditional on procurement activity: `Race`, `Gender`, `OPO`, and continuous `Age`.

```
## # A tibble: 4 x 4
##   variable method                                          statistic   p.value
##   <chr>    <chr>                                               <dbl>     <dbl>
## 1 Race     Pearson's Chi-squared test                          126.  3.52e- 27
## 2 Gender   Pearson's Chi-squared test with Yates' continuit~    29.7 4.93e-  8
## 3 OPO      Pearson's Chi-squared test                          516.  2.30e-109
## 4 Age      Welch Two Sample t-test                             103.  0
```

All variables tested showed statistically significant differences in procurement rates:

- **Race** was significantly associated with procurement status ($x^{2} = 126$, $p < 0.001$), suggesting that procurement rates vary across racial and ethnic groups.
- **Gender** also showed a significant difference ($x^{2} = 29.7$, $p < 0.001$), although the magnitude of the association was smaller.
- **Organ Procurement Organization (OPO)** region was strongly associated with procurement likelihood ($x^{2} = 516$, $p < 0.001$), indicating considerable variation in outcomes across geographic regions.
- **Age**, treated as a continuous variable, was significantly different between procured and non-procured groups (Welch $t = 103$, $p < 0.001$), with younger patients more likely to have organs procured.

These results support the presence of disparities in procurement outcomes related to race, gender, region, and age, motivating further multivariable modeling to quantify these differences while adjusting for potential confounding.

3.1.3 Logistic Regression

```
log_reg_model <- stats::glm(procured ~ Race + Gender + OPO, data = obj1_abt, family = binomial)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.0733048	0.0342624	-76.268104	0.0000000	0.0685198	0.0783695
RaceHispanic	1.3468028	0.0381810	7.797953	0.0000000	1.2497622	1.4515502
RaceOther / Unknown	0.8706965	0.0590022	-2.346721	0.0189394	0.7748362	0.9765194
RaceWhite / Caucasian	1.1734948	0.0302799	5.283587	0.0000001	1.1061759	1.2455952
GenderM	1.1153456	0.0219386	4.975901	0.0000006	1.0684655	1.1644221
OPOOPO2	0.6704818	0.0412144	-9.699493	0.0000000	0.6181792	0.7265866
OPOOPO3	0.6205724	0.0456070	-10.461394	0.0000000	0.5671213	0.6781594
OPOOPO4	0.8498368	0.0298171	-5.456965	0.0000000	0.8015663	0.9009546
OPOOPO5	1.2726847	0.0342640	7.037382	0.0000000	1.1898686	1.3609242
OPOOPO6	0.6475777	0.0354920	-12.242669	0.0000000	0.6039110	0.6940658

After controlling for all covariates in the model:

- **Race**: Compared to Black patients (reference group), Hispanic patients had **higher odds** of organ procurement (OR is approx. 1.35), while patients categorized as Other/Unknown had **lower odds** (OR is approx. 0.87). White/Caucasian patients also had significantly higher odds (OR is approx. 1.17), suggesting potential racial disparities in procurement outcomes even after adjustment.
- **Gender**: Male patients had **higher odds** of procurement compared to females (OR is approx. 1.12), consistent with patterns observed in the unadjusted analysis.
- **OPO Region**: Substantial geographic variation was observed. Compared to OPO1 (reference), patients in OPO5 had the **highest odds** of procurement (OR is approx. 1.27), while those in OPO2, OPO3, and OPO6 had **significantly lower odds** (ORs ranging from 0.62 to 0.67). These differences may reflect regional variation in practices, infrastructure, or consent processes.

All associations were statistically significant ($p < 0.05$), and none of the confidence intervals included the null value (OR = 1.0). This indicates meaningful adjusted differences in organ procurement across race, gender, and OPO.

Specifically, Hispanic patients, White patients, male patients, and those in OPO5 had significantly higher odds of organ procurement compared to their respective reference groups. In contrast, patients in OPOs 2, 3, 4, and 6 had significantly lower odds.

Figure 2 displays the adjusted odds ratios (ORs) and 95% confidence intervals from the logistic regression model evaluating associations between demographic/geographic characteristics and the likelihood of organ procurement.

3.2 Objective 2: Evaluating the Influence of Clinical Variables on Procurement Success

3.2.0.1 Goal To evaluate whether specific medical and laboratory parameters, such as blood gas values, chemistry panels, complete blood counts, and hemodynamic indicators, are associated with successful organ procurement, and to quantify the strength and direction of these associations using univariate logistic regression models.

For this analysis, the following datasets are used:

- Arterial blood gas measurements
- Complete blood count lab events
- Chemistry lab test results
- Fluid input/output documentation

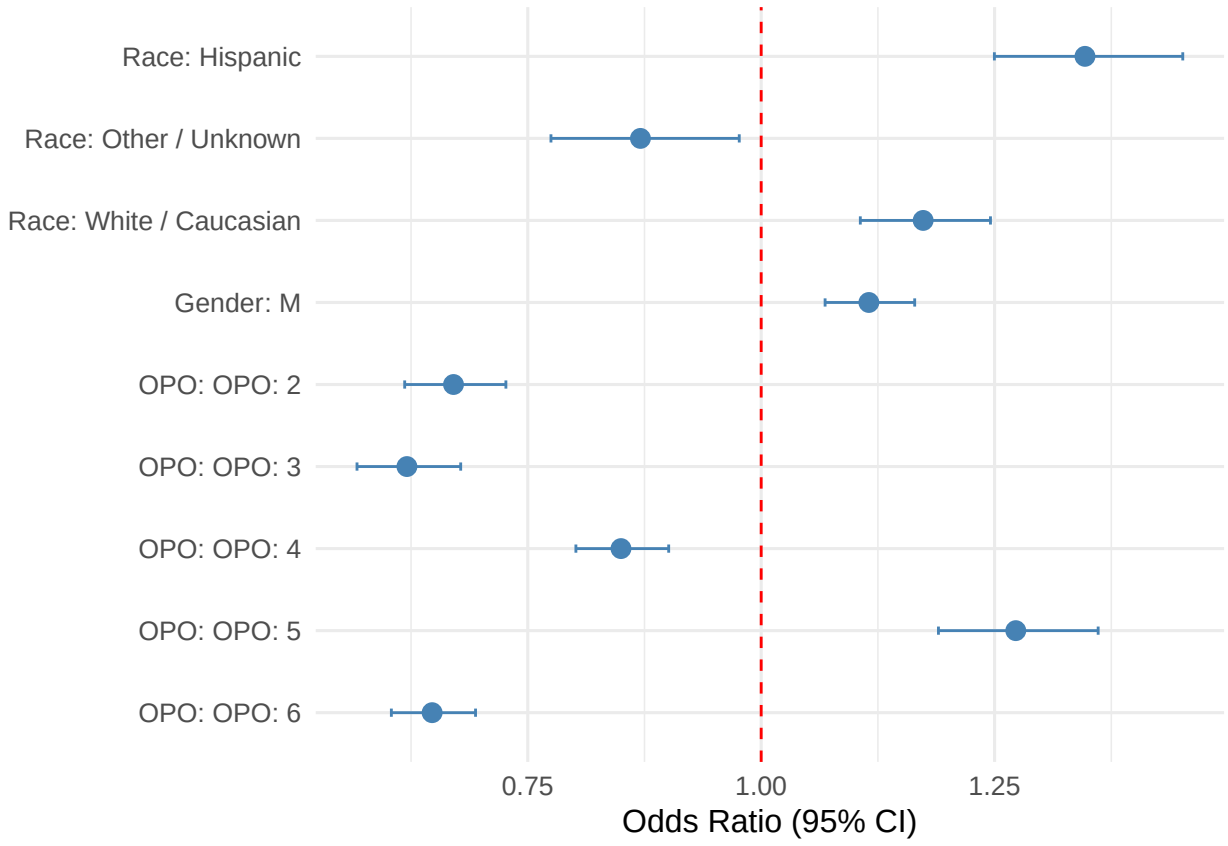

Figure 2: Odds Ratios from Logistic Regression Model

- Hemodynamic monitoring values

For each measurement, we take the patient's **latest** value. And then all latest labs information are joined together to create a one `all_labs_info` dataframe.

Outcome

- `procured` (binary): 1 = organ procured, 0 = not procured.

Since not all patients had every laboratory test completed, the dataset size decreased substantially after joining all lab datasets: from an initial 133,101 patients to 10,843 patients with at least one lab result recorded. Importantly, some patients may have only partially completed a given panel, for instance, a patient might have a White Blood Cell (WBC) count but lack values for Red Blood Cell (RBC) count or Hemoglobin (Hgb). Such patients are still retained in the `all_labs_info` dataframe, provided they have any available lab values.

This approach allows us to maximize sample size while preserving partial clinical information, ensuring that we do not exclude potentially informative cases due to isolated missingness.

Since there are still a lot of missing values, however, further processing is required to run the logistic regression model. I will further filter out variables with more than 50% missingness:

After this, we are now at 10,843 observations with 44 usable variables. Additionally, due to the massive imbalance of the `fluid_type`, I also had to remove that column, which leaves us with 43 variables.

Statistical Modeling Approach:

To identify clinical laboratory variables associated with successful organ procurement, we use a **manual backward stepwise logistic regression** approach. We begin with a full model that includes all available lab-based predictors from each patient's most recent arterial blood gas (ABG), complete blood count (CBC), chemistry, hemodynamic, and fluid balance data.

At each step, we remove the variable that contributes the least to the model, typically the one with the highest p-value or smallest effect size. After each removal, we refit the model and compare it to the previous version using **likelihood ratio tests (ANOVA)** or changes in the **Akaike Information Criterion (AIC)**. We continue this process until all remaining variables are either statistically significant or clinically relevant.

This approach yields a more parsimonious model, reduces overfitting, and helps isolate the strongest clinical predictors of procurement success.

3.2.1 Multivariable Logistic Regression Identifying Clinical Predictors of Procurement

To evaluate the association between clinical variables and the likelihood of organ procurement, I performed a multivariable logistic regression with backward stepwise elimination. The final model retained significant predictors after adjusting for potential collinearity and missingness.

Table 3 shows that several variables were independently associated with increased odds of procurement. These included higher **arterial pH** ($OR = 13.1$, 95% CI: 1.75–98.0, $p = 0.012$), **white blood cell count** ($OR = 1.05$ per unit increase, 95% CI: 1.02–1.09, $p < 0.001$), **hemoglobin** ($OR = 1.45$, 95% CI: 1.04–2.05, $p = 0.033$), **sodium** ($OR = 1.04$, 95% CI: 1.03–1.07, $p < 0.001$), **albumin** ($OR = 1.44$, 95% CI: 1.11–1.88, $p = 0.007$), and **magnesium** ($OR = 1.58$, 95% CI: 1.09–2.34, $p = 0.020$).

In contrast, higher values of **hematocrit** ($OR = 0.85$, 95% CI: 0.76–0.94, $p = 0.004$), **platelet count** ($OR = 0.99$ per unit increase, 95% CI: 0.993–0.997, $p < 0.001$), **potassium** ($OR = 0.74$, 95% CI: 0.58–0.96, $p = 0.021$), and **serum creatinine** ($OR = 0.90$, 95% CI: 0.84–0.96, $p = 0.002$) were negatively associated with procurement.

Ventilator mode was also significant: compared to the reference group, **pressure-controlled ventilation (PC)** was associated with increased odds ($OR = 1.49$, 95% CI: 1.01–2.24, $p = 0.049$), whereas **CPAP mode**

Table 3: Final Logistic Regression Model: Clinical Predictors of Procurement Success (Odds Ratios)

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.000	7.998	-3.335	0.001	0.000	0.000000e+00
OPOOPO2	2.883	0.215	4.932	0.000	1.910	4.436000e+00
OPOOPO3	1.306	0.584	0.458	0.647	0.458	4.761000e+00
OPOOPO6	1.772	0.397	1.443	0.149	0.852	4.093000e+00
abg_ventilator_modeAPRV	432129.727	386.508	0.034	0.973	0.000	
abg_ventilator_modeBiPAP	0.000	882.743	-0.018	0.986		1.452288e+71
abg_ventilator_modeCMV	1.435	0.476	0.759	0.448	0.597	3.932000e+00
abg_ventilator_modeCPAP	0.118	1.053	-2.033	0.042	0.012	8.180000e-01
abg_ventilator_modeOther	0.618	0.327	-1.471	0.141	0.333	1.205000e+00
abg_ventilator_modePC	1.494	0.204	1.966	0.049	1.007	2.245000e+00
abg_ventilator_modePRVC	1.096	0.247	0.371	0.711	0.683	1.804000e+00
abg_ventilator_modeSIMV	0.888	0.289	-0.409	0.682	0.513	1.604000e+00
PH	13.122	1.026	2.509	0.012	1.748	9.800100e+01
TV	1.001	0.001	2.117	0.034	1.000	1.003000e+00
WBC	1.055	0.015	3.539	0.000	1.025	1.087000e+00
Hgb	1.453	0.175	2.129	0.033	1.039	2.054000e+00
Hct	0.846	0.058	-2.902	0.004	0.755	9.440000e-01
Ptl	0.995	0.001	-5.748	0.000	0.993	9.970000e-01
Lymp	1.036	0.014	2.497	0.013	1.008	1.066000e+00
Sodium	1.045	0.010	4.470	0.000	1.025	1.065000e+00
K	0.743	0.129	-2.309	0.021	0.577	9.570000e-01
Creatinine	0.898	0.035	-3.124	0.002	0.839	9.610000e-01
Calcium	1.162	0.108	1.396	0.163	0.983	1.440000e+00
Mg	1.578	0.197	2.319	0.020	1.086	2.341000e+00
SGOTAST	0.999	0.000	-3.861	0.000	0.999	1.000000e+00
Albumin	1.444	0.135	2.714	0.007	1.108	1.885000e+00
TotalBili	0.932	0.039	-1.816	0.069	0.869	1.014000e+00
AverageHeartRate	1.014	0.004	3.255	0.001	1.006	1.023000e+00

was associated with decreased odds ($OR = 0.12$, 95% CI: 0.012–0.82, $p = 0.042$). Procurement likelihood varied significantly by OPO, with **OPO2** exhibiting a markedly higher odds of procurement ($OR = 2.88$, 95% CI: 1.91–4.44, $p < 0.001$) relative to the reference group.

These findings suggest that a combination of metabolic, hematologic, and respiratory parameters, as well as operational region, play important roles in procurement success.

4 Conclusions

This analysis of over 130,000 decedents from the ORCHID dataset reveals clear and statistically significant disparities in organ procurement across racial, gender, age, and geographic lines. Younger age, male sex, and residence within certain Organ Procurement Organizations (notably OPO5) were all associated with higher odds of organ procurement. After adjusting for covariates, both Hispanic and White decedents had significantly greater odds of organ procurement compared to Black decedents, indicating persistent racial inequities even when controlling for observable clinical and demographic factors.

Geographic disparities were also pronounced, with some OPO regions demonstrating procurement rates nearly double those of others, suggesting variation in operational efficiency, consent practices, or systemic access. Importantly, fields such as cause, mechanism, and circumstances of death were often incomplete for non-procured cases, raising the possibility of documentation bias and reinforcing the need for standardization and completeness in clinical reporting.

These findings underscore the urgent need for policy-level interventions aimed at standardizing referral practices, improving donor identification workflows, and addressing structural barriers to equitable organ donation. I believe that future work should focus on disentangling institutional, clinical, and societal drivers of these disparities and evaluating targeted interventions across under-performing regions and demographic groups.

5 Discussion

The results of my project have important implications for the donor-eligible population represented in the ORCHID dataset. The observed disparities in organ procurement rates, especially by race, age, gender, and geographic region (proxied by OPO), highlight structural and systemic inequities that may limit access to life-saving transplant opportunities. For example, the significantly lower procurement rates among Black decedents, even after adjusting for demographic and geographic factors, suggest the influence of unmeasured social, institutional, or consent-related barriers. Likewise, substantial variation in procurement performance across OPO regions indicates a lack of consistency in donor identification and recovery practices, with some regions demonstrating markedly lower efficiency or success rates.

If I were to repeat my study, several improvements could enhance its robustness and validity. First, more comprehensive adjustment for clinical variables, such as comorbidities, hemodynamic stability, or documented contraindications to procurement, would allow for a more nuanced understanding of whether disparities are clinically justified or reflect systemic bias. Second, future analyses would benefit from linkage to hospital- and OPO-level operational data, which could show some light on institutional differences in referral timing, staffing, consent protocols, and donor management practices. Additionally, integrating qualitative or socio-cultural data, such as language preference or next-of-kin refusal rates, may help contextualize demographic disparities.

Several potential biases may affect these findings. The most prominent is **documentation bias**—key variables such as cause, mechanism, and circumstances of death were often missing or marked as "unknown" in non-procured cases. This likely reflects a post-referral or post-evaluation data entry process, where full clinical documentation is completed only after a patient is deemed a viable donor (i.e., the donor has passed away). Consequently, analyses that include these fields may suffer from **informative missingness**, where the absence of data is correlated with the outcome itself. In addition, there may be **selection bias** if referral

practices differ systematically across hospitals, regions, or patient populations, potentially underrepresenting certain groups in the donor evaluation process.

In summary, while the findings point to real and actionable disparities in organ procurement, they must be interpreted in the context of the limitations presented above. Future study on the ORCHID dataset should aim to improve data completeness, incorporate richer clinical context, and examine upstream processes, such as referral timing and evaluation criteria, to more precisely target opportunities for reform.

6 Appendix

List of Tables

List of Figures