KIPA2024_01 Simulation Analysis Report

Report for the Data Request from the OPTN Kidney and Pancreas Committees

Date:
November 25, 2024
Prepared by:
Tim Weaver, Nick Wood, Josh Pyke
Data Request ID:
KIPA2024_01
Timeline:
Committee met: March 8, 2024
Request made: March 19, 2024
Analysis plan submitted: April 2, 2024
Analysis report to be submitted: November 13, 2024

This report presents simulation results and analysis. The "Data Request" section is the request from the committees including overall goals and specific research questions of interest. The "Analysis Plan" section is the analysis plan we submitted in response to the request and describes how we arrived at the results in the "Simulated Results" section. We largely stuck to this plan except where noted in footnotes with the label "CTAP" (change to analysis plan).

1 Executive Summary

The purpose of this request was to select simulation submodels that will be used to answer the Kidney and Pancreas Committees' "Research Questions" in future simulation requests. To that end, we designed and assessed various collections of submodels (CSMs) in accordance with the "Analysis Plan". Each figure in the "Simulated Results" section corresponds to a research question, and each figure contains data on a metric that is intended to answer that research question. The far-left point on each figure represents the true, historical value for that metric, while each remaining point represents the mean simulated value of that metric for a CSM. The ability of each CSM to answer a research question was qualitatively assessed based on how closely the simulated and historical data matched. The selected CSM was the one which was judged to best replicate historical data across all of the research questions.

- *Kidney*
  - The CSM labeled "CSM: A" was judged to best replicate historical kidney data.
  - This CSM slightly over-estimated the utilization of deceased donor kidneys both overall and by various subgroups; however, in most cases it correctly captured trends in kidney utilization across sub-groups.
  - The inclusion of utilization modeling in the simulation did not clearly cause any deficiencies in other simulated kidney metrics compared to a transplant-only simulation.
- *Pancreas*
  - All CSMs showed limited ability to replicate historical pancreas data for most research questions.
  - This was true regardless of whether utilization modeling was incorporated into the simulation.

  - Consequently, the Pancreas Committee should pursue alternative, non-simulation, methods for evaluating proposed allocation policies.

The selected CSM which will be used in subsequent requests is the one labeled "CSM: A".

2 Submodel (Scenario) Naming Key

Information about the CSMs shown in the "Simulated Results" section are given here.

- CSM: A
  - Includes a center-level covariate and allocation-related metrics in the acceptance models. An organ was considered not utilized for transplant after 1,000 offers.
- CSM: B
  - Includes a center-level covariate and allocation-related metrics in the acceptance models. An organ was considered not utilized for transplant after 100 offers.
- CSM: C
  - Includes a center-level covariate and allocation-related metrics in the acceptance models. Organ utilization was determined using a pre-placement model with characteristics of candidates at the top 25 sequence numbers on the match.
- CSM: D
  - Includes neither a center-level covariate nor allocation-related metrics in the acceptance models. Organ utilization was determined using a pre-placement model with characteristics of candidates at the top 25 sequence numbers on the match.
- CSM: E
  - Includes neither a center-level covariate nor allocation-related metrics in the acceptance models. An organ was considered not utilized for transplant after 100 offers.

- CSM: F
  - Includes neither a center-level covariate nor allocation-related metrics in the acceptance models. An organ was considered not utilized for transplant after 1,000 offers.
- CSM: G
  - Includes a center-level covariate and allocation-related metrics in the acceptance models. Organ utilization was determined using a pre-placement model with no information about candidates on the match run.
- CSM: H
  - Includes neither a center-level covariate nor allocation-related metrics in the acceptance models. Organ utilization was determined using a pre-placement model with no information about candidates on the match run.
- CSM: Tx-Only
  - Includes a center-level covariate and no allocation-related metrics in the acceptance models. Organ utilization not modeled.

3 Data Request

3.1 Background

The Organ Procurement and Transplantation Network (OPTN) Kidney and Pancreas Committees are currently working on adopting the continuous distribution framework for kidney, pancreas, kidney-pancreas, and pancreas islets allocation. At the direction of the OPTN Board of Directors, all organ-specific committees are required to consider utilization[1] in relation to the adoption of a continuous distribution framework.

The study design for the most recent Scientific Registry of Transplant Recipients (SRTR) kidney-pancreas simulation analysis, KIPA2023_01, did not consider utilization as a modeled outcome. The cohort of donors used as input was restricted to only those who had an organ transplanted historically; for this document, call this a "transplanted only" simulation design.

To include the utilization question as a modeled simulation outcome in the kidney-pancreas context, new simulation methods and models will need to be developed and assessed. This research has not been undertaken, and it is still unknown if utilization can be credibly modeled as a simulated outcome while maintaining high credibility across other important analysis outcomes.

3.2 OPTN Strategic Goal

Increase equity in access to transplants.

3.3 Request: Develop and assess models required for simulation analysis that includes utilization related outcomes, while maintaining high credibility across other important metrics, in preparation for future continuous distribution simulation studies

Historically, simulation studies have been a useful tool for policy-making, and this request will inform the Kidney and Pancreas Committees if simulation analysis will also be a useful tool for analyzing the utilization question in the context of continuous distribution.

The Kidney and Pancreas Committees are requesting analysis to determine if simulation can be credibly used to answer the questions in the "Research Questions" section

[1] In broad terms, utilization refers to the split between the number of organs that were transplanted and the number of organs that were not transplanted.

below. New research questions related to utilization are included for both kidney and pancreas; all other research questions are the same as those in the KIPA2023_01 data request.

Following analysis there are three categories of potential outcomes:

1. Utilization can be confidently added while maintaining credibility in other outcomes
    - Updated models will be available to simulate both utilization and other metrics for future continuous distribution simulation studies
2. Utilization cannot be confidently added while maintaining credibility in other outcomes
    - Updated models will be available to run a "transplanted only" simulation design for future continuous distribution simulation studies
3. The results are ambiguous, where the addition of utilization makes some other outcomes less reliable
    - The pros and cons of including utilization will need to be weighed for future continuous distribution simulation studies

3.4 Research Questions

The following research questions are those that will be of interest to the committees in a future data request that would include a range of potential continuous distribution scenarios. For the purpose of the current data request, KIPA2024_01, the potential policies do not need to be defined, only the form of the research questions. The models being built in KIPA2024_01 will be targeted to specifically address these questions; changes to the research questions may require rebuilding models.

3.4.1 Kidney

Non-Use

- **KI-NU 1:** How do the proposed policies impact non-use of donor kidneys overall and by the following?
    - **KI-NU 2:** kidney donor profile index (KDPI)

  - **KI-NU 3:** kidney donor risk index (KDRI) (rao)
  - The factors used to calculate KDPI:
    * **KI-NU 4:** Donor age
    * **KI-NU 5:** Height
    * **KI-NU 6:** Weight
    * **KI-NU 7:** History of hypertension
    * **KI-NU 8:** History of diabetes
    * **KI-NU 9:** Cause of death
    * **KI-NU 10:** Serum creatinine
    * **KI-NU 11:** Donation after circulatory death (DCD) status

- **KI-NU 12:** How do the proposed policies impact sequence number of the final acceptor?
- **KI-NU 13:** How do the proposed policies impact cold ischemic time? What additional cold ischemic time is attributable to transport logistics?
  - Cold ischemic time at acceptance
  - Cold ischemic time at transplant
- **KI-NU 14:** What is the impact of the policies on organ transport logistics?

Patient Access

- **KI-PA 1:** Do the proposed policies maintain the high level of access that pediatric candidates receive in the current system?
- **KI-PA 2:** Do the proposed policies maintain a high level of access for the extremely highly sensitized (calculated panel-reactive antibody [cPRA] 99.9+)?
  - Under currently policy, the highly sensitized have very high access to transplant; do the proposed policies result in reduced access for the highly sensitized (cPRA 98-99.9; using the buckets used in the previous addendum report) and overall lower disparities in access across cPRA groups?
- **KI-PA 3:** Do the proposed policies have those with the highest qualifying times undergoing transplant at a rate equal to or higher than current policy?
- **KI-PA 4:** Does patient access differ by OPTN region (transplant rates by OPTN region)?
- **KI-PA 5:** How does median qualifying time at transplant differ between proposed policies?

- **KI-PA 6:** Do the proposed policies impact the distribution of KDPI by estimated post-transplant survival (EPTS)? In other words, are low-EPTS patients appropriately prioritized for low-KDPI kidneys?

Placement Efficiency

- **KI-PE 1:** On average, how far are organs traveling?
- **KI-PE 2:** What is the distribution of travel distance?
- **KI-PE 3:** Are higher KDPI kidneys traveling shorter distances? In other words, is the increased donor modifier having the intended effect?
- **KI-PE 4:** When organs travel farther are they traveling farther to reach vulnerable populations (i.e., pediatrics, extremely highly sensitized)?

Candidate Biology

- **KI-CB 1:** Do the proposed policies maintain access for blood type O and type B candidates? Committee expressed that decreased access for type B and type O candidates would not be tolerable.
  - Do the proposed policies result in fewer disparities in access to transplant across blood types?
- **KI-CB 2:** How do the proposed policies impact the percent of recipients by DR mismatches (0, 1, or 2)?

Posttransplant Outcomes

- **KI-PT 1:** Do the proposed policies result in decreased graft failure and higher survival (short and long term)?
- **KI-PT 2:** Do the proposed policies balance longevity matching and qualifying time? In other words, are we able to have candidates with EPTS of 0-20% undergo transplant with low-KDPI kidneys without dropping their access while still having those with the longest qualifying times undergo transplant?

Other

- Do the proposed policies help diminish any disparities in access to transplant for subpopulations:
  - Sex
  - Race
  - Ethnicity
  - Age
  - Rural/urban
  - Geography
  - cPRA
  - Blood type
  - EPTS
  - Medical urgency
  - Time on dialysis groups
  - Safety net candidates
- The proposed policies aim to balance priority for patient access groups but may inadvertently result in decreased access for some subpopulations in an effort to prioritize others. Are there any unintended consequences on waitlist outcomes (additional time waiting, access to transplant, higher cumulative incidence of death, etc.) for any subpopulations:
  - Sex
  - Race
  - Ethnicity
  - Age
  - Rural/urban
  - Geography
  - cPRA
  - Blood type
  - EPTS
  - Medical urgency
  - Time on dialysis groups
  - Safety net candidates

3.4.2 Pancreas

Non-Use

Goal: Analyze the impact of proposed policies on pancreas utilization and identify ways to improve pancreas utilization.

- **KPPA-NU 1:** How do the proposed policies impact utilization of deceased donor pancreata, overall and by donor characteristics:
  - **KPPA-NU 1.1:** age
  - **KPPA-NU 1.2:** body mass index (BMI)
  - **KPPA-NU 1.3:** DCD status?
- **KPPA-NU 2:** How do the proposed policies impact non-use of deceased donor pancreata, overall and by donor characteristics (age, BMI, DCD status)?
- **KPPA-NU 3:** How do the proposed policies impact pancreas recovery rates?
- **KPPA-NU 4:** How do the proposed policies impact sequence number of the final acceptor?
- **KPPA-NU 5:** How do the proposed policies impact the timing of final acceptance relative to donor recovery (final acceptance pre- versus post-operation)?
- **KPPA-NU 6:** How do the proposed policies impact cold ischemic time:
  - At acceptance (overall, and separately for KP versus PA)?
  - At transplant (overall, and separately for KP versus PA)?
- **KPPA-NU 7:** How do the proposed policies impact allocation by center aggressiveness (e.g., the distribution of pancreata accepted by more aggressive versus less aggressive centers), overall and separately for KP versus PA?

Placement Efficiency

Goal: Maintain or reduce KP/PA travel distances relative to the current system (using travel distance as a proxy for anticipated impact on pancreas utilization).

- **KPPA-PE 1:** What is the distribution of organ travel distance (assess separately for KP and PA)?
- **KPPA-PE 2:** When KP/PA travel farther, are they doing so to reach highly sensitized candidates, pediatric candidates, and/or candidates with long qualifying times?

Candidate Biology

Goal: Equitable access to transplant across cPRA groups (to the extent possible).

- **KPPA-CB 1:** How does access to transplant for highly sensitized candidates (cPRA 80-97%; cPRA 98-100%) compare with access under the current system?
  - How does access to transplant compare across cPRA groups?
- **KPPA-CB 2:** How does access to transplant by candidate blood type compare with access under the current system (expect no change given no ABO attribute but would like to confirm)? Ideally look at this separately for KP and PA since they have different blood type screening rules (this stratification would be new).

Patient Access

Goal: (1) Increase access to transplant for pediatrics and prior living donors (note: we recognize that OASim cannot model prior living donors). (2) Maintain similar candidate waiting times relative to the current system.

- **KPPA-PA 1:** How does overall access to KP versus PA transplant compare with access under the current system? (e.g., would we expect KP transplants to increase and PA to decrease?)
- **KPPA-PA 2:** How does access to transplant for pediatric candidates compare with access under the current system?
- **KPPA-PA 3:** How does access to transplant by candidate qualifying time compare with access under the current system?
  - Do candidates with the highest qualifying times receive transplants at a rate similar to with current policy? Higher than with current policy?
  - Ideally look at this separately for KP and PA, and would like to look at both qualifying time and time on the waiting list for KP (since KP qualifying time includes time on dialysis prior to listing).
  - How does median qualifying time at transplant differ between proposed policies (separately for KP versus PA)?

Other

- Do the proposed policies result in any new/unintended disparities in access to transplant for any of the following subpopulations by:
  - Geography
  - Age

- Race
- Ethnicity
- Sex

4 Analysis Plan

For this data request, we will be building a range of utilization and acceptance simulation submodels, assessing them in isolation and together, and determining the best combination of the submodels in order to answer the research questions from the "Research Questions" section above. All other submodels will be the same as in KI2022_01[2].

4.1 Cohort

The cohort made available for the training and assessment of new simulation submodels will be all kidney and pancreas candidates who were active from March 15, 2020, through March 15, 2023, and all recovered organs from the same period[3]. This period was chosen to correspond with 1 year of the "KAS" allocation policy and 2 years of the current "KAS250"[4] allocation policy (since March 15, 2021), and allow for 3 full years of donated organs from which to sample.

4.2 Submodel Building and Selection Outline

We describe a range of potential utilization models below in the "Potential Utilization Models" section and a range of potential acceptance models in the "Potential Acceptance Models" section.

Each of these individual potential submodels (IPSMs) will be built and assessed as standalone statistical models using standard statistical evaluation techniques. If, after this step, any of the IPSMs exhibit clear deficiencies, they will be eliminated from further consideration.

The IPSMs that passed the last stage of analysis will now move onto the operational validation (OV) stage, where individual utilization and acceptance submodels are considered together as CSMs. Simulations will be run on a partition of the cohort for a subset

[2]Where appropriate, existing submodels will be updated with the new cohort data; however, no new techniques or algorithms will be explored for these submodels. See the "KI2022_01 Submodels" section in the Appendix for details on these submodels.

[3]CTAP: The cohort used for simulation analysis was all organs recovered *for transplant*. There are many pancreata recovered for research that are not included in the simulation cohort.

[4]The current allocation rules for kidney-pancreas allocation (OPTN Policy 11.5.A) give organ procurement organizations (OPOs) a choice between two pathways for kidney-pancreas donors. However, the simulator must follow a deterministic allocation order for all donors. At the March 15, 2022, meeting of the OPTN OPO Committee, OPO representatives indicated that current practice generally follows the pathway of offering both kidney and pancreas to the complete kidney-pancreas match run before offering the kidney to the kidney-alone match run. Accordingly, all simulated analysis for this request will also follow this pathway.

of the potential CSMs[5] and compared to historical results in order to determine which potential CSM is the best of the options.

1. For each figure described below in the "Operational Validation Outline" section, we will qualitatively determine the "best" CSM.
    - In cases where a single CSM is not best for all factor levels, then the research question should be examined. For example, if the grouping factor being considered were age group, and different CSMs were closest for different age ranges, the CSM that is closest for the pediatric group should be prioritized because this feature is specifically mentioned in the research questions.
2. After considering each figure in isolation, we will consider how the "winning" CSMs across the metrics/groupings stack up to help determine the overall selected CSM.
    - If no clear winner exists, we will fall back to the simplest of the set of CSMs.

Of course, this could play out in a huge number of ways, and we anticipate there will be either a clear winner or a set of CSMs that perform comparably.

Following the selection of a CSM, the overall reliability of the simulation submodels as a whole will be characterized; for each figure in the "Primary Assessment Metrics Based on Research Question" section below, the selected CSM will be classified based on how it compared to the historic cohort for that metric in relation to the other potential CSMs that were considered[6]. This characterization will be used during analysis of future simulation studies to provide context for the simulated results.

4.2.1 Donor Dataset Partitioning

As a part of the submodel building and selection process, the donor dataset will be randomly partitioned into three disjoint sets:

- 60% training,
- 20% testing, and
- 20% (operational) validation[7].

[5] For more details on the selection of the CSM subset, see the "Specific CSMs" section.

[6] For more information on SRTR simulation methodology, see the attached reference document: "Background and Methodology in Simulation Analysis: OASim".

[7] For KIPA2023_01, the dataset partitioning was only performed on the group of adult kidney donors because it was determined the other donor groups (pediatric kidney, kidney-pancreas, and pancreas) did not have sufficient sizes. For this study we plan to perform partitioning on all four groups, however this decision may be reassessed for the pediatric kidney, kidney-pancreas, and pancreas groups. If it is determined these group sizes are not large enough, the models will be trained on the entire cohort for each group.

Each donated organ also has a corresponding match run–the specific ordered candidates who were offered the organ historically. For the training and testing steps, the partitioning of the donors will also extend to each donor's corresponding match run; that is, all data on the candidates who saw an offer for the organ will be available for the training and testing processes[8]. The IPSMs described in the "Potential Utilization Models" and "Potential Acceptance Models" sections will each individually use the training donor dataset for the model building stage, and then the testing set for standard statistical evaluation of each IPSM in isolation.

4.2.2 OV Cohort Construction

The final set of partitioned donors will be used for the OV stage of analysis where each CSM will be evaluated in comparison to each other. The OV process will run simulations using each CSM and the simulated results compared to historical results. In this context the partitioning will not extend to each donor's corresponding match run; for the OV analysis, novel match runs will be created using randomization of the donor arrivals; see the "Donor Arrival Generation" section for details on this submodel.

The OV simulations for each CSM will be constructed to be over four non-overlapping subsets of the entire simulation cohort, described in the "Cohort" section above; the OV period start dates will be March 15, 2020; March 15, 2021; March 15, 2022; and January 15, 2023*[9].

Let:

- V: the 20% validation donor cohort
- N_V: number of donors in V
- N_p: number of test periods.

For each donor in V, randomly assign to one of the N_p test periods. With this, let:

- V_p: validation cohort for test period, p
- N_{Vp}: number of donors assigned to test period, p

For each start date, the OV period end date will be determined to be where the number of historical donors, N_{hp}, is equal to the number of test donors, N_{Vp}. For each test

[8]With this framing, an individual candidate may appear in both the training and testing datasets because an individual candidate may receive multiple offers over the course of their listing. Note, however, that any time-varying fields for the candidates will apply at the time of the arrival of the specific donor in question.

[9]CTAP: The start dates used in the simulation analysis were December 1, 2020; March 15, 2021; March 15, 2022; and January 15, 2023. With corresponding end dates of January 30, 2021; May 5, 2021; May 5, 2022; and March 5, 2023.

period cohort validation cohort, V_p, the donor arrivals will be randomized as described in the "Donor Arrival Generation" section, bounded by the pre-defined start dates and end dates determined by the size of N_{Vp}.

4.3 Utilization Mechanism

The utilization mechanism is a mechanism by which an organ can go unused in the simulation. In this request we will explore two broad styles of utilization mechanisms: pre-placement mechanisms and peri-placement mechanisms.

4.3.1 Pre-Placement (Utilization) Mechanism Background

The pre-placement mechanism is a new submodel in OASim that occurs after the match run is generated but before the placement mechanism. In this request, we will use the pre-placement mechanism to explore various utilization models. When an organ becomes available for transplant, the utilization model(s) first determines whether that organ will be utilized for transplant. If no, the organ is not offered to anyone and is immediately determined to not be utilized for transplant. If yes, the simulation proceeds to the placement mechanism whereby candidates are offered the organ until it is accepted. In this scenario, it is assumed that *someone* will accept the organ. If, by chance, every candidate on the match run declines the organ, then this would be considered a model artifact, not an unused organ.

4.3.2 Peri-Placement Utilization Mechanism Background

A peri-placement utilization mechanism is a utilization mechanism that occurs during the placement mechanism. If the organ is not accepted by the time some condition is met, that organ will not be utilized for transplant.

4.3.3 Potential Utilization Models

In this request, we will explore a variety of pre-placement and peri-placement utilization models. For pre-placement utilization models, we will use one or more logistic regression models. Donor factors contribute heavily to whether an organ is utilized, and therefore those factors will play a key role in the utilization mechanism. The donor cohort will remain constant across any simulated policies, therefore the only way utilization can vary from policy to policy is via the match-run ordering. Since the earliest portion of the match run is likely most relevant to the question of whether an organ is utilized, we will

use summary data about the candidates at the the top of the match run[10] to inform the model(s). Specifically we will explore models summarizing data about candidates at the top 10, 25, and 50 sequence numbers on the match run. We will additionally include a model that does not include any information about the match run as a reference. Each of these models will be built independently for the utilization of kidneys and pancreata.

For peri-placement utilization models, we will explore several conditions during the placement mechanism that will indicate the organ is not used.

1. The organ is not utilized after all candidates on the match run decline.
2. The organ is not utilized after a number of candidates has declined the organ equal to the 95th percentile of offer number at acceptance in the cohort used to train the placement acceptance models.
3. The organ is not utilized after a number of transplant programs declines the organ for at least one of its candidates equal to the 95th percentile of center number at acceptance in the cohort used to train the placement acceptance models.

4.4 Placement Mechanism

The placement mechanism is a submodel that determines who (if anyone) on a match run will accept a deceased donor organ for transplant. As in previous requests, we will use an "accept/decline" style placement mechanism; that is, an organ will be offered to candidates sequentially on the match run, and for each offered candidate a probability of acceptance is calculated to determine whether that candidate accepts or declines the offer. The probability of acceptance will be determined based on one or more logistic regression models.

For future simulation requests, the same placement mechanism will be used for all proposed policies. This introduces an important assumption: the accept/decline behavior of candidates is invariant across policies (i.e., the probability of acceptance under different allocation policies is reasonably approximated by the same logistic regression model[s]). This assumption is likely not true in practice. However, the degree to which this assumption is violated in our simulations will depend on the degree of allocation change under consideration and on what variables we allow to inform these models.

Since offer acceptance behavior depends, at least in part, on the allocation system in effect, the logistic regression model used to calculate the probability that a candidate will accept an organ will be trained using match-run data from the baseline allocation system. We represented each individual accept/decline decision made by a candidate

[10] Including information about difference in center offer acceptance behavior. See "Center Variability Background".

on a match run as a record in a logistic regression model. The decision to accept or decline likely depends on characteristics of both the donor and the candidate. The SRTR database provides a large number of possible donor and candidate characteristics that could be used to inform our model. However, we need to be careful to not overfit to the baseline scenario, given our assumption that behavior does not change across policies.

The offers that are used to train these logistic regression models will depend on the type of utilization mechanism that is being tested. For pre-placement utilization mechanisms, it is assumed that any organ that makes it to the placement mechanism will be accepted for transplant, and therefore only offers of transplanted organs will be used to train the offer acceptance models. For peri-placement utilization mechanisms, all organs (both those that will be utilized and those that will ultimately not be utilized) make it to the placement mechanism, and therefore offers of all organs (both transplanted and not transplanted) will be used to train the offer acceptance models.

4.4.1 Center Variability Background

Center variability attempts to characterize the notion that acceptance behavior varies widely across centers. Center variability can particularly play a role in whether an organ is utilized for transplant. If a center is more (less) likely to accept a given kind of organ, then proposed policies can make it more (less) likely that the organ is utilized if candidates at those centers appear early in the match run. Note, this does not mean the results of the simulation analysis are meant to be interpreted at the center level; these features are being accounted for in an attempt to make the aggregate metrics described below closer to the historical results.

Center variability is included in potential acceptance models via two possible covariates: offer acceptance ratio (OAR) and center-level covariate. The OAR is a model external to this simulation analysis based on the SRTR program-specific report (PSR) models. The OARs for each center were calculated based on the results of the external PSR model as applied to the simulation cohort; this factor in various forms is included for potential inclusion. The second is a model that simply includes a center-level covariate.

4.4.2 Allocation-related Metrics Background

Allocation-related metrics are those that describe a potential transplant recipient's location on the match run. In this request, we will consider three allocation-related metrics:

1. Offer number – A potential transplant recipient's offer number is one plus the number of non-bypassed potential transplant recipients with a lower sequence

number.

2. Center number – For the potential transplant recipient at offer number x, the center number is the number of unique transplant centers represented by all potential transplant recipients with offer number less than or equal to x.
3. Center rank – A potential transplant recipient's center rank is their relative priority among non-bypassed potential transplant recipients at the same transplant center. The potential transplant recipient with the lowest offer number at a transplant center has a center rank of one, the potential transplant recipient with the second-lowest offer number at a transplant center has a center rank of two, etc.

These metrics are particularly important for the peri-placement utilization mechanisms, as they provide a natural way in which reordering the match run can affect whether an organ is utilized. Offer number and center number are surrogates for time in the allocation process, and as time progresses offers are less likely to be accepted due to time constraints. This in combination with center variability in offer acceptance behavior can lead to an organ not being utilized. Center rank can be an indication of readiness for transplant (i.e., potential transplant recipients at the top of a centers list are more likely to be ready for transplant than those lower on the list), of center-level decline (i.e., once a center declines for a certain number of its potential transplant recipients it is more likely that it will decline for all of them), and of decision theory (i.e., determining whether to accept now or wait for a future offer), all of which can affect whether an organ is utilized for transplant.

The introduction of these allocation-related metrics into the placement mechanism does pose potential problems regarding the invariance assumption. For example, in the past we have excluded offer number as a predictor due to empirical evidence that the meaning of offer number can change from policy to policy. However, it is unclear whether the meaning of offer number changes after additionally accounting for the other allocation-related metrics of center number and rank. Furthermore, if the inclusion of these metrics in the placement model enables us to adequately model utilization, then we will need to consider the trade-off of potentially overfitting to a specific allocation policy versus being able to model utilization.

4.4.3 Potential Acceptance Models

Given the introduction of the utilization mechanism, in this request we will explore various acceptance models that account for center variability and allocation-related metrics, both of which can play a role in whether an organ is utilized for transplant. Specifically,

we will consider three ways of accounting for center variability in the potential acceptance models. Each of the types of models below will be assessed with and without accounting for allocation-related metrics:

- No center variability metrics
  - This is a reference case to show whether the inclusion of center variability/allocation-related metrics meaningfully improves outcomes
- All OARs
  - Each center's overall OAR is included as a covariate as well as an OAR covariate based on subsets of hard-to-place organs
- Center-level covariate

Each of the six potential models above will be built independently for these four subgroups:

- kidney and 18 years or older at listing,
- kidney and younger than 18 years at listing,
- kidney-pancreas, and
- pancreas.

Furthermore, each will be built three times, once to be paired with the peri-placement utilization models, once to be paired with the pre-placement utilization models, and once to be paired with simulations that do not model non-use.

4.5 Allocation Process Time Modeling

Several research questions pertain to how long it takes the allocation process to progress to a certain point: in particular, whether acceptance happens prior to cross-clamp (KPPA-NU 5), how much cold ischemic time is on the organ at acceptance and at transplant (see KI-NU 13, KPPA-NU 6), and questions surrounding transport logistics (i.e., time from acceptance to transplant [KI-NU 14]). To predict these points in time, we will build three linear regression models to be applied post hoc. The first will predict the time at which acceptance occurs relative to cross-clamp, and we will use as predictors the offer number and center number at acceptance[11]. The second will predict the cold ischemic time at

[11] See "Allocation-related Metrics Background"

transplant, and we will use as predictors the offer number and center number at acceptance as well as distance. The third will predict the time from acceptance to transplant, which will use logistical factors such as distance as predictors.

4.6 Operational Validation Outline

The following outline will be used to determine the CSM that does the best job of answering the research questions, as described in the "Submodel Building and Selection Outline" section.

4.6.1 Specific CSMs

Every potential utilization model from the "Potential Utilization Models" section can be paired with any potential acceptance model from the "Potential Acceptance Models" section[12]; there are seven potential utilization models and six potential acceptance models, giving 42 CSMs for simulations with non-use as a modeled outcome. Additionally for comparison, the six potential acceptance models will each be assessed in a scenario that does not model non-use as an outcome; in this case, the input donor dataset for simulation will be restricted to the subset that was transplanted historically, the "transplanted only" simulation study design that was used for KIPA2023_01.

This is a large number of combinations to assess. Following the model building stage, the number of potential utilization and acceptance models will be pared down based on the diagnostic results, outlined in the "Penalized Regression Model Building Report Outline" section below, in order to reach a manageable number of combinations to compare.

4.6.2 Use of Post Hoc Models

There are a number of post hoc models in this design–that is, models that are applied to simulated results after running the simulation rather than models that are a part of the simulation processing. The models described in the "Allocation Process Time Modeling" section and the "Posttransplant Models" section are post hoc models. These models are not considered across a range of options, and are not directly involved with the OV selection process. Instead, figures derived from these models will be used only to look for obvious model deficiencies; these figures are marked with a † symbol below.

[12]The utilization and acceptance model combinations will be paired based on the utilization type (pre-placement, peri-placement, or no non-use modeling) as discussed in the "Potential Acceptance Models" section.

4.6.3 Cohort Summary Tables

- Table 1: Candidate cohort description, nonsimulated. Kidney, columns: Adult, pediatrics, all. This results of this table will influence the specific groupings for the figures below.
  - Age groups (at start)
  - cPRA groups (at start)
  - Qualifying time groups (at start)
  - OPTN region
  - Blood type
- Table 2: Candidate cohort description, nonsimulated. Pancreas, columns: PA, PK.
  - Age groups (at start)
  - cPRA groups (at start)
  - Qualifying time groups (at start)
  - Blood type
- Table 3: Donor cohort description, nonsimulated, columns: KI, PA.
  - Utilized versus not-utilized
  - KDPI
  - KDRI
  - Age
  - Height
  - Weight
  - History of hypertension
  - History of diabetes
  - Cause of death
  - Serum creatinine
  - DCD status

A range of potential CSMs will be considered, and the best as measured by the following metrics will be selected.

4.6.4 Primary Assessment Metrics Based on Research Question

Figures below are as usual from the reports. The figures described in this section will be used as the primary metrics for selecting a CSM.

The corresponding research question is indicated in parenthesis for each figure below.

Overall Figures by Organ:

- Percent of organs not used (KI-NU 1, KPPA-NU 1)
- Unadjusted transplant rate (KPPA-PA 1)
- Median sequence number at acceptance (KI-NU 12, KPPA-NU 4)
- Median cold ischemic time at acceptance† (KI-NU 13, KPPA-NU 6)
- Median cold ischemic time at transplant† (KI-NU 13, KPPA-NU 6)
- Median time from acceptance to transplant† (KI-NU 14)
- Median qualifying time at transplant (KI-PA 5)
- Median travel distance (KI-PE 1, KPPA-PE 1)
- Distribution of travel distance (KI-PE 2, KPPA-PE 1)
- Percent accepted prior to cross-clamp (KPPA-NU 5)
- Percent of recipients at centers with OARs in top 25% (KPPA-NU 7)
- Percent of recipients whose organ traveled more than 250 nautical miles with cPRA >99.9% (KI-PE 4, KPPA-PE 2)
- Percent of pediatric recipients whose organ traveled more than 250 nautical miles (KI-PE 4, KPPA-PE 2)

Kidney Alone:

Percent of Kidneys Not Used by Donor Characteristics:

- KDPI (KI-NU 2)
- KDRI (KI-NU 3)
- Age (KI-NU 4)
- Height (KI-NU 5)
- Weight (KI-NU 6)
- History of hypertension (KI-NU 7)
- History of diabetes (KI-NU 8)
- Cause of death (KI-NU 9)
- Serum creatinine (KI-NU 10)
- DCD status (KI-NU 11)

Adjusted[13] Transplant Rate by Candidate Characteristics:

[13]Adjusted rates as described in the attached reference document: "Time-to-Event Rate Calculations in Simulation Analysis: OASim". All covariates listed here will be included in the adjusted rate model.

- Age (KI-PA 1)
- cPRA (KI-PA 2)
- Qualifying time (KI-PA 3, KI-PT 2[14])
- OPTN region (KI-PA 4)
- Blood type (KI-CB 1)

Other Simulated Metrics

- Median travel distance by donor KDPI (KI-PE 3)
- Percent of recipients by 0, 1, or 2 DR mismatches (KI-CB 2)
- Median KDPI by EPTS (KI-PA 6, KI-PT 2)

Posttransplant

- 1-year graft failure percent by recipient age† (KI-PT 1)
- 10-year graft failure percent by recipient age† (KI-PT 1)

Pancreas Alone:
(Assess the group sizes to determine if KP versus PA split is possible)

Percent of Pancreas Not Used by Donor Characteristics:[15]

- Age (KPPA-NU 1.1)
- BMI (KPPA-NU 1.2)
- DCD status (KPPA-NU 1.3)

Adjusted Transplant Rate by Candidate Characteristics:

- cPRA (KPPA-CB 1)
- Blood type (KPPA-CB 2)
- Age (KPPA-PA 2)
- Qualifying time (KPPA-PA 3)

[14]Research question KI-PT 2 is partially addressed by the combination of the "adjusted transplant rate by qualifying time" and "median KDPI by EPTS" figures.

[15]Within the simulation there is no distinction between nonuse (proportion of organs recovered for transplant that were not transplanted) and nonutilization (proportion of organs from all donors that were not transplanted) because we are currently not modeling recovery. Therefore we are simply calculating the percentage of organs (kidney or pancreas) that were fed into the simulation (see "Cohort") that were not transplanted. Consequently, we see no distinction between research questions KPPA-NU 1.1-1.3 and KPPA-NU 2, and research question KPPA-NU 3 is out of scope.

4.6.5 Primary Assessment Metrics Based on Modeling/Assumptions

The Committee's research questions all pertain to the potential impact of proposed continuous distribution policy scenarios. For those simulations, the same CSMs will be used across all proposed policies. This introduces what we refer to as the invariance assumption, which is to say we are assuming that each individual submodel is able to perform well regardless of the allocation policy in effect. This assumption applies to all submodels but is most relevant to the placement and utilization mechanisms, where we are assuming that behavior is reasonably approximated by the same logistic regression models regardless of policy. This may not be true in practice. This section will evaluate the validity of this assumption by comparing several metrics across the policy change from KAS to KAS250. In particular, for each era we will compare:

- Non-use overall and stratified by:
  - KDPI
  - KDRI
  - Donor age
  - Height
  - Weight
  - History of hypertension
  - History of diabetes
  - Cause of death
  - Serum creatinine
  - DCD status
- Sequence number at acceptance

4.6.6 Secondary Assessment Metrics Based on Research Questions

These metrics will not be directly used to determine the best CSM. They will be used as a check for any noticeable deficiencies in a way similar to post hoc models, see "Use of Post Hoc Models".

Kidney
Unadjusted transplant rate, as well as cumulative incidence of death, by:

- Sex
- Race

- Ethnicity
- Age (above)
- Rural/urban
- Geography (above)
- cPRA (above)
- Blood type (above)
- EPTS
- Medical urgency
- Time on dialysis groups
- Safety net candidates

Offer number and center number at acceptance, by:

- KDPI

Pancreas

Unadjusted transplant rate by:

- Geography
- Age (above)
- Race
- Ethnicity
- Sex

Table 1: Characteristics for the entire simulation candidate cohort

Characteristic	Adult KI N = 346080[1]	Pediatric KI N = 5740[1]	Kidney-Pancreas N = 7833[1]	Pancreas N = 2817[1]
Age at Listing				
0-<18	0 (0%)	5,740 (100%)	14 (0.2%)	229 (8.1%)
18-<35	35,073 (10%)	0 (0%)	1,955 (25%)	718 (25%)
35-<50	93,346 (27%)	0 (0%)	4,015 (51%)	1,332 (47%)
50-<65	152,555 (44%)	0 (0%)	1,828 (23%)	516 (18%)
65+	65,106 (19%)	0 (0%)	21 (0.3%)	22 (0.8%)
Sex				
Female	132,106 (38%)	2,168 (38%)	3,538 (45%)	1,428 (51%)
Male	213,974 (62%)	3,572 (62%)	4,295 (55%)	1,389 (49%)
Blood Type				
A	93,626 (27%)	1,890 (33%)	2,353 (30%)	1,089 (39%)
AB	8,749 (2.5%)	180 (3.1%)	229 (2.9%)	114 (4.0%)
B	56,926 (16%)	839 (15%)	1,422 (18%)	340 (12%)
O	186,779 (54%)	2,831 (49%)	3,829 (49%)	1,274 (45%)
cPRA[2]				
0-60%	281,682 (81%)	4,871 (85%)	6,387 (82%)	2,191 (78%)
>60-80%	18,479 (5.3%)	236 (4.1%)	413 (5.3%)	140 (5.0%)
>80-98%	21,988 (6.4%)	250 (4.4%)	531 (6.8%)	230 (8.2%)
>98-99.5%	6,160 (1.8%)	91 (1.6%)	180 (2.3%)	82 (2.9%)
>99.5-99.9%	6,079 (1.8%)	167 (2.9%)	120 (1.5%)	89 (3.2%)
>99.9-100%	11,692 (3.4%)	125 (2.2%)	202 (2.6%)	85 (3.0%)

[1] Values are given as number (percentage).
[2] Determined at the later of listing date or simulation start.

5 Simulation Cohort

Tables 1 and 2 show characteristics of the candidate cohort, and Table 3 shows characteristics of the donor cohort.

Table 2: Waiting list characteristics for the entire simulation candidate cohort

Characteristic	Adult KI N = 346080[1]	Pediatric KI N = 5740[1]	Kidney-Pancreas N = 7833[1]	Pancreas N = 2817[1]
OPTN Region				
1	20,443 (5.9%)	154 (2.7%)	269 (3.4%)	130 (4.6%)
2	40,969 (12%)	721 (13%)	1,094 (14%)	400 (14%)
3	41,781 (12%)	439 (7.6%)	786 (10%)	166 (5.9%)
4	34,670 (10%)	426 (7.4%)	430 (5.5%)	120 (4.3%)
5	77,286 (22%)	1,412 (25%)	1,118 (14%)	291 (10%)
6	8,318 (2.4%)	236 (4.1%)	273 (3.5%)	31 (1.1%)
7	28,035 (8.1%)	630 (11%)	1,311 (17%)	691 (25%)
8	11,238 (3.2%)	249 (4.3%)	275 (3.5%)	93 (3.3%)
9	27,899 (8.1%)	582 (10%)	660 (8.4%)	422 (15%)
10	17,094 (4.9%)	265 (4.6%)	503 (6.4%)	234 (8.3%)
11	38,347 (11%)	626 (11%)	1,114 (14%)	239 (8.5%)
Qualifying Time (Years)[2]				
0-1	61,582 (18%)	1,813 (32%)	2,246 (29%)	648 (23%)
>1-2	59,758 (17%)	1,146 (20%)	2,051 (26%)	468 (17%)
>2-5	140,498 (41%)	1,817 (32%)	2,648 (34%)	958 (34%)
>5-10	71,320 (21%)	754 (13%)	791 (10%)	576 (20%)
>10	12,922 (3.7%)	210 (3.7%)	97 (1.2%)	167 (5.9%)

[1] Values are given as number (percentage).

[2] Determined at the later of listing date or simulation start.

Table 3: Characteristics for the entire simulation donor organ cohort. KI includes counts for both left and right kidneys.

Characteristic	KI N = 15120[1]	PA N = 733[1]
Age		
0-<18	795 (5.3%)	139 (19%)
18-<35	4,103 (27%)	485 (66%)
35-<50	4,689 (31%)	100 (14%)
50-<65	4,647 (31%)	9 (1.2%)
65+	886 (5.9%)	0 (0%)
Sex		
Female	5,737 (38%)	208 (28%)
Male	9,383 (62%)	525 (72%)
BMI		
<20	1,047 (6.9%)	101 (14%)
20-<25	3,905 (26%)	332 (45%)
25-<30	4,438 (29%)	239 (33%)
30-<35	2,998 (20%)	46 (6.3%)
35-<40	1,484 (9.8%)	13 (1.8%)
40+	1,248 (8.3%)	2 (0.3%)
Blood Type		
A	5,529 (37%)	276 (38%)
AB	557 (3.7%)	9 (1.2%)
B	1,663 (11%)	79 (11%)
O	7,371 (49%)	369 (50%)
KDPI		
0-20%	2,957 (20%)	494 (67%)
>20-35%	2,196 (15%)	144 (20%)
>35-85%	7,615 (50%)	95 (13%)
>85-100%	2,352 (16%)	0 (0%)

[1] Values are given as number (percentage).

6 Simulated Results

6.1 Overall Figures by Organ

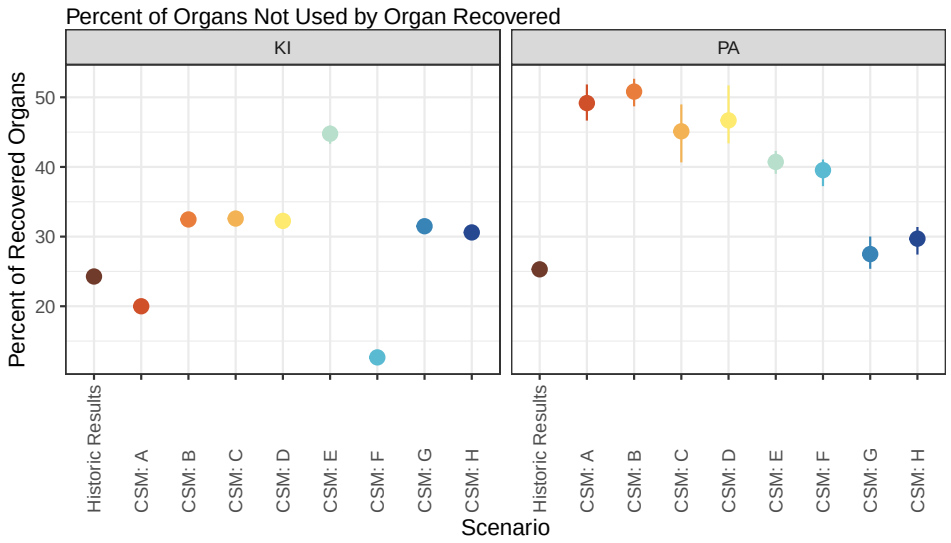

Figure 1: Percent of Organs Not Used by Organ Recovered. Includes all organs that were recovered for transplant.

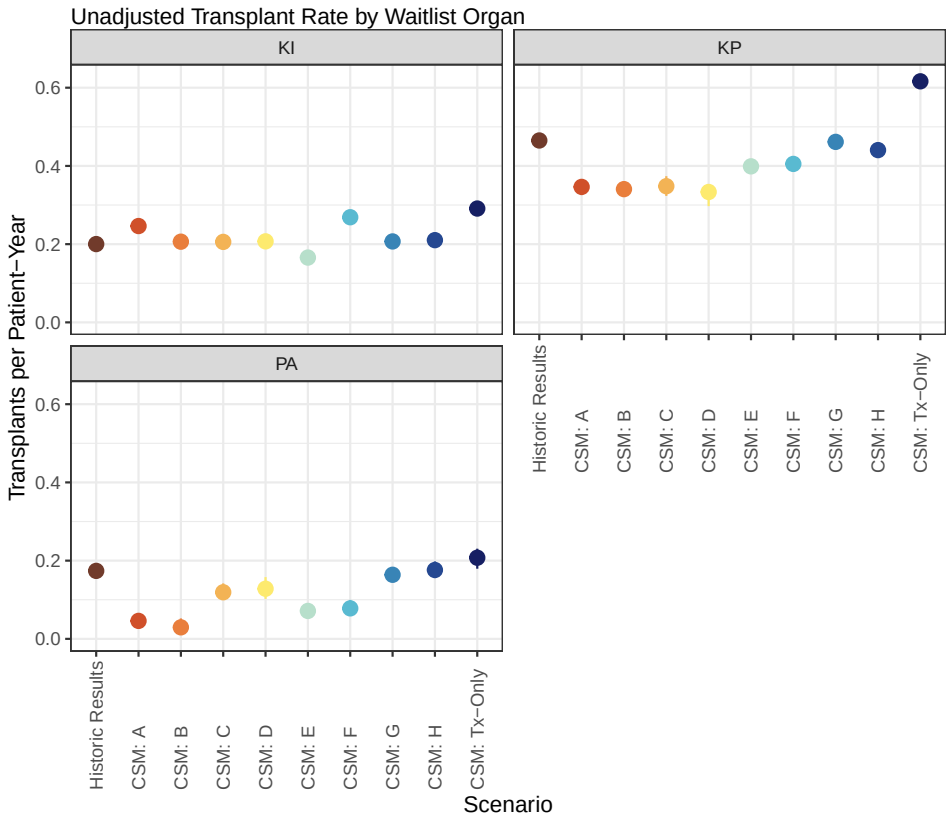

Figure 2: Unadjusted Transplant Rate by Waitlist Organ. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

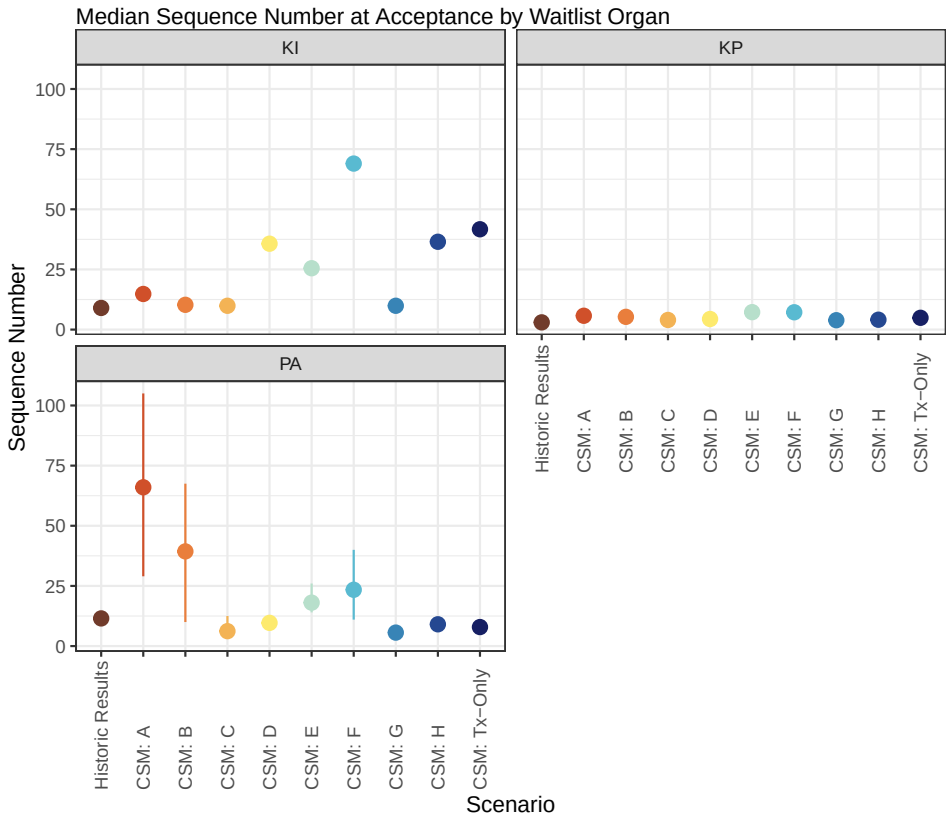

Figure 3: Median Sequence Number at Acceptance by Waitlist Organ.

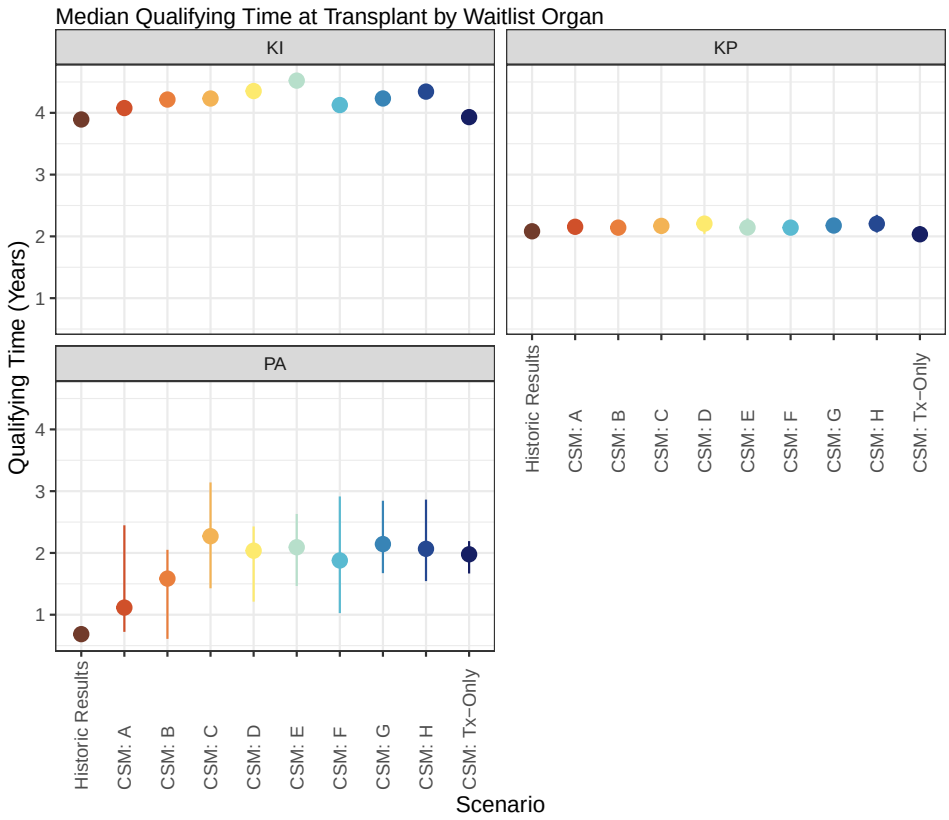

Figure 4: Median Qualifying Time at Transplant by Waitlist Organ. Qualifying time is time in years from qualifying date to transplant date.

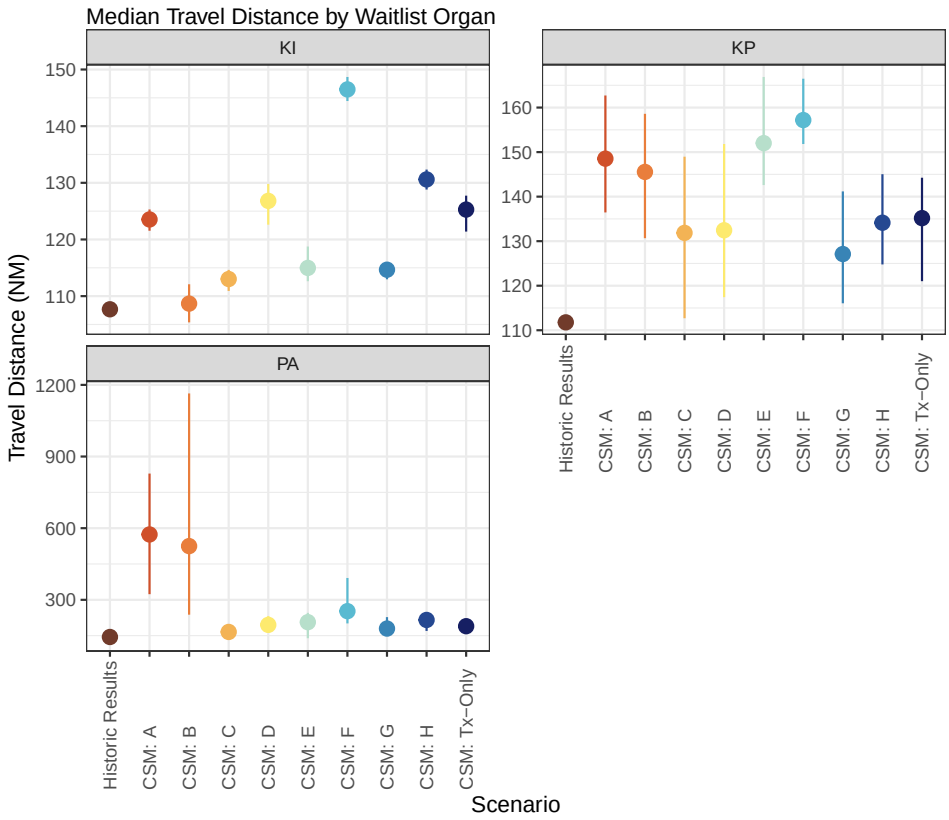

Figure 5: Median Travel Distance by Waitlist Organ. Travel distance is between the donor hospital and the transplant center, in nautical miles.

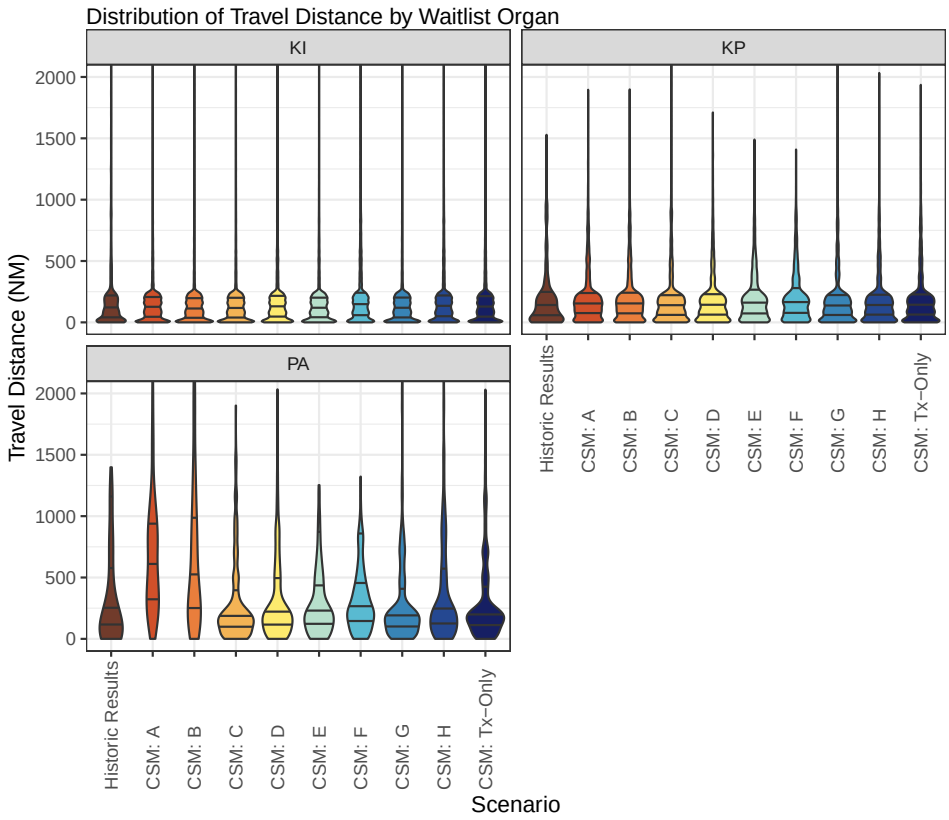

Figure 6: Distribution of Travel Distance by Waitlist Organ. Travel distance is between the donor hospital and the transplant center, in nautical miles.

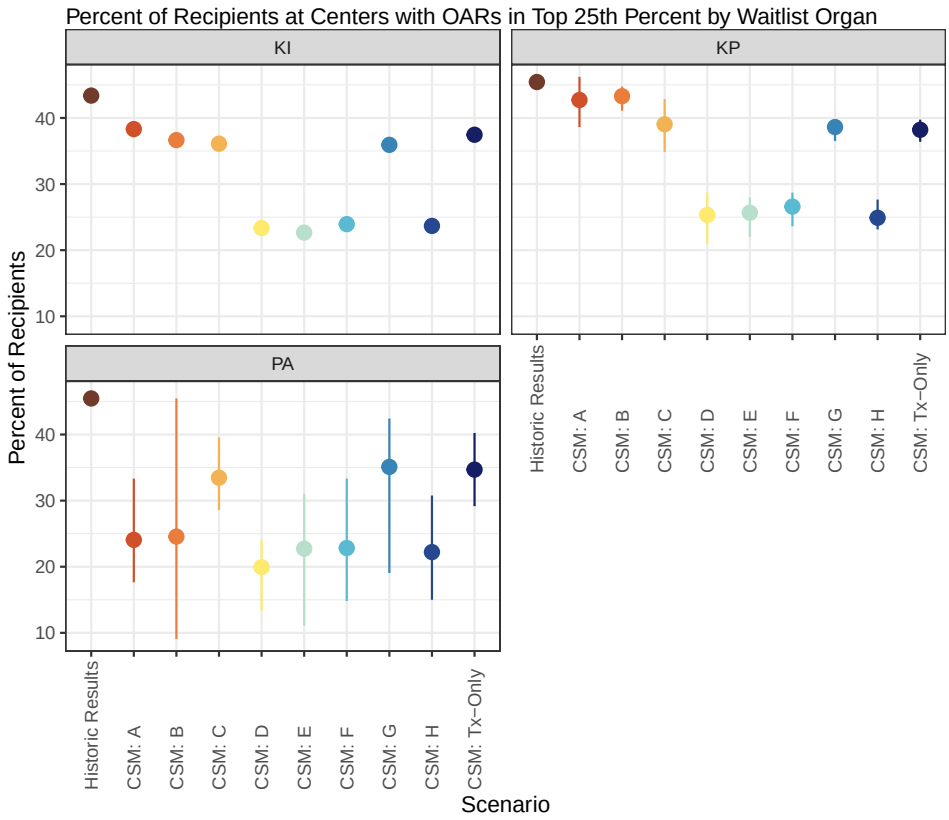

Figure 7: Percent of Recipients at Centers with OARs in Top 25th Percent by Waitlist Organ. OAR: offer acceptance ratio.

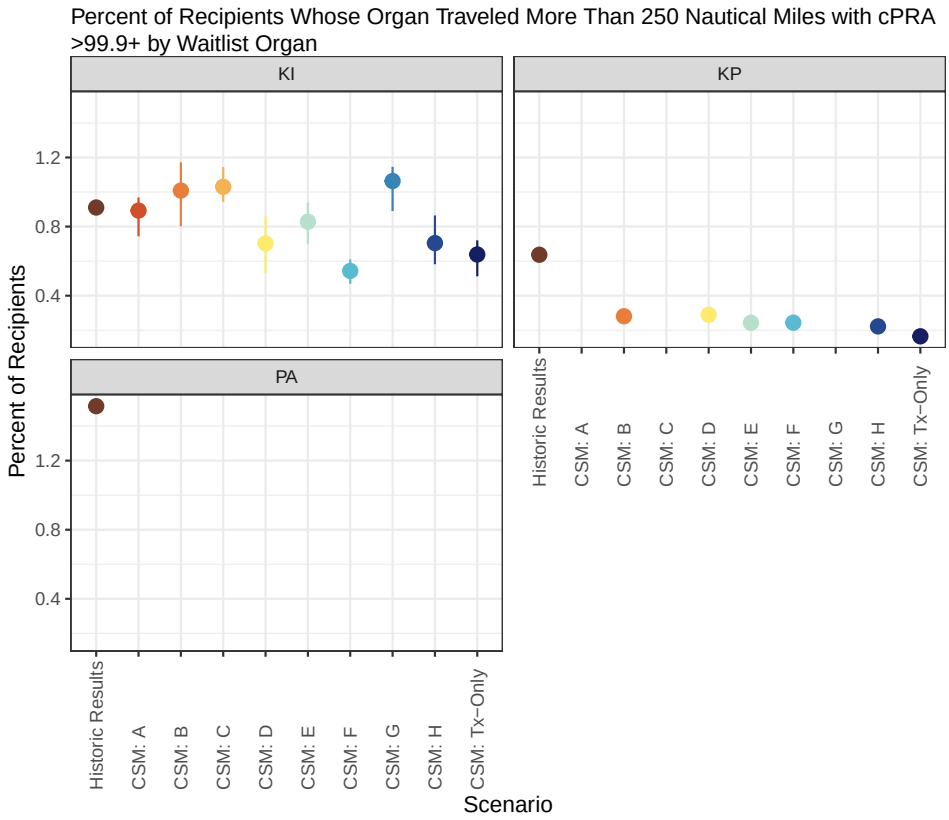

Figure 8: Percent of Recipients Whose Organ Traveled More Than 250 Nautical Miles with cPRA >99.9+ by Waitlist Organ. Travel distance is between the donor hospital and the transplant center, in nautical miles.

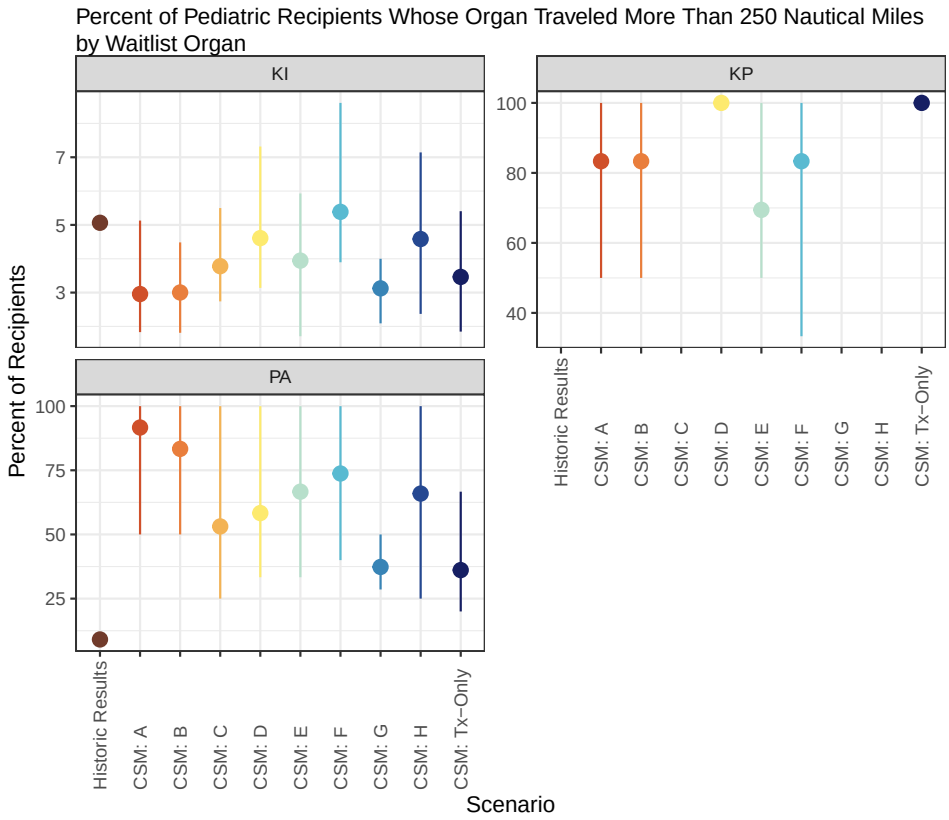

Figure 9: Percent of Pediatric Recipients Whose Organ Traveled More Than 250 Nautical Miles by Waitlist Organ. Travel distance is between the donor hospital and the transplant center, in nautical miles.

6.2 Kidney Alone

6.2.1 Percent of Kidneys Not Used by Donor Characteristics

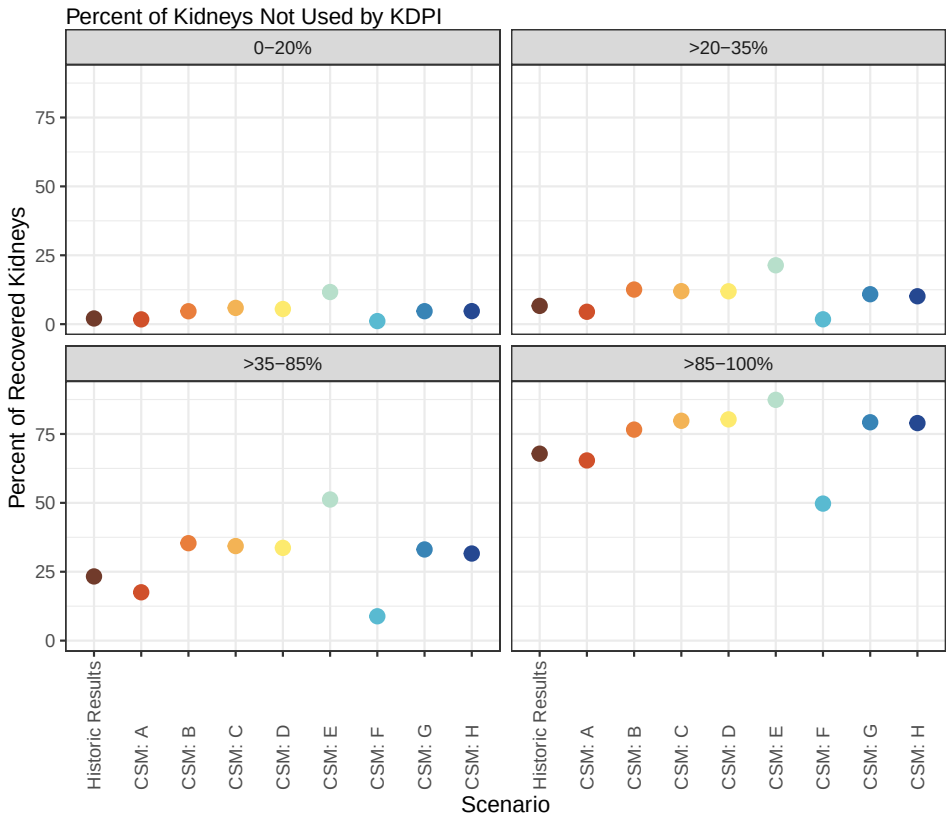

Figure 10: Percent of Kidneys Not Used by KDPI. Includes all kidneys that were recovered for transplant.

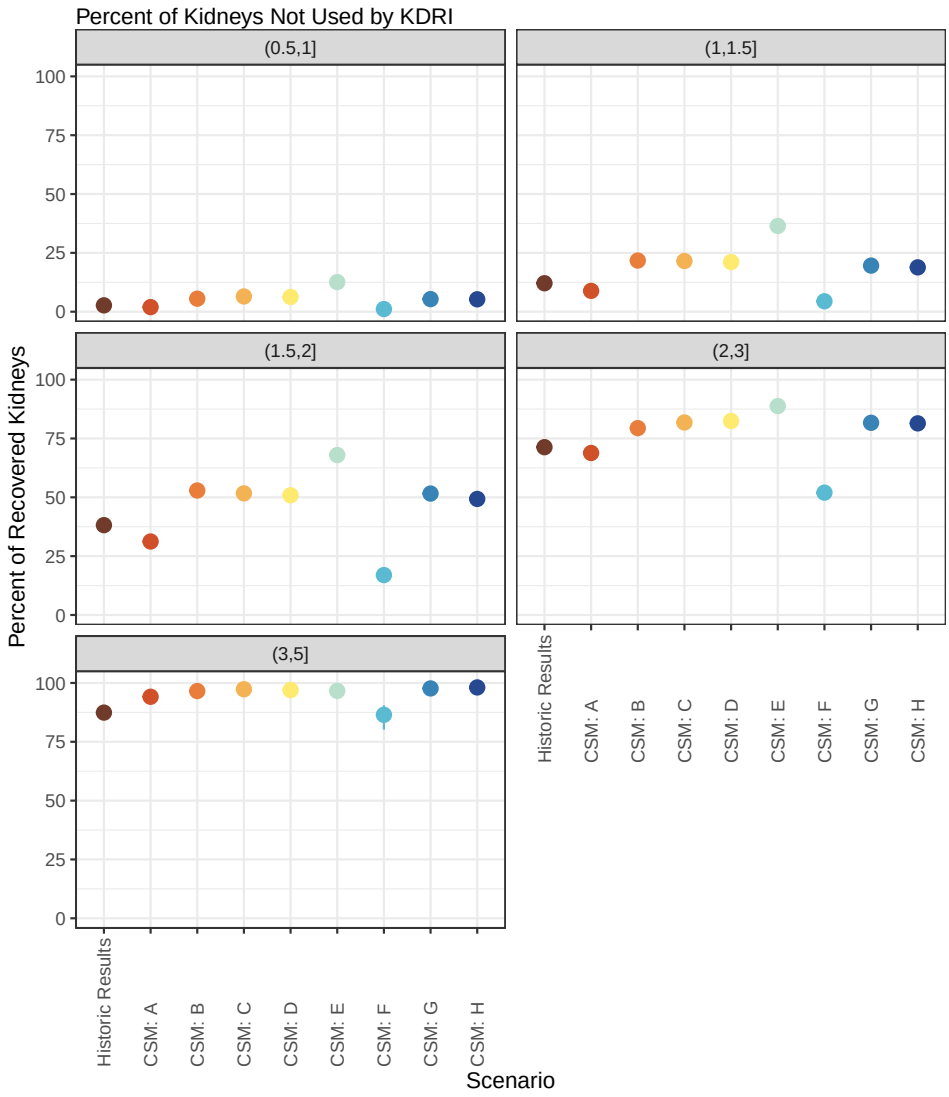

Figure 11: Percent of Kidneys Not Used by KDRI. Includes all kidneys that were recovered for transplant.

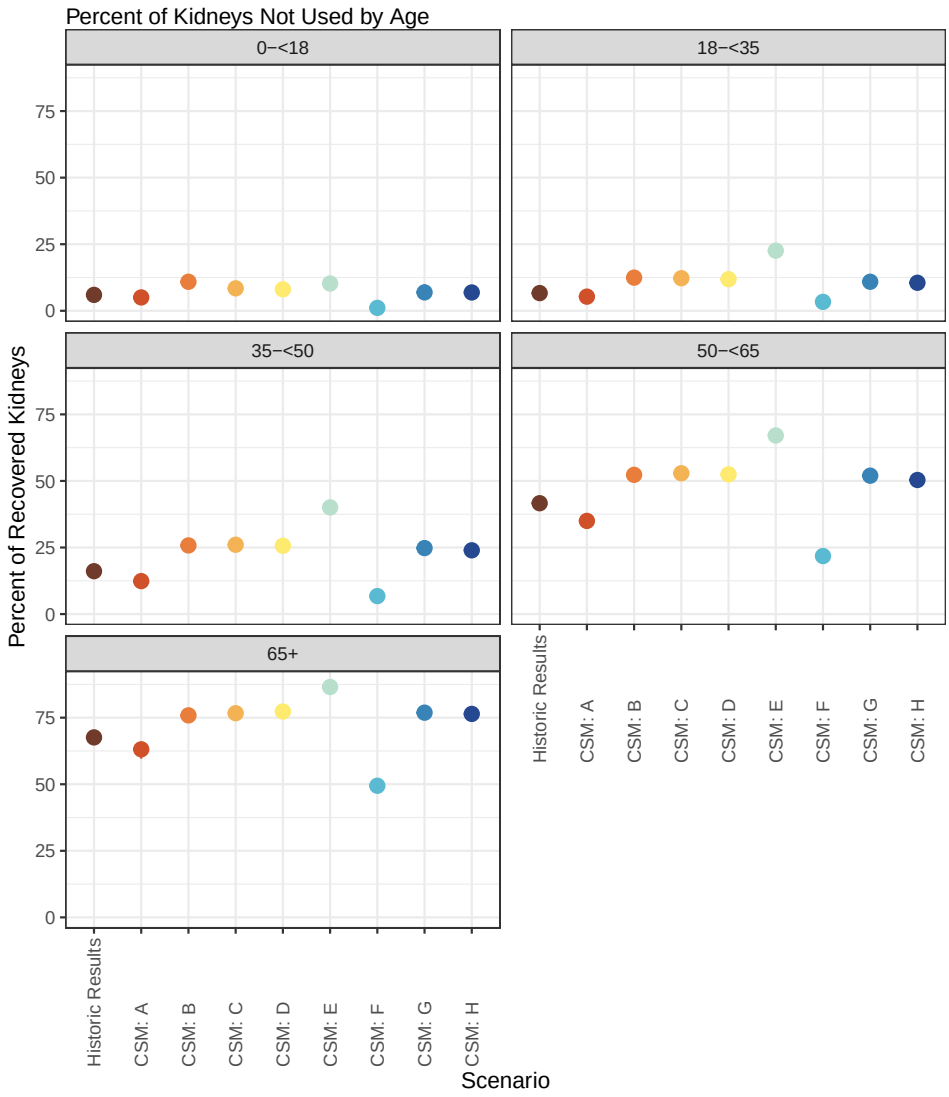

Figure 12: Percent of Kidneys Not Used by Age. Includes all kidneys that were recovered for transplant.

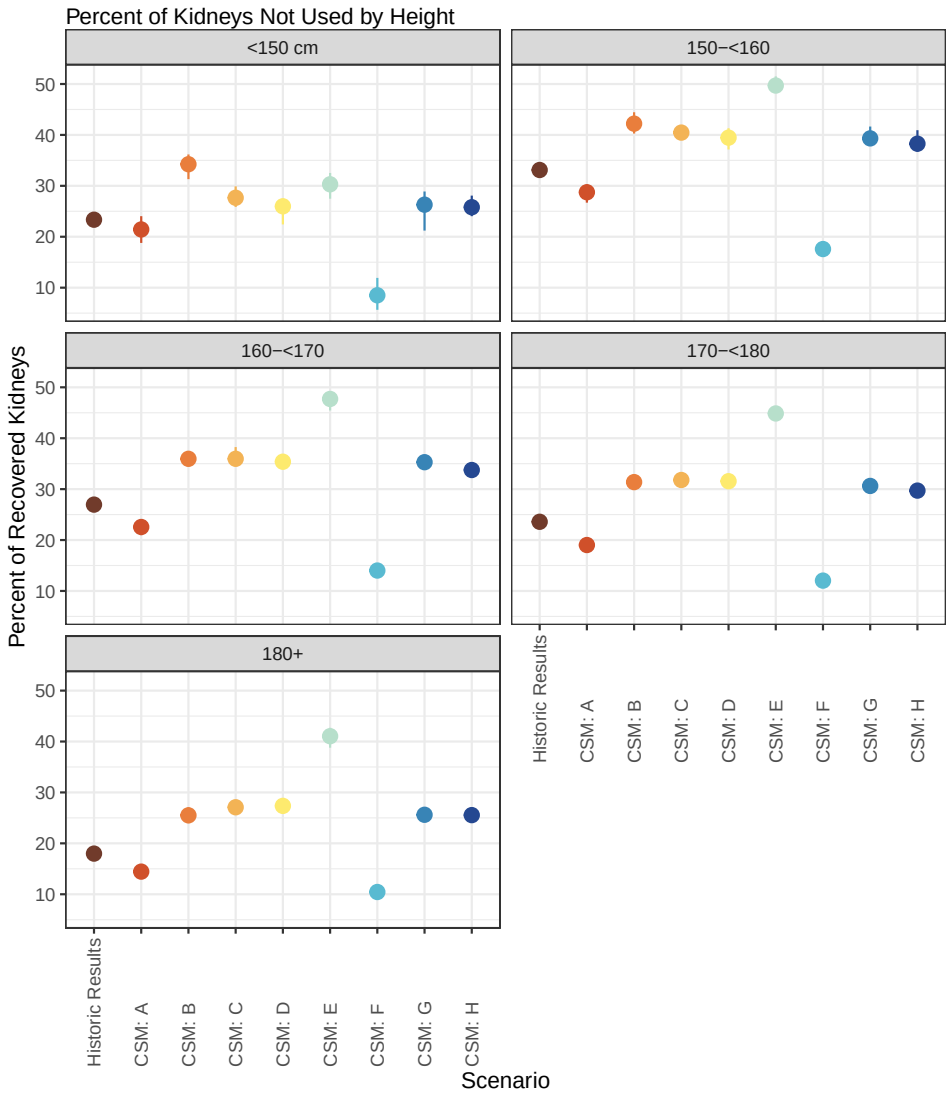

Figure 13: Percent of Kidneys Not Used by Height. Includes all kidneys that were recovered for transplant.

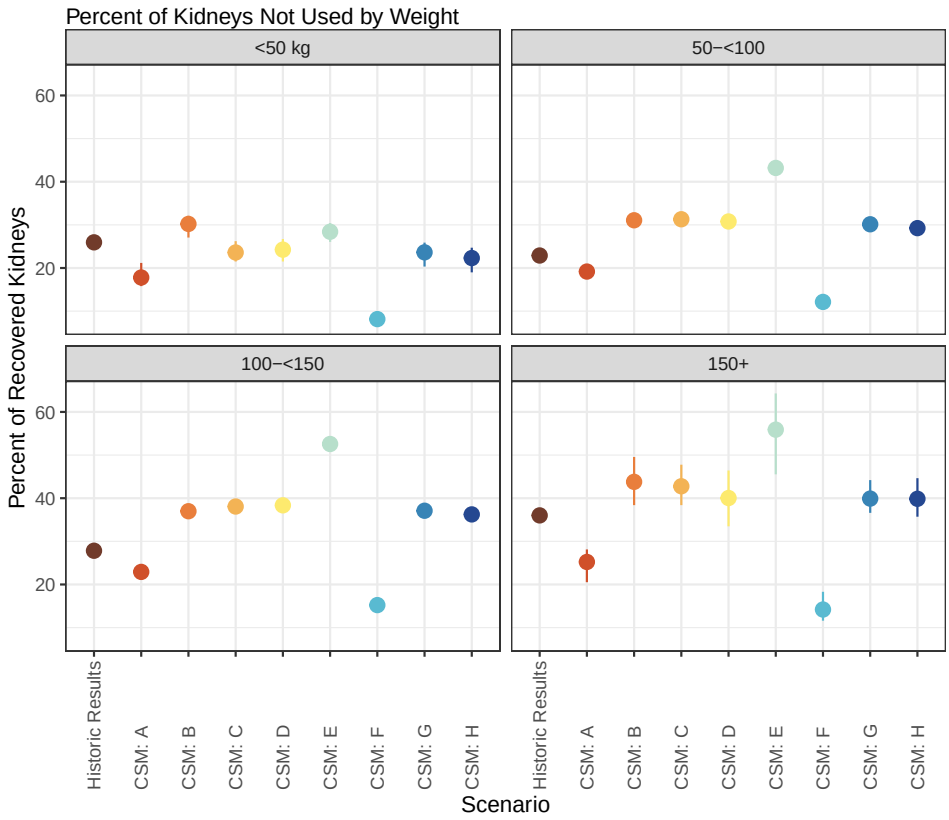

Figure 14: Percent of Kidneys Not Used by Weight. Includes all kidneys that were recovered for transplant.

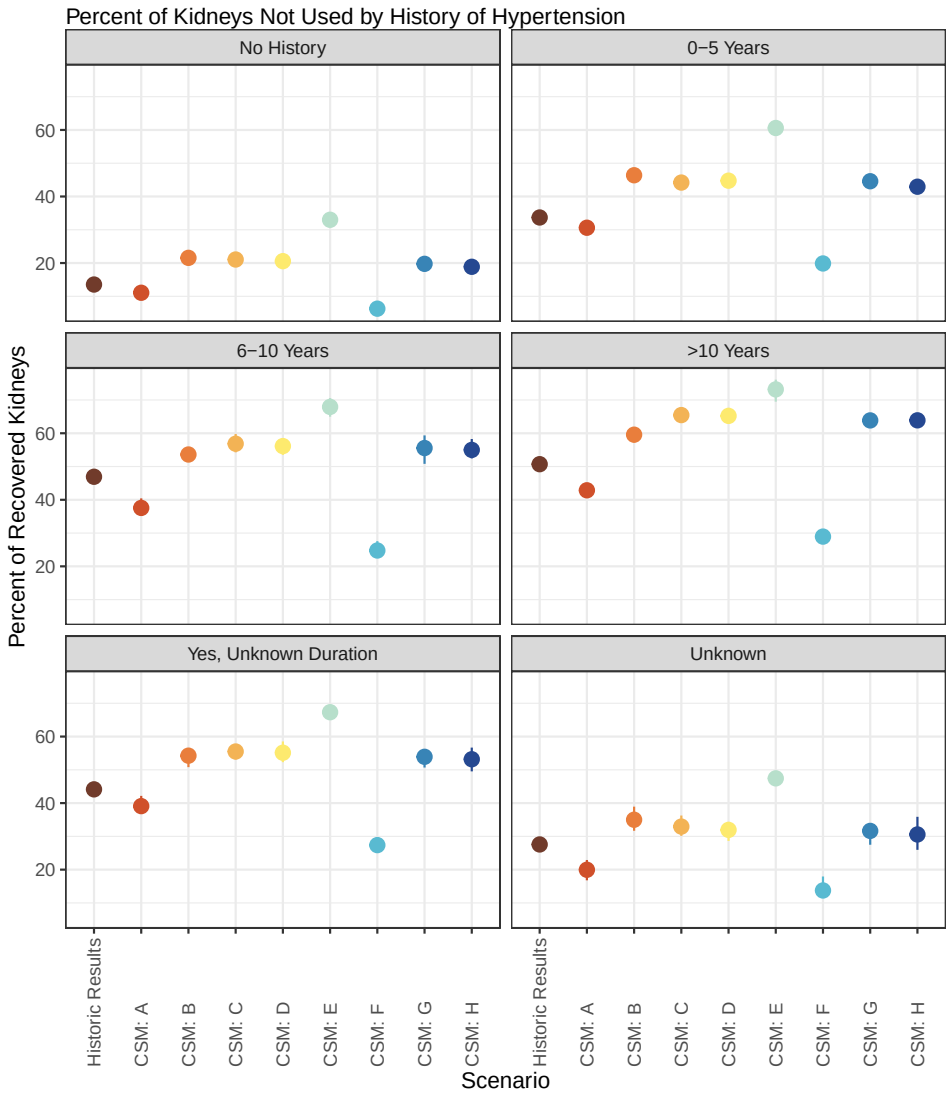

Figure 15: Percent of Kidneys Not Used by History of Hypertension. Includes all kidneys that were recovered for transplant.

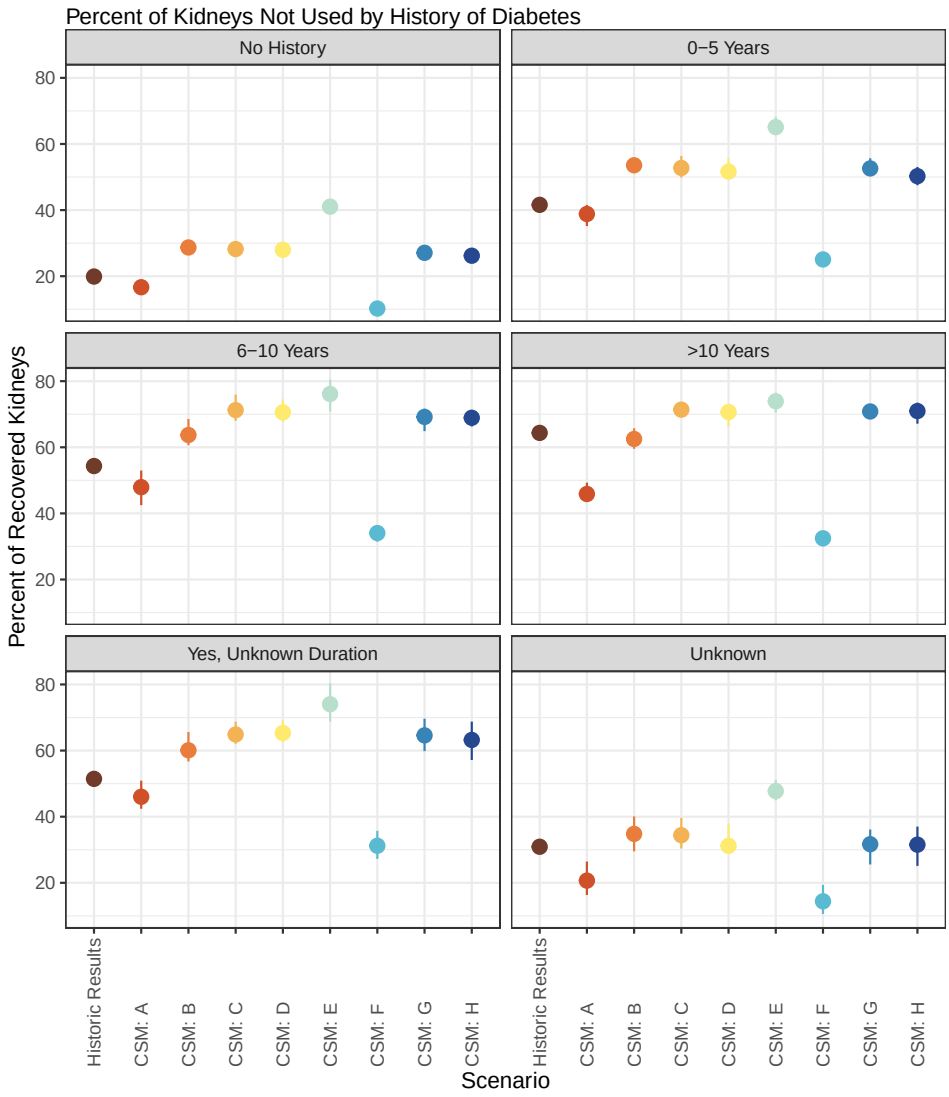

Figure 16: Percent of Kidneys Not Used by History of Diabetes. Includes all kidneys that were recovered for transplant.

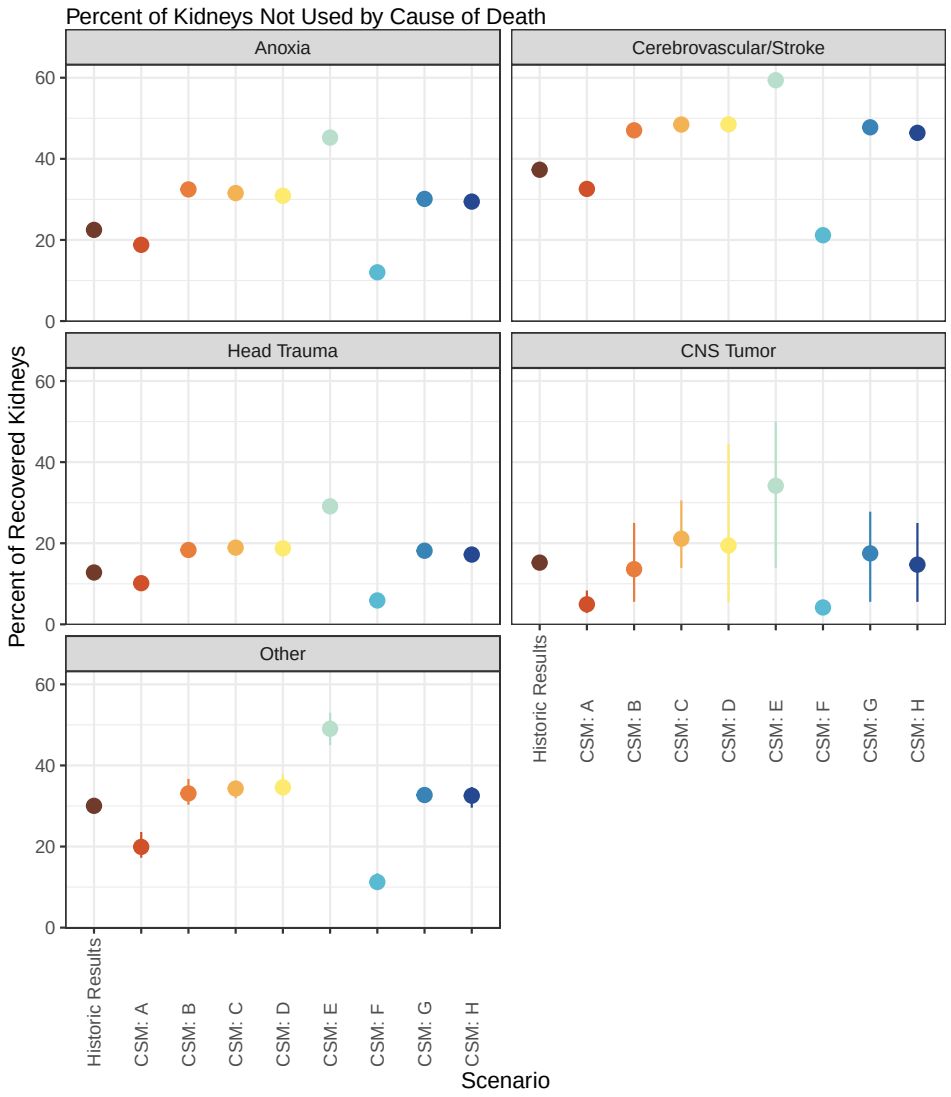

Figure 17: Percent of Kidneys Not Used by Cause of Death. Includes all kidneys that were recovered for transplant.

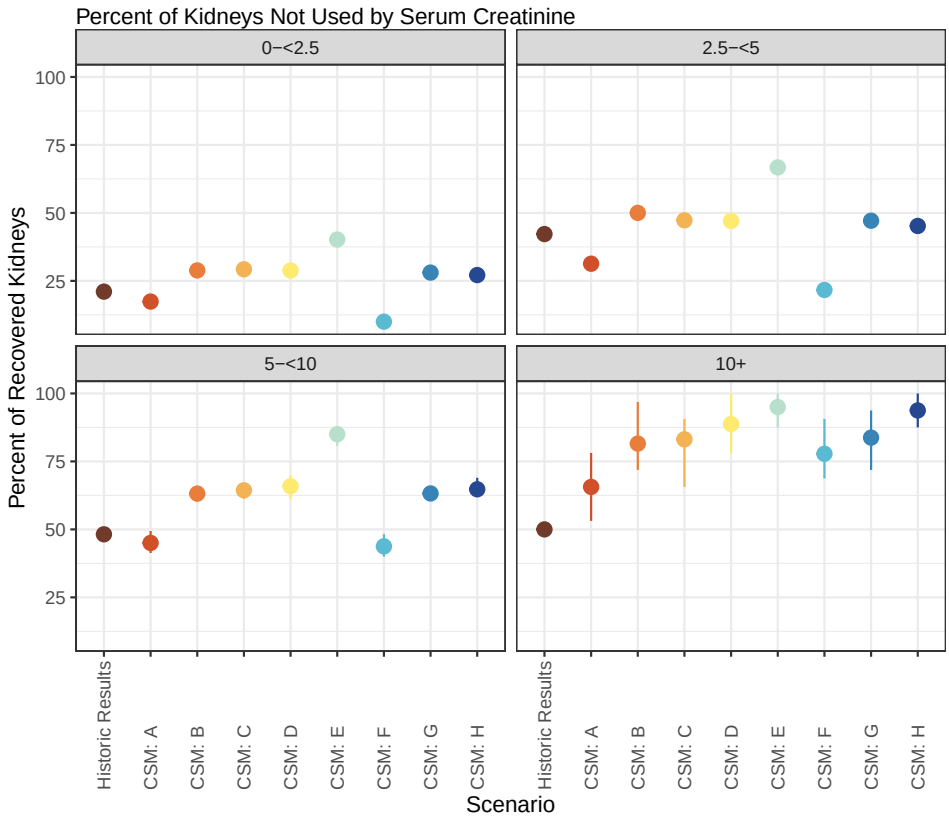

Figure 18: Percent of Kidneys Not Used by Serum Creatinine. Includes all kidneys that were recovered for transplant.

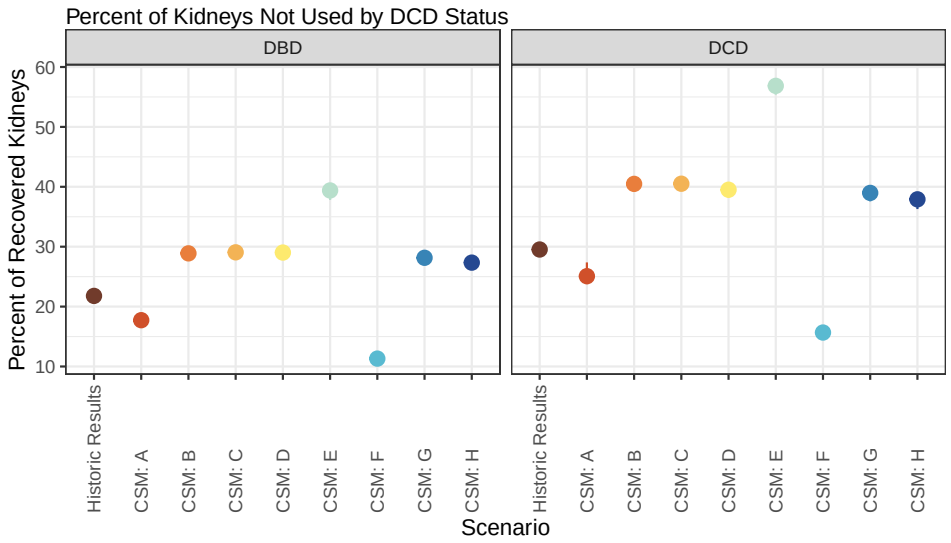

Figure 19: Percent of Kidneys Not Used by DCD Status. Includes all kidneys that were recovered for transplant.

6.2.2 Kidney Adjusted Transplant Rate by Candidate Characteristics

6.2.2.1 Overall Model Diagnostics For kidney an adjusted rate model was fit at each scenario/iteration and the overall model diagnostics for each scenario are shown in Figure 20. The adjusting factors are age at listing, cPRA, qualifying time, OPTN region, and blood type. These figures are meant to be a check of the overall rate model quality, before using those models to calculate by group rates using standarization. These figures are not meant to be used for scenario selection directly but rather to look for potential deficiencies in the simulated outcomes of the scenarios.

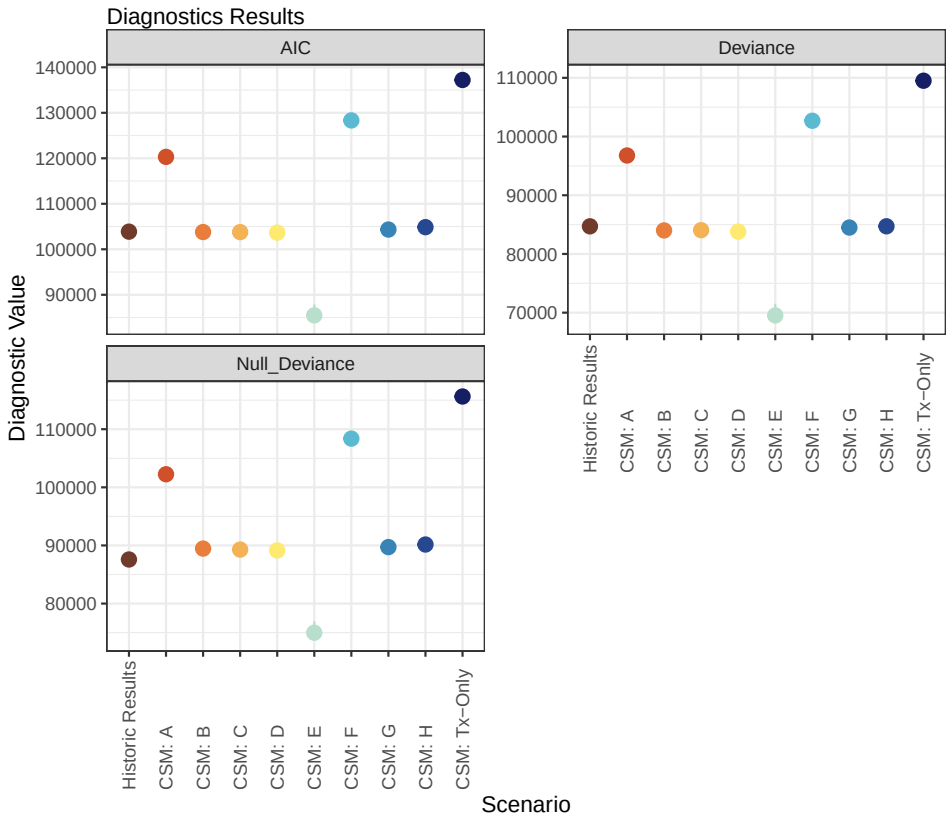

Figure 20: Diagnostics Results. Kidney adjusted transplant rate model with adjusting factors: age at listing, cPRA, qualifying time, OPTN region, and blood type.

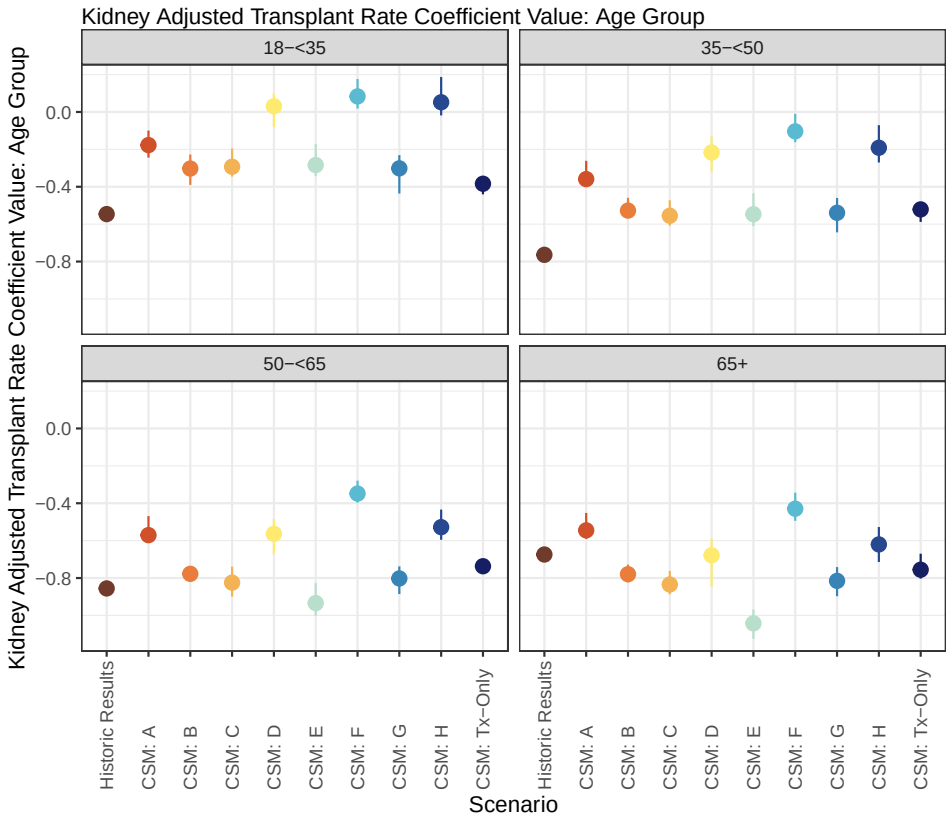

Figure 21: Adjusted Rate Model Results.

6.2.2.2 Coefficient-Level Diagnostics The following figures show the coefficient estimate and P-value at the covariate level for each of the adjusting factors for further detail about the quality of the adjusted rate models. For example, a scenario that calculates both positive and negative coefficient values across iterations might be an indication that the resulting rates calculated for that scenario should not be trusted.

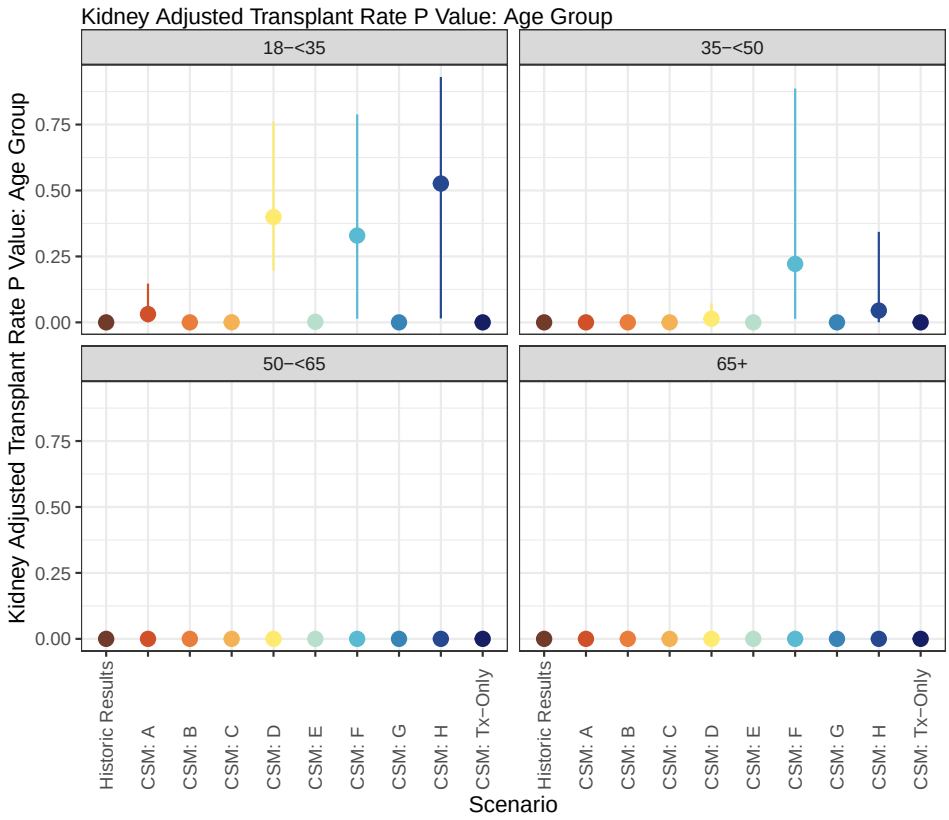

Figure 22: Adjusted Rate Model Results.

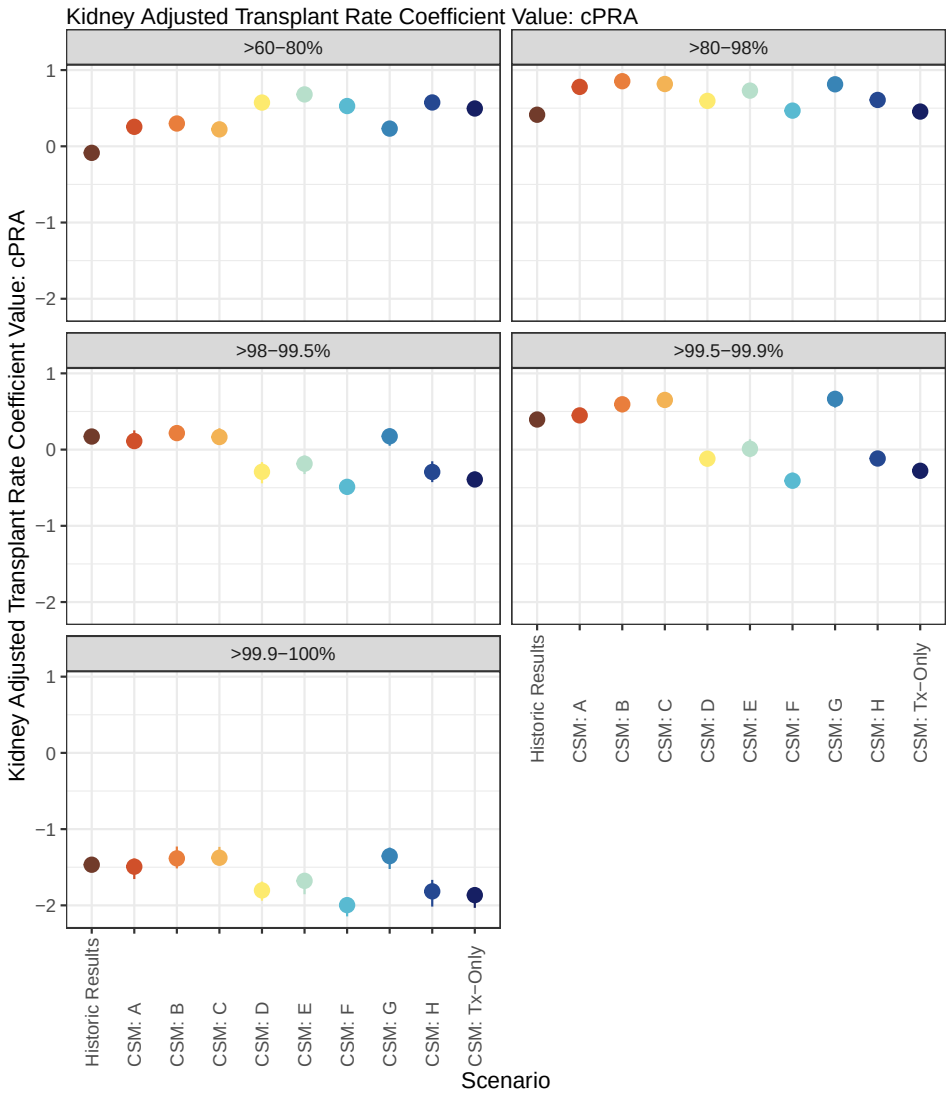

Figure 23: Adjusted Rate Model Results.

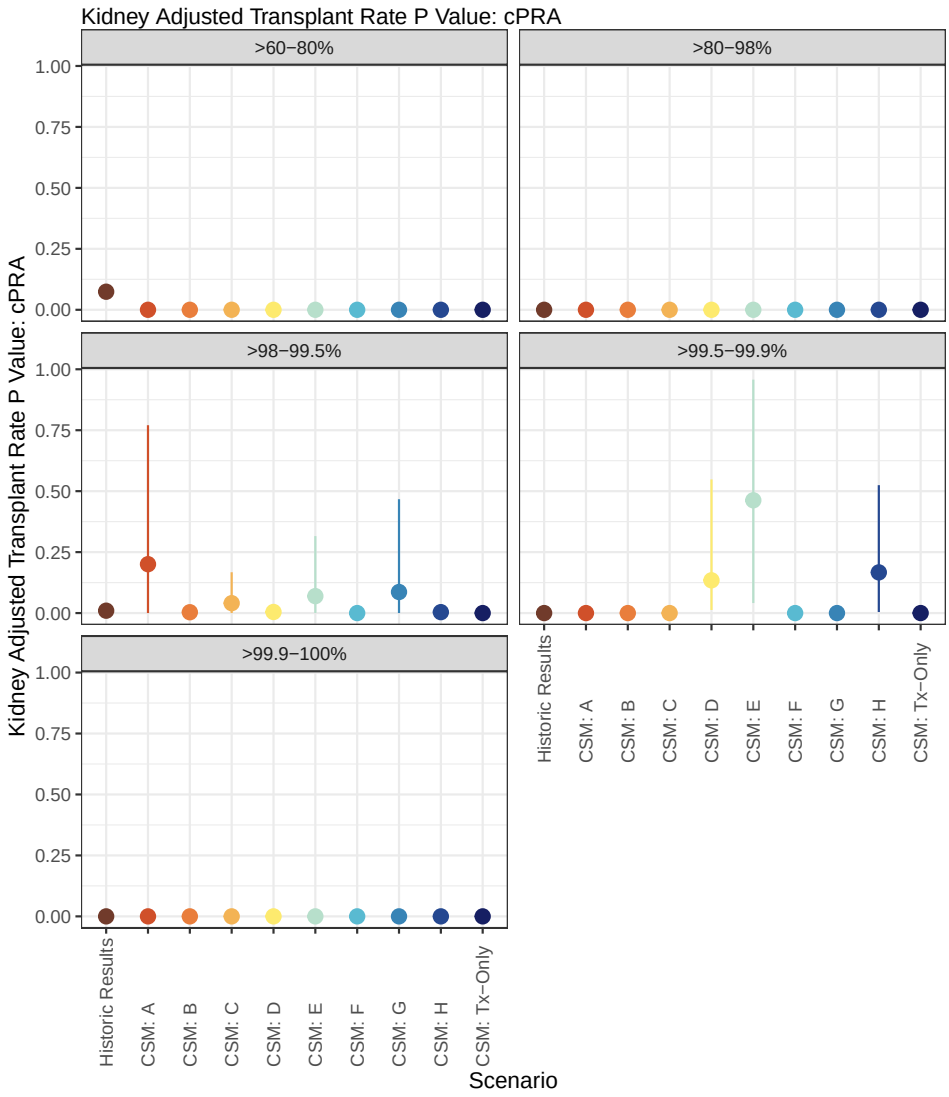

Figure 24: Adjusted Rate Model Results.

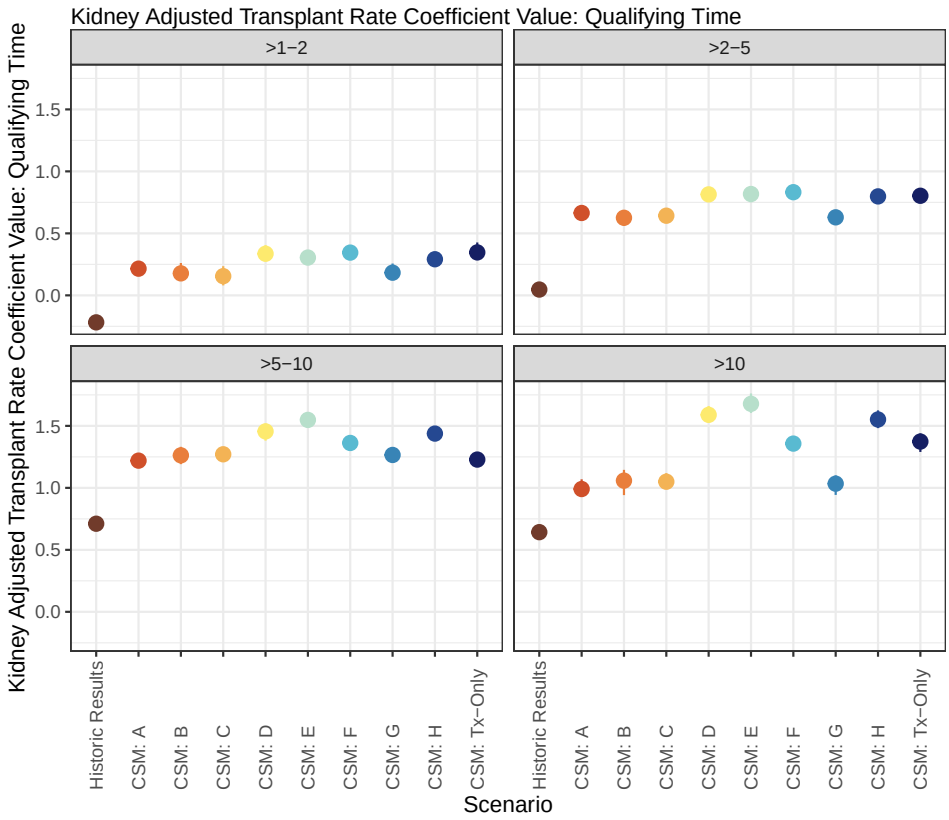

Figure 25: Adjusted Rate Model Results.

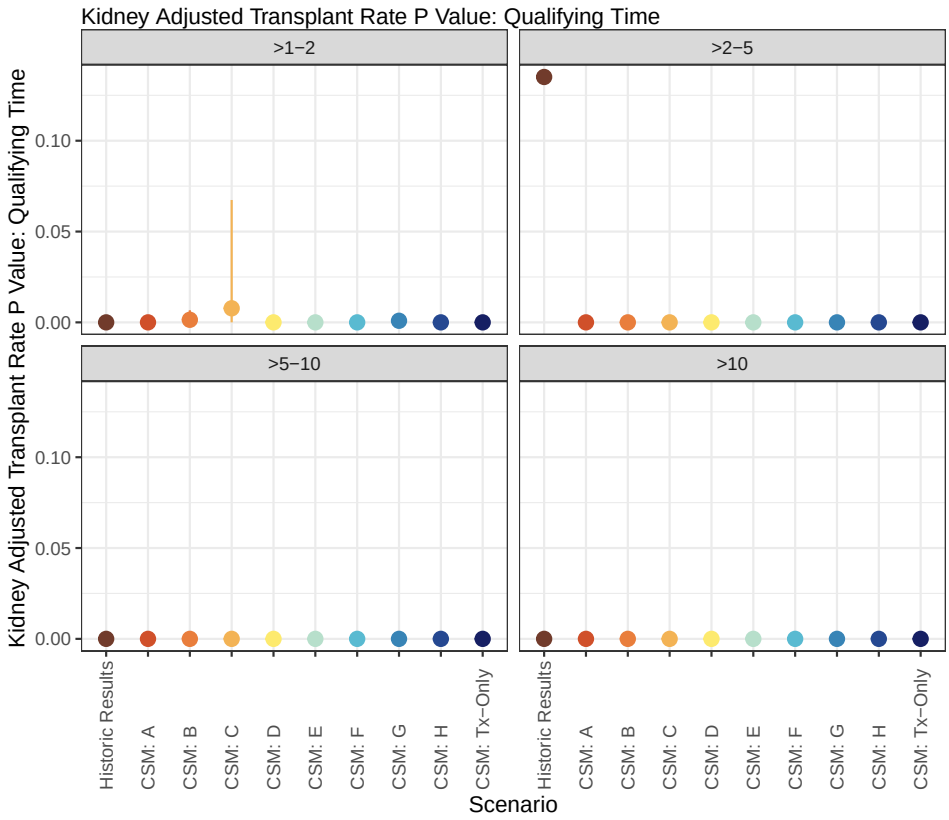

Figure 26: Adjusted Rate Model Results.

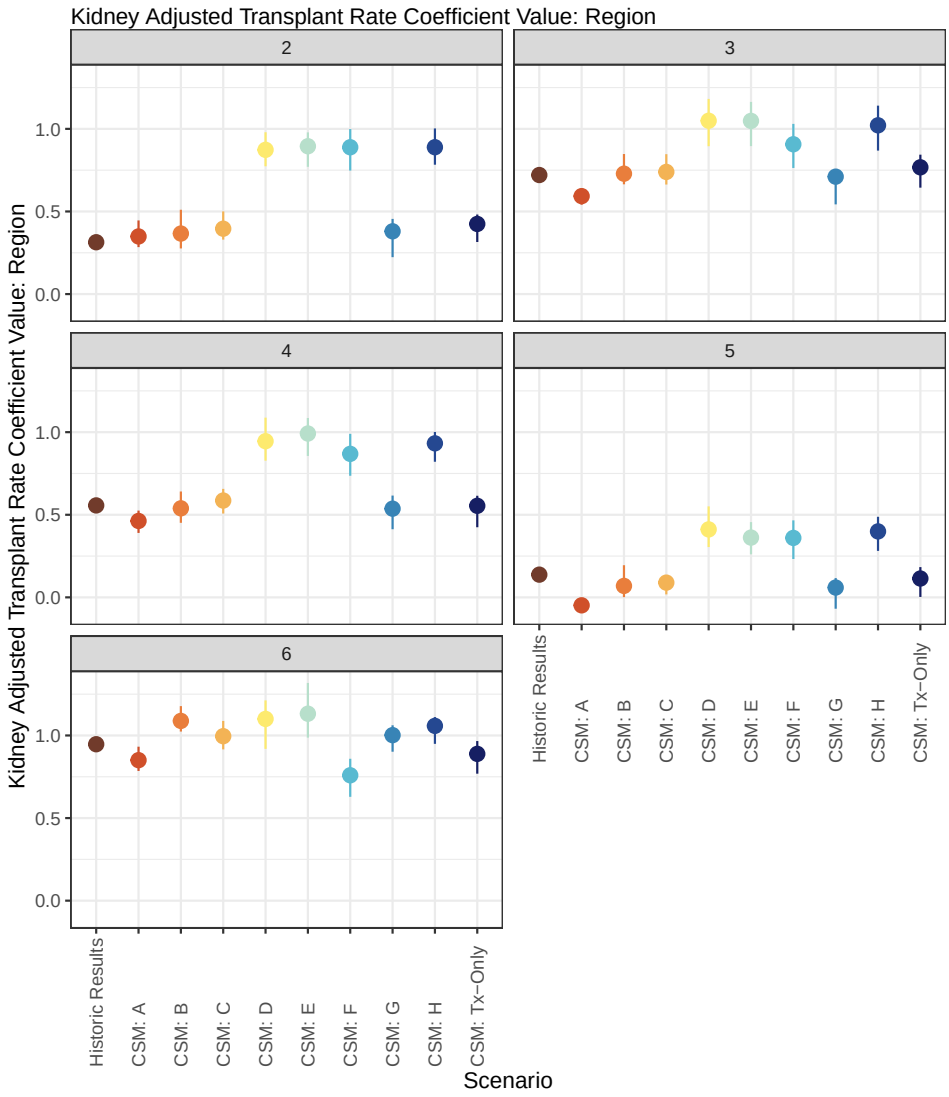

Figure 27: Adjusted Rate Model Results.

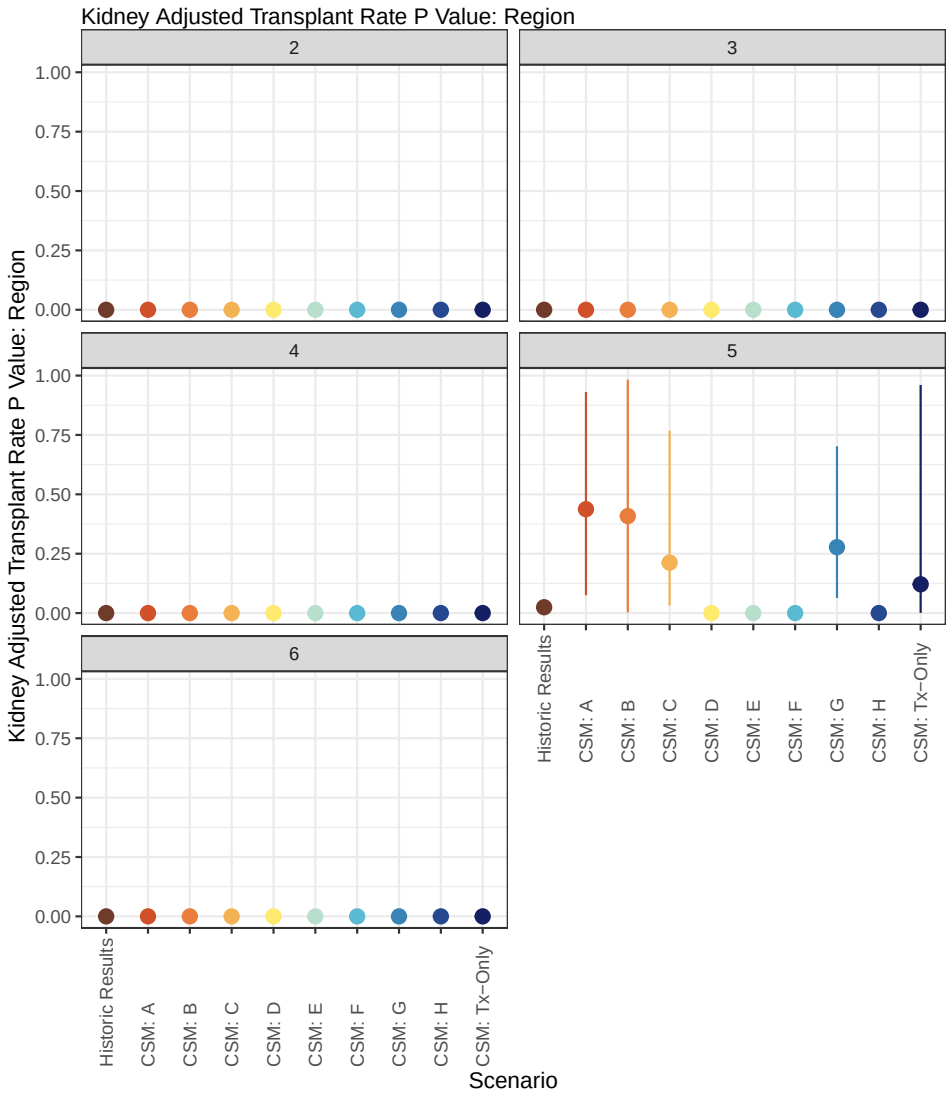

Figure 28: Adjusted Rate Model Results.

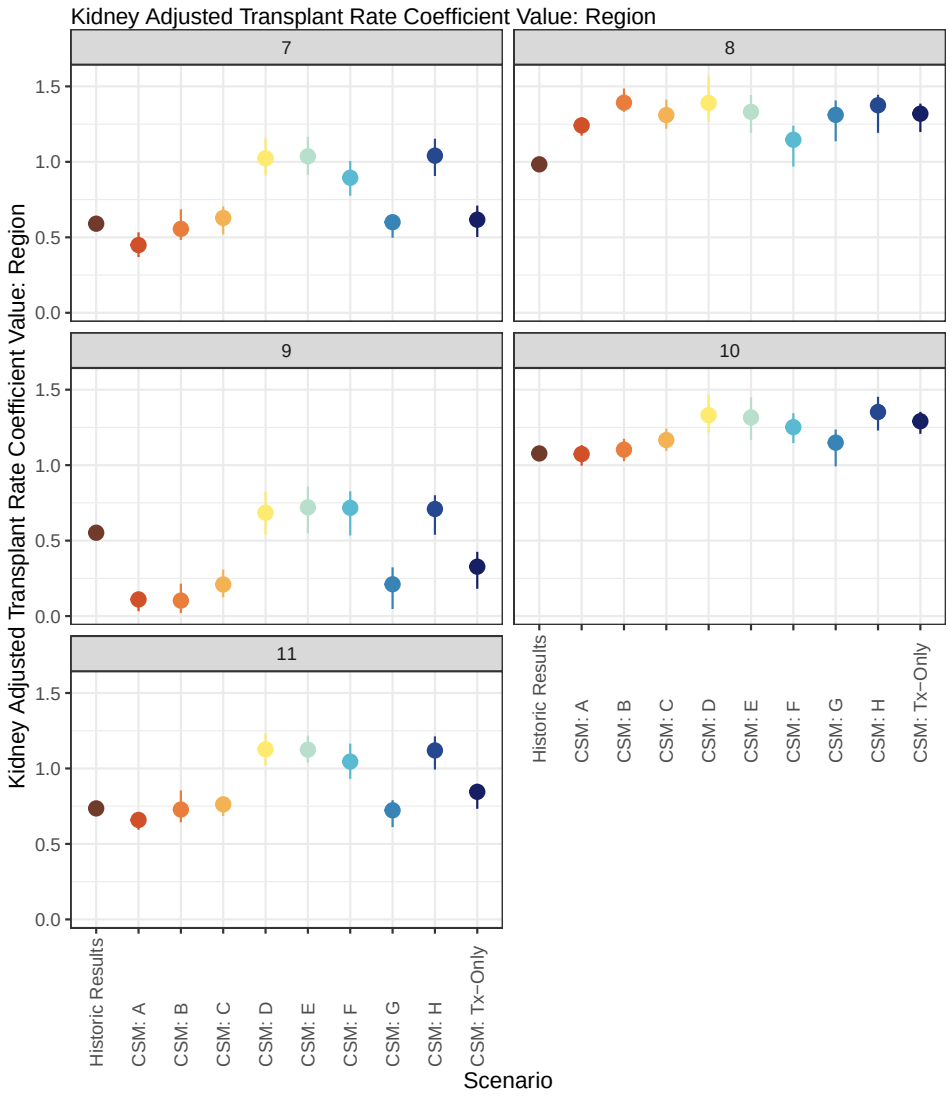

Figure 29: Adjusted Rate Model Results.

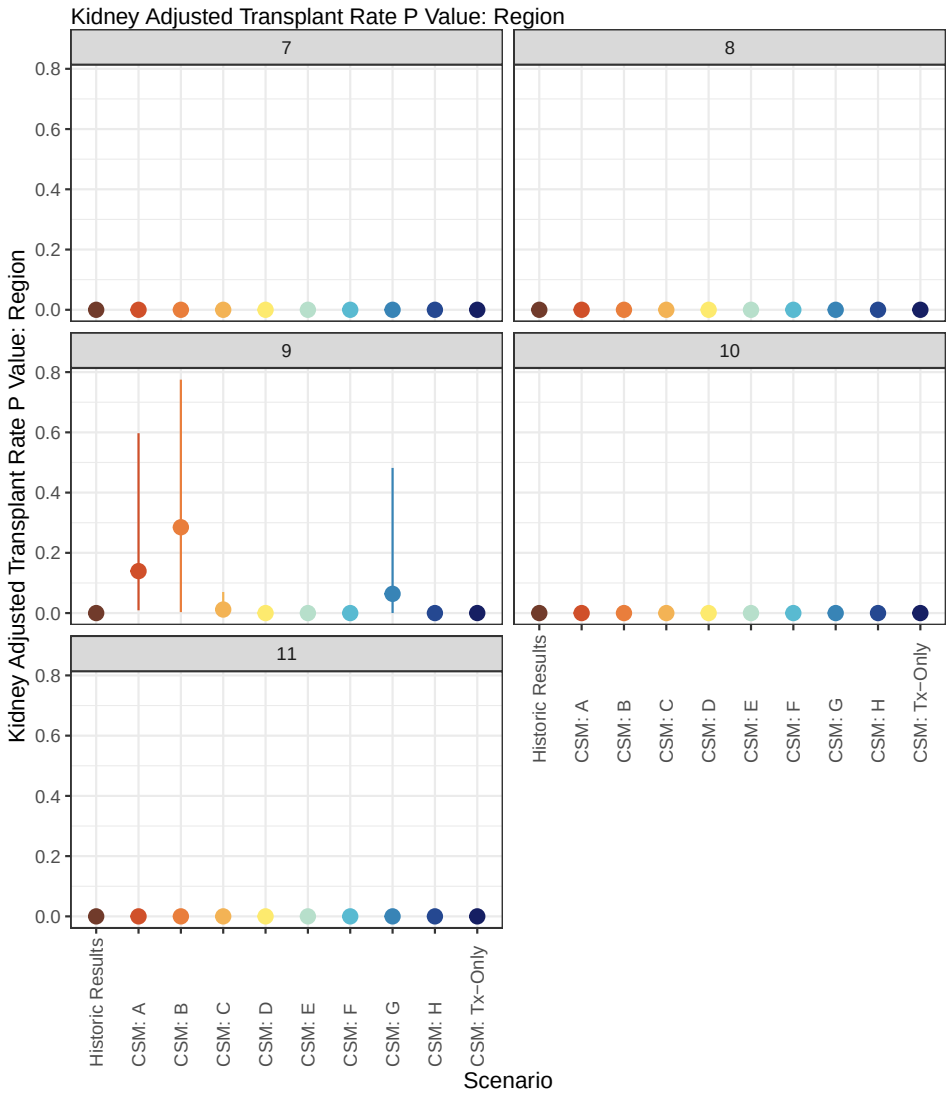

Figure 30: Adjusted Rate Model Results.

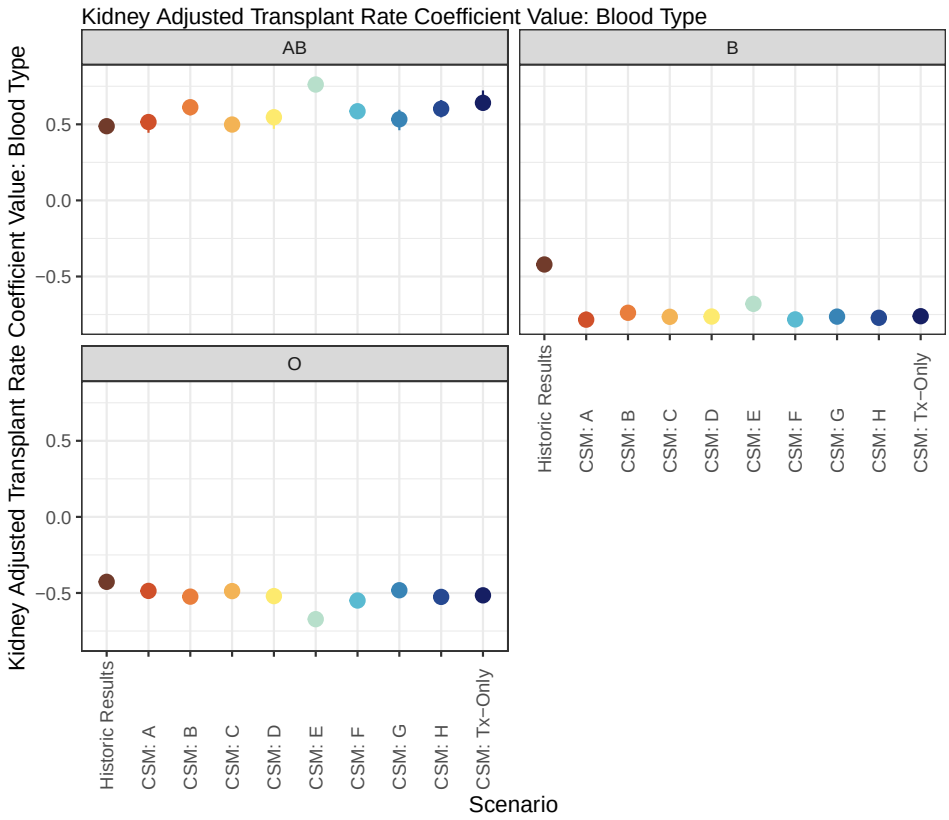

Figure 31: Adjusted Rate Model Results.

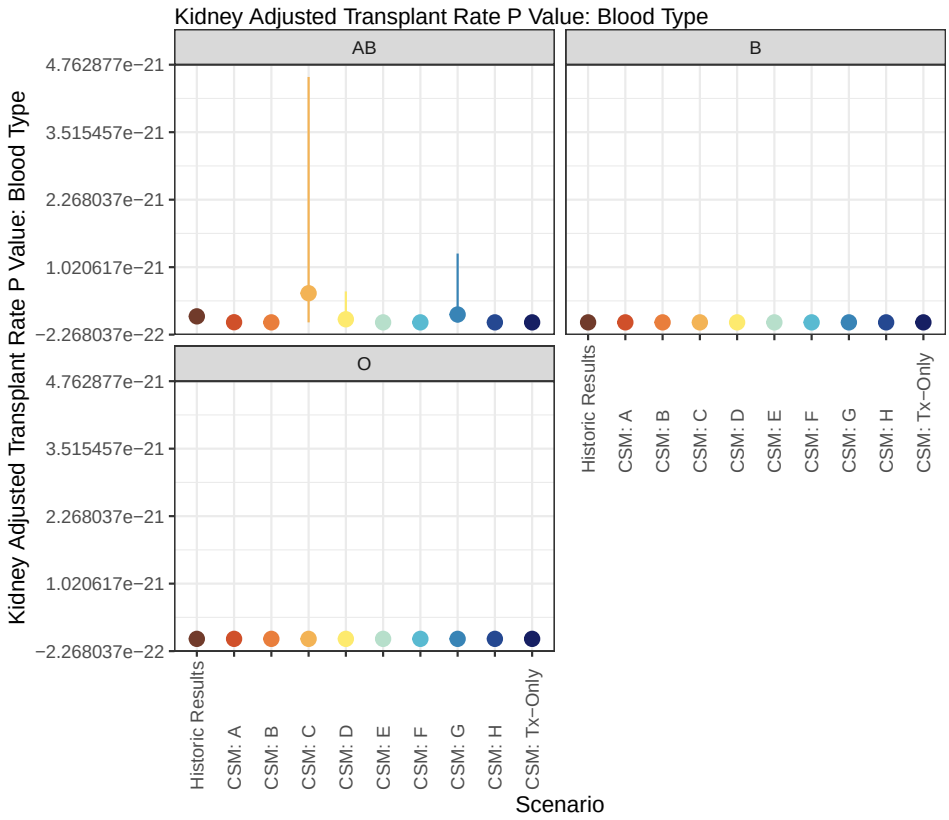

Figure 32: Adjusted Rate Model Results.

6.2.2.3 Kidney Adjusted Transplant Rates Basic diagnostics of the adjusted rate models were shown in the last section. Here the adjusted rate models are used to calculate adjusted rates via the standardization process.

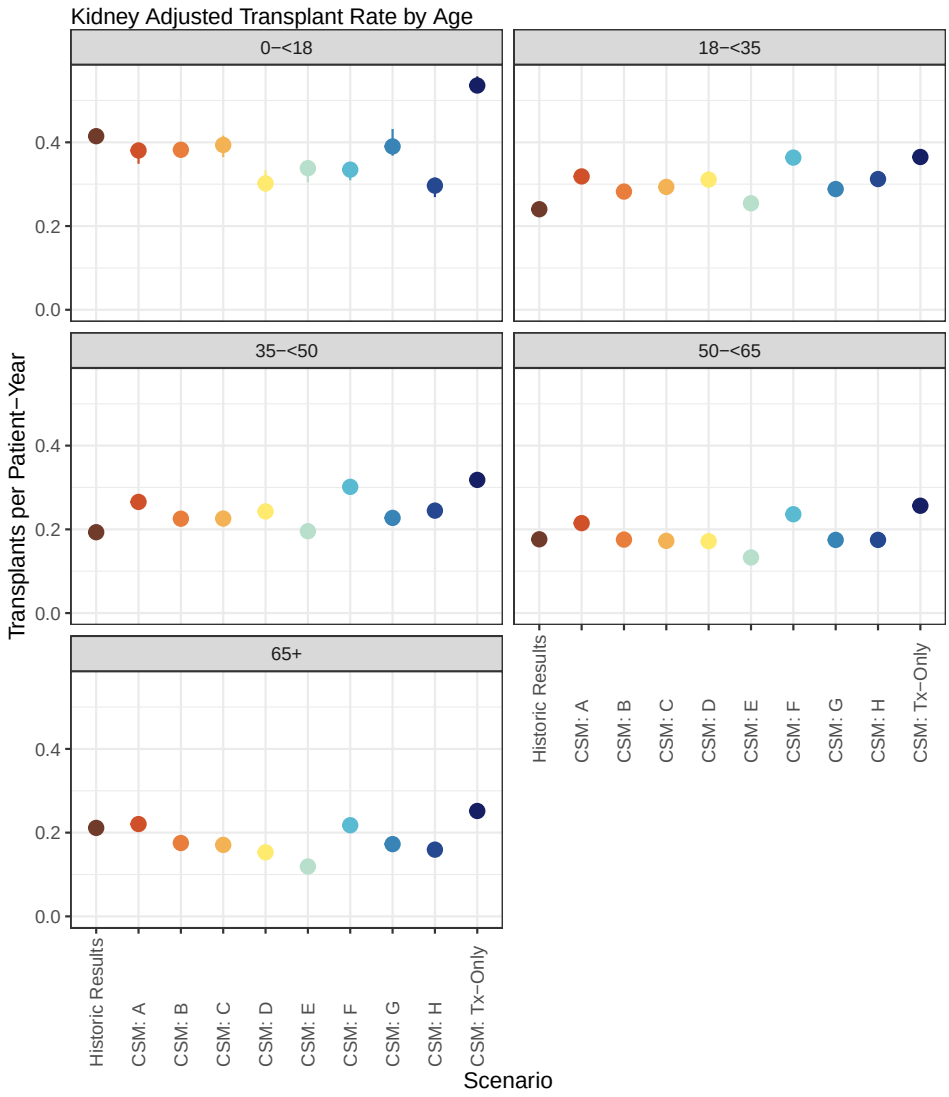

Figure 33: Kidney Adjusted Transplant Rate by Age. Kidney adjusted transplant rate model with adjusting factors: age at listing, cPRA, qualifying time, OPTN region, and blood type. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end. Age at listing.

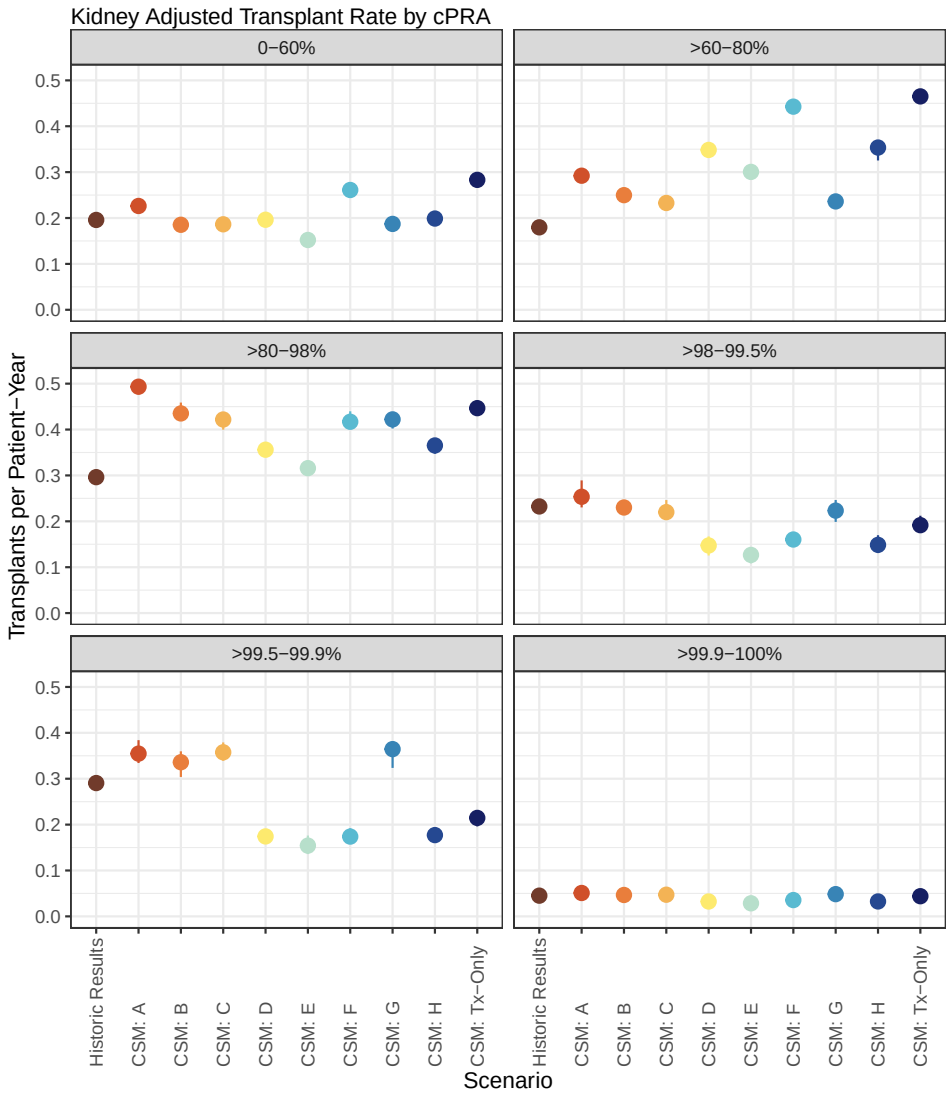

Figure 34: Kidney Adjusted Transplant Rate by cPRA. Kidney adjusted transplant rate model with adjusting factors: age at listing, cPRA, qualifying time, OPTN region, and blood type. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end. cPRA is at the last value the candidate had prior to the simulation start or their value at listing.

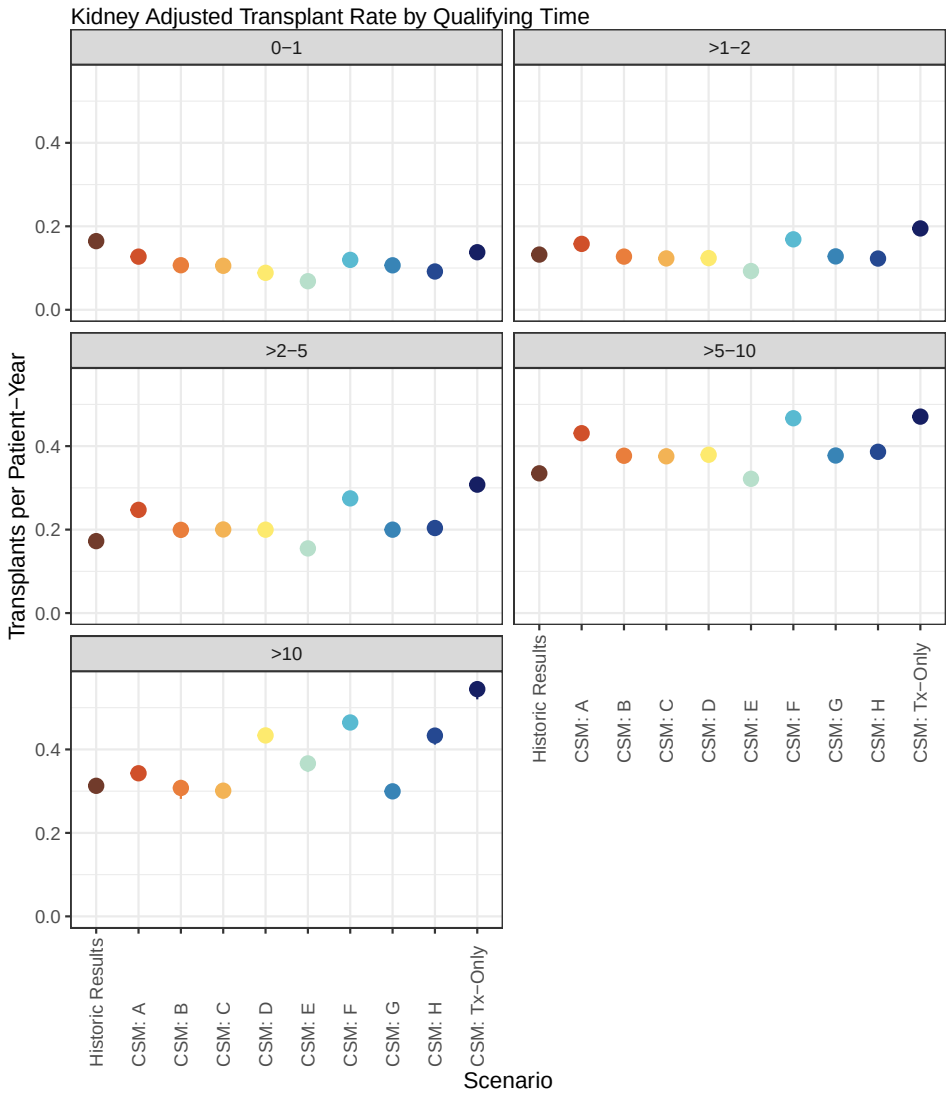

Figure 35: Kidney Adjusted Transplant Rate by Qualifying Time. Kidney adjusted transplant rate model with adjusting factors: age at listing, cPRA, qualifying time, OPTN region, and blood type. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end. Qualifying time is time in years from qualifying date to the start of their simulation period or 0 if listed during the simulation period.

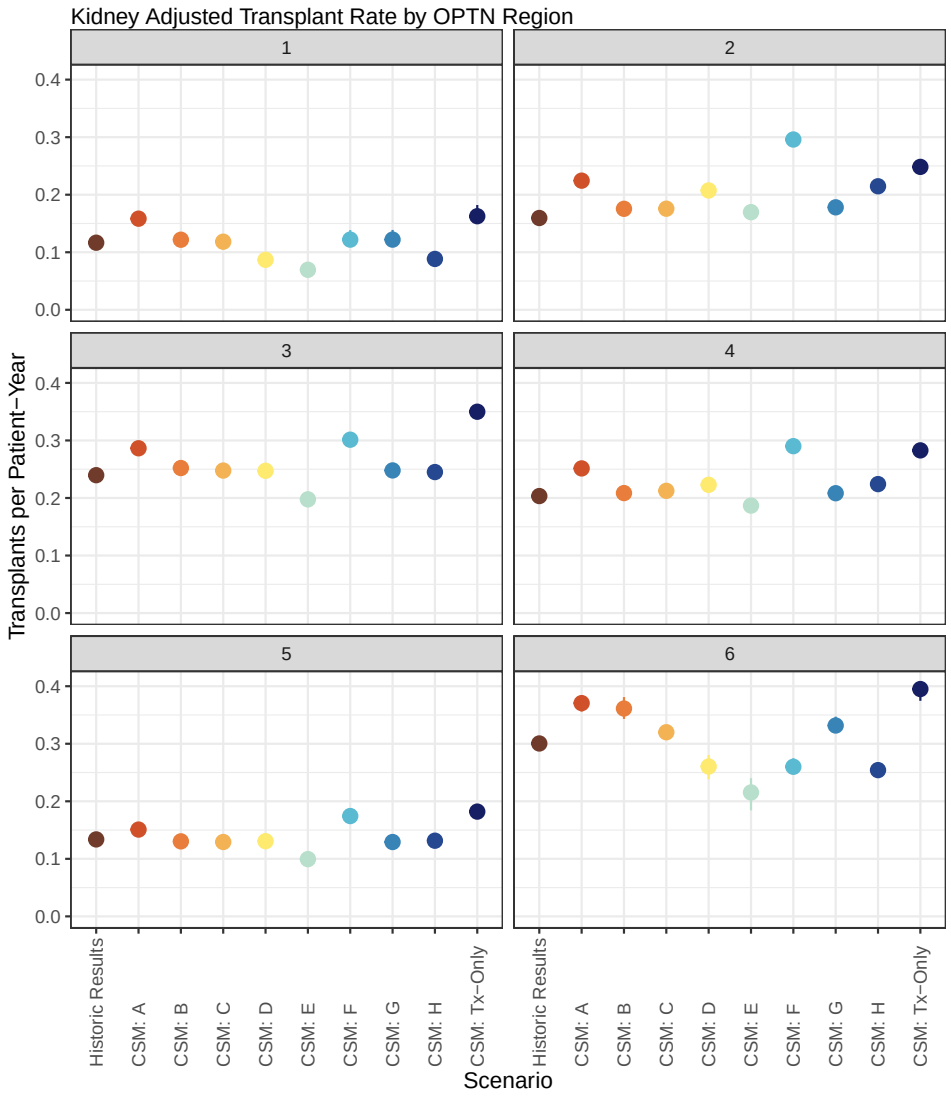

Figure 36: Kidney Adjusted Transplant Rate by OPTN Region. Kidney adjusted transplant rate model with adjusting factors: age at listing, cPRA, qualifying time, OPTN region, and blood type. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

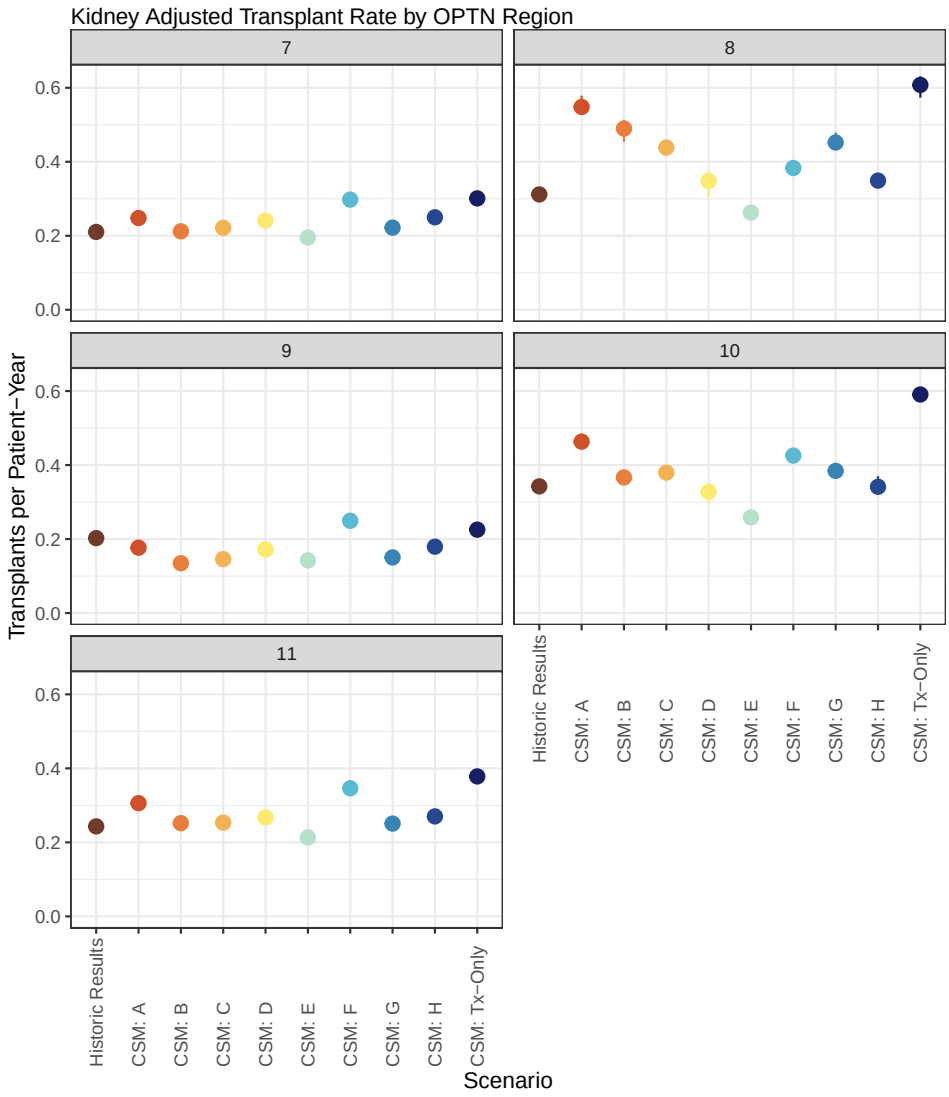

Figure 37: Kidney Adjusted Transplant Rate by OPTN Region. Kidney adjusted transplant rate model with adjusting factors: age at listing, cPRA, qualifying time, OPTN region, and blood type. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

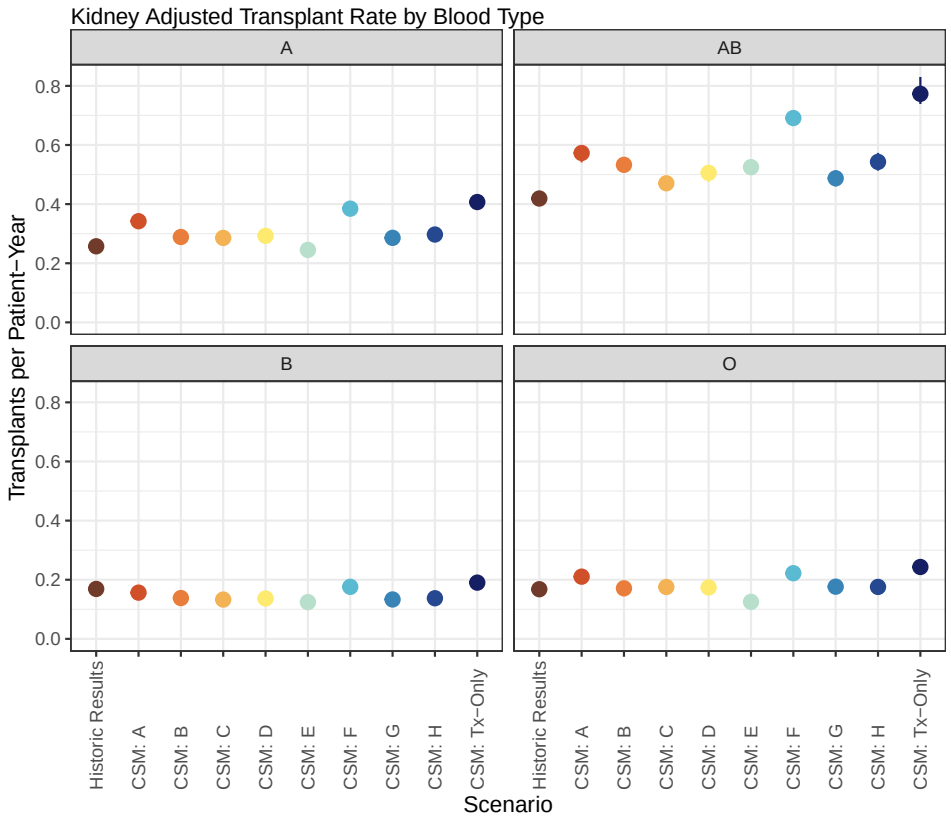

Figure 38: Kidney Adjusted Transplant Rate by Blood Type. Kidney adjusted transplant rate model with adjusting factors: age at listing, cPRA, qualifying time, OPTN region, and blood type. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

6.2.3 Other Simulated Metrics

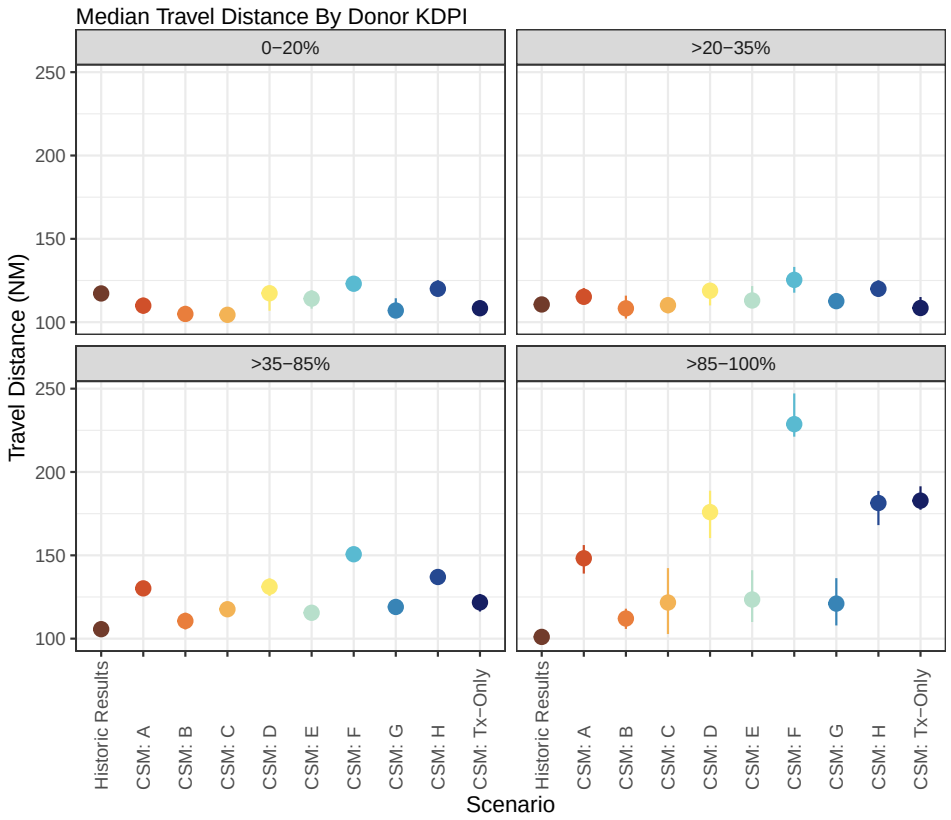

Figure 39: Median Travel Distance By Donor KDPI. Travel distance is between the donor hospital and the transplant center, in nautical miles.

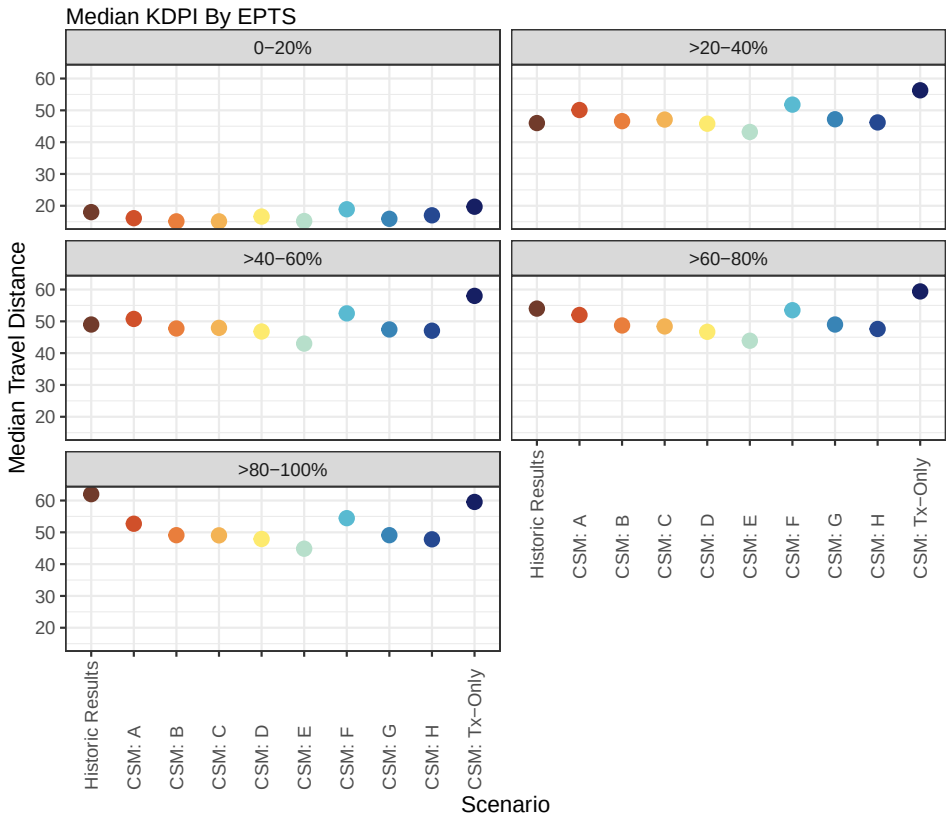

Figure 40: Median KDPI By EPTS.

6.2.4 Posttransplant

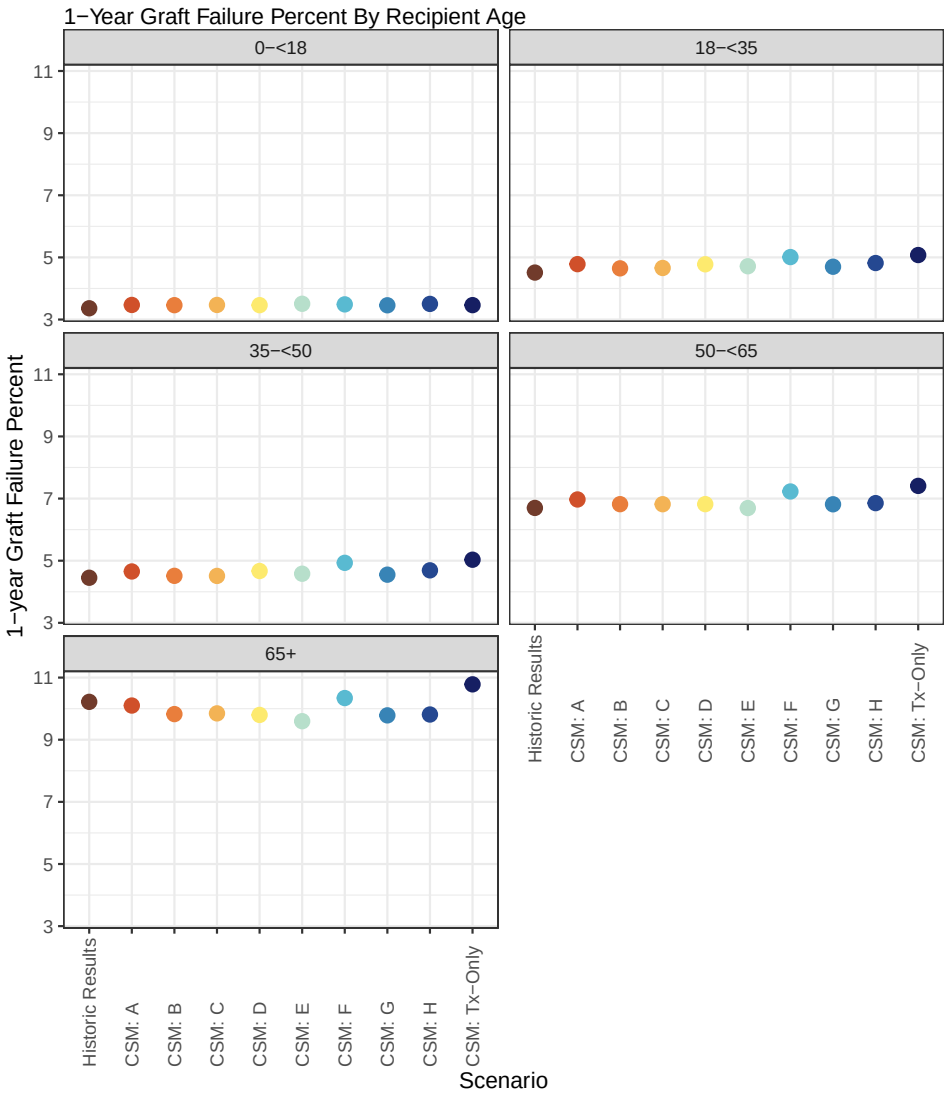

Figure 41: 1-Year Graft Failure Percent By Recipient Age.

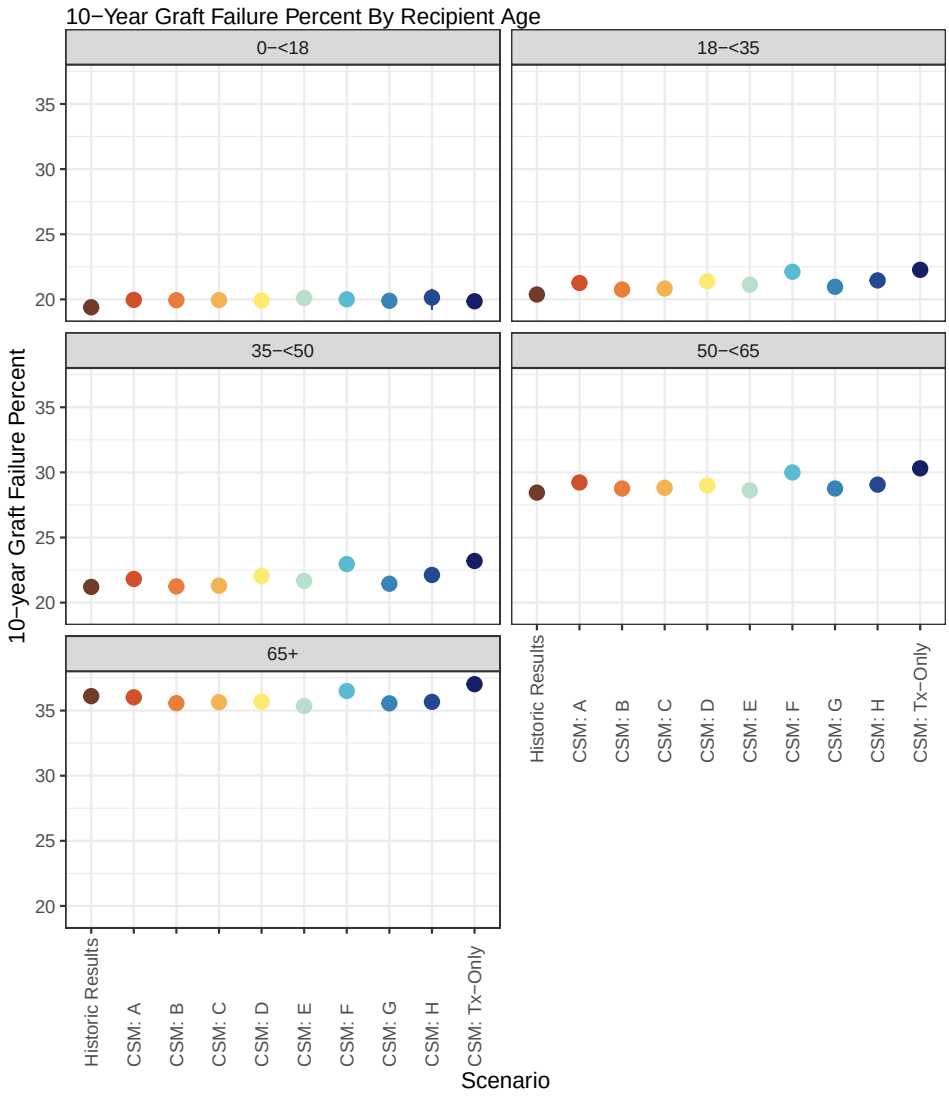

Figure 42: 10-Year Graft Failure Percent By Recipient Age.

6.3 Pancreas Alone

6.3.1 Percent of Pancreas Not Used by Donor Characteristics

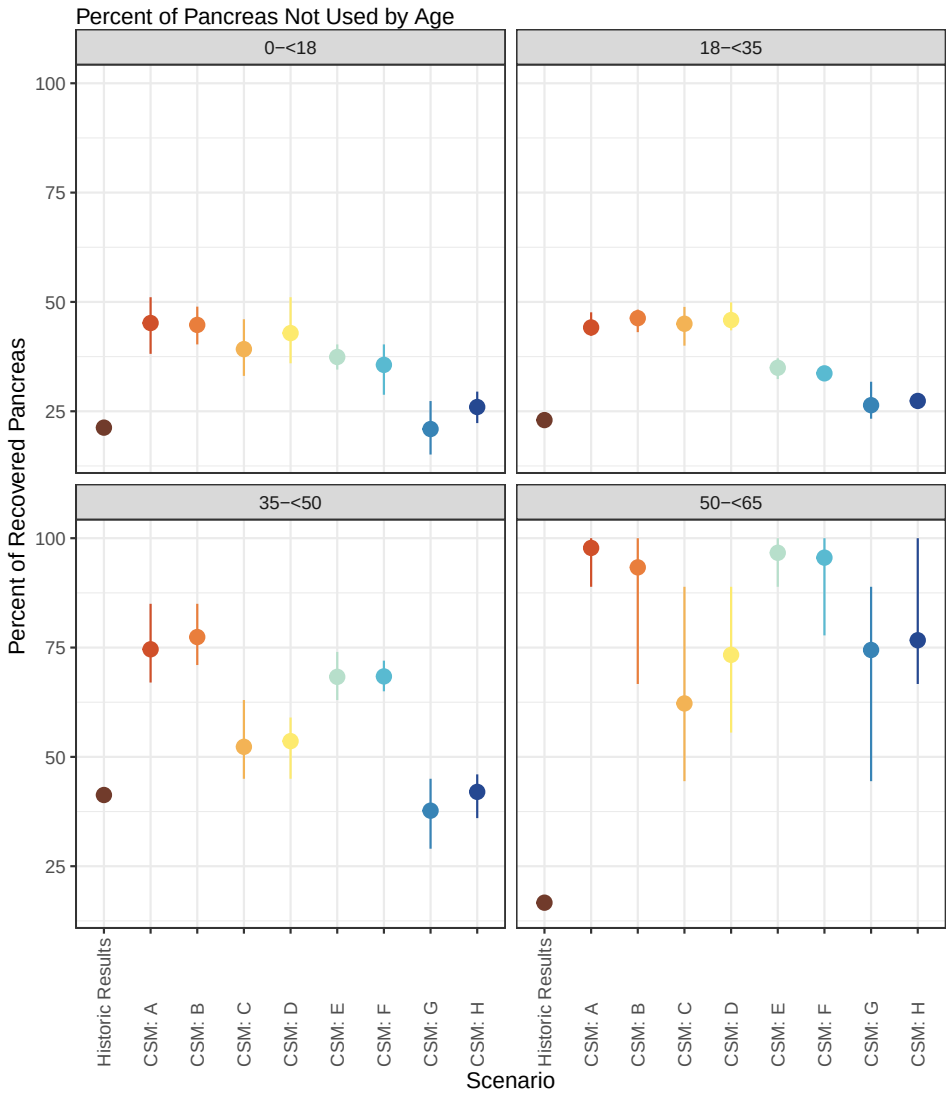

Figure 43: Percent of Pancreas Not Used by Age. Includes all pancreas that were recovered for transplant.

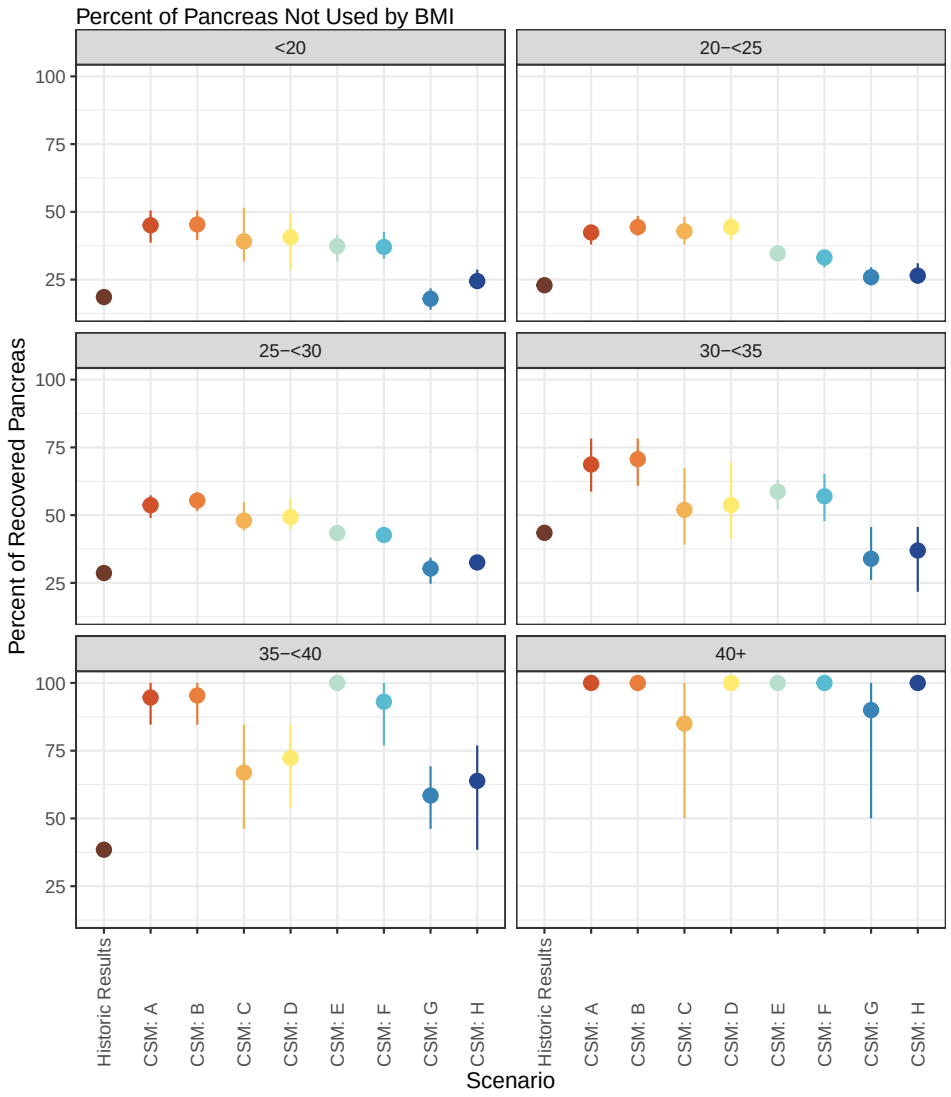

Figure 44: Percent of Pancreas Not Used by BMI. Includes all pancreas that were recovered for transplant.

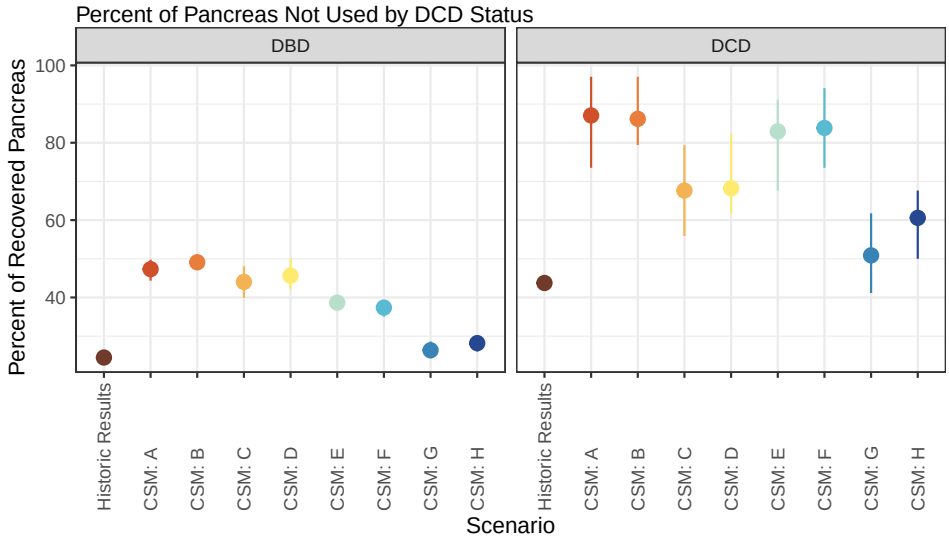

Figure 45: Percent of Pancreas Not Used by DCD Status. Includes all pancreas that were recovered for transplant.

6.3.2 Pancreas Adjusted Transplant Rate by Candidate Characteristics

For pancreas an adjusted rate model was fit at each scenario/iteration and the overall model diagnostics for each scenario are shown in Figure 46. The adjusting factors are age at listing, cPRA, qualifying time, and blood type. These figures are meant to be a check of the overall rate model quality, before using those models to calculate by group rates using standarization. These figures are not meant to be used for scenario selection directly but rather to look for potential deficiencies in the simulated outcomes of the scenarios.

Given the results below, in particular in the coefficient level diagnostics (Figure 47 to Figure 54), it is clear this adjusted rate model is not a good fit to the data for the historical results. For nearly every covariate level, the historical coefficient estimate was essentially zero–indicating the adjusting factors are not associated with the observed waitlist outcomes. Given this, we will not be presenting adjusted rates for pancreas.

6.3.2.1 Overall Model Diagnostics

6.3.2.2 Coefficient Level Diagnostics

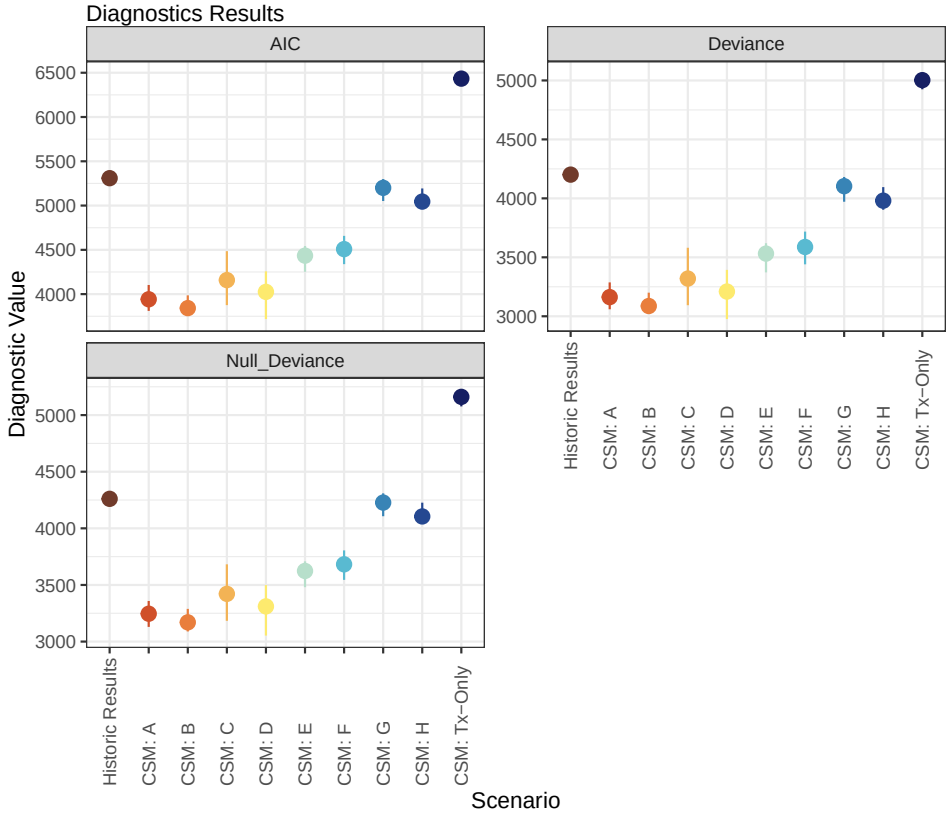

Figure 46: Diagnostics Results.

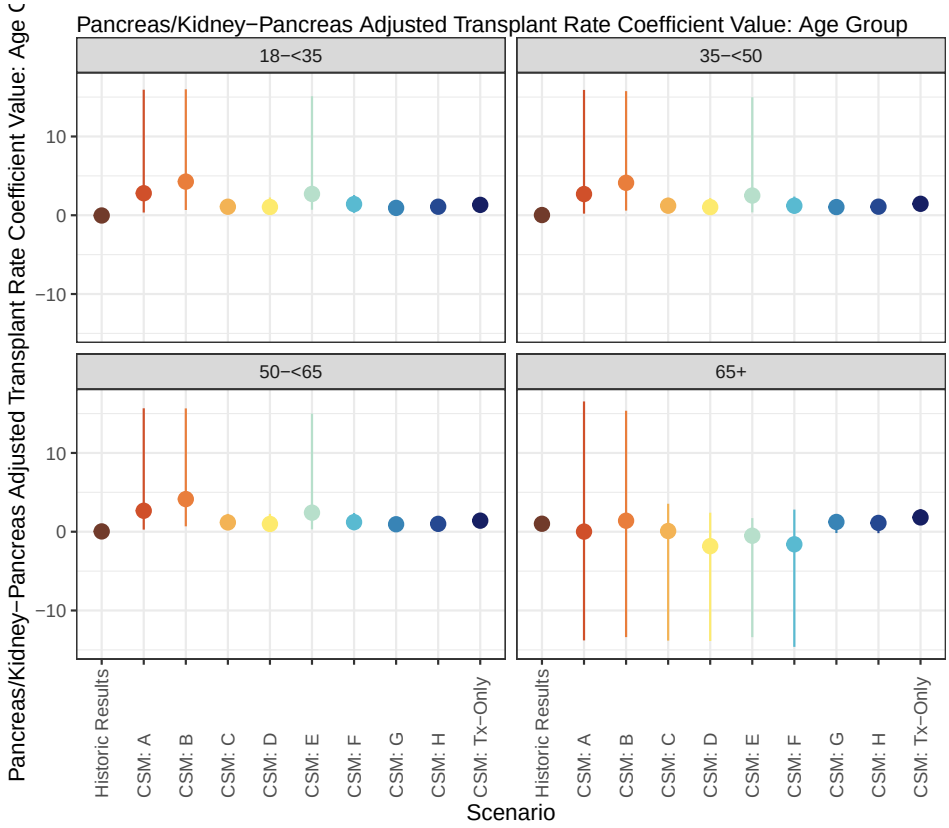

Figure 47: Adjusted Rate Model Results.

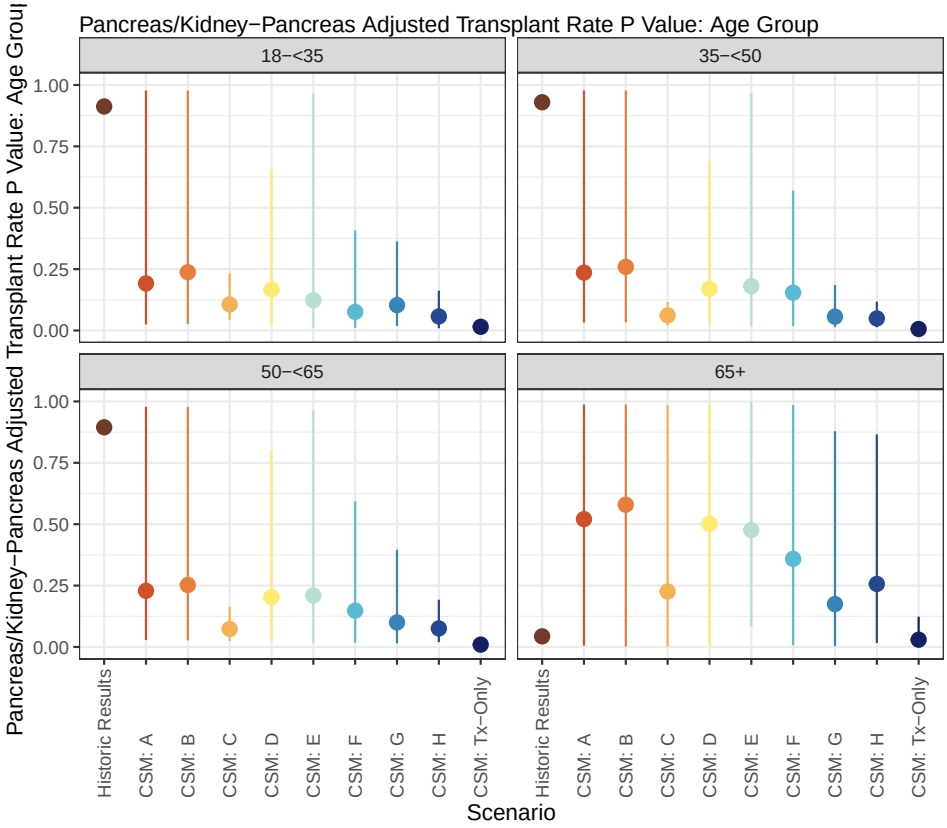

Figure 48: Adjusted Rate Model Results.

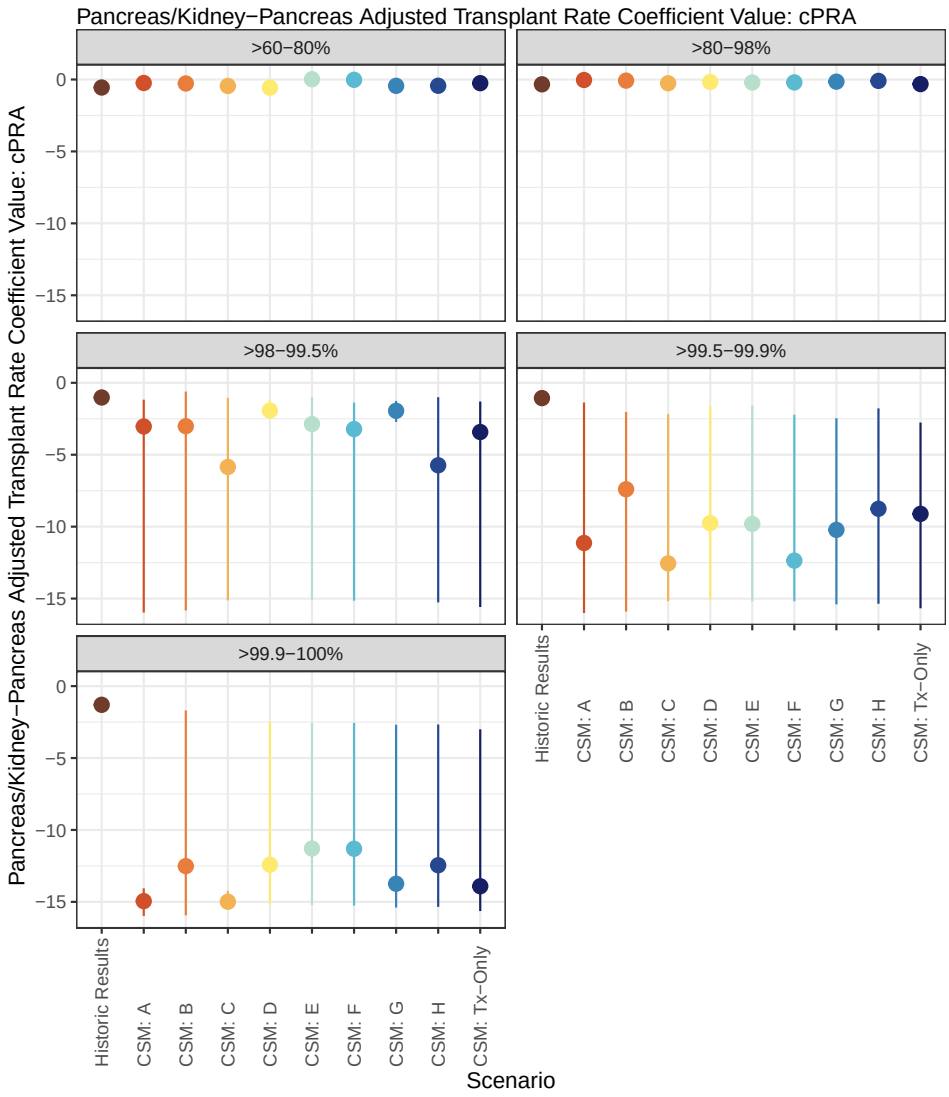

Figure 49: Adjusted Rate Model Results.

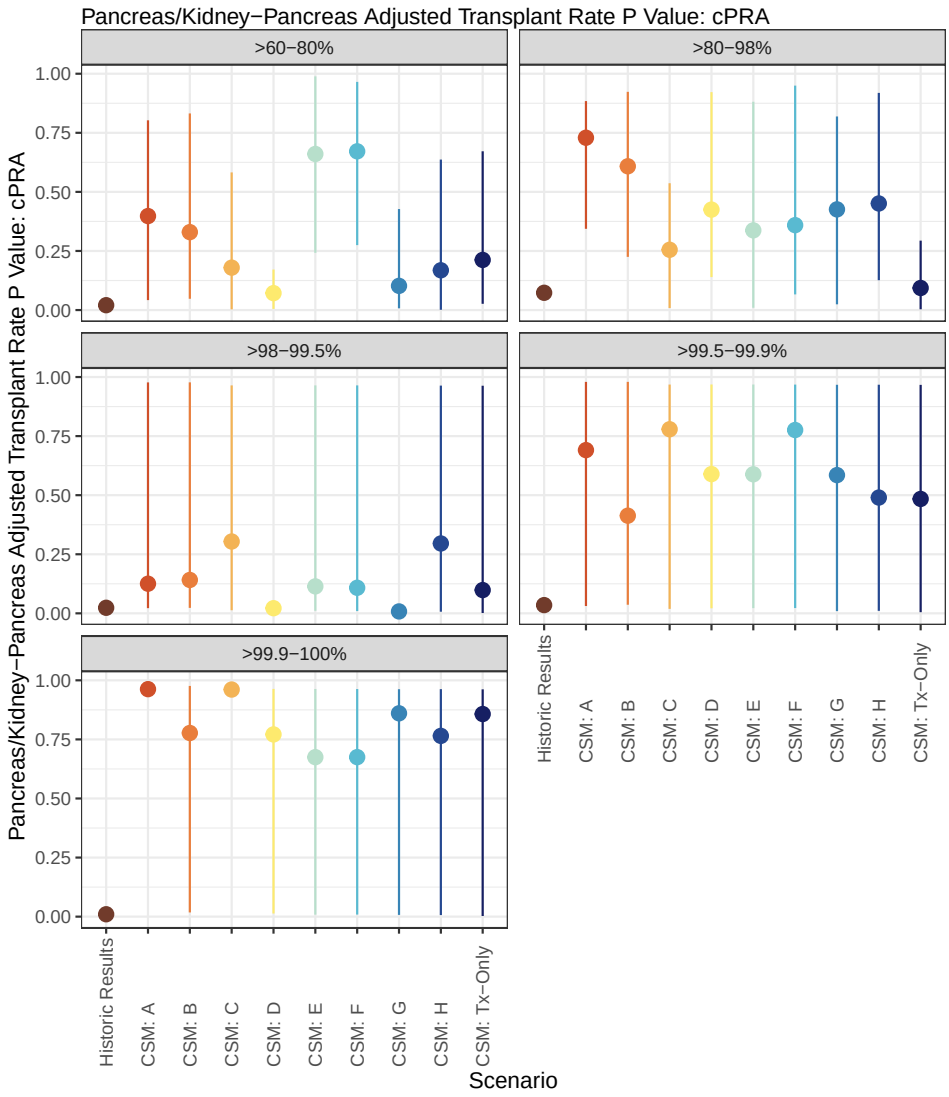

Figure 50: Adjusted Rate Model Results.

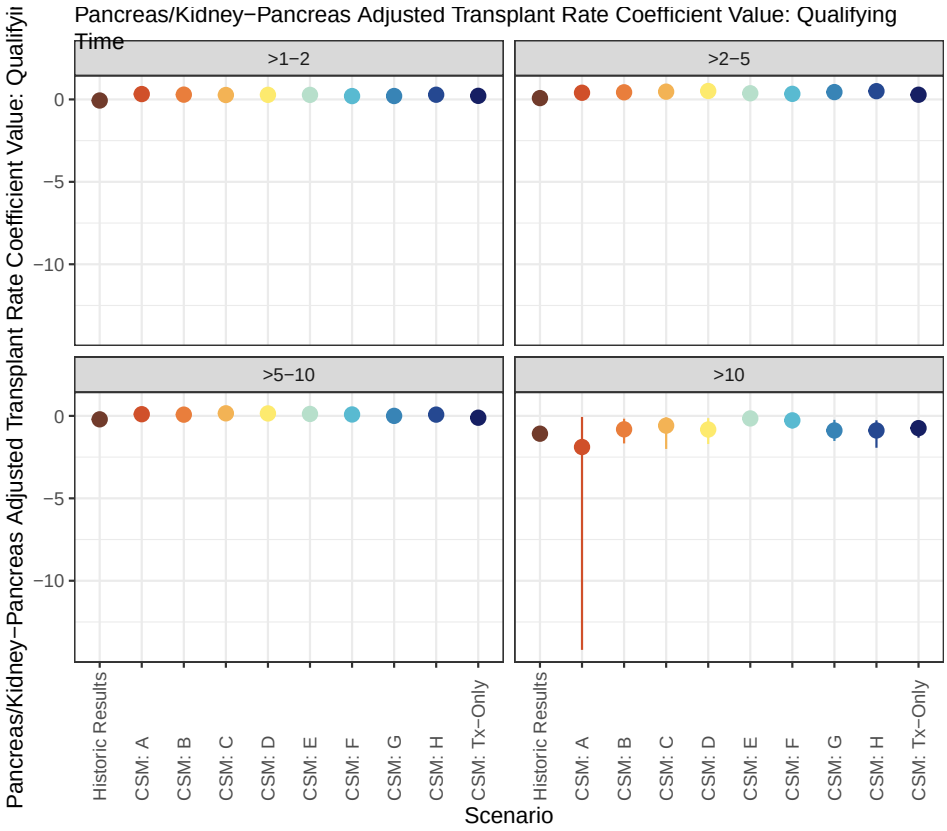

Figure 51: Adjusted Rate Model Results.

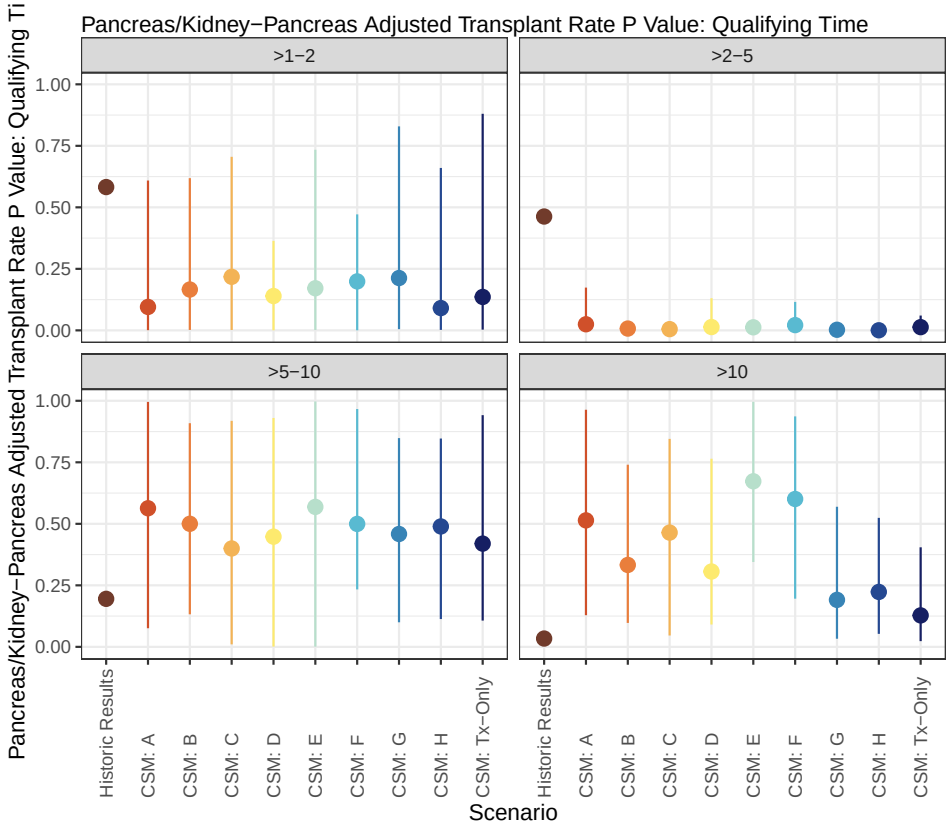

Figure 52: Adjusted Rate Model Results.

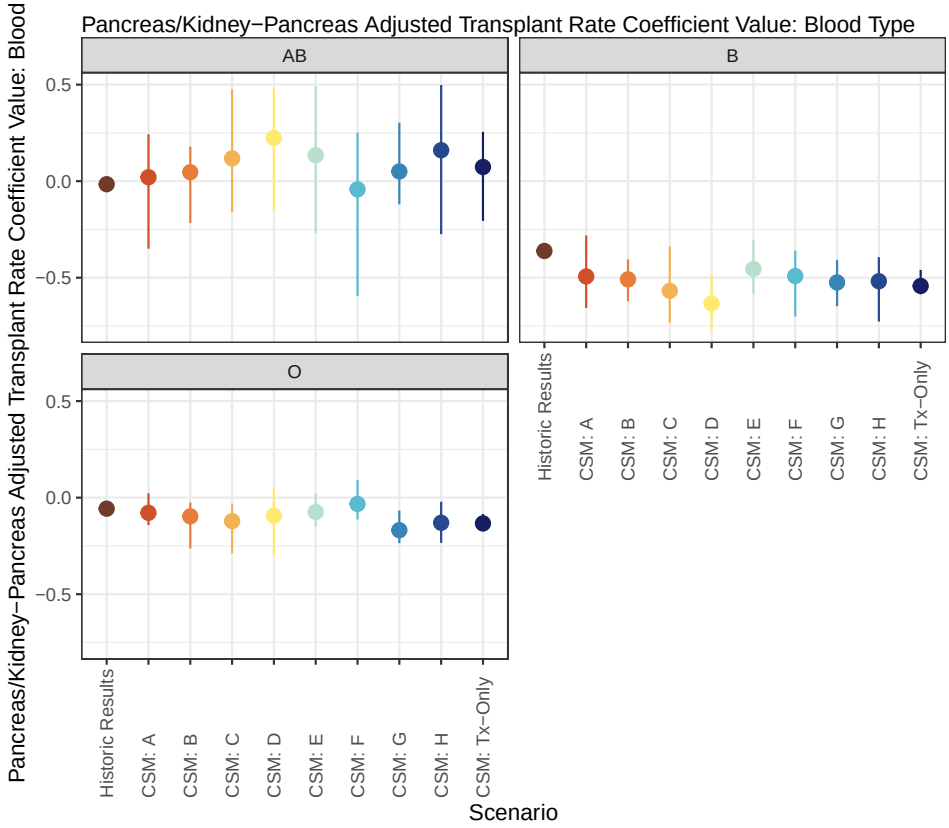

Figure 53: Adjusted Rate Model Results.

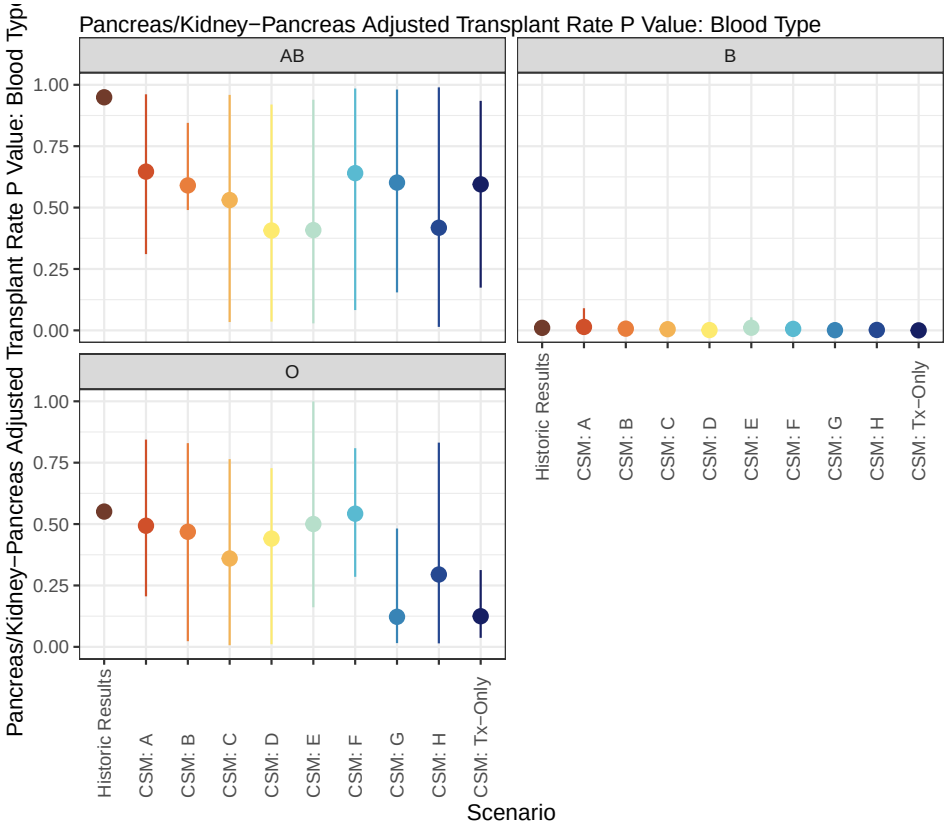

Figure 54: Adjusted Rate Model Results.

6.3.2.3 Pancreas/Kidney-Pancreas Adjusted Transplant Rates As noted in "Pancreas Adjusted Transplant Rate by Candidate Characteristics", the given adjusting factors were not good predictors for the waitlist outcomes and in turn we are not presenting adjusted transplant rates for pancreas.

6.4 Primary Assessment Metrics Based on Modeling/Assumptions

6.4.1 Kidney

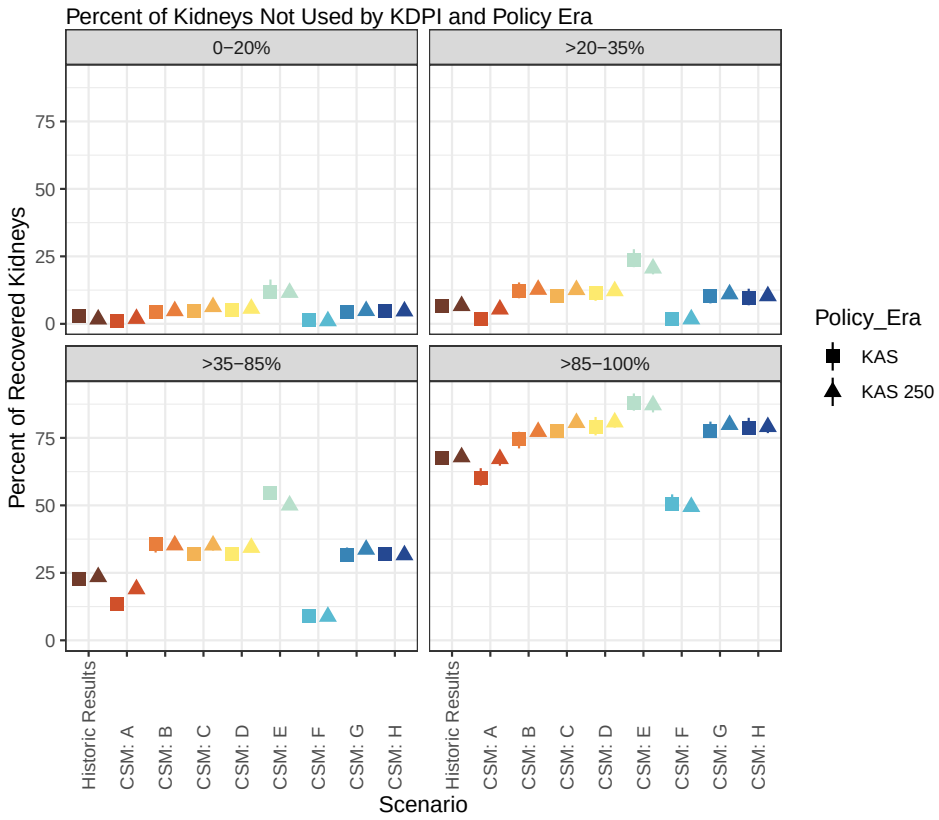

Figure 55: Percent of Kidneys Not Used by KDPI and Policy Era. Includes all kidneys that were recovered for transplant.

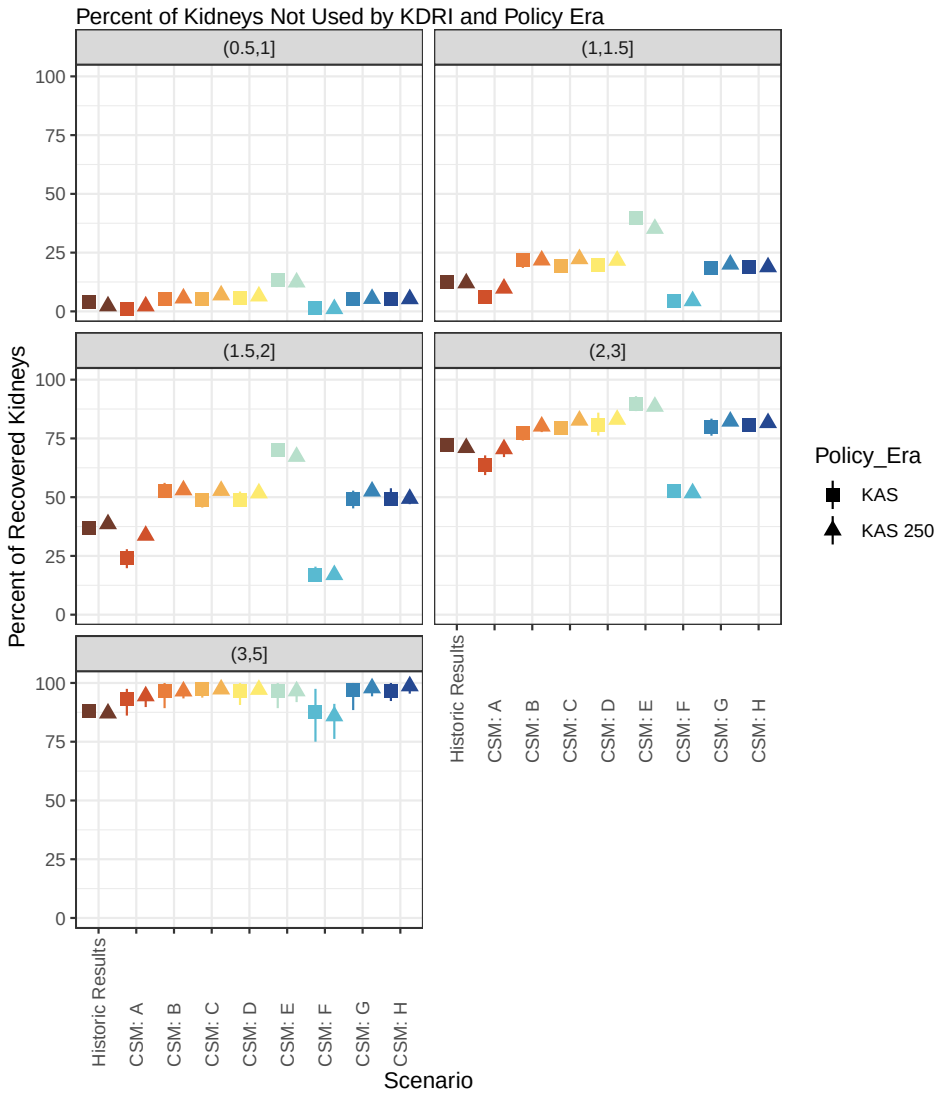

Figure 56: Percent of Kidneys Not Used by KDRI and Policy Era. Includes all kidneys that were recovered for transplant.

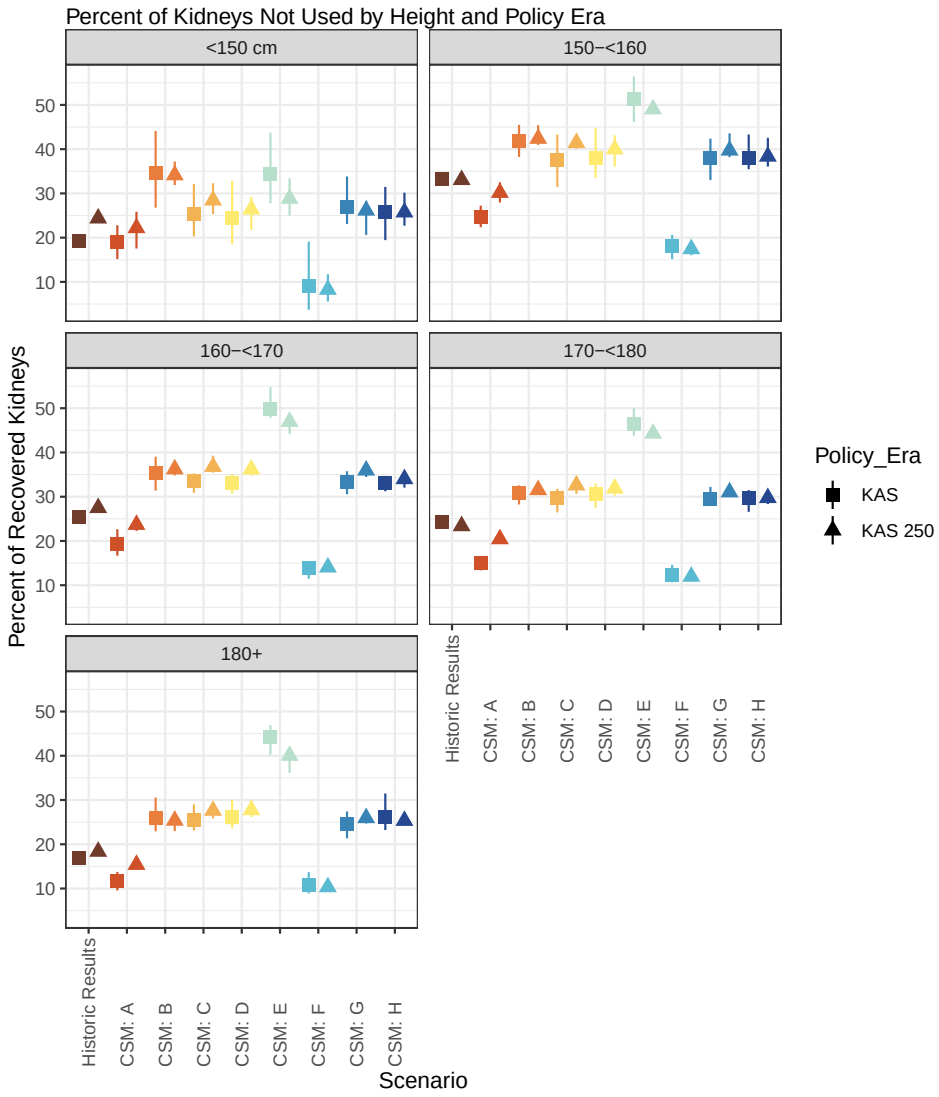

Figure 57: Percent of Kidneys Not Used by Height and Policy Era. Includes all kidneys that were recovered for transplant.

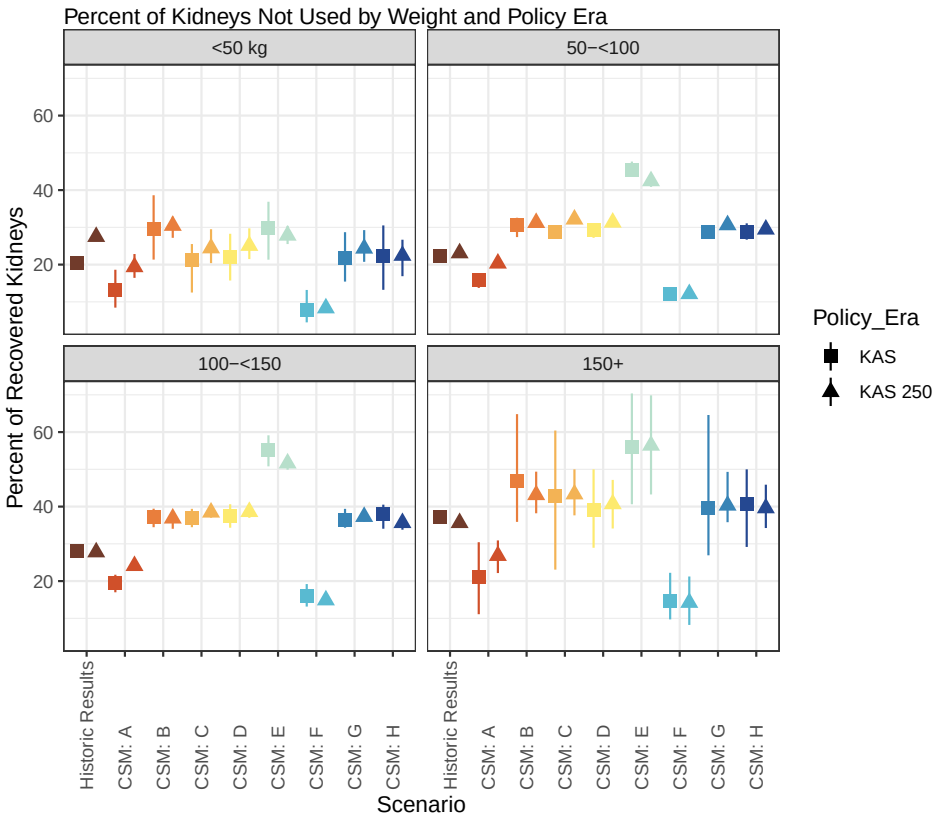

Figure 58: Percent of Kidneys Not Used by Weight and Policy Era. Includes all kidneys that were recovered for transplant.

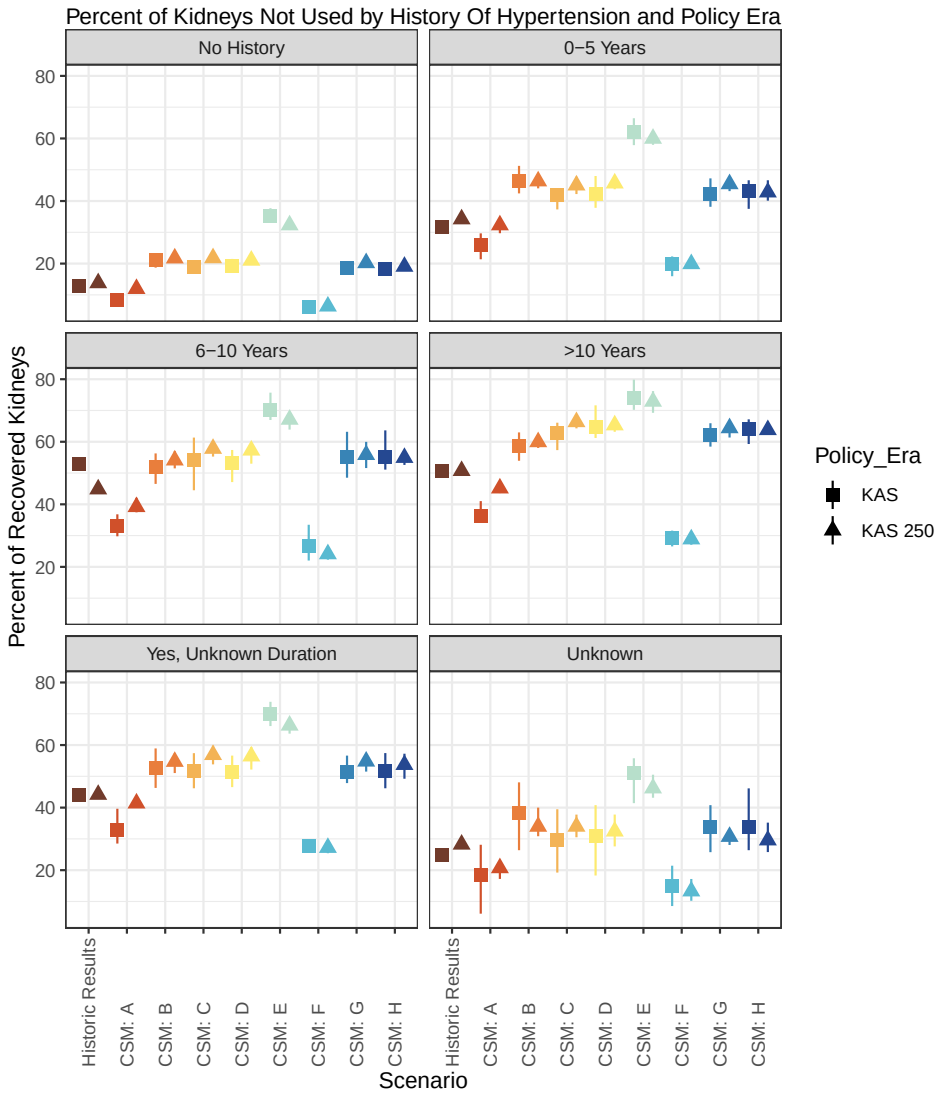

Figure 59: Percent of Kidneys Not Used by History Of Hypertension and Policy Era. Includes all kidneys that were recovered for transplant.

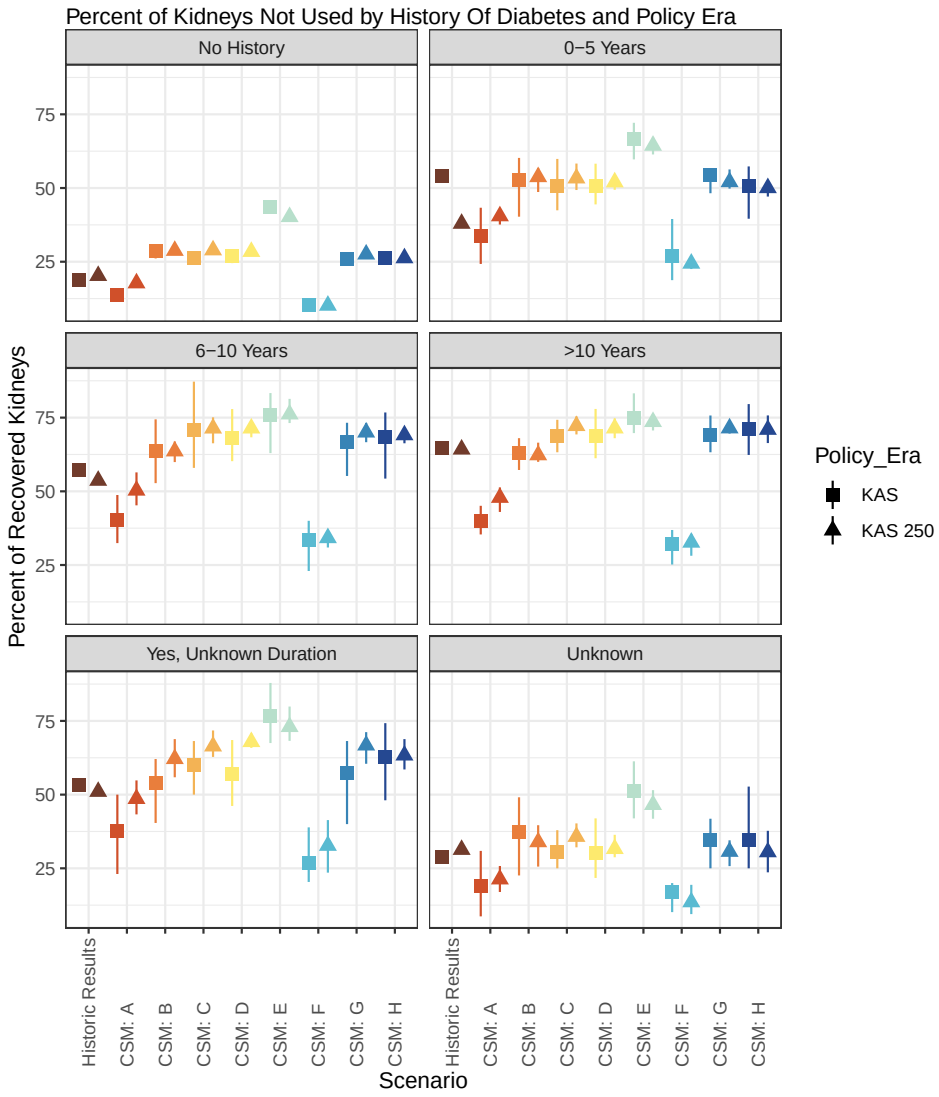

Figure 60: Percent of Kidneys Not Used by History Of Diabetes and Policy Era. Includes all kidneys that were recovered for transplant.

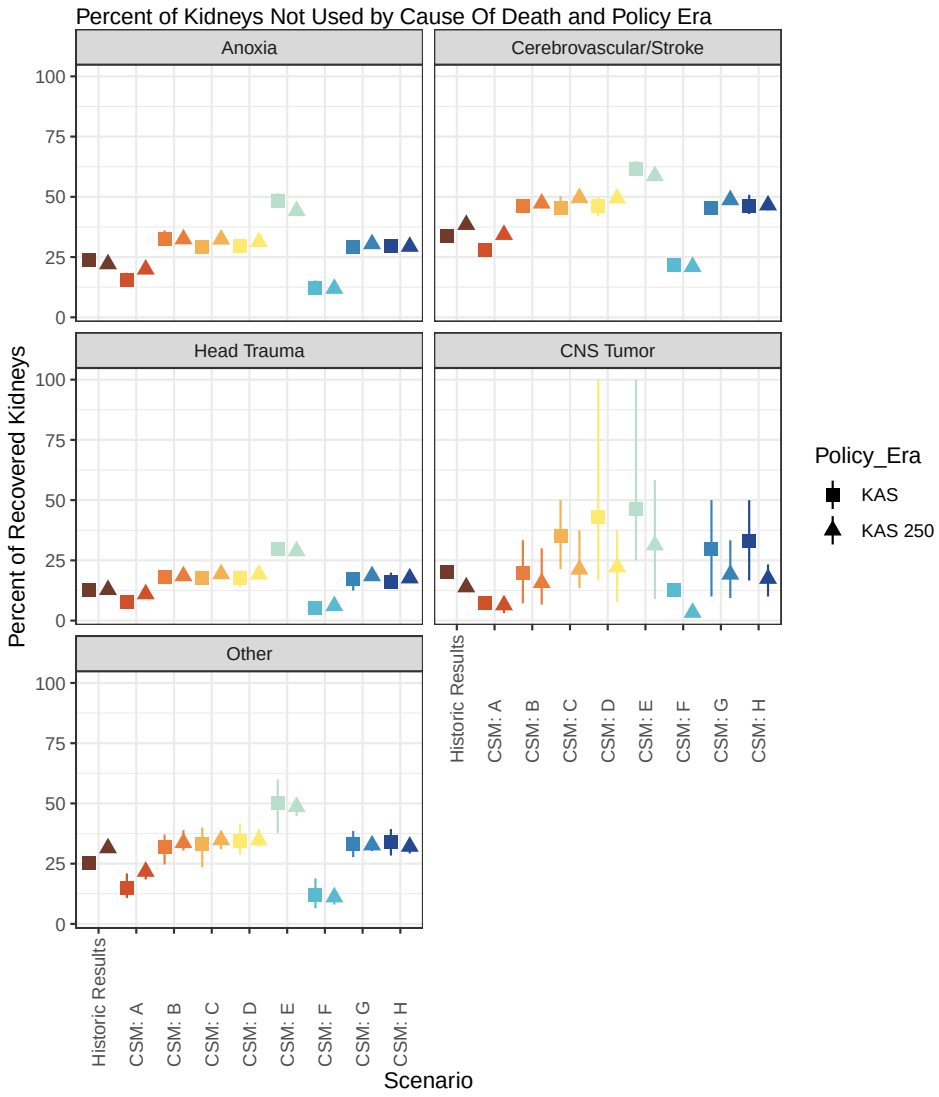

Figure 61: Percent of Kidneys Not Used by Cause Of Death and Policy Era. Includes all kidneys that were recovered for transplant.

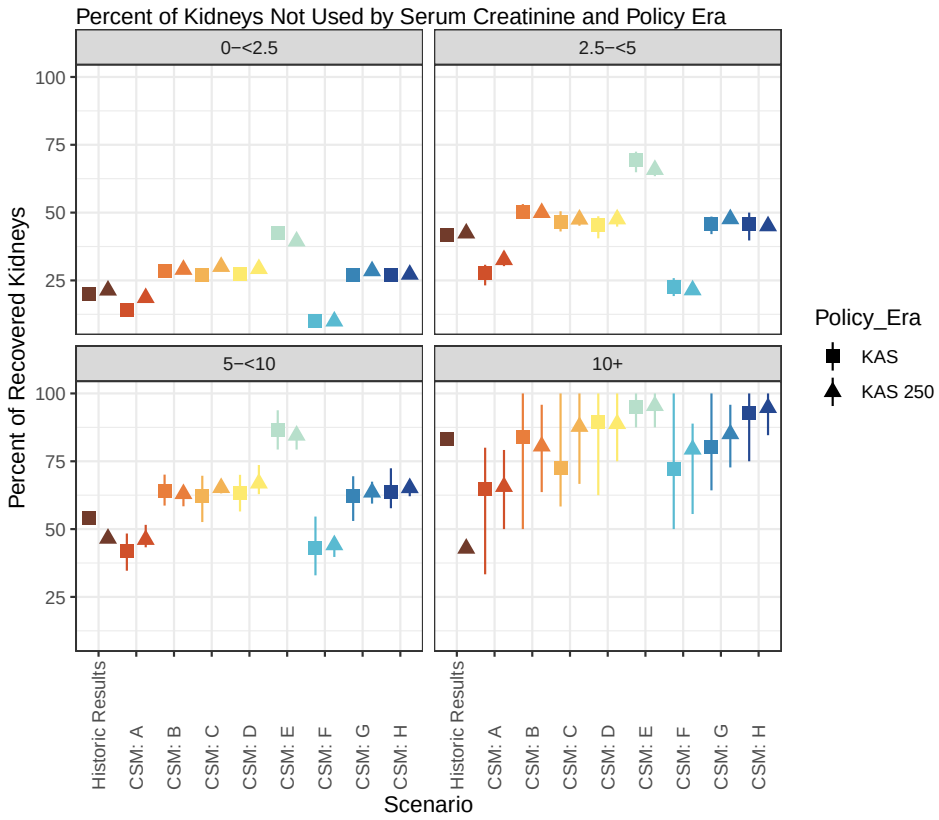

Figure 62: Percent of Kidneys Not Used by Serum Creatinine and Policy Era. Includes all kidneys that were recovered for transplant.

6.4.2 Pancreas

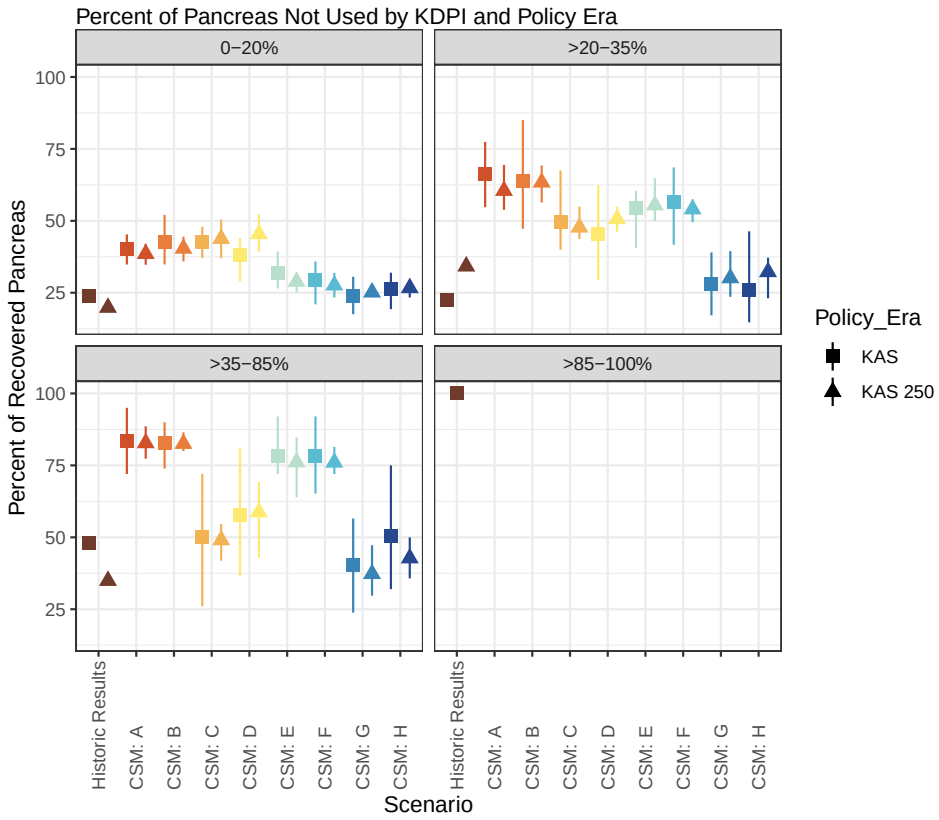

Figure 63: Percent of Pancreas Not Used by KDPI and Policy Era. Includes all pancreas that were recovered for transplant.

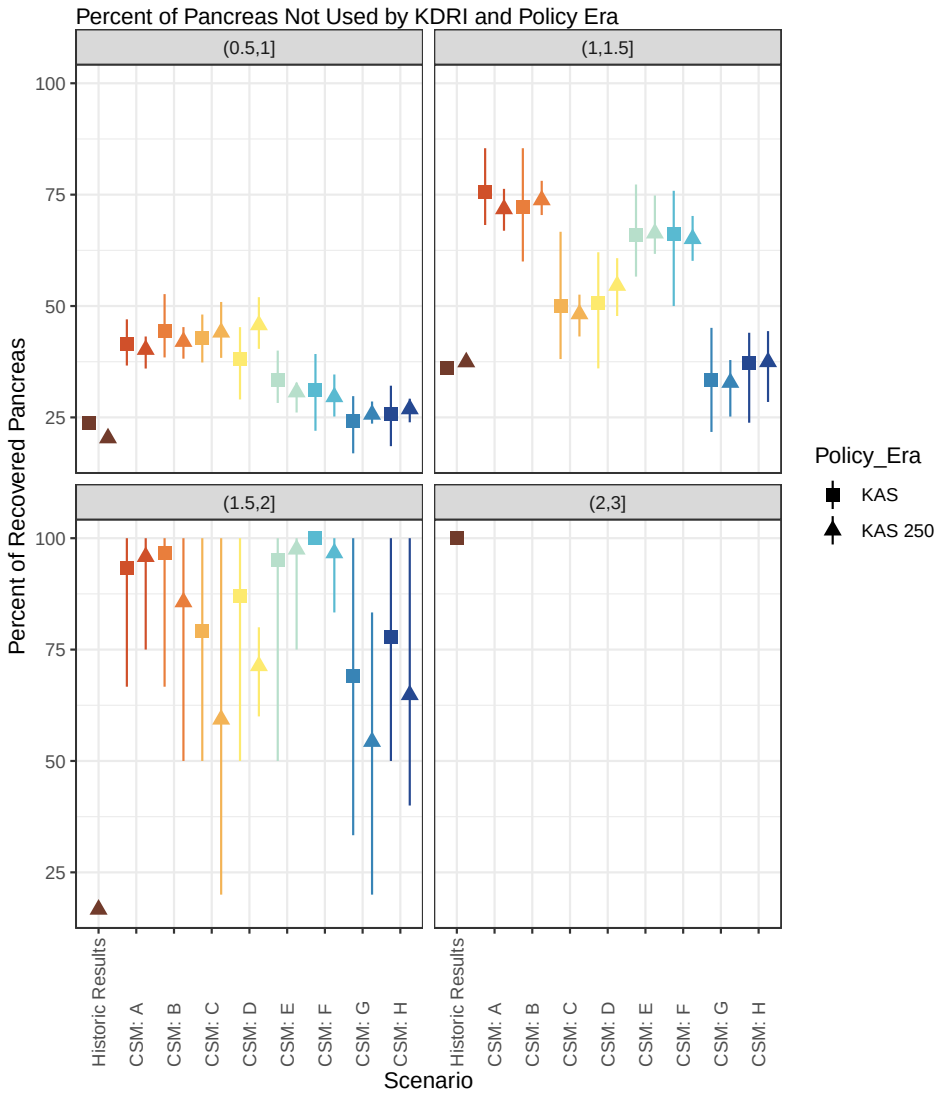

Figure 64: Percent of Pancreas Not Used by KDRI and Policy Era. Includes all pancreas that were recovered for transplant.

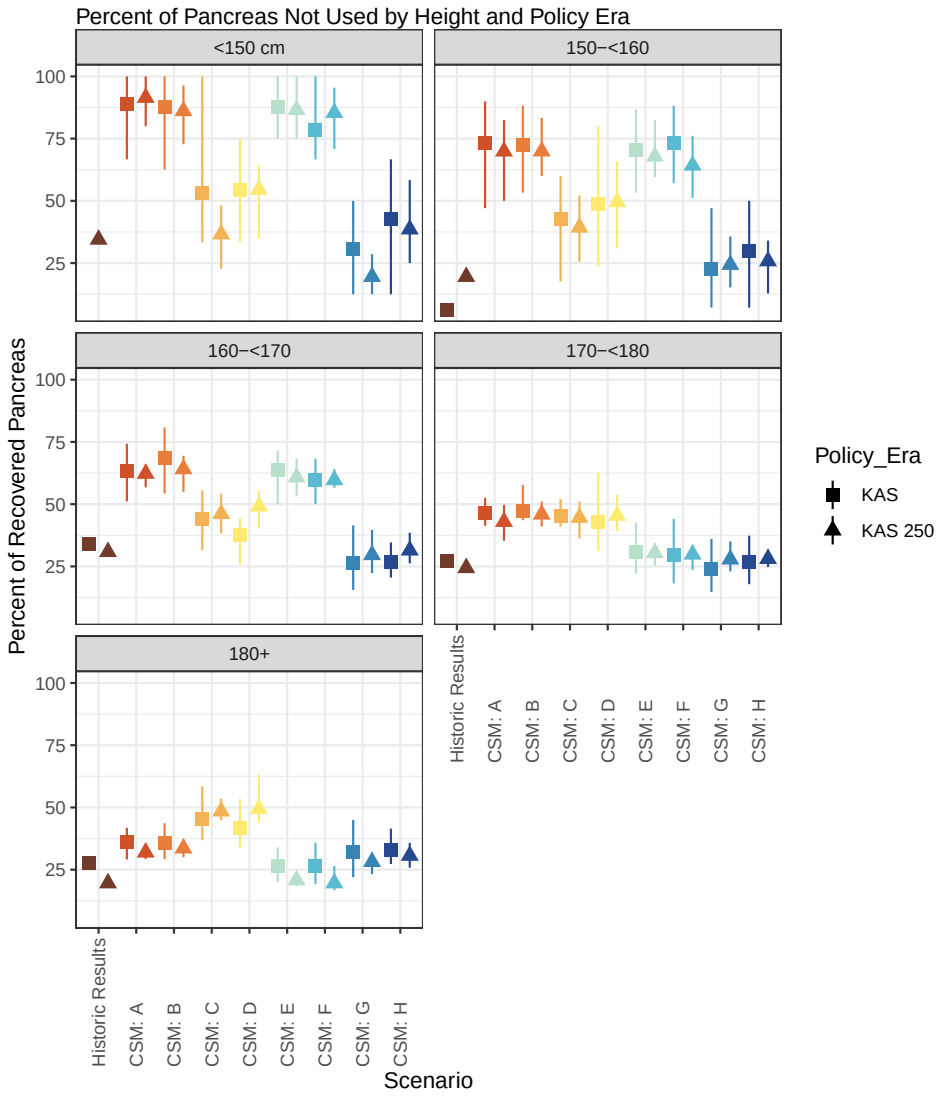

Figure 65: Percent of Pancreas Not Used by Height and Policy Era. Includes all pancreas that were recovered for transplant.

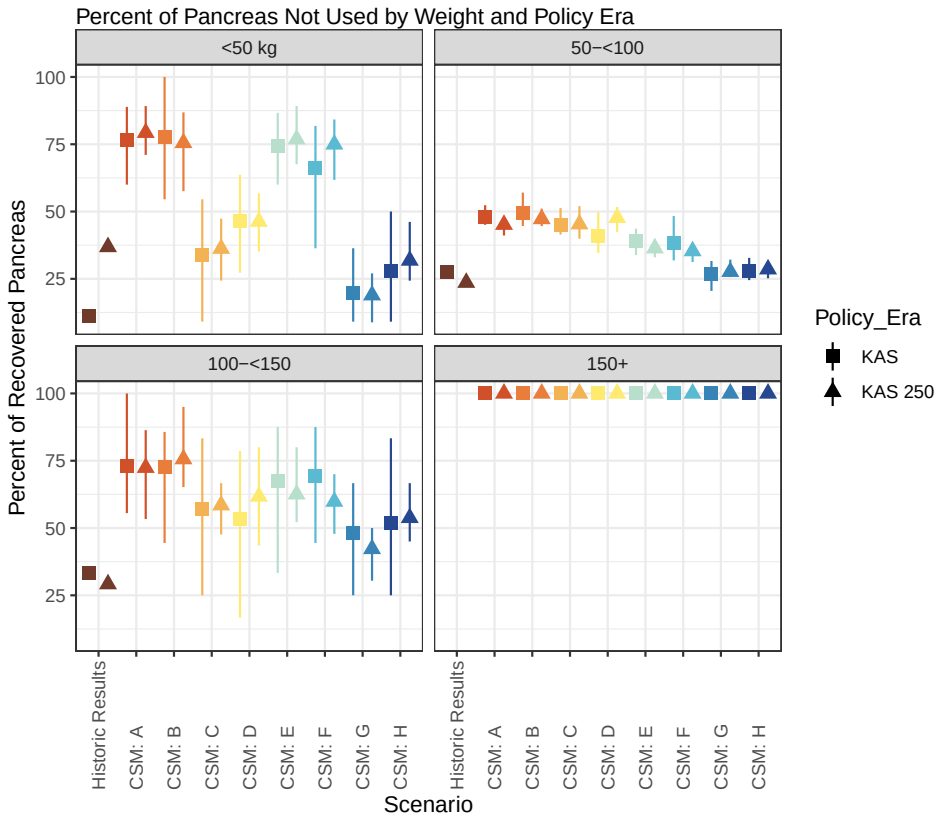

Figure 66: Percent of Pancreas Not Used by Weight and Policy Era. Includes all pancreas that were recovered for transplant.

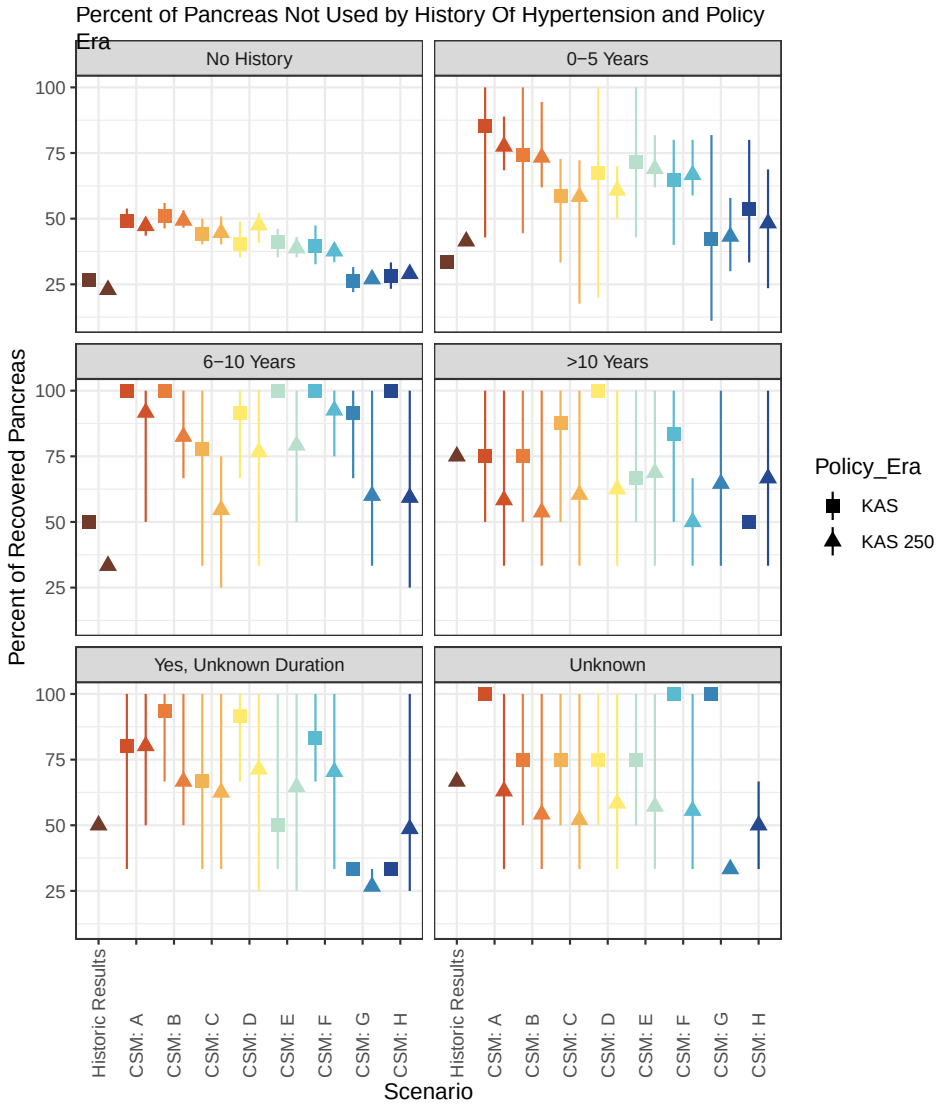

Figure 67: Percent of Pancreas Not Used by History Of Hypertension and Policy Era. Includes all pancreas that were recovered for transplant.

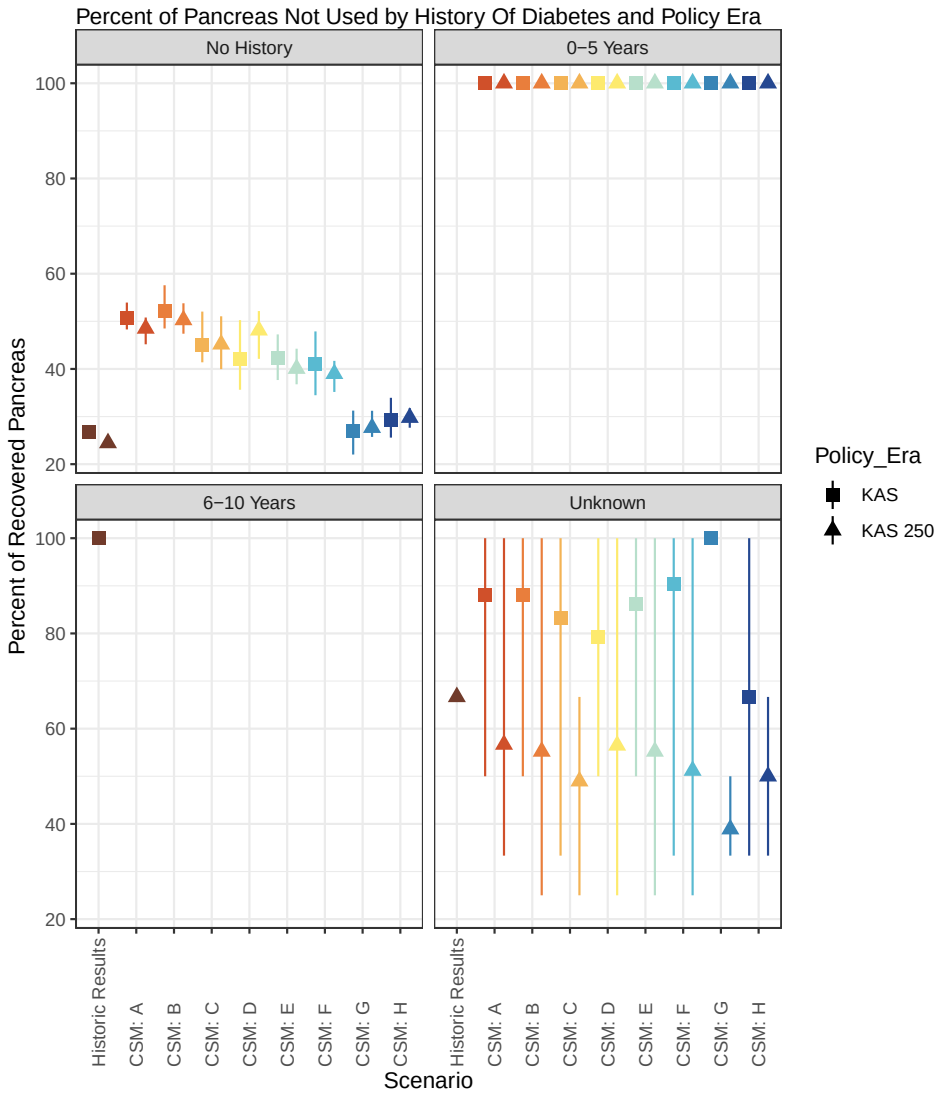

Figure 68: Percent of Pancreas Not Used by History Of Diabetes and Policy Era. Includes all pancreas that were recovered for transplant.

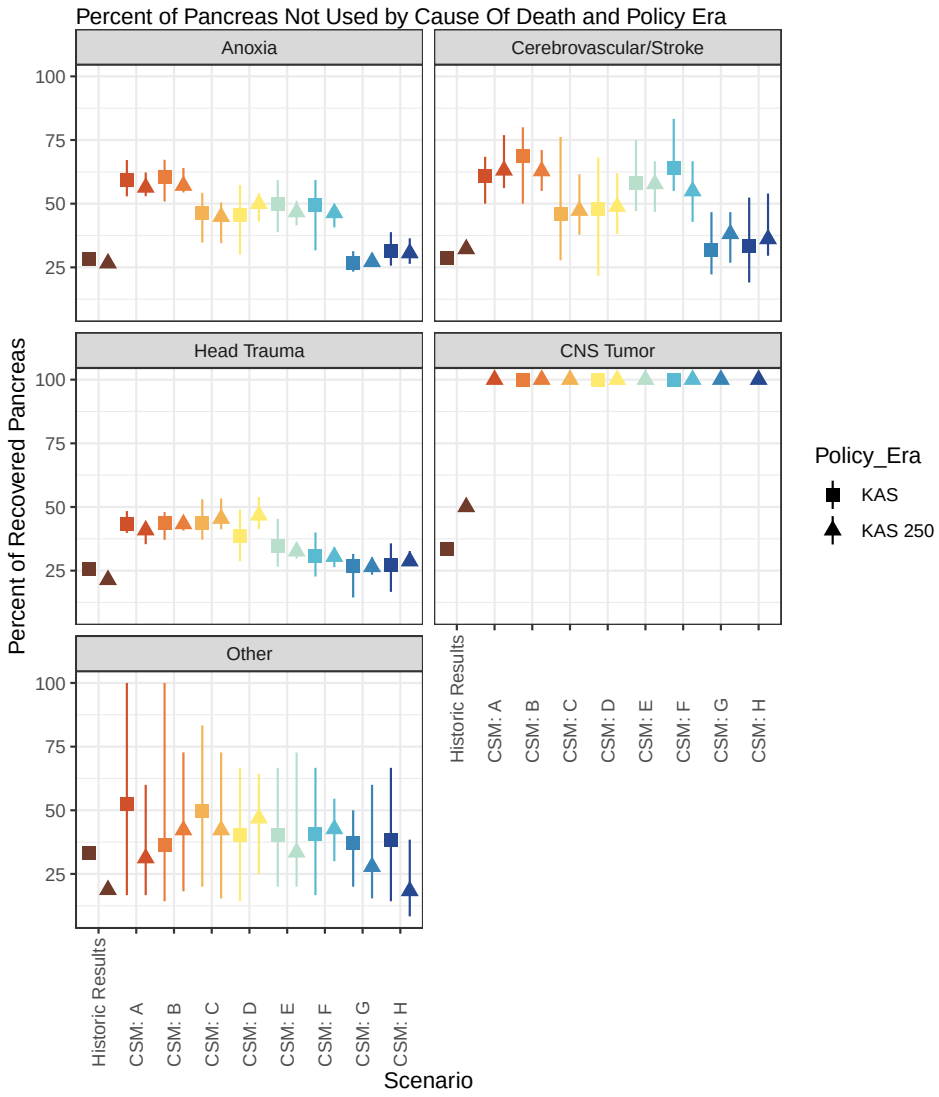

Figure 69: Percent of Pancreas Not Used by Cause Of Death and Policy Era. Includes all pancreas that were recovered for transplant.

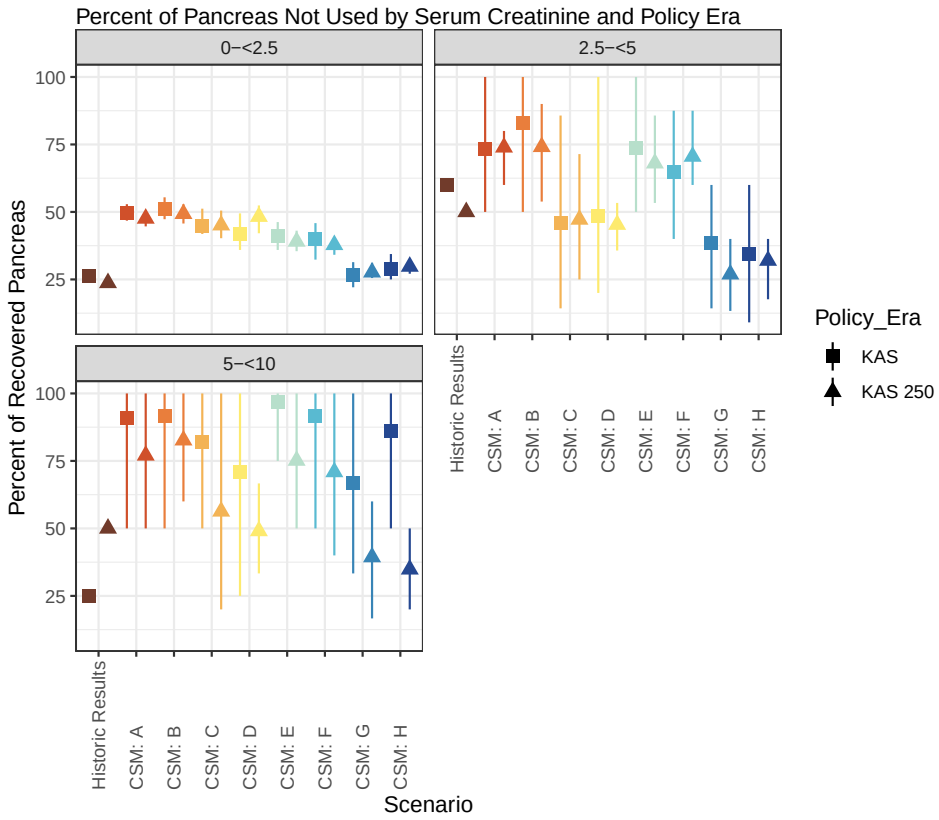

Figure 70: Percent of Pancreas Not Used by Serum Creatinine and Policy Era. Includes all pancreas that were recovered for transplant.

6.5 Secondary Assessment Metrics Based on Research Questions

6.5.1 Kidney Unadjusted Transplant Rates and Cumulative Incidence

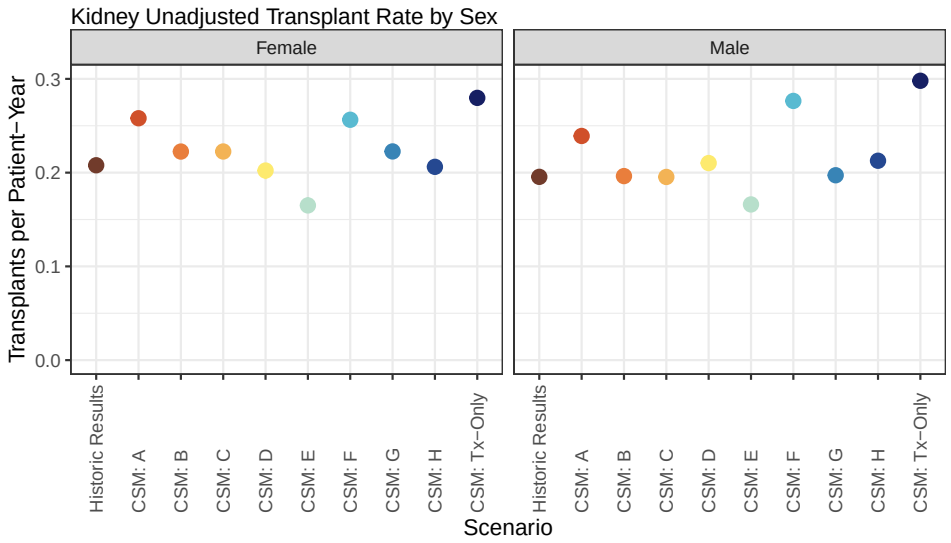

Figure 71: Kidney Unadjusted Transplant Rate by Sex. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

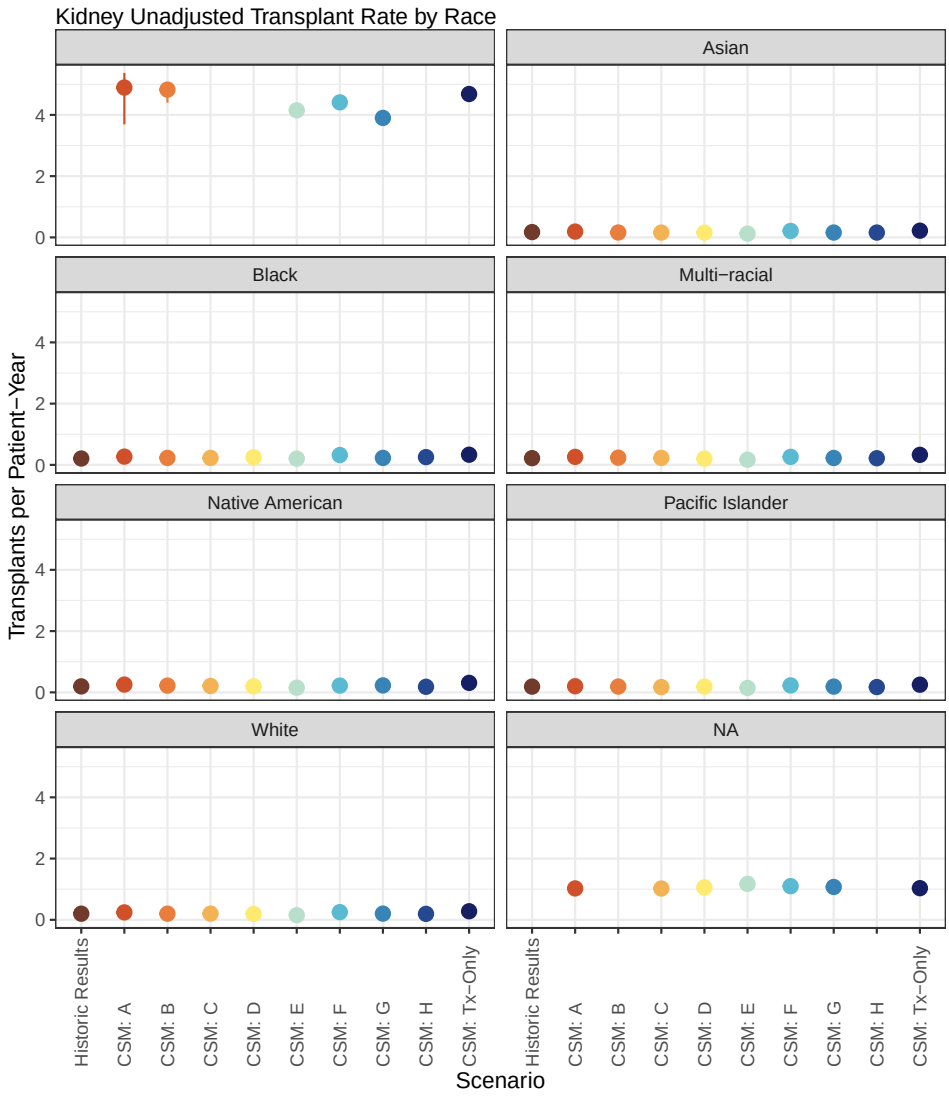

Figure 72: Kidney Unadjusted Transplant Rate by Race. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

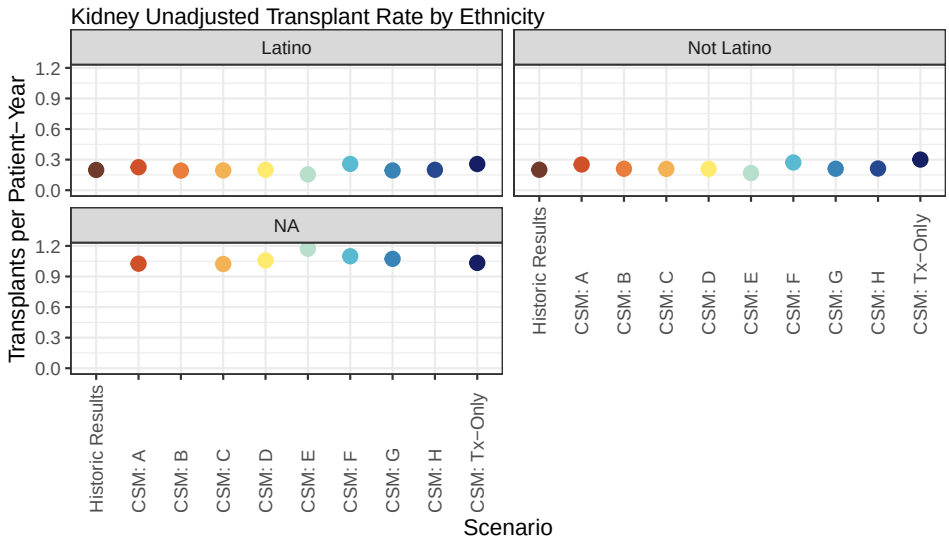

Figure 73: Kidney Unadjusted Transplant Rate by Ethnicity. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

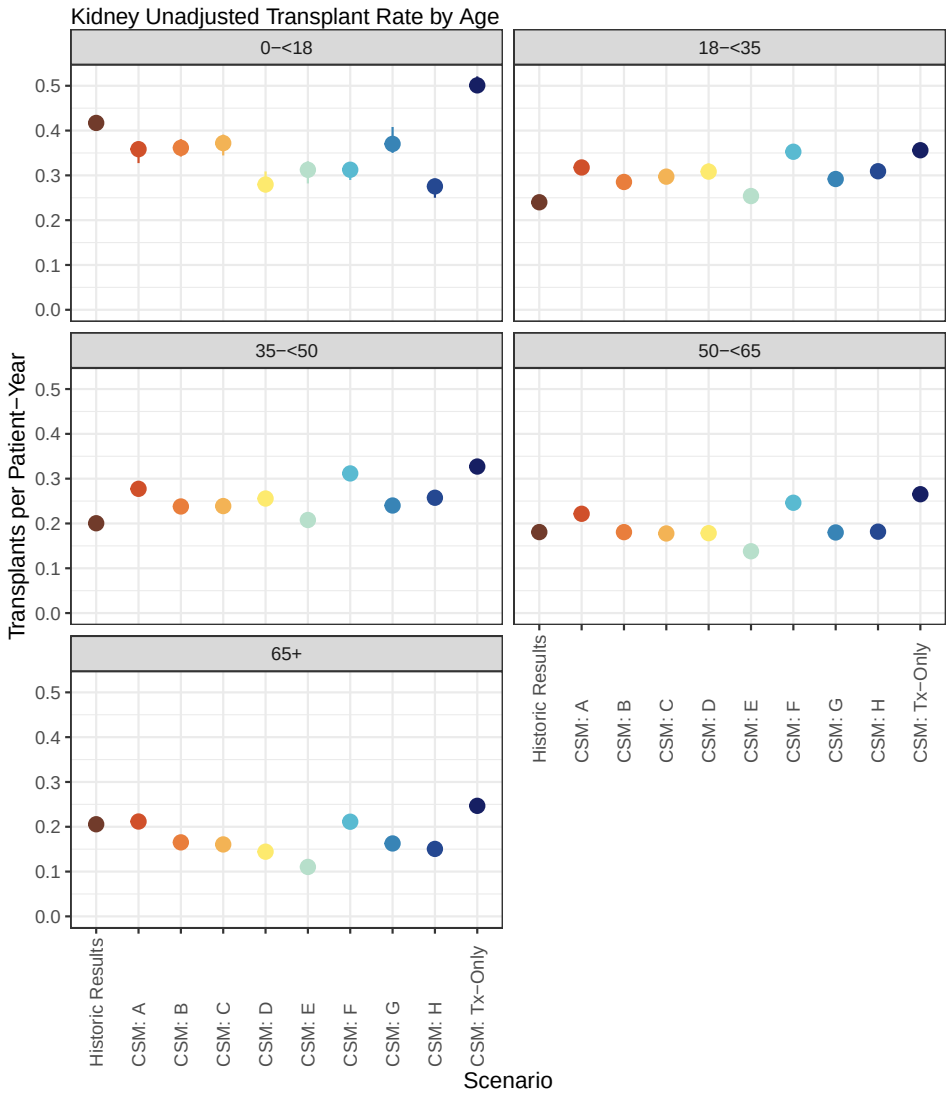

Figure 74: Kidney Unadjusted Transplant Rate by Age. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

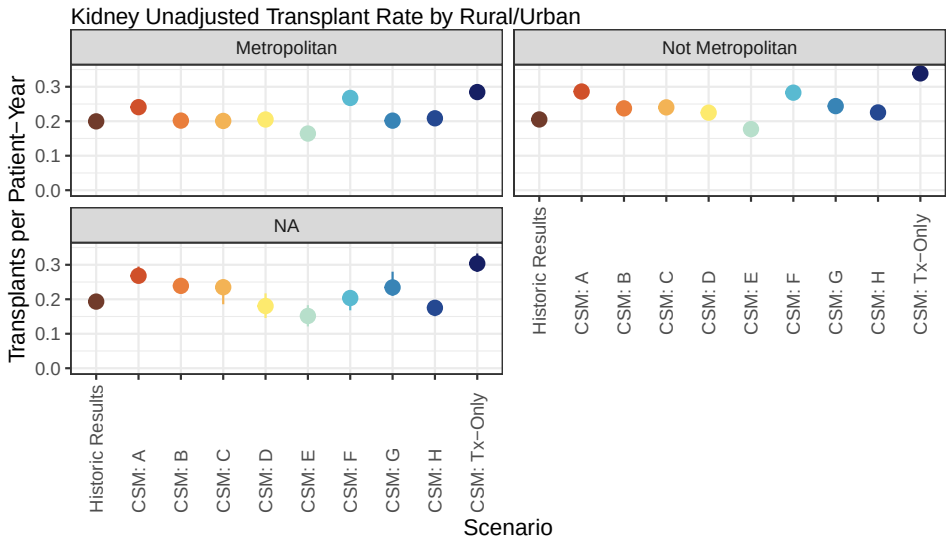

Figure 75: Kidney Unadjusted Transplant Rate by Rural/Urban. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

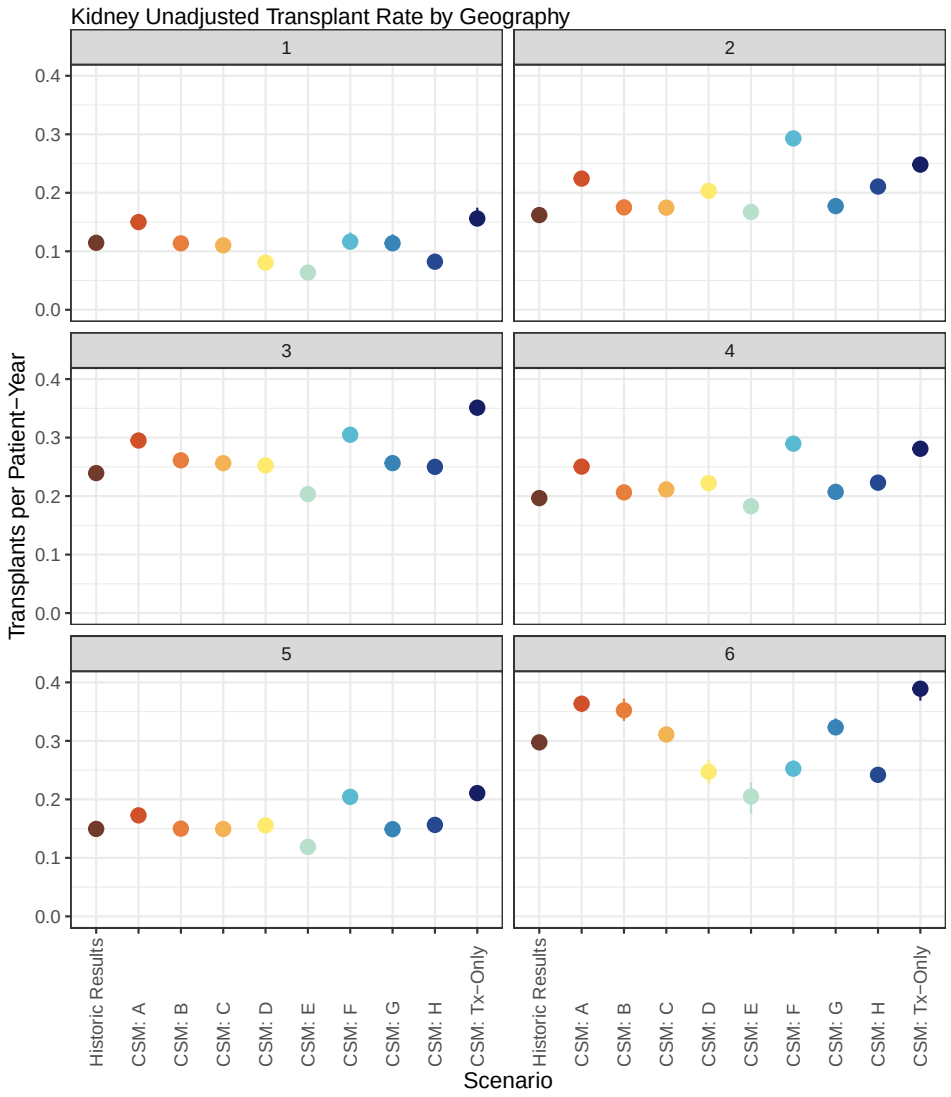

Figure 76: Kidney Unadjusted Transplant Rate by Geography. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

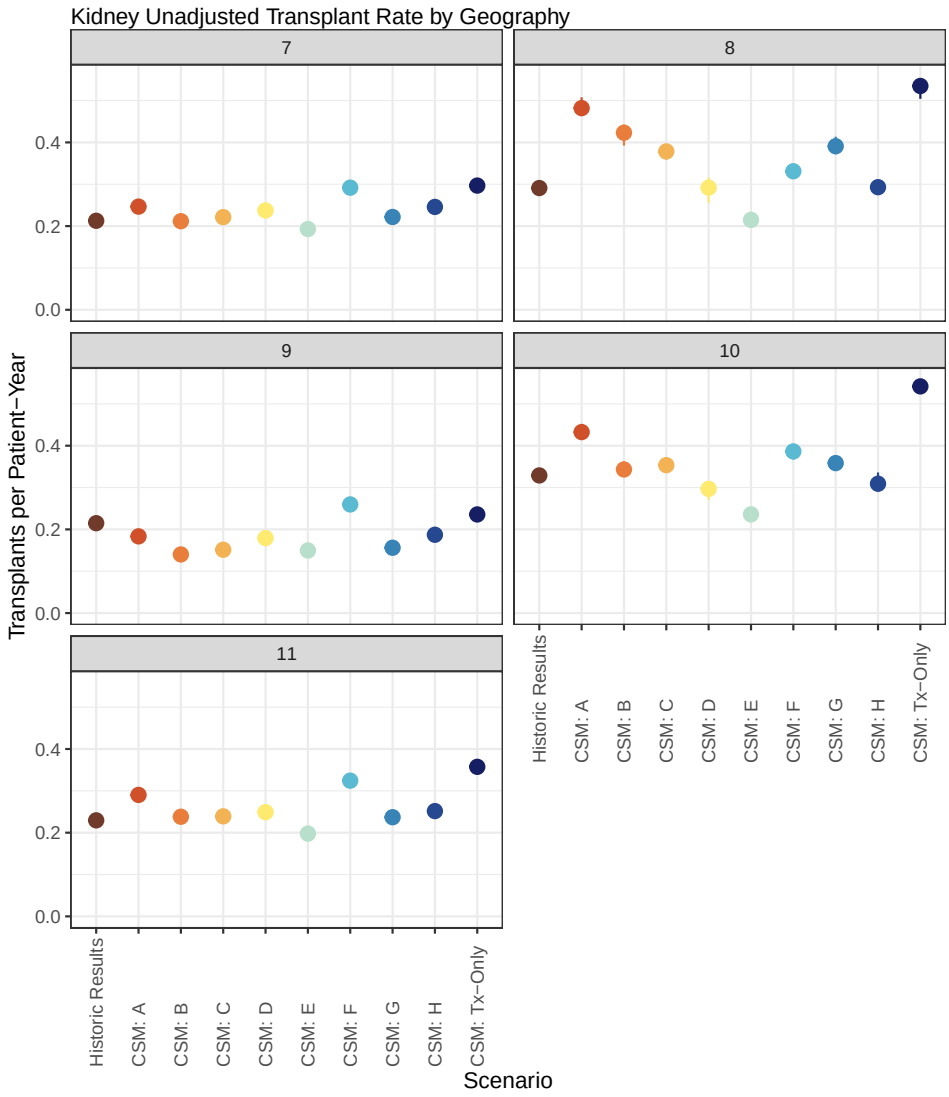

Figure 77: Kidney Unadjusted Transplant Rate by Geography. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

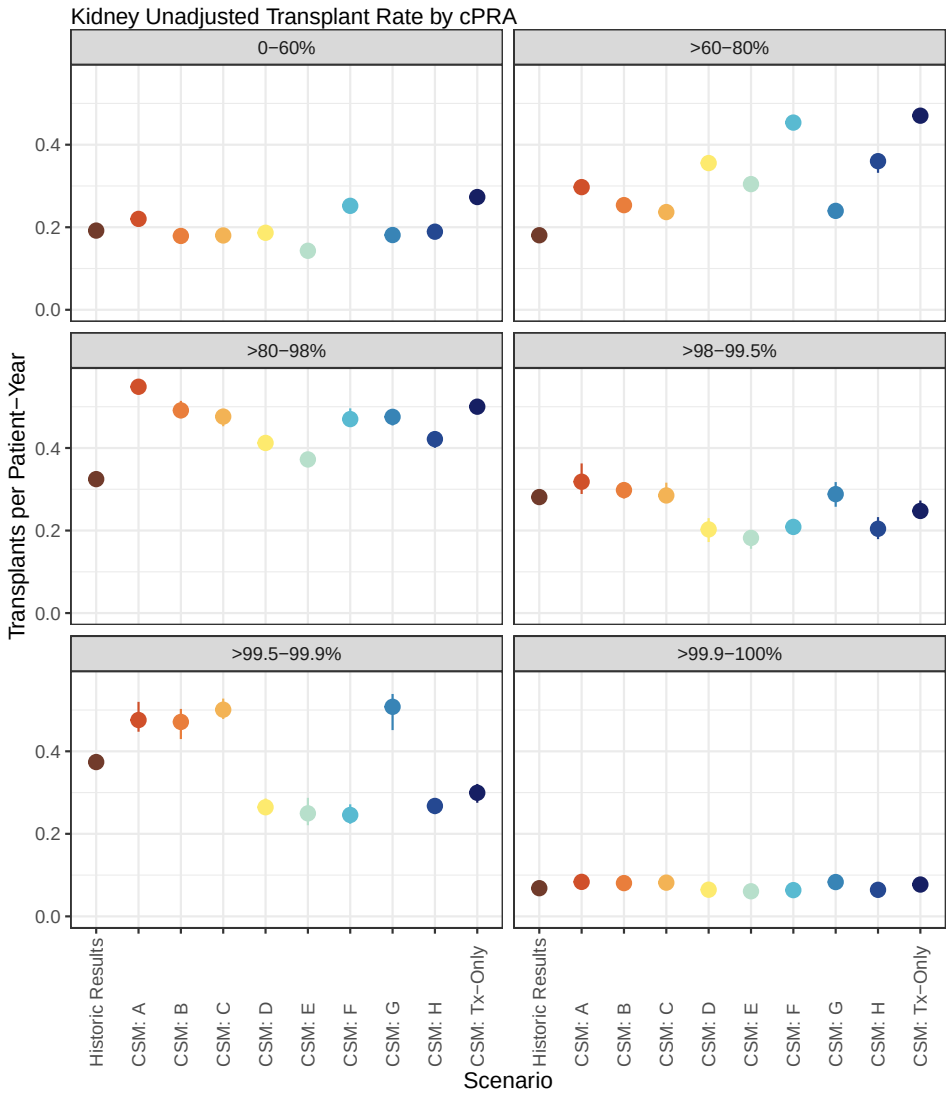

Figure 78: Kidney Unadjusted Transplant Rate by cPRA. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

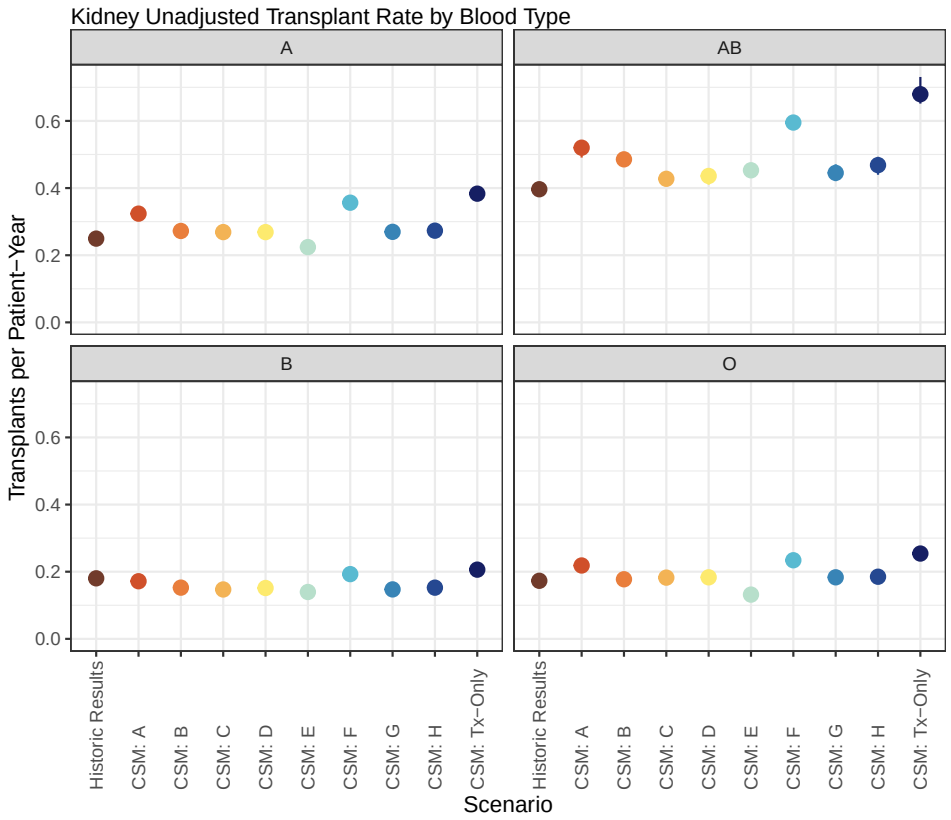

Figure 79: Kidney Unadjusted Transplant Rate by Blood Type. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

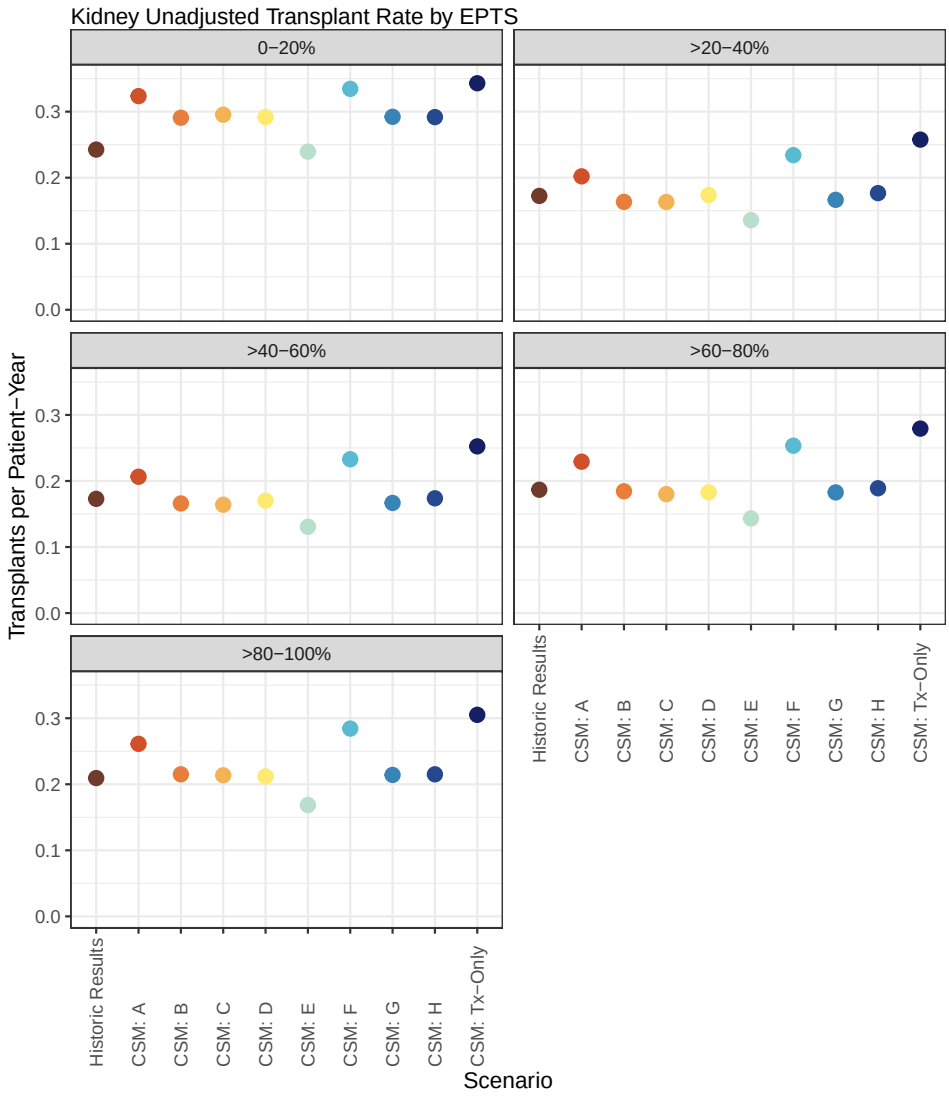

Figure 80: Kidney Unadjusted Transplant Rate by EPTS. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

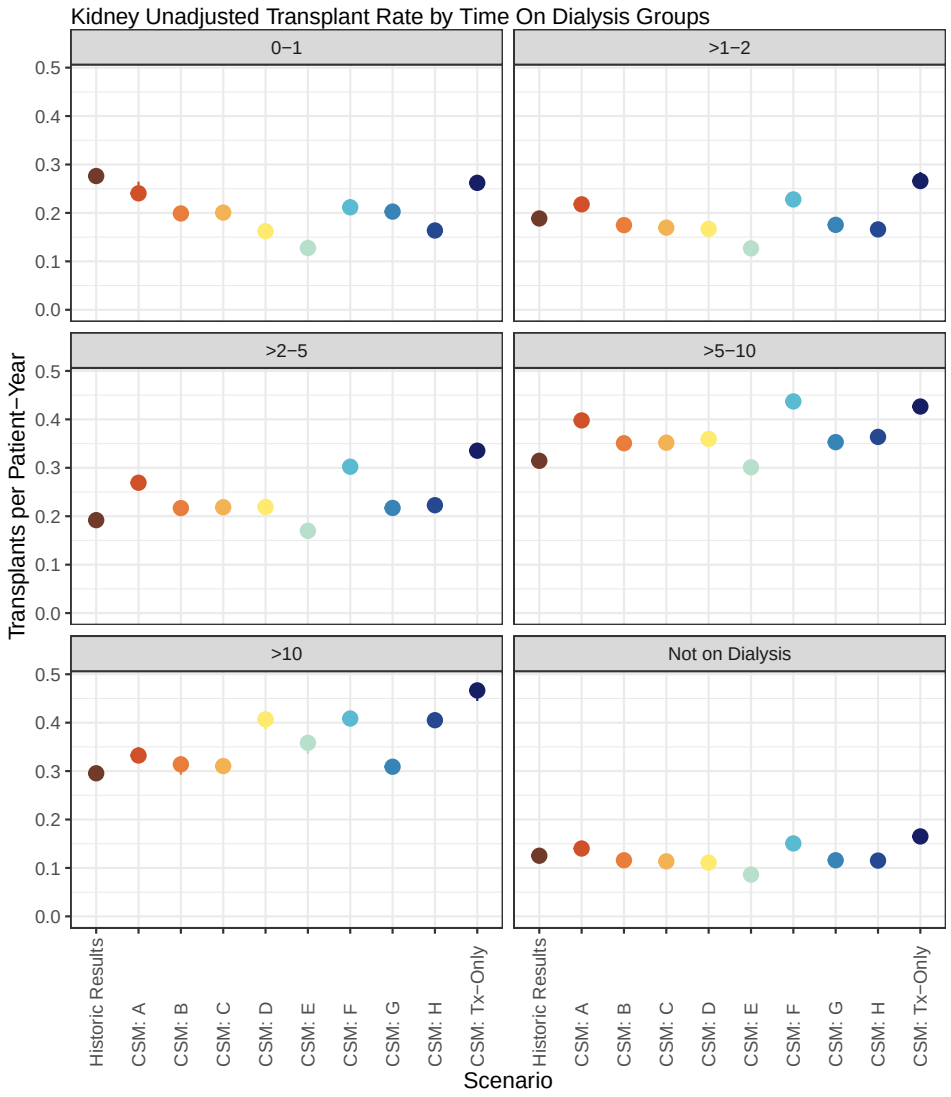

Figure 81: Kidney Unadjusted Transplant Rate by Time On Dialysis Groups. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

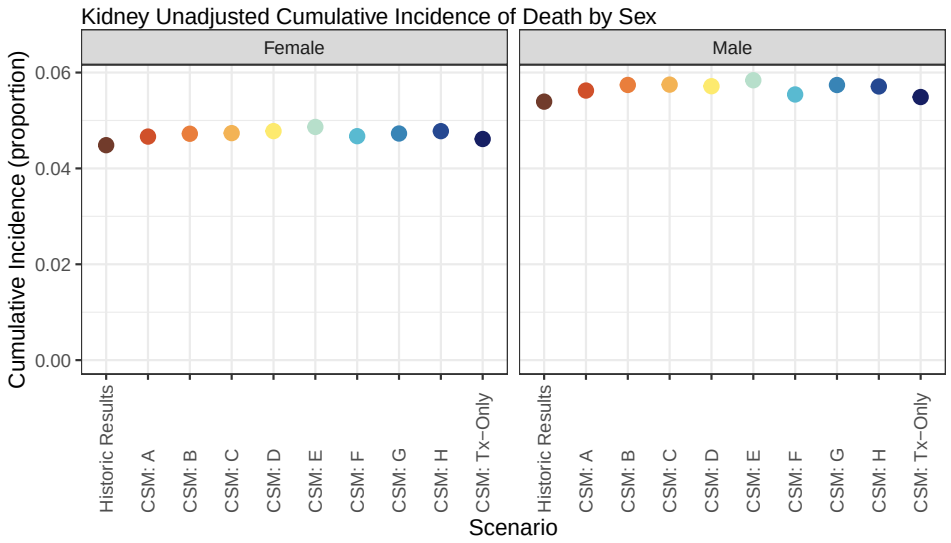

Figure 82: Kidney Unadjusted Cumulative Incidence of Death by Sex. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

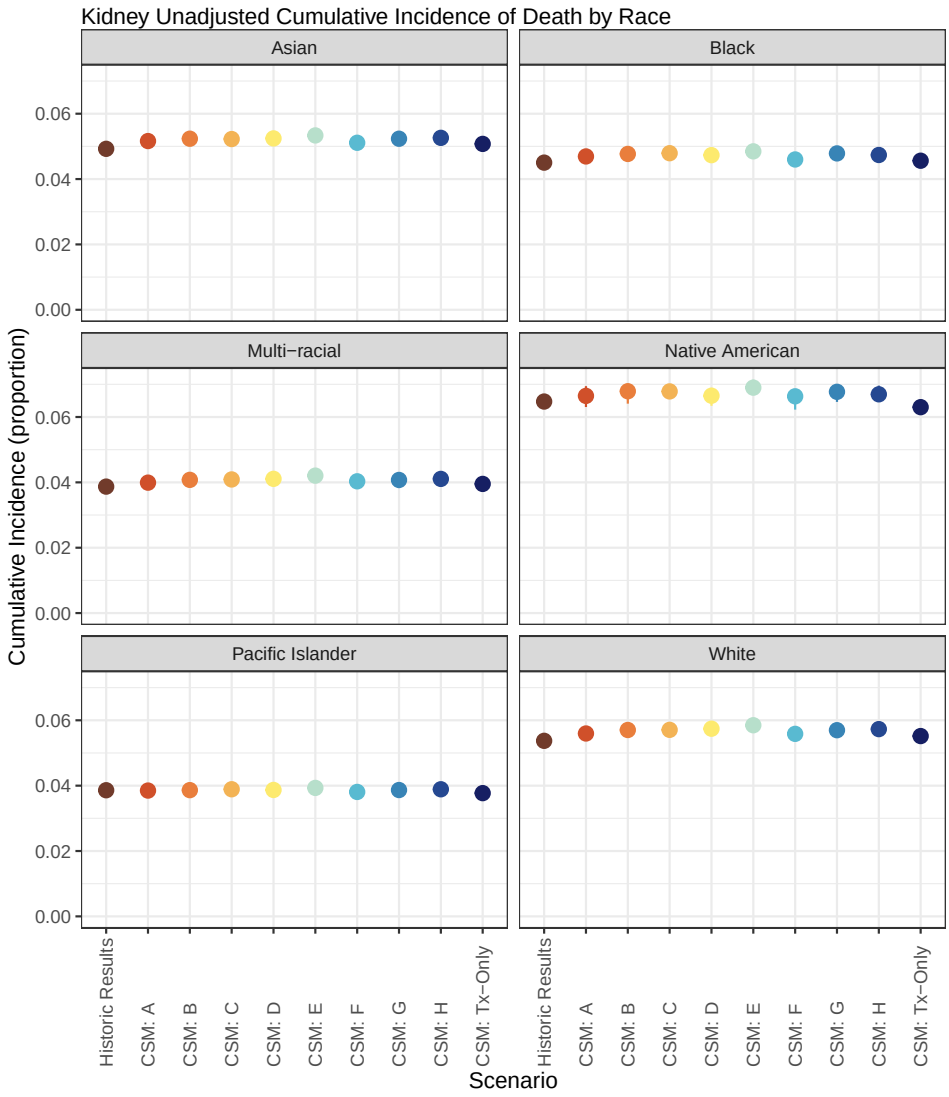

Figure 83: Kidney Unadjusted Cumulative Incidence of Death by Race. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

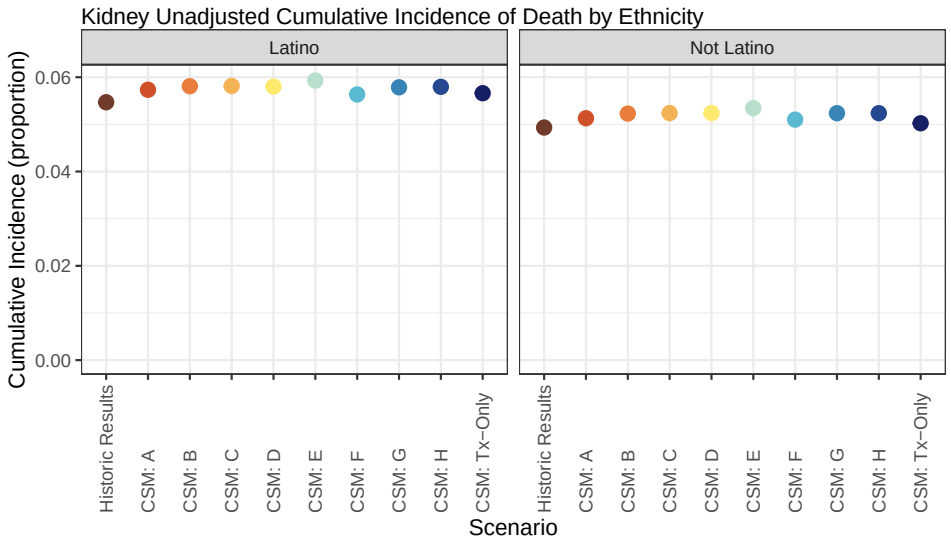

Figure 84: Kidney Unadjusted Cumulative Incidence of Death by Ethnicity. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

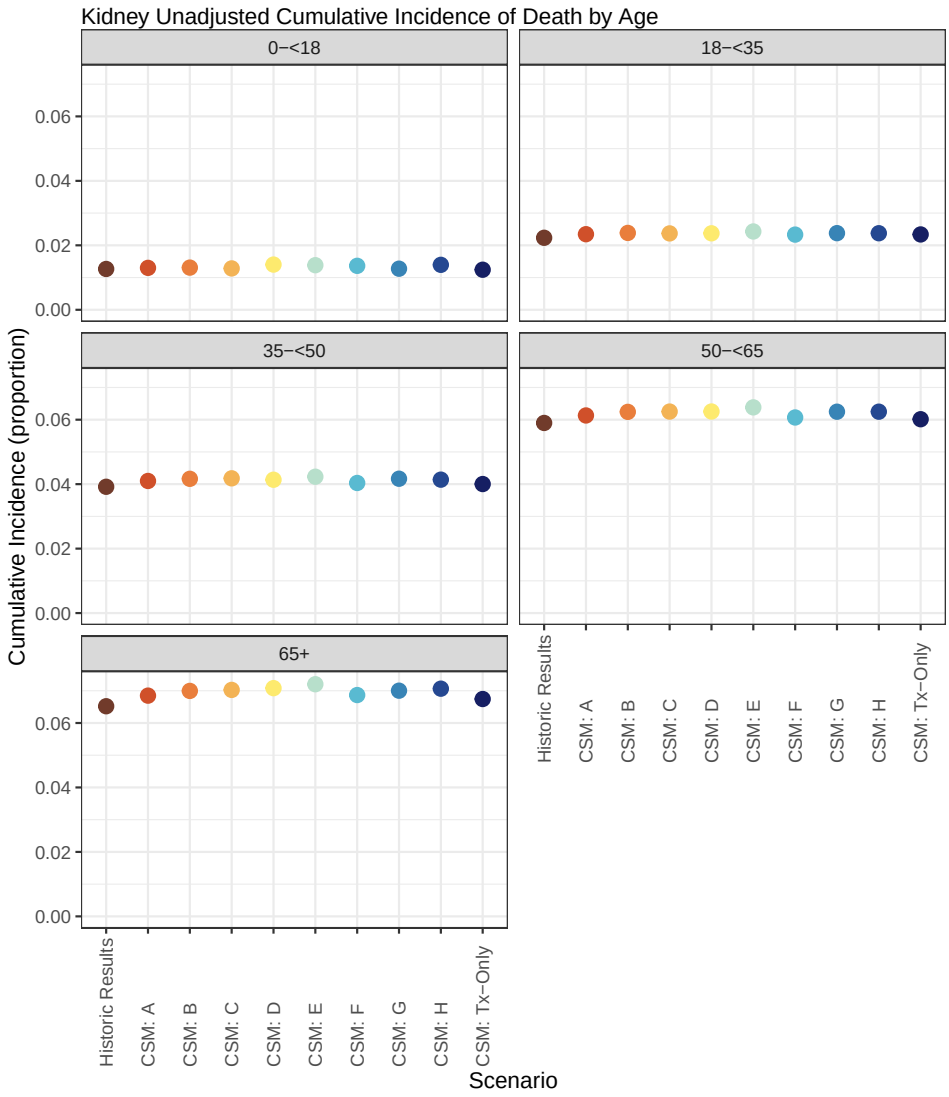

Figure 85: Kidney Unadjusted Cumulative Incidence of Death by Age. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

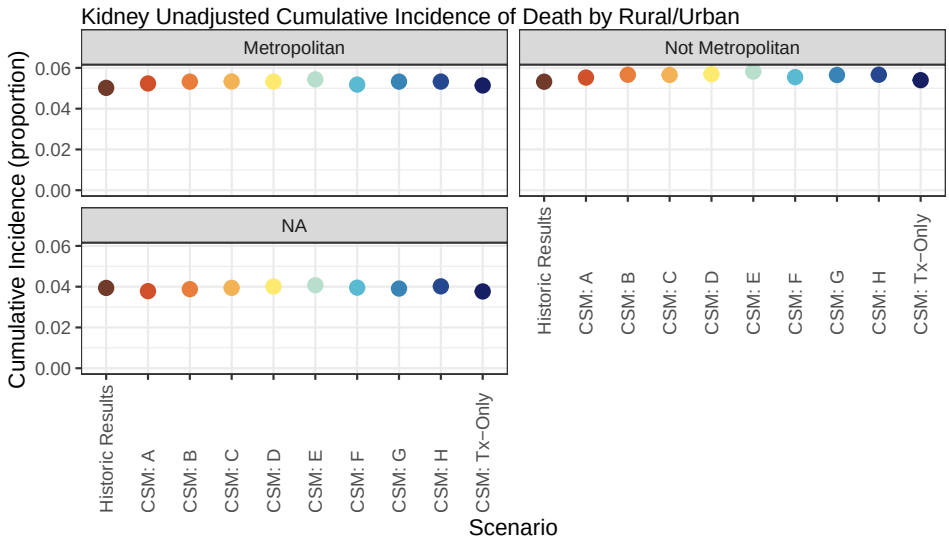

Figure 86: Kidney Unadjusted Cumulative Incidence of Death by Rural/Urban. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

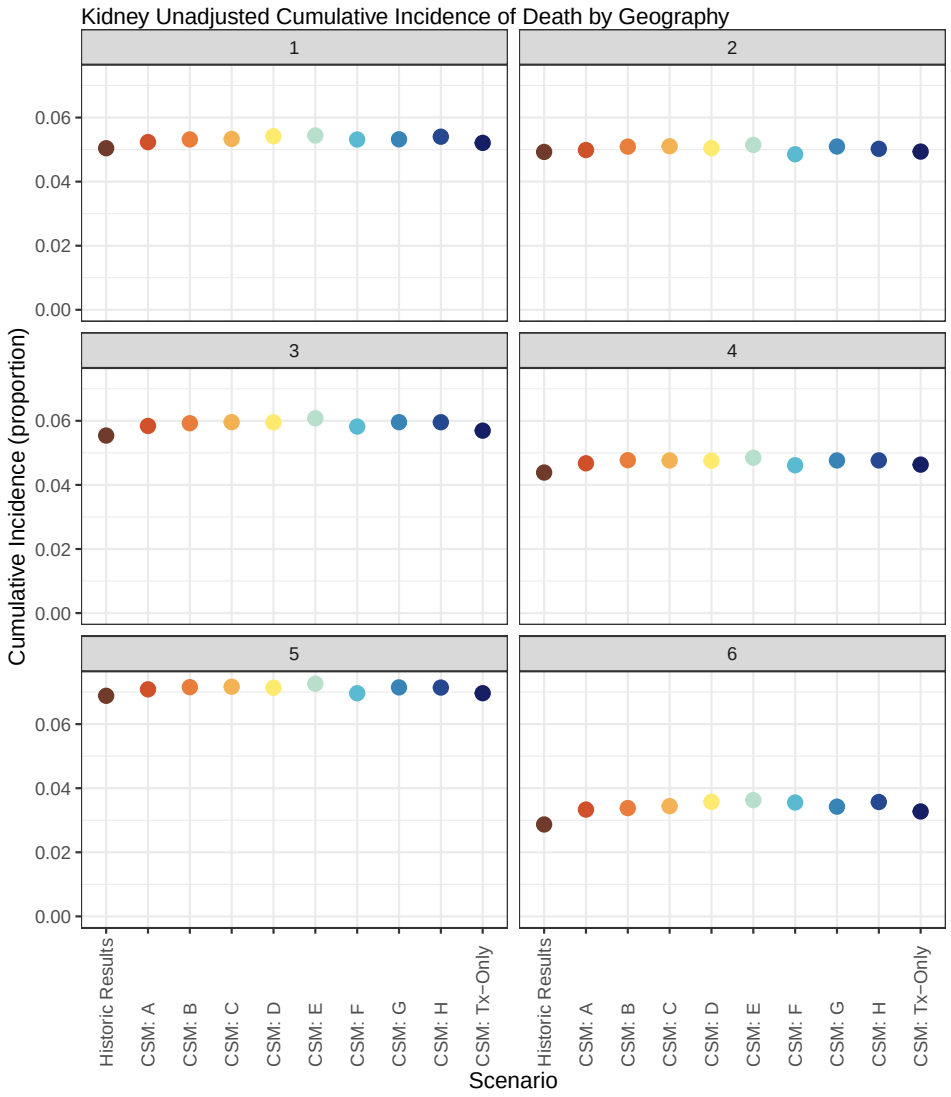

Figure 87: Kidney Unadjusted Cumulative Incidence of Death by Geography. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

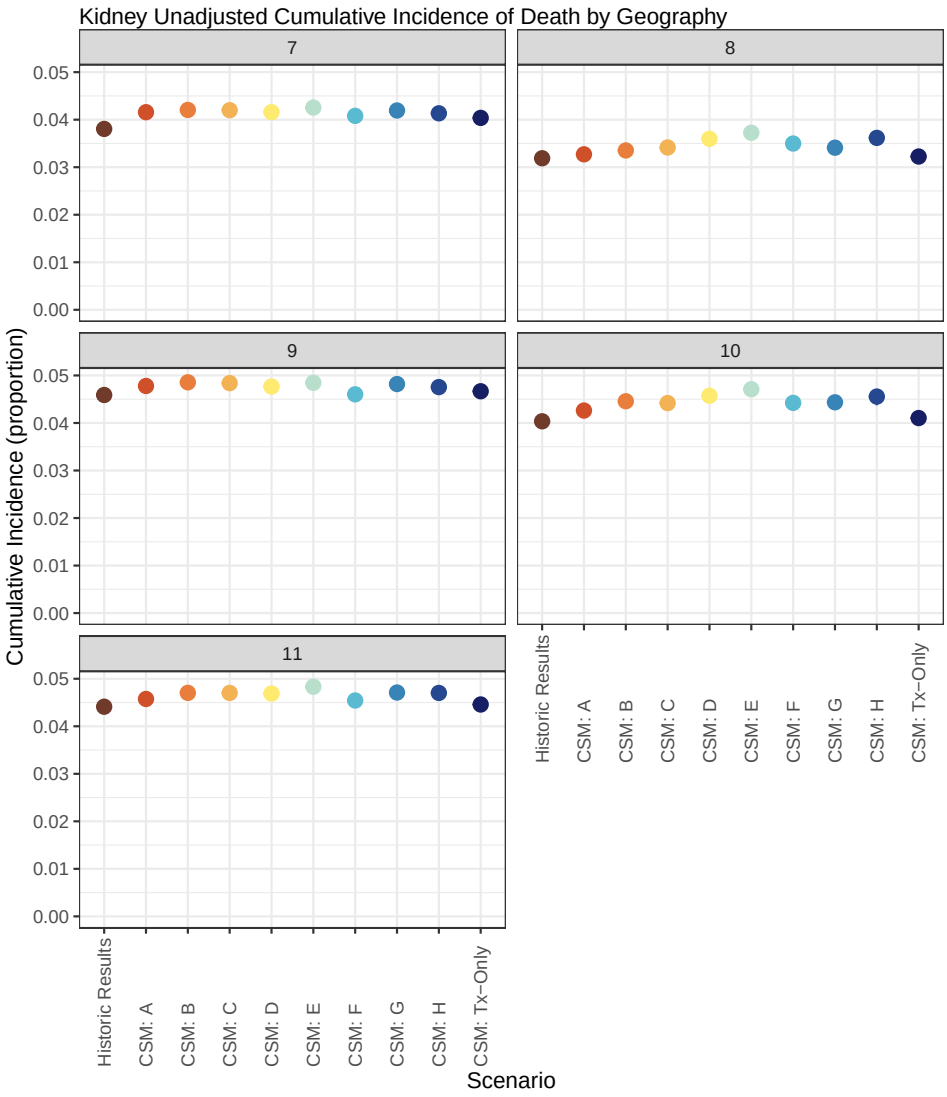

Figure 88: Kidney Unadjusted Cumulative Incidence of Death by Geography. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

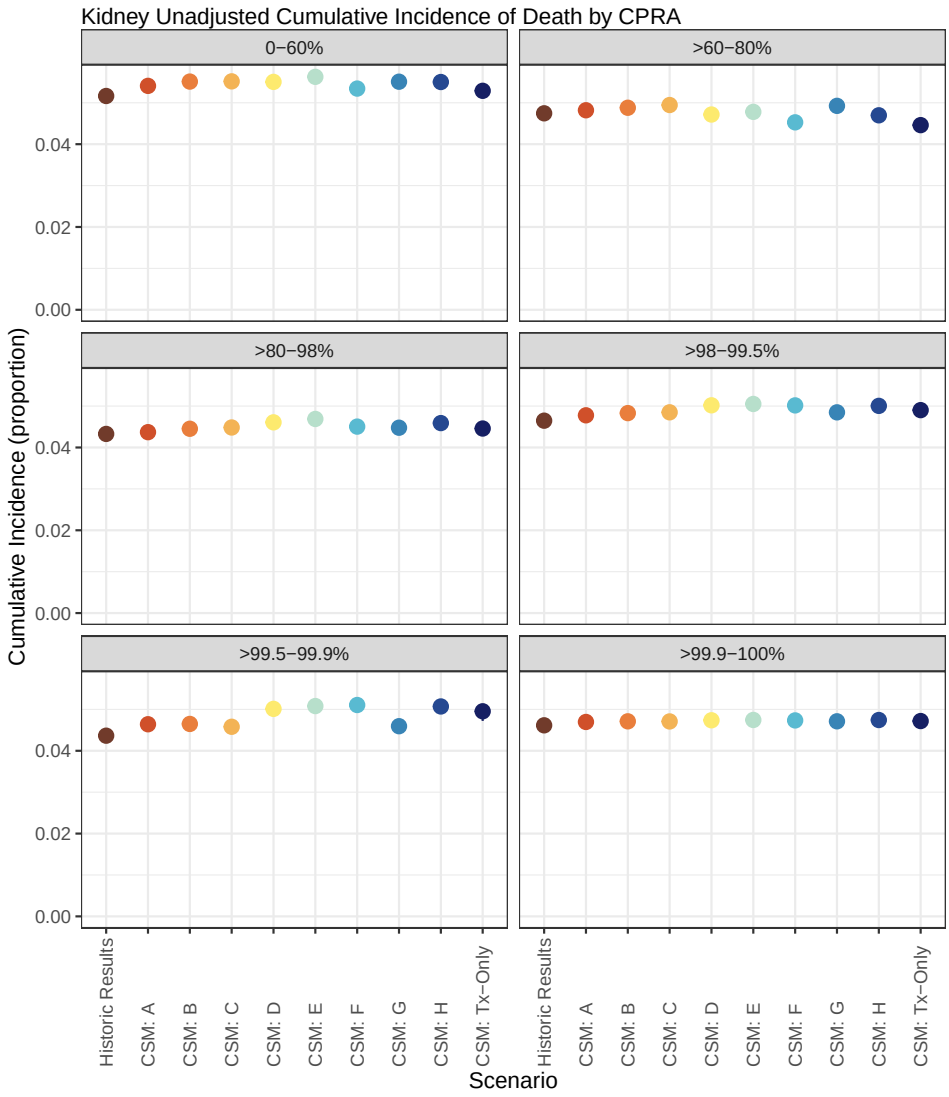

Figure 89: Kidney Unadjusted Cumulative Incidence of Death by CPRA. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

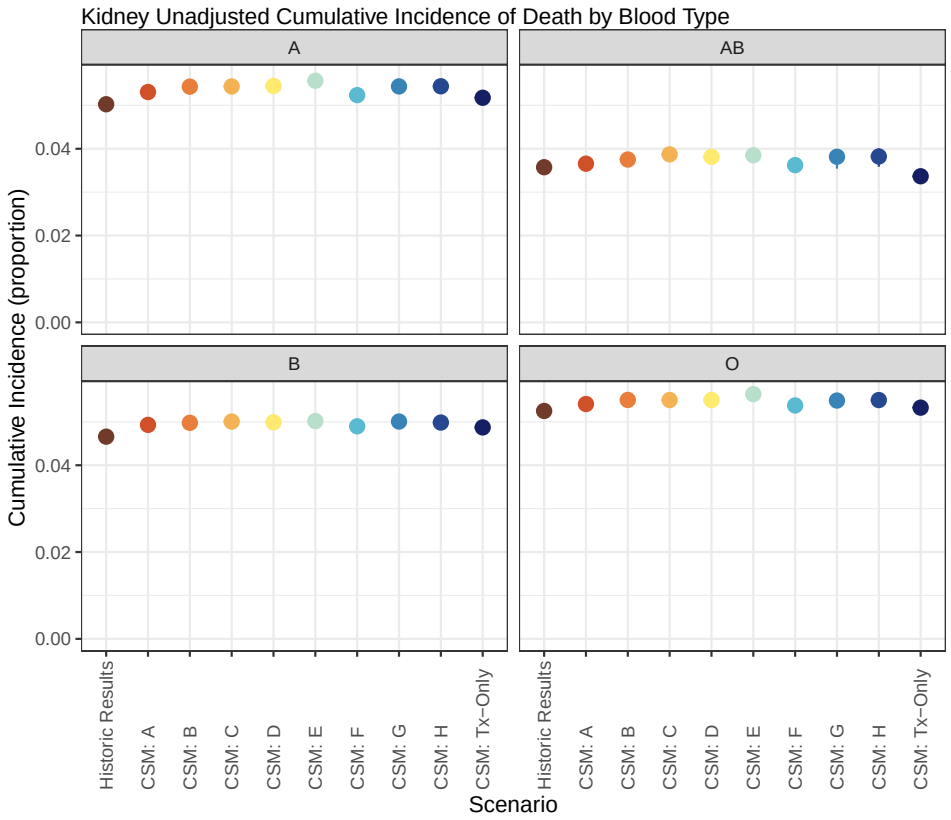

Figure 90: Kidney Unadjusted Cumulative Incidence of Death by Blood Type. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

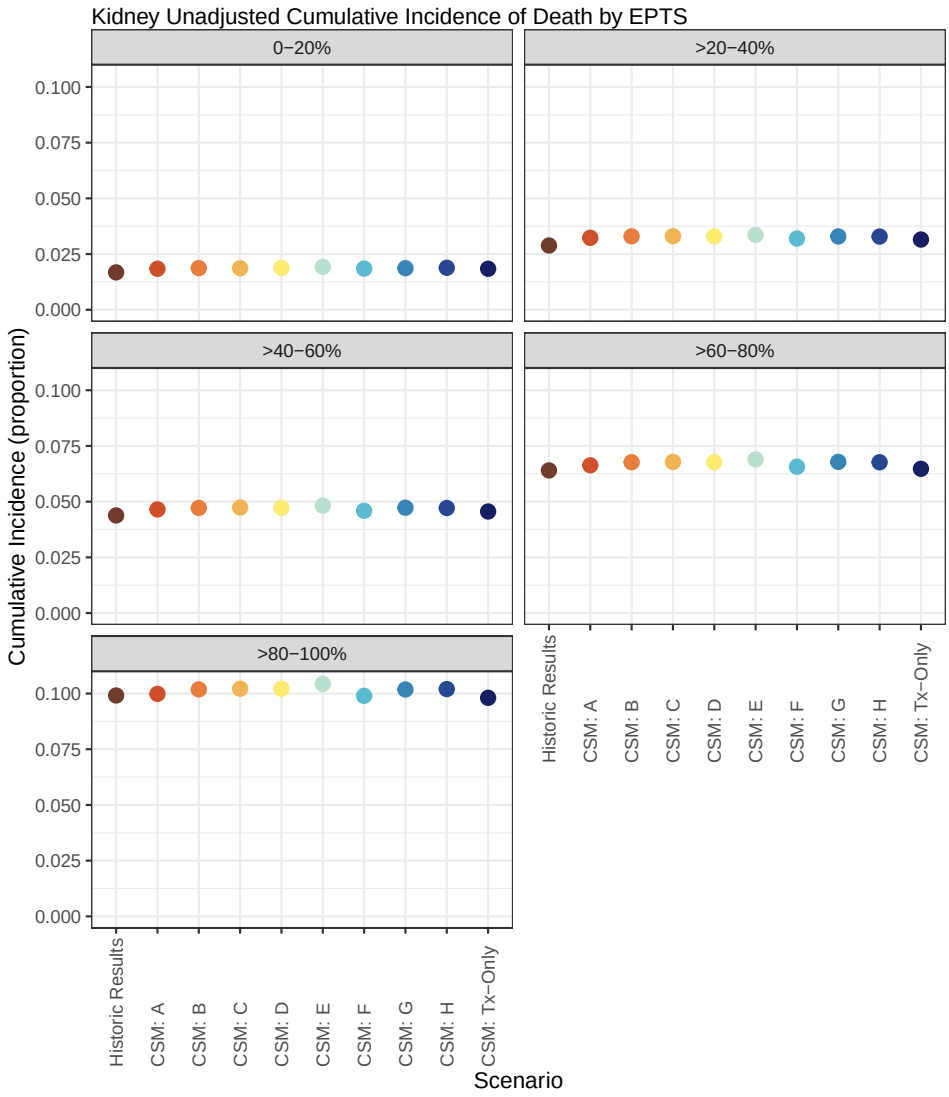

Figure 91: Kidney Unadjusted Cumulative Incidence of Death by EPTS. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

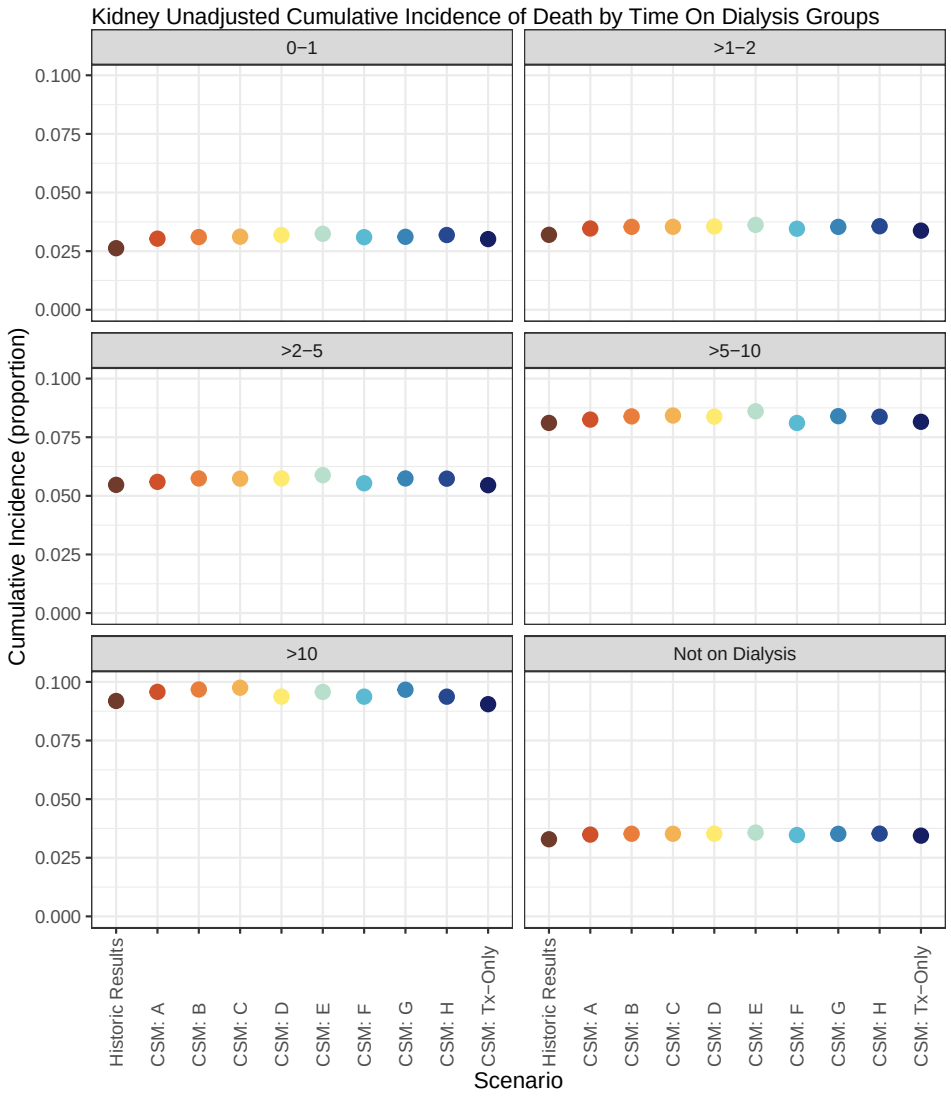

Figure 92: Kidney Unadjusted Cumulative Incidence of Death by Time On Dialysis Groups. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

6.5.2 Kidney Match-Run Results

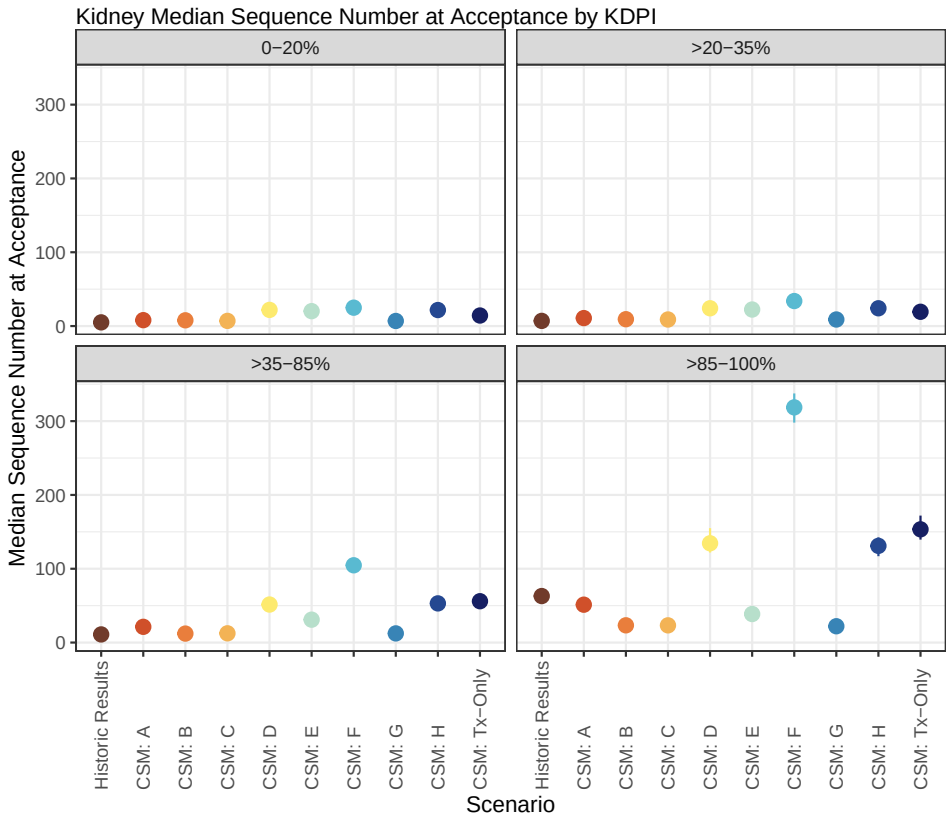

Figure 93: Kidney Median Sequence Number at Acceptance by KDPI.

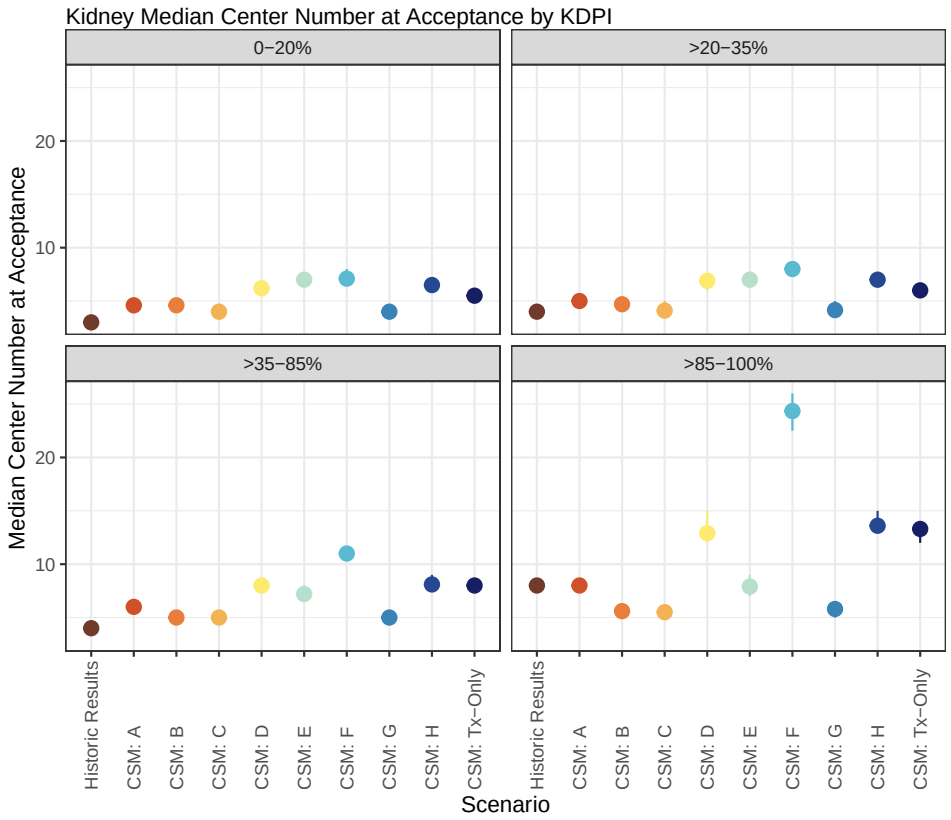

Figure 94: Kidney Median Center Number at Acceptance by KDPI.

6.5.3 Pancreas/Kidney-Pancreas Unadjusted Transplant Rates

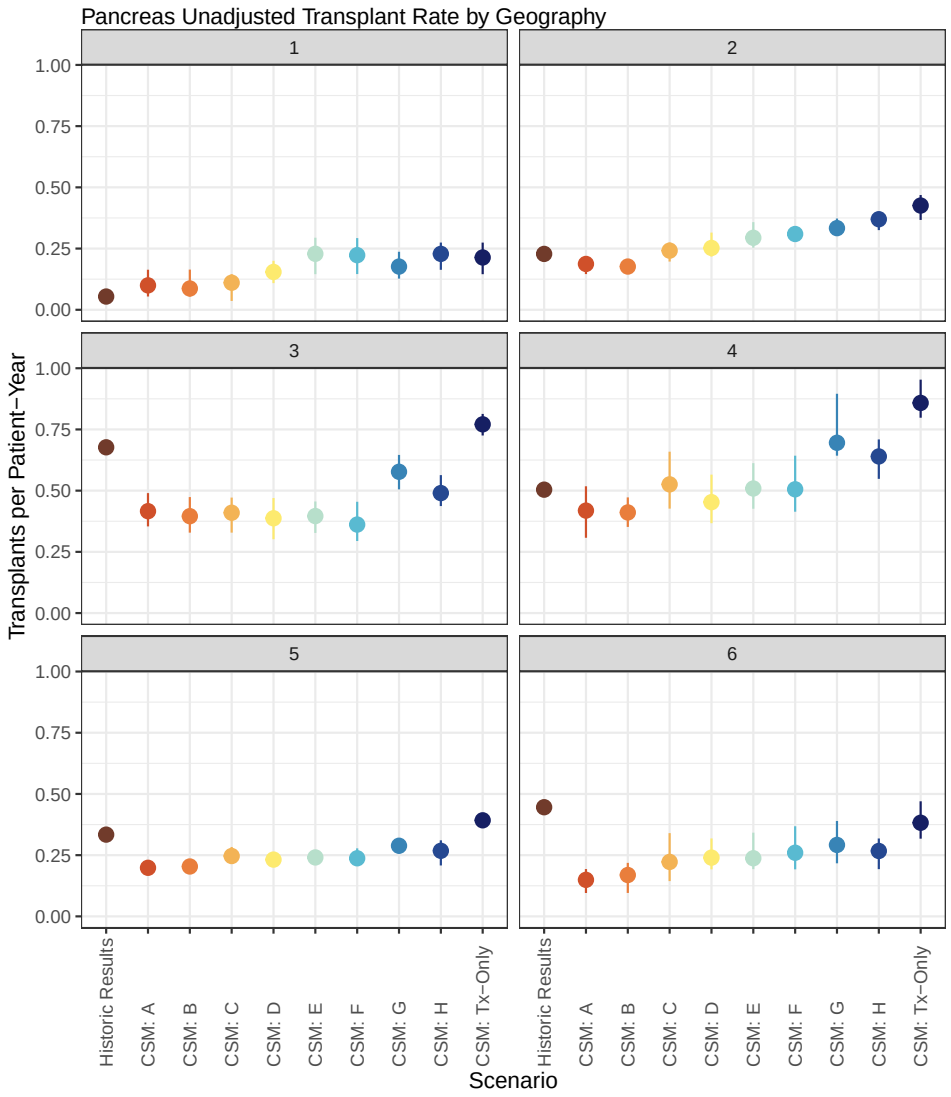

Figure 95: Pancreas Unadjusted Transplant Rate by Geography. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

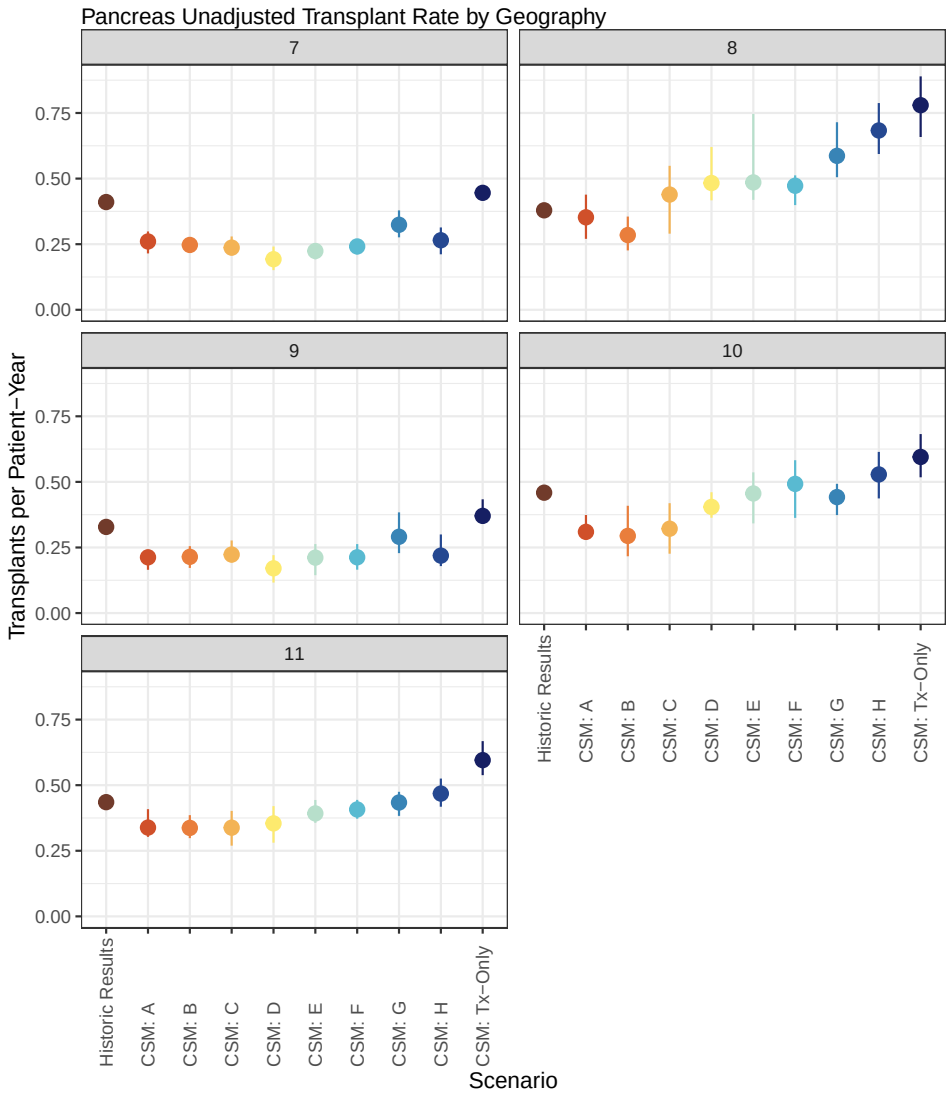

Figure 96: Pancreas Unadjusted Transplant Rate by Geography. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

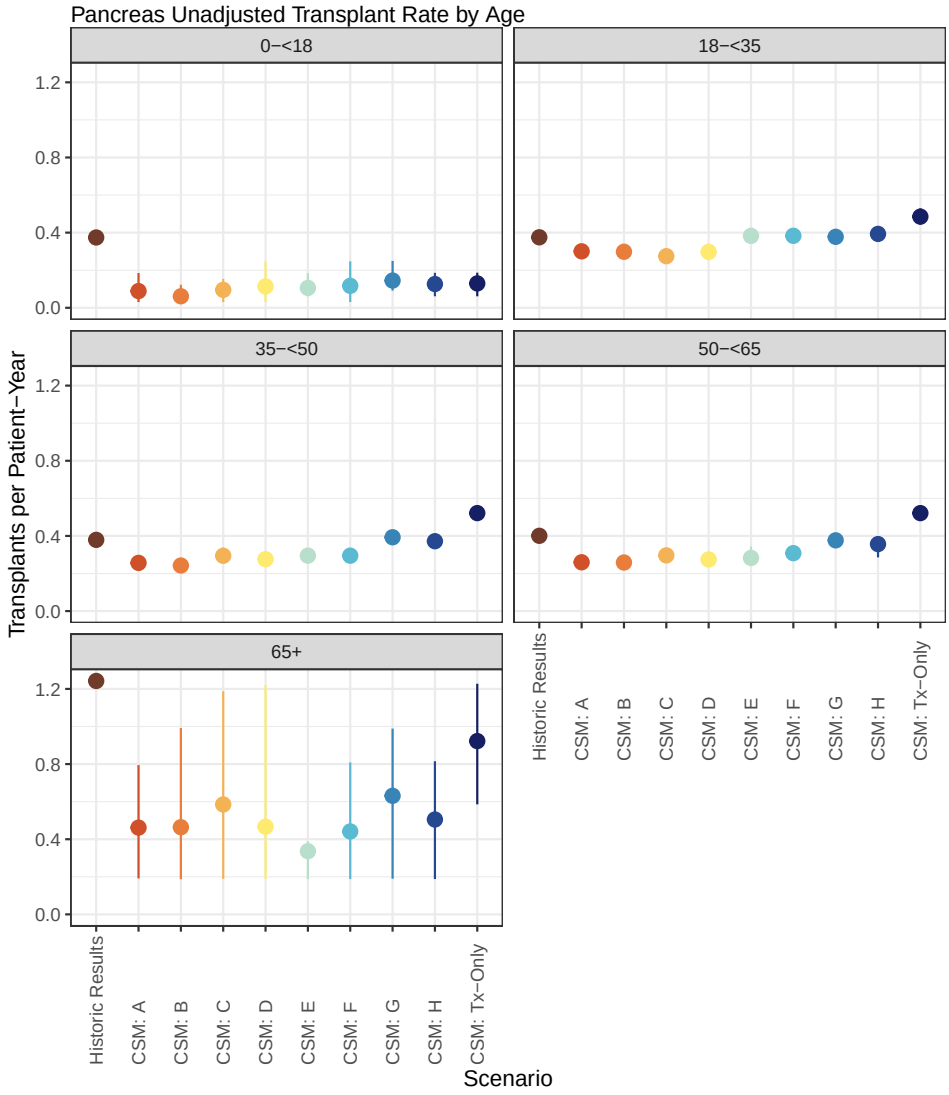

Figure 97: Pancreas Unadjusted Transplant Rate by Age. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

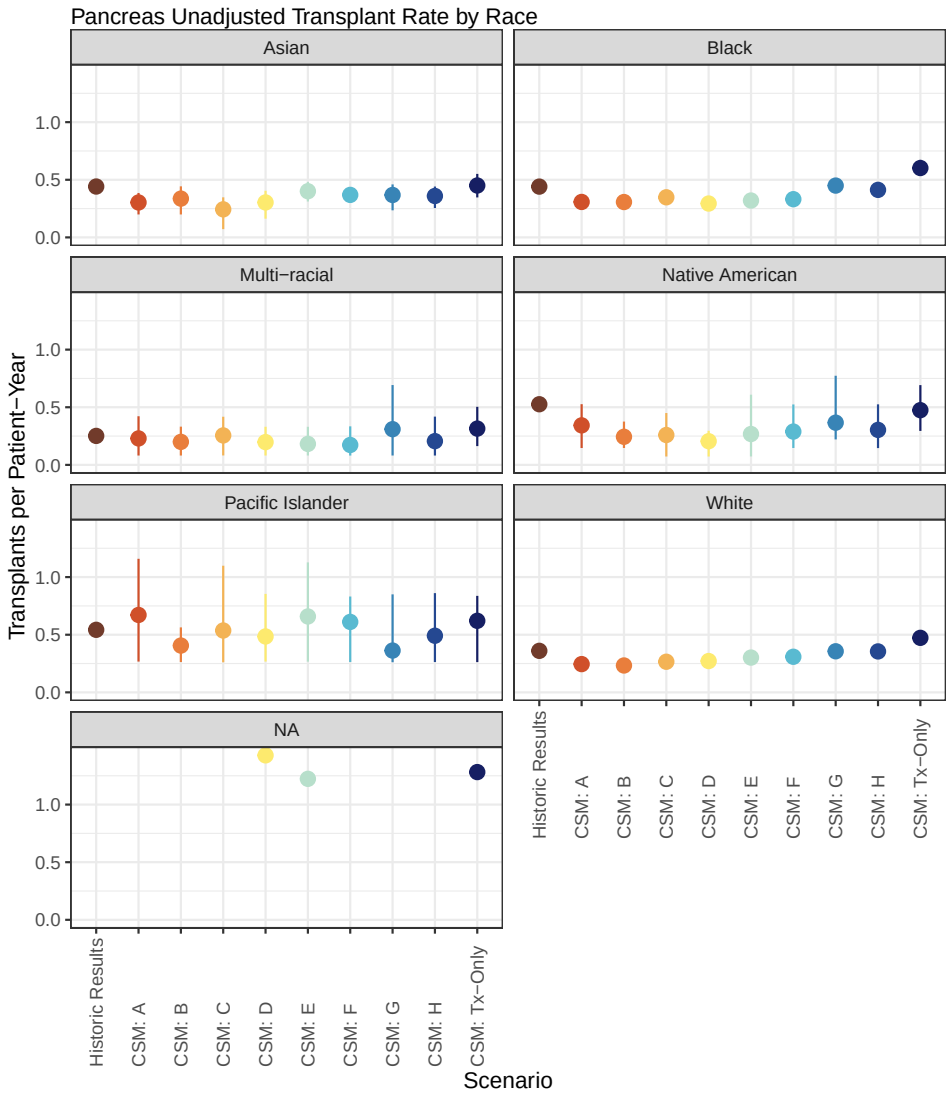

Figure 98: Pancreas Unadjusted Transplant Rate by Race. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

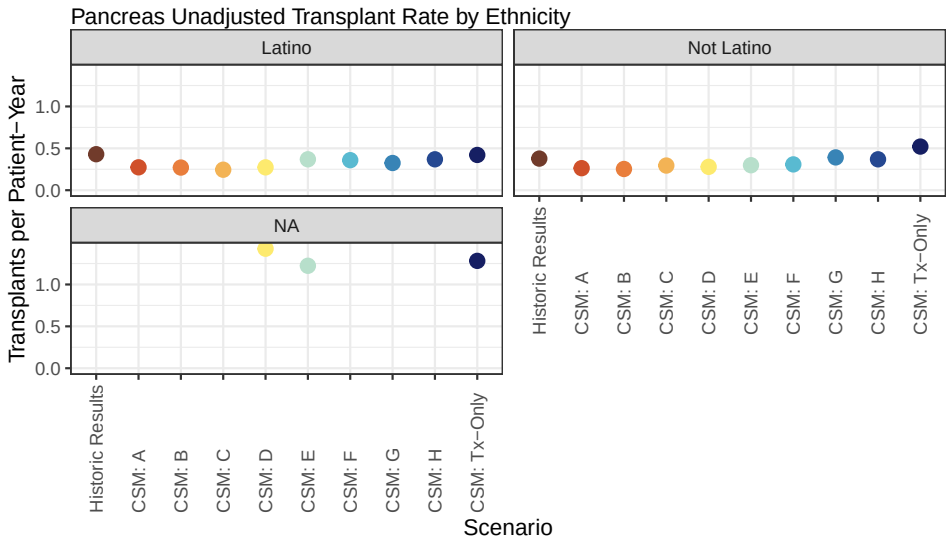

Figure 99: Pancreas Unadjusted Transplant Rate by Ethnicity. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

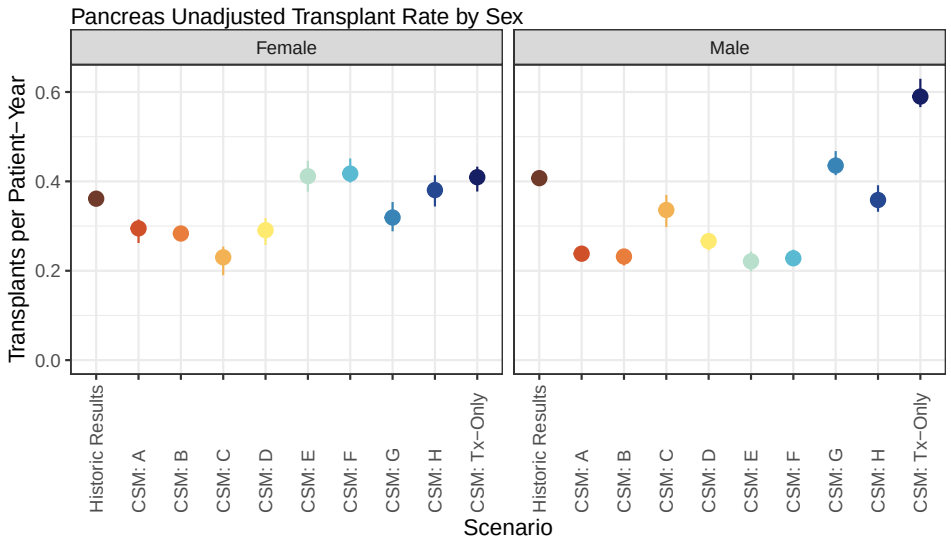

Figure 100: Pancreas Unadjusted Transplant Rate by Sex. Person-time is calculated in days for all candidates from the start of their simulation period (simulation start or listing) to transplant, waitlist removal, or simulation end.

7 Appendix

7.1 Penalized Regression Model Building Report Outline

All models described in the "Potential Utilization Models" and "Potential Acceptance Models" sections will be built following the same penalized regression process and will share the common outline below for model building diagnostics.

7.1.1 Background

7.1.2 Dataset Description

- For each subgroup:
  - Distribution of calculated values by dataset
    - For example, offer distance
  - Additional background figures

7.1.3 Model Building

Methods description.

Diagnostic Figures

- For each subgroup:
  - Figure 1: Penalized regression solution paths by dataset
  - Figure 2: Cross-validation paths by dataset
  - Figure 3: Predicted probability by dataset
  - Figure 4: Calibration by dataset
  - Figure 5: ROC curves by dataset

7.2 KI2022_01 Submodels

7.2.1 History Generation

Transplant recipients in the historical cohort do not have a complete history from the standpoint of the simulation. Through simulation we hope to create novel match runs, so transplant recipients require a waitlist history for the simulation period after their transplant: a model of what would have happened had they not received a transplant.

Histories were generated for candidates who underwent transplant with an organ from a deceased donor allocated through the OPTN process. Living donor recipients and those who underwent transplant in another country did not have histories generated; in the simulation they were removed from the list at their time of transplant like any other removal. The availability of living donors and foreign transplants are external to the simulated system.

Each listing for recipients who were listed at multiple centers was treated independently. Each will have a history generated based on the last records available for the listing of each center; although this is the same individual, the value for each of their records at the two (or more) centers are not *required* to match.

There are two time-varying fields important for allocation policies in this simulation analysis: cPRA and EPTS.

cPRA is updated when the candidate's transplant center enters new unacceptable antigen information. For this history generation model, we assumed that candidates who received a transplant had an already advantageous cPRA value and so their transplant programs did not make any updates to their unacceptable antigen information. That is, the recipients keep their cPRA value at transplant.

Raw EPTS is calculated as:

```
Raw EPTS = 0.047 * max(Age - 25, 0) -
0.015 * Diabetes * max(Age - 25, 0) +
0.398 * Prior Solid Organ Transplant -
0.237 * Diabetes * Prior Organ Transplant +
0.315 * log(Years on Dialysis + 1) -
0.099 * Diabetes * log(Years on Dialysis + 1) +
0.130 * (Years on Dialysis = 0) -
0.348 * Diabetes * (Years on Dialysis = 0) +
1.262 * Diabetes
```

For the purpose of calculating EPTS in a generated patient history, we assumed Diabetes and Prior Organ Transplant statuses do not change. Given this, the only values

that changed were Years on Dialysis and Age. The Raw EPTS was simply calculated every day of the simulation period posttransplant.

Waitlist removal was modeled with a matching algorithm. An attempt was made to match each transplant recipient to a candidate who did not receive a transplant during the cohort period; potential candidates were removals from the list or those who were still waiting. The matching was based on:

- Kidney:
  - Gender
  - Age at listing +/- 5 years of transplanted candidate
  - Waitlist organ
  - At least 80% of the waiting time as the transplanted candidate
- Pancreas and Kidney-Pancreas
  - Gender
  - Age at listing +/- 10 years of transplanted candidate
  - Waitlist organ

After matching to create a group of potential candidates and checking that there were at least 10 unique candidates, a single candidate and date on their waitlist history that met the criteria were randomly selected. This sampled waitlist history was then applied to the transplant recipient at their transplant date. If this sampled removal history was:

- still waiting on the list historically, then the transplant recipient did not have a generated removal;
- removed historically, but not within the remaining simulation period, then the transplant recipient did not have a generated removal; or
- removed within the remaining simulation period, then the transplant recipient had a generated removal of the same reason as the selected candidate.

After the matching algorithm, there may be recipients who could not be matched based on too few matching records. These recipients will be assumed to remain on the list for the entire simulation period.

7.2.2 Donor Arrival Generation

Novel simulated match runs are created in part via randomization of the donated organ arrivals. We used a sampling approach to create different simulation iterations based on donor arrival date. All donors were sampled as follows:

- The donor arrival dates were sampled without replacement; reshuffling the donor arrival dates. This was used for four simulation iterations, and was intended to closely match the historical record.
- The donor arrival dates were sampled with replacement. This was used for three iterations, and was intended to broaden the range of possible match runs.
- Donor arrival dates were sampled uniformly from the entire cohort period. This was used for three iterations, and was intended to create more variability. This sampling scheme may "smooth out" trends for donor arrival.

7.2.3 Posttransplant Models

Each simulation produces a unique group of patients who undergo transplant, some of whom may not have yet received a transplant in reality. To represent posttransplant outcomes in these simulated groups of transplant recipients, predicted probabilities at 1 year and 10 years posttransplant of all-cause graft-failure and of death after transplant were estimated with Cox proportional hazards survival models.

Patients who underwent transplant between January 1, 2007, and November 2, 2021, were included in the cohort to fit the survival models. Patients were followed until the earliest of graft failure, death, or November 2, 2021. Patients who did not experience death or graft failure were assumed alive until November 2, 2021, even if their date of last follow-up was prior to November 2, 2021. Living donor transplants were excluded.

Separate models were fit for four different outcomes:

1. Kidney graft failure, including patient death. This outcome is defined as the earliest of death, relisting, retransplant, resuming dialysis, or center-reported graft failure.

2. Kidney recipient death.

3. Pancreas graft failure, including patient death. This outcome is defined as the earliest of death, relisting, or retransplant as there has only recently been a consistent OPTN definition of graft failure.

4. Pancreas recipient death.

The model cohort was split into an 80% training dataset and a 20% validation dataset. Elastic net Cox proportional hazards models with alpha of 0.99999 were fit with the 80% training dataset for variable selection. Variables identified from program-specific reports as predicting graft failure or death, and additional variables hypothesized to be associated with these outcomes, were included. Models were be stratified on demographic or clinical predictors with evidence of violating the proportional hazards assumption to the extent possible. Continuous variables were transformed with linear splines. After variable selection with the elastic net models, center-level random effects were estimated with a Cox proportional hazards frailty model with an offset for the linear predictor from the elastic net model.

The linear predictor from the elastic net models and the center-level random effect were used to predict the probability of an outcome at 1 year and 10 years posttransplant. For each model, strata-specific baseline cumulative hazards at 1 year and 10 years posttransplant were estimated, multiplied by the patient-level linear predictor and center-level random effect, and transformed to a probability of event at 1 year or 10 years posttransplant for each patient. Using the 20% validation dataset, the sum of the probabilities of events from the models was compared to the observed number of events to estimate a multiplier for adjusting the baseline hazard. For example, if there were 120 predicted events, but only 100 observed events, the multiplier (or divisor) is 1.2, and each individual probability is divided by 1.2 to bring the baseline percents closer to those observed in reality.

For the simulated transplants, individual-level probabilities were estimated using the model parameters and divided by the multipliers. These individual-level probabilities were averaged across population subgroups of interest and multiplied by 100 to get the predicted percent of patients within a subgroup that would experience the event by 1 year or 10 years posttransplant.

Background and Methodology in Simulation Analysis: OASim

Date:
November 27, 2023
Authors:
Tim Weaver, Josh Pyke

1 Introduction

OASim offers a robust set of tools that can be used to simulate many aspects of the organ allocation system (OAS) as it processes through a sequence of donors and candidate events. The framework of OASim may also be applicable to other systems that involve a queue along with rules for sorting the queue, but we investigate only matters related to organ allocation.

Given the robust nature of the software, a wide range of research questions can be investigated. Here we discuss possibilities and considerations when designing simulation studies.

2 Background: The Organ Allocation System

The OAS includes all aspects of the process of allocating donated organs to individuals who are waiting to receive a transplant. There are two main populations of people involved in this process: those waiting for an organ and those who have donated an organ (see Figure Methods 1). In the figure, those who go on to wait for an organ are shown on the left track and those who donate an organ after death are shown on the right.

Starting from the general population, some individuals will develop disease that has the possibility of transplant as a treatment. Of these individuals, some will go on to receive treatment for the condition. As a part of their treatment, some may be referred for evaluation for transplant; of these individuals, some will ultimately be listed for transplant (that is, they will be added to an organ transplant waiting list). The population of individuals who have been listed and are waiting for a transplant are referred to as “candidates.”

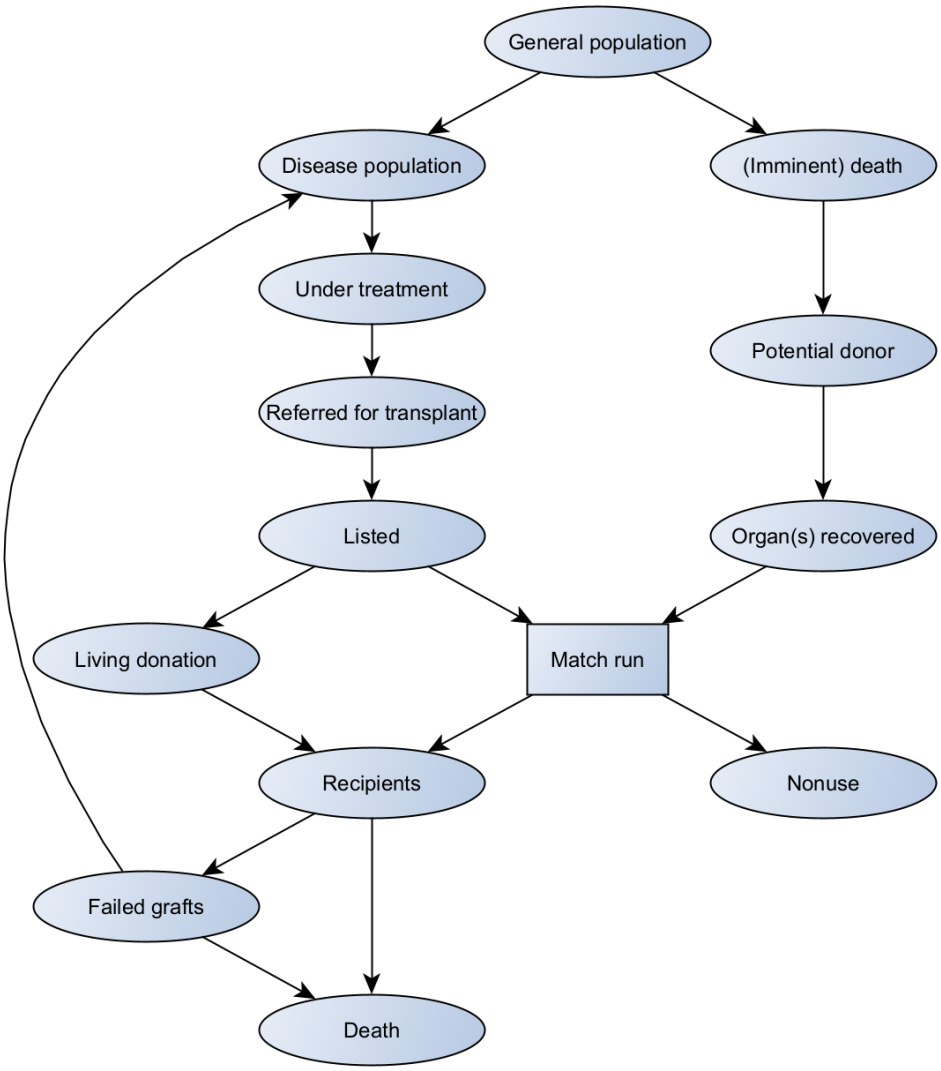

Figure Methods 1: Entire Transplant Process

Starting again from the general population, some individuals will be in a state of imminent death from disease or injury. Of these, some are at a hospital and able to be evaluated for potential donation of their organs. Of those individuals who have at least one organ that is a viable option for deceased donor transplant, some will have agreed to donation and go on to have the organ(s) recovered in preparation for allocation to an individual waiting for an organ (ie, a candidate).

At this point in the process, a match run (MR) can be performed; this is the sorting of a group of candidates into priority order for a given donated organ. The process sorts the candidates based on an allocation policy defined by Organ Procurement and Transplantation Network (OPTN) committees under the Final Rule. After the sorted results of the MR have been determined, the organ is offered to each candidate in order. The first candidate on the list (in concert with their treatment center) then has the option to accept the donated organ. If they choose to accept the organ, their transplant can proceed. If they decline the donated organ, it will be offered to the next candidate on the sorted MR list. If all candidates on the MR have declined the organ, it will not be used (nonuse)[16].

For some organs (eg, kidney), living donation is an additional pathway to transplant. In this process, candidates do not take part in MRs for donated organs but instead make arrangements for a living donor transplant.

Following transplant of either a deceased or living donor organ, individuals transition from being transplant candidates to "recipients," and their new organ is referred to as a "graft." This population may live with a functioning graft until death caused by injury or any other disease. These recipients may even have organ failure in other organs and repeat this same process for the additional organ. This same population may develop disease of their transplanted organ (failed graft). Those with graft failure have now transitioned back to the disease population at the top of the figure. Those who have rejoined the disease population may repeat the transplant candidate process as described above; in this case they are referred to as "re-listed candidates."

2.1 The Portion Modeled by OASim

The entire OAS is a broad system, and OASim focuses on a section of the overall processes (see Figure Methods 2). The focus of OASim is on studying the *allocation* of donated organs with the MR process being the main step where this occurs. To this end, the simulation process starts at the point where candidates and recovered organs have

[16]Note: Declines by all candidates is not the sole reason for nonuse of a donated organ, but for simplicity it is the most relevant reason for the simulated environment.

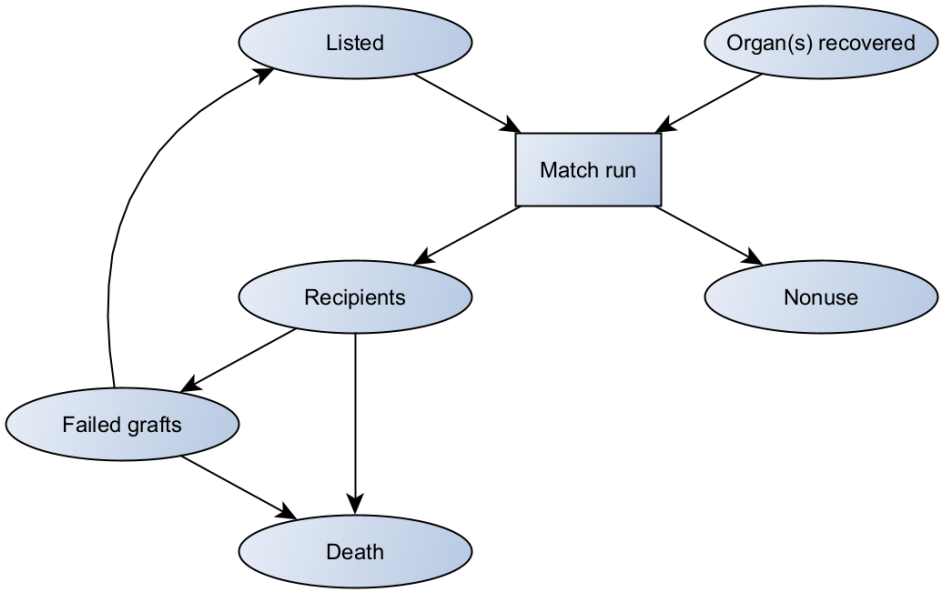

Figure Methods 2: OASim Domain

already been identified; or, said another way, OASim takes as input a population of candidates and recovered organs.

3 Organ Allocation Simulation

3.1 What Is OASim?

This section of the document is based on sections 1-3 in the entry "Computer Simulations in Science" from the *Stanford Encyclopedia of Philosophy (Winter 2019 Edition)*. Here we summarize and describe the frameworks laid out in the encyclopedia entry specifically in terms of the OAS and OASim to answer this question: What is OASim?

3.1.1 A Narrow Definition

In a narrow sense, OASim is a computer program that uses step-by-step methods to explore the approximate behavior of the OAS (or more generally any system with a queue, sorting rules, and events that trigger the sorting). Given the state of the OAS at some initial time t, OASim uses a set of rules and instructions to calculate the state at t+1; from the state at t+1 it uses the rules to calculate the state at t+2, and so on. The step-by-step processing is a natural choice for the OAS because the system is largely recorded as a sequence of discrete events (eg, a candidate visits a clinic and has lab values updated, a candidate applies for and receives an exception, a donated organ arrives). The algorithm produces a numerical history of the evolution of the system's state where the resultant "data" are meant to mimic a numerical history of the actual OAS.

From a user's point of view, this would be a situation where they have installed the OASim software and created all of their own input data and configuration files.

3.1.2 A Broad Definition

A broader definition of OASim may refer to the entire process of *an OASim study*; it is a comprehensive method for studying the OAS. In this framing, the narrow definition above is only a part of OASim, with all inputs and parameter settings that are processed by the computer program along with the presentation and study of the simulated data making up the rest of OASim. From a user's point of view, this would be a situation where they have installed the OASim software along with a set of input files that represent more aspects of the OAS (Figure Methods 2). As an example, the population of candidates is represented by a dataset derived from historical records, and a statistical distribution is used to randomize when the candidates arrive.

This comprehensive method for studying the OAS may include:

- Choosing and accessing appropriate models to represent different components

of the OAS (eg, statistical models of candidate acceptance decisions or input data randomization models)
- Implementing the components of the OAS as a computer program:
  - the statistical and data models associated with representing the OAS in a mathematical setting
  - parameters that represent the allocation rules of the OAS
  - instructions that control the computer program as it progresses through the sequence of events
- Running the "simulation" in the narrow sense of the definition above in order to create the simulated data
- Presenting and analyzing the resultant "data" to draw conclusions about the system

3.2 Types of Simulation in OASim

OASim allows for multiple types of simulation to be implemented. Here we discuss different simulation types and how they may be present in an OASim investigation.

Equation-Based
We do not believe that any components of OASim can be described as equation-based.

Agent-Based
In an agent-based simulation, each individual is modeled and has their own set of rules that govern their behavior. This is the only type of simulation that is guaranteed in every OASim study; donors and candidates are modeled at the individual (agent) level and have a set of rules that govern how they interact. Each candidate is represented as a dataset and each row in that dataset represents an event related to that candidate; as the simulation progresses, the population of candidates is updated one at a time for each candidate and each of their events. Similarly, each donor is represented as a row of a dataset and are processed one at a time; as the simulation processes a donor event, an MR is created and can be offered to candidates at the level of an individual candidate and individual donor.

Multiscale
Multiscale simulations combine models at different scales of description. An OASim simulation may incorporate models that operate at different scales. As has already been

discussed, modeling is required at the individual candidate and donor levels of the system, but OASim allows for a broad range of calculation and there is potential for models to apply to groups of candidates–say, at the transplant-center level.

Monte Carlo
Monte Carlo simulations use randomness to calculate properties of a system, but the randomness itself is not under investigation. OASim offers a number of tools for stochasticity, and this may be an important feature of a simulation study design. For example, by randomizing the arrivals of the candidates in an OASim dataset, novel MRs can be created for a simulation. OASim also offers random number generators along with a rich expression syntax, so the options for introducing Monte Carlo techniques in a simulation study are very expansive.

3.2.1 Models in OASim

The types of simulation described above are implemented in OASim via models of the OAS (see Figure Methods 3[17]). As mentioned, all OASim designs will involve some agent-based elements; models for randomized arrivals, as well as calculation of the MR, certainly happen at the individual agent level. Other models may be based on elements derived from the population level; the placement mechanism or history generation may be of this type. Stochasticity may be introduced into virtually any element of the OASim design; the randomized arrival model is a direct example of this type.

3.3 Why Simulate Organ Allocation

Heuristic Purposes
In this view of simulations, the point is to help understand the operation of the system, either for a broader public audience or for researchers within the transplant discipline.

Prediction
Here simulations are run in order to create data that we do not have access to. This is anticipated to be the main use for the OASim system. Simulating a change in allocation policy is a question of this type; the only data we have access to are historical records based on the existing allocation policy. Because of the nature of the policy, experimentation is not an option, so simulation may be used as a means to create scenarios within

[17] Note, this framing of models is not the only way the system could be described; it is representative of our understanding of most analyses of the OAS. This framing is shown to help describe the simulation system and OASim and represent our understanding of both, but it is not meant to say this is the only framing possible.

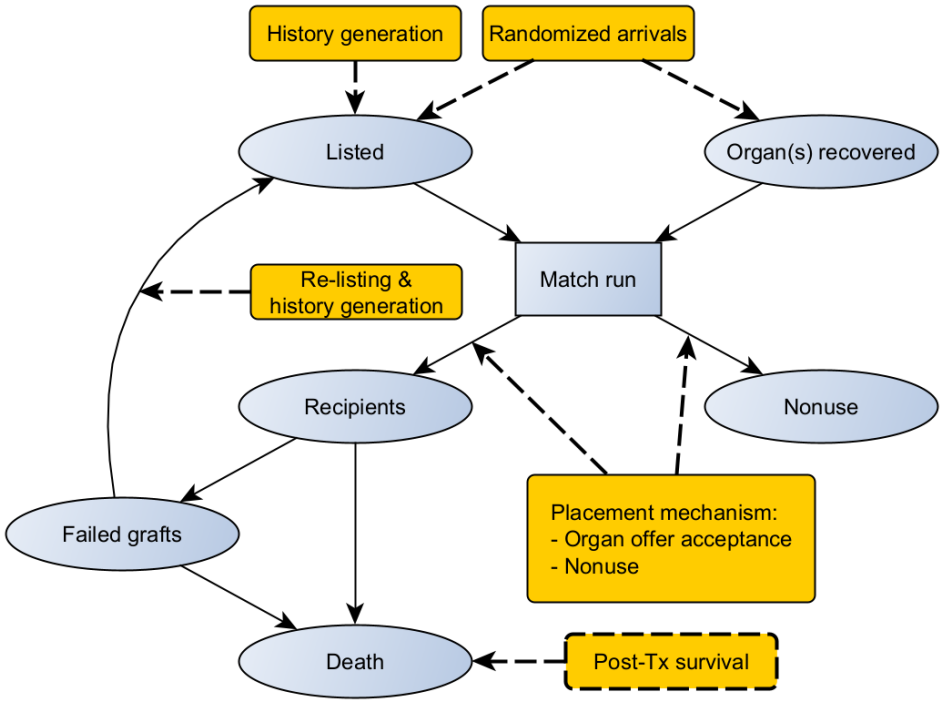

Figure Methods 3: Annotated OASim Domain. Tx, transplant.

an OAS that cannot be observed in reality.

Understand the System and Its Behavior

Simulation studies of this type could be used to help understand how a current state of the OAS arrived, or what impacted the state. For example, if there is a variation in some metric across the country, a range of simulations could be run to help understand which components would contribute to this outcome.

4 Types of Questions

The OASim framework is robust and allows a wide range of features of the OAS to be modeled and, in turn, allows for a wide range of questions to be investigated. In this section we discuss a number of questions that might be investigated using OASim. This is not meant to be an exhaustive list, and these techniques may, of course, be implemented together.

4.1 Past Simulation Studies of Organ Allocation

Historically, simulation studies of organ allocation undertaken by OPTN committees and the Scientific Registry of Transplant Recipients (SRTR) have used the simulated allocation models (SAMs) software. Here we will briefly describe features of historical simulation study designs, because they are well known to researchers in the field, are fairly constrained in scope and we anticipate future OASim studies will incorporate many of these features.

The main questions asked historically have related to changes in allocation policy. To address these questions, the studies have compared simulations run under different rules for organ allocation while keeping all other input data and settings the same between simulations. Data used to create models and run the simulation were drawn from SRTR, and the data available in the registry set boundaries on the domain of the simulation analysis. With this framework, the potential allocation rules are compared against the current allocation policy under a framework that is *meant to mimic a given historical period* as closely as possible. The simulations were backwards looking and created a predictive type of analysis. However, because the data conditions of the simulation were trained and tuned to historical data, the predictions are of a counterfactual nature and best described as predictions of what *would have* happened in the historical era under different (counterfactual) allocation policies.

An important component of the historical SAMs studies involved creating simulated data results across a range of potential historical possibilities. This was achieved by way of creating multiple input candidate and donor datsets sampled from historical data. The sampling was done to create randomized arrival times for both candidates and donors in order to create novel MRs that did not actually occur historically. This framework requires an assumption that the arrival of candidates and donors does not depend on the characteristics of the individuals; their arrivals are essentially random and so the reordering is thought to create valid counterfactual MRs. This randomization process was repeated a number of times to create a range of input datasets; these can be called

"iterations." The simulation results using the range of input dataset iterations were then treated as if they were sampled from a larger distribution (of hypothetical potential MRs) and summary values were averaged across the iterations[18].

This randomization of candidate and donor arrivals can be described as introducing a Monte Carlo aspect to the simulation, because the randomization of the arrivals is not under investigation in its own right. The stochasticity was only introduced in order to help calculate summary values (between different allocation policies) across hypothetical datasets.

Each allocation policy scenario under review for the study took the same randomized datasets described above as inputs. Another component of the SAMs models that introduced a stochastic element was the "acceptance model," which was used to determine which (if any) candidate on the MR received a donated organ. This was implemented in such a way where an acceptance probability in (0, 1) was calculated based on a formula of a single candidate and donor characteristics based off of statistical modeling of the data cohort; a uniform variable was drawn to determine if the candidate "accepted" the organ. The sampling from this uniform distribution in this type of modeling of the OAS introduced an additional range in the simulated outcomes; given the same input data and settings, including a candidate and donor for an MR, different simulation scenario runs may have different outcomes for the "same" acceptance question[19].

The current allocation policy, or the policy that was current during the timeframe of the data cohort, had a number of uses in this overall simulation framework. The first was as a "tuning" target for the component models within the broader OASim framework. Simulated runs of the historical time period were performed and certain outcome metrics compared against those calculated against the historical dataset, and modification of the acceptance model was used to bring the summary measures closer to those seen historically. The second was as a comparison group for the historical predictions, with the results often being limited in interpretation to directional changes only. That is, the simulated results of the alternative allocation policies were not compared to historical results. The simulated results of the different allocation policies were only compared between each other and the simulated results of the current allocation policy.

[18]Note: Even though the results were treated as if they were from a larger hypothetical distribution and the results averaged, very few assumptions of statistical distribution were made, and thus formal statistical testing was not undertaken.

[19]For a single scenario, multiple runs could be guaranteed to return the same results for a single MR by setting a random seed.

4.2 Varying Data Conditions

The previous section described the overall logic behind the questions asked and methods used to investigate features of the OAS in past simulations using the SAMs. The following sections describe additional questions that may be asked with the broader set of OASim tools.

We have described a study design that attempted to mimic as closely as possible a specific era in history in the simulation input dataset and compared across allocation policies[20]. In a design of this nature, prediction about the future would not be appropriate unless the future state of candidates and donors was assumed to not be changing in relation to the data cohort period. This assumption is often not valid. A study question related to future prediction would need to model some range of possibilities for the future state of the system. A study question interested in prediction could be phrased along these lines: "Given the current rules for allocation, *what is likely to happen* under potential future data conditions?"

Under a research question of this type, the range of simulated outcomes might come from varying the listing and donation trends of the candidates and donors: what happens if the listing trends do not change from today? what if the rate at which candidates list increases while the rate of donation decreases? and so forth. The details of creating the modeled (future) datasets would be the responsibility of the researcher; using the randomized arrivals framework from the last section with some sort of oversampling could be one method of achieving an input dataset that is appropriate for prediction.

4.3 Placement Mechanism Differences

Another component of the OAS that can be modeled within OASim is the mechanism of placing an organ with a candidate after an MR has occurred. The MR creates an ordered list of candidates, with each candidate having the possibility of accepting the organ. A stochastic process based on statistical modeling has already been applied, but a much simpler method would be to simply allocate the organ to the candidate who appeared first on the MR. A placement mechanism of this sort might not be best for broad inference or prediction but may be useful to a researcher to help establish a possible range of values; a simulation of the "first in list" placement might be used to access questions about who is prioritized under a given allocation policy.

Variations in placement mechanism might not be the main source of simulated vari-

[20] Note: Even though the input datasets were designed to mimic the historical data, they are still models and some elements of the datasets do not reflect every aspect of the historical data perfectly.

ability in a study but could aid in giving confidence that the simulation study (in the broad sense) is valid and the results can be used to make inferences about the real world.

Another example that has been discussed in historical simulation analysis has been related to travel distance after some allocation policy change. Under (simulated) policies that prioritize travel differently than historical policy, it is often wondered if acceptance decisions will change. This could be modeled as part of the range of simulated outcomes and become another iteration that the analysis can examine. A study with this design might use the historical acceptance model as a middle ground, with a model that highly values low travel distance as one extreme and another model that does not as the other extreme.

4.4 Simulate at the Extremes

The previous example illustrated what can be an important simulation technique: setting parameter values to extremes in order to simulate as wide a range of outcomes as possible. This approach will not be appropriate for all research questions but should be considered, in particular for situations where there is some a priori idea of what might happen (ie, acceptance decisions around travel might adjust after a policy change).

4.5 Any Component Model Can Be Varied

The above examples taken together show that any subcomponent of OASim that introduces a model also introduces the possibility for simulations across a range of the model's parameters.

5 Verification, Validation, and Credibility

This section outlines Sargent's "Verification and Validation of Simulation Models" paper in terms of OASim and the OAS. The entire text can be found here. All quotations in this section come from this Sargent paper. A key point that will be repeated is that all verification and validation are only valid with respect to a given research question. In this section it will be important to make a distinction between OASim in the narrow sense (ie, restricting to a computer program) and the broad sense (ie, referring to an overall OASim study).

OASim studies will likely be used to aid in investigations on the OAS as well as for decision making, whether by researchers, SRTR, or OPTN committees. A key question will of course be whether the study and the results of the study are "correct."

Verification

"Ensuring that the computer program of the computerized model and its implementation are correct."

Under this definition, verification is a technical aspect of the overall process. In the narrow sense of OASim, verification is purely a software development task; it ensures that the software can correctly process a sequence of candidate and donor events. In the broader sense of an overall study, verification ensures that results from data modeling are correctly translated into instructions for OASim.

Validation

"Substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model."

An OASim study should be developed for a specific question (or type of questions), recall the "Types of Questions" section, and its validity determined with respect to that question. For example, an OASim study designed for prediction of (potential) future trends in the OAS would need validation related to future listing and donation trends; an OASim study designed to address counterfactual predictive questions would require validation related to comparison with historical records. Further, validity for one purpose does not (generally) imply validity for another. Validity needs to be determined within some acceptable range that is determined by the accuracy required of the study results to make inferences.

Credibility

"Developing in (potential) users the confidence they require in order to use a model and in the information derived from that model."

Careful documentation of all logical and analytic steps of verification and validation of an OASim study will be required to provide the information needed by users to evaluate the study for credibility.

5.1 Basic Approaches

Sargent outlines four decision-making approaches to simulation study validation.

1. The simulation development team determines validity
2. The user(s) are heavily involved with development team in deciding the validity
3. Independent verification and validation
    - Third party (independent) of both developers and users
    - Useful when simulation involves multiple teams
4. Scoring model: Sargent does not recommend this approach

In the context of OASim, the SRTR biostatisticians and software developers can be thought of as the simulation development team and the SRTR biostatisticians, independent researchers, and OPTN committees could be thought of as the users. There will likely not be a place for independent verification and validation that is undertaken by someone who would not be considered a researcher and thus fall under the "user" label.

5.2 Paradigm

Sargent's paradigm for computer simulation verification and validation is shown in Figure Methods 4.

The *problem entity* for an OASim study is the OAS under investigation. The problem entity may include historical records of organ allocation, or if the research question involves generating data[21], may include counterfactual or future predictive situations. The *conceptual model* is the collection of all mathematical/logical/verbal representations of the OAS problem entity developed for a particular study. The conceptual model(s) could include models of where in a sorted MR an organ is allocated, models of historical candidate records, models of donor arrival trends, etc. The *computerized model* is the conceptual model implemented as a computer program. "The conceptual model is developed through an *analysis and modeling phase*, the computerized model is developed

[21] Often a main reason why a simulation is undertaken.

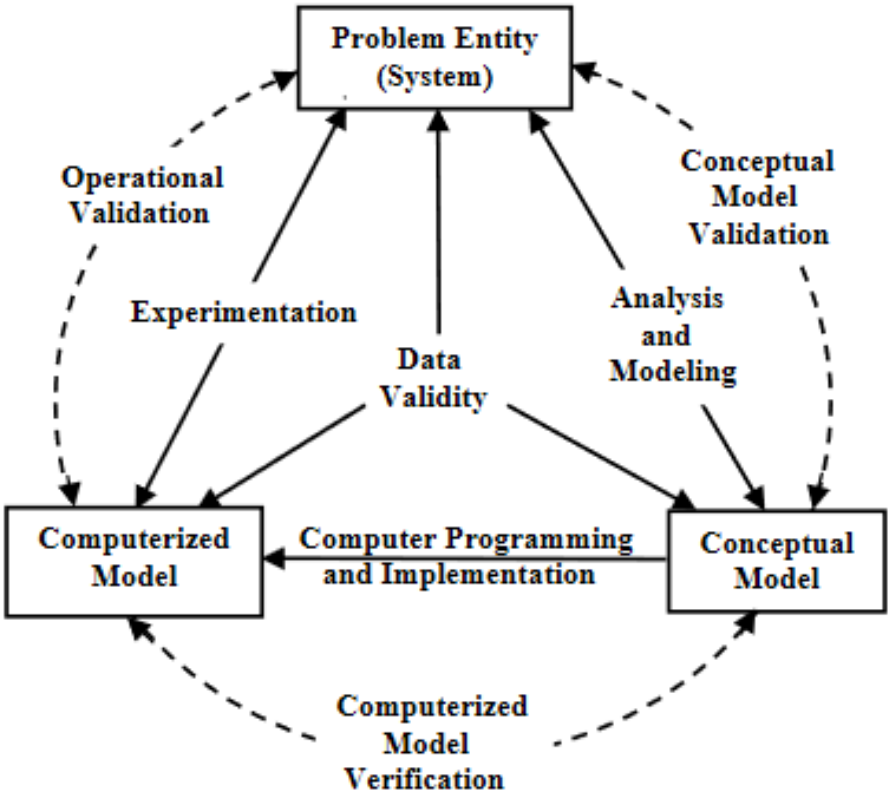

Figure Methods 4: Simplified Version of the Modeling Process

through a *computer programming and implementation phase*, and inferences about the problem entity are obtained by conducting computer experiments on the computerized model in the *experimentation phase*."

5.3 Conceptual Model Validation

The process of conceptual model validation is used to determine if[22] "1) the theories and assumptions underlying the conceptual model are correct and 2) the model's representation of the problem entity and the model's structure, logic, and mathematical and causal relationships are 'reasonable' for the intended purpose of the model."

5.3.1 In the Narrow Sense

Recall that OASim operates as an agent-based simulation (ie, each candidate and donor is represented), where it processes a sequence of events. Within this framework is the assumption that the OAS under investigation can be represented as a sequence of discrete events; "time" is not important in and of itself, it is only used as a way to order the sequence of events. This basic assumption will be present in all OASim studies because it is built into the OASim software and needs to be considered when determining if an OASim study will be informative for a given research question.

5.3.2 In the Broad Sense

The conceptual model for an OASim study will (almost always) be made up of a number of submodels. Some examples that have been discussed include stochastic models of candidate (or donor) arrivals, history generation for transplanted candidates, and statistical models for placement of a donated organ to a position on the MR.

Examinations of the theories and assumptions underlying each model need to be performed using mathematical analysis and statistical methods with respect to data from the OAS under investigation. In the case of a statistical model for organ placement for example, statistical methods for model fit should be utilized that are appropriate for the model. This might include partitioning the data into training and test sets. Model assumptions such as independence of observations should be tested.

The result of this step in conceptual model validation will be a collection of analysis results that have validated each submodel on its own terms with respect to data from the OAS. Each of these analyses should essentially be stand-alone models of which validation

[22] Sargent uses "determine that"

can be interpreted using only data from the OAS and without making any reference to simulation or how the submodels will be used within the simulation study.

Following validation of each submodel in isolation, the research question of the OASim study is considered; each submodel along with their relationship to each other (ie, the overall model) are evaluated to determine if they are reasonable and correct for the specific research question. "This should include determining if the appropriate detail and aggregate relationships have been used for the model's intended purpose, and also if appropriate structure, logic, and mathematical and causal relationships have been used." For example, consider a study concerned with detailed changes around allocation policy in the short term compared to a study that was interested in predictions around long-term trends in waitlist size; the former might require a detailed history for each candidate, whereas in the latter study a simple model that only includes a single "listing" record per candidate might be appropriate. The submodels are then examined together to ensure that the precision required overall and by model can be achieved when the submodels are combined. Consider again the predictive study of long-term waitlist size; a detailed statistical organ placement model may be incompatible with the simple "listing" only candidate history model, so in this case a first-in-list placement mechanism may be sufficient for prediction of overall waitlist size. To further validate that the collection of submodels function together as expected, individual entities can be "traced" through the models. This involves examining how a candidate is recorded throughout the course of the simulation (not the realized values but the form the values would take).

5.4 Computerized Model Verification

The process of computerized model verification "is primarily concerned with determining that the simulation functions (e.g., the time-flow mechanism, pseudo random number generator, and random variate generators) and the computerized (simulation) model have been programmed and implemented correctly." OASim provides a special-purpose simulation language created using the higher level programming language C#. It was designed, developed, and implemented using modern software engineering techniques including object-oriented design, structured programming, and program modularity. The computerized model verification process is narrow in scope and focused on technical details related to implementation. The modular nature of OASim allows for the implementation of the distinct components of the OAS described in Figure Methods 3 to be examined in isolation.

There are two basic approaches to computerized model verification: static testing and dynamic testing. Static testing involves analysis of the OASim input files without ac-

tually running the program. This may involve structured code reviews to avoid errors in implementation, comparisons between computerized implementation and the conceptual model representation to ensure the models have been translated correctly into a computer readable format, and examination of the structural relationships between the implemented submodels to ensure they accurately represent the intended relationships. Dynamic testing involves running the program under different conditions and examining calculated quantities to ensure they produce the expected results.

Comparisons in computerized model verification are quantitative in nature as there are predefined correct values for the results of the operations (eg, unit testing). Internal calculated values may also be examined during the program's run (ie, "debugging"). Comparisons between independent programming of the processes can also be used to ensure correct implementation. These methods may be applied at the level of individual calculation or aggregations may be used in cases of large numbers of comparisons. Finally, "[i]t is necessary to be aware while checking the correctness of the computer program and its implementation that errors found may be caused by the data, the conceptual model, the computer program, or the computer implementation."

5.5 Operational Validation

Operational validation involves running OASim with the submodels validated in the previous steps to determine if "the simulation model's output behavior has the accuracy required for the model's intended purpose over the domain of the model's intended applicability." Here the models are considered in tandem and so output behavior that does not behave as expected could be caused by any submodel or overall data quality issues. In this step the circular nature of the paradigm in Figure Methods 4 comes into play. Deficiencies in output behavior should lead back to the submodels so remedies may be considered. If improvements can be made the process repeats. However, if on the other hand the submodels have been built as accurately as possible, the deficiencies in output behavior may be unavoidable; in this case, the discrepancies should be noted and used to put limits in interpretation of simulation inferences. As with the prior validation steps, the research question of interest needs to inform the operational validation. The metrics that will be used to make inferences about the problem entity should be used to examine the model behavior in the validation steps.

The data comparisons made during operational validation need to be carefully considered. If the OAS problem entity is observable, then direct comparisons between simulated and real-world output behavior can be made. Consider a study concerned with changes in allocation policy that is modeling a historical period (a what *would have* hap-

pened type of question); in this case a simulation(s) could be run to try and closely mimic the real-world historical data. Important metrics can then be compared between the generated and historical data using standard mathematical and statistical techniques.

However, the OAS problem entity is often not observable. Recall that simulation studies are often undertaken to generate data that is not available. It is anticipated that OASim studies will often be performed for this reason; the research question will be interested in prediction, either of the future or counterfactual situations from the past. In this case there is no real-world data available from the OAS that can be directly used for comparison. Exploring model output behavior across a range of input values can indicate if the OASim results are directionally correct as well as if the magnitudes of changes are reasonable. For example, with all other parameters held constant, would using an input dataset with a 10% higher donor arrival rate lead to more simulated transplants along with a smaller waitlist size?

5.6 Commentary

A key distinction in Figure Methods 4 that should be emphasized is a clear separation between the conceptual model and the computerized model. The relationship between the two models is also important. Notice the arrow between the two models (computer programming and implementation) is the only one in the diagram that is uni-directional; in other words, the conceptual model directly defines the form of the computerized model, but the computerized model should not have influence on the conceptual model.

With any quantitative study it is easy to lose the distinction between the two models (conceptual and computerized) and the conceptual model ends up being contained "in the code"; in this case the problem entity (types of questions posed or even framing and vocabulary of discussions) can end up being influenced not by theory or analysis but by software implementation choices and details. This is an inappropriate direction for the inferences to take. This distinction is especially important when a simulation is being used to generate data that are unavailable.

6 References

- Winsberg, Eric, "Computer Simulations in Science", The Stanford Encyclopedia of Philosophy (Winter 2019 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2019/entries/simulations-science/. Accessed 11/27/2023.
- Sargent, Robert G., "Verification and Validation of Simulation Models", Proceedings of the 2011 Winter Simulation Conference, S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, (eds.), URL = https://www.informs-sim.org/wsc11papers/016.pdf. Accessed 11/27/2023.

Time-to-Event Rate Calculations in Simulation Analysis: OASim

Date:
November 27, 2023
Authors:
Tim Weaver, Grace Lyden

1 Introduction

This document describes calculation of rates in a time-to-event setting for Scientific Registry of Transplant Recipients (SRTR) simulation analysis using the OASim software. Prior simulation analysis has presented "unadjusted" rate calculations, and in upcoming studies we will be incorporating "adjusted" rate calculations into the analysis. Here we describe in detail both calculations, how each calculation is applied in the simulation context, and, finally, how the methods are incorporated into a broader simulation study design.

For this discussion, we will only consider "rates" to be values that are expressed as the number of events per unit time; other values called rates (eg, percentages) will not be considered rates here. At all points in this discussion, the terms "hazard" and "rate" could be used interchangeably.

2 Background

Rates in the setting of a transplant waiting list are a competing risk time-to-event problem. All observations contribute some amount of time until either censoring or one event; the events are mutually exclusive and absorbing states.

2.1 Subscript Definitions

Each observation, indexed by i, can have one and only one outcome, z, or no outcome (censored observation). In the context of waitlist rates, each candidate is an observation and the outcomes are transplant, death, still waiting (censored), and possibly removal

for other reasons, which may be treated as a terminal event or a censoring event. Each observation also has a set of covariate levels, indexed by j as in $x_{i,j}$, eg, blood type = A or age-at-listing in [18,35); we will also use the term "subgroup" to refer to a specific covariate level, $x_{i,j}$.

$$\begin{aligned} Observation : i \in (1, 2, ...N) \\ Covariate\ level : j \in (1, 2, ..., J) \\ Outcome : z \in (1, 2, ...Z) \end{aligned} \tag{1}$$

2.2 Metrics

Each observation contributes time until an event or censoring, t_i, to the rate calculation. Each observation can contribute up to one count, $I_{i,z}$, to the total count of outcome z.

$$\begin{aligned} Indicator\ variable\ for\ observation\ i\ with\ outcome,\ z : I_{i,z} \\ Time\ to\ (mutually\ exclusive)\ event\ or\ censoring\ for\ observation\ i : t_i \end{aligned} \tag{2}$$

In cases where a specific subgroup is being considered, the subscript j is used.

$$\begin{aligned} I_{i,z}\ where\ x_{i,j} = 1 : I_{i,j,z} \\ t_i\ where\ x_{i,j} = 1 : t_{i,j} \end{aligned} \tag{3}$$

2.3 Covariates

Each observation i has a set of covariates levels that are combined into an overall design matrix, $\mathbf{X}$.

$$\begin{aligned} 1 \times J\ vector\ of\ covariates\ for\ observation\ i :\ \mathbf{x}_i \\ N \times J\ design\ matrix :\ \mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ ... \\ \mathbf{x}_N \end{pmatrix} \end{aligned} \tag{4}$$

Note that, in this framing, $\mathbf{X}$ is a "wide" matrix of 0/1 indicator values for each subgroup, j.

3 Rate Calculations

3.1 Empirical Calculations

Below is a simple "unadjusted" rate calculation. This type of unadjusted rate calculation has been used in past simulation analysis as well as the waitlist rates presented in the SRTR Annual Data Report. Note that the rate for outcome type q, h_q, takes into account the time contributed by patients who experience other outcome types z, $z \neq q$.

This calculation can also be applied to entire cohorts, in which case the covariate subscript j is unnecessary.

$$\begin{aligned} N_{j,z} &= \sum_i I_{i,j,z} \\ Total\ Time_j = T_j &= \sum_i t_{i,j} \\ Rate_{j,z} = h_{j,z} &= \frac{N_{j,z}}{T_j} \\ &= \frac{\sum_i I_{i,j,z}}{\sum_i t_{i,j}} \end{aligned} \tag{5}$$

3.1.1 Commentary

This framing is an "average" rate across some period of observation. In particular, it is "averaged" over the observation window; that is, the rate is not a function of time. However, this could be a matter of interpretation; you do not *need* to take this "averaged" view and just interpret the calculated rate value as the observed rate for the period.

This is similar to calculating an arithmetic mean over some population, say, the average height of a group of people in a room. In this case, the average height of that group of people could be treated as just an empirical value, where you do not need to invoke the idea of a larger population or "target" value that the calculated mean is estimating[23]. The average is just the average. It can be useful to take extra steps like assuming some larger population and probability distribution in order to do inference, but those would be additional steps beyond the simple arithmetic mean calculation.

[23]The lack of a "target" could be disputed. Even without invoking the idea of a larger population for height, there could be a theoretical value for the mean of the group of people that our estimate deviates from because of measurement error. But that is not important here.

3.2 Modeled Rates

Regression models for rates can be framed as a survival proportional hazards model, or Poisson regression (with offset) when estimating constant rates. Both of these regression models are "adjusted" in the sense that they can accommodate multiple covariates.

For our simulation analysis, we are calculating overall (non time-varying) rates for the entire simulation period; that is, we are calculating *constant* rates (see the "Applications to Simulation" section for more details). Given this, Poisson regression is an appropriate choice and will be discussed here.

3.2.1 Poisson Rate Regression Model

Poisson regression estimates an average rate across the observation period; or, the model assumes a constant event rate. The estimates that come out of the fitted model are for the coefficient vectors, θ_z for each h_z. The point of this model is to estimate the effect ($\theta_{j,z}$) of each covariate in $\mathbf{X}$, adjusting for other factors, on the rate. Note that the events indexed by z (other than censoring) are still treated as competing events in this framing.

$$\mathbf{h}_z(\mathbf{X}) = exp(\mathbf{X}\theta_z) \tag{6}$$

3.2.2 Calculating a Rate

The results of the regression model above are estimates of the coefficient vectors, $\hat{\theta}_z$. To arrive at an estimate of a rate, some input dataset is required, $\mathbf{W}$. Given some input dataset, $\mathbf{W}$, rate estimates can be directly calculated from the Poisson regression results as $\hat{\mathbf{h}_z} = exp(\mathbf{W}\hat{\theta_z})$[24].

3.2.3 Calcuated Rates via Standardization

To calculate a rate for each subgroup level within a covariate based on the Poisson regression results, we will use standardization methods using the entire cohort as a reference population in order to calculate marginal rates. The adjusted rates are counterfactuals and are interpreted as the rate for each subgroup of interest had that group had the same distribution of all other adjusting factors as the overall population.

[24]This is an additional reason for choosing the Poisson model over the proportional hazards model. The proportional hazards model would have required an additional step of estimating a "baseline hazard."

Steps are described at the scenario/iteration level, for each subgroup of interest j, eg, blood type = A or a binned continuous variable like age at listing in [18, 35).

1. Create a reference design matrix, $\mathbf{W}_j$, for the subgroup of interest, j, by setting every record in column $\mathbf{x}_j$ of $\mathbf{X}$ to the subgroup of interest.

$$\mathbf{W}_j \equiv \mathbf{X},\ s.t.\ x_{i,j}\ is\ set\ to\ 1,\ \forall\ i \tag{7}$$

2. Calculate vector of rates for this subgroup via the Poisson regression results, $\hat{\theta}_z$, for the outcome of interest, z.

$$\hat{\mathbf{h}}_{j,z} = exp(\mathbf{W}_j \hat{\theta}_z) \tag{8}$$

3. Average the vector of rates (across the observed distribution of all other adjusting factors) to arrive at a single marginal rate value for subgroup j.

$$\bar{\hat{h}}_{j,z} = \frac{1}{N} \sum_i \hat{h}_{i,j,z} \tag{9}$$

These steps apply for a single subgroup level j, eg, blood type = A. To create a set of rates for the overall blood type metric, the calculation would be repeated for each blood type; the calculations are linked by the common regression coefficients for the outcome, $\hat{\theta}_z$.

4 Applications to Simulation

The simulation problem introduces additional layers of complexity for the presentation and analysis of rates. In historic (and presumably upcoming) simulation studies, the comparison is across simulated scenarios (different allocation policies), where the metrics being compared are an average value over multiple stochastic iterations. The transplant rate is a fundamental metric for all studies, where it is the key metric for analysis and, in turn, is the key metric for all simulation model building.

The methods discussed for creating adjusted models in the "Modeled Rates" subsection are generally used to make statements about different subgroups within a single cohort; inferences are made across a range (or levels) of a single covariate, or between covariates. The key point is that the inferential comparisons are generally made within a single model, not generally across models as is the case in the simulation setting.

In simulation analysis, we are concerned with calculating rates for subgroups of interest over the simulation period; the rates presented are an attribute of the overall organ allocation system, not necessarily individuals on the list. Additionally, unless there is a specific mechanism to indicate the rate(s) vary over the simulation, there is no need to model the rates as time varying. Given these considerations, we present the average rates over the simulation cohort period; ie, we present a constant rate h_z, not a time-varying rate $h_z(t)$.

4.1 Historic Analysis: Unadjusted Rates (Empirical Calculations)

In recent SRTR simulation analysis, rates were calculated with the formulas described in the "Empirical Calculations" subsection. The simple calculation there was applied to a single cohort; the calculations were repeated in each different subgroup/stratification group. Figure Rates 1 is an example from the KIPA2023_01 data request. In this figure, each panel (facet) has repeated the rate calculation independently for each organ group (KI, KP, PA). In the figure, only a single outcome, $z = transplant$, is shown, while the corresponding rates for the other complementary outcomes besides transplant (death, removal, etc) are not.

Each dot (and bars) represents an aggregate measure(s) across 10 simulated iterations: the mean, minimum (min), and maximum (max). That is, for each simulated scenario, the rate calculation was calculated for the cohort group for each iteration independently, and then these simulation/iteration level rates were aggregated to mean/min/max.

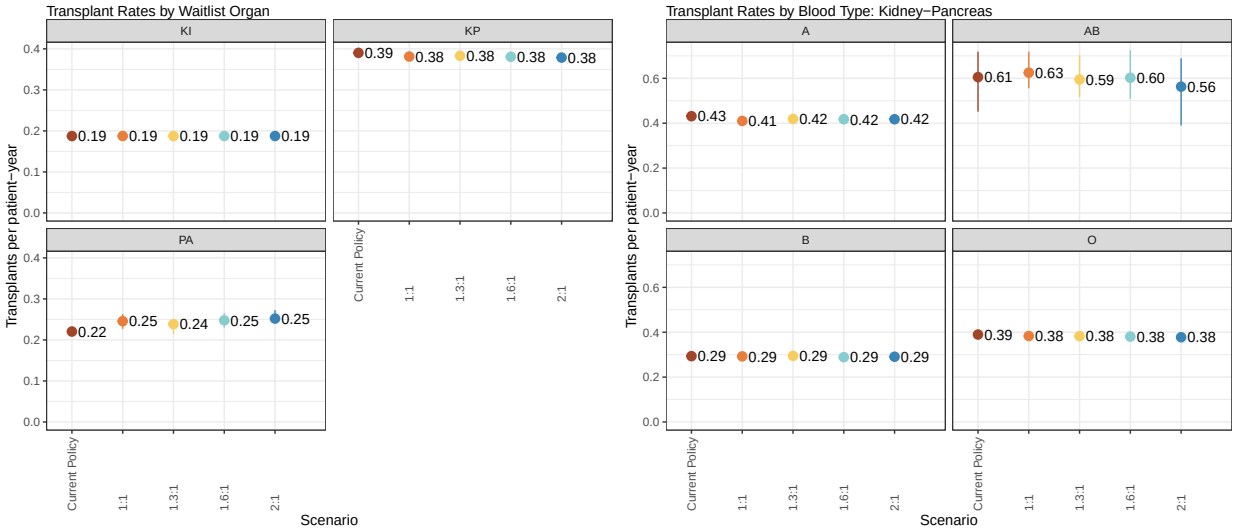

Figure Rates 1: Example Rate Figures from KIPA2023_01

4.1.1 Commentary

An independent calculation for each simulation/iteration is an important feature of the design that should be maintained regardless of the rate formula being used; the model used for an adjusted rate should also be calculated (ie, refit) at the simulation/iteration level and then those results aggregated (via some method) across the simulated iterations.

This framing/calculation does not need to be interpreted as a model. As discussed above, these rates are/were presented in a purely declarative way; they are not presented as "estimates" of some theoretical value, rather they are presented only as descriptive measures of what occurred in the simulated results. The parsimony of this framing is a benefit. No assumptions are needed or made, and the results are easily interpretable. Since there are no assumptions made about the calculations, no diagnostics are needed at the simulation/iteration level[25]. Furthermore, there is a nice correspondence between the figures in the SRTR Annual Data Report.

[25]We could apply "guardrails" by using rules for minimum group size, or other similar checks. Note that these type of steps are not diagnostics in that they are not meant to assess model quality in light of the modeling assumptions, because there are no assumptions.

5 Adjusted Rates in Simulation

Here we describe specific steps needed to calculate adjusted rates in a simulation study.

5.1 Definition of $\mathbf{X}$

The definition of which factors will be "adjusted" for (ie, which covariates will be included in $\mathbf{X}$) is a primary step in the simulation analysis study design. Because waitlist rates will be a primary analysis metric, they also need to be a primary metric for submodel building and operational validation (OV), and, in turn, need to be defined prior to any model building.

The a priori-defined set of covariates that are included in the model could be arrived at in a number of ways. This could be based on committee, SRTR senior staff, and statistical inputs, or could be an empirical study to determine important adjusting factors. This should largely correspond to the set of figures planned for the final analysis report.

5.2 Aggregation Across Simulation Iterations

After rates have been calculated for each outcome and subgroup, $\bar{\hat{h}}_{j,z}$, at the scenario/iteration level, the rates need to be aggregated across the simulation iterations. In prior simulation studies, this has been a simple arithmetic mean across the simulated iterations. This will likely be the choice for future studies as well, with summary of the min and max value across iterations.

5.3 Diagnostics Across Simulation Iterations

Diagnostics for the regression models need to be considered, but the problem is complicated in the simulation setting. A separate regression model(s) needs to be fit for each scenario/iteration, and there are corresponding diagnostic measures at this level. However, it is likely not practical to evaluate each model in isolation. Given this, the simulation study design needs to specify some method to "aggregate" the evaluation of the specific model fits across iterations.

5.4 Operational Validation

The OV process can encompass a broad range of approaches that holistically evaluate the performance of a simulation study. Given this, we will consider the limited OV scenario where the "best" collection of submodels (COS) is selected from a range of potential

COSs[26]. Here the adjusted (transplant) rate is the primary assessment metric comparing the different COSs to a historic baseline "target" across a range of grouping factors.

Under this study design, the assessment metric itself (adjusted rate) is a modeled value, and so the quality of the model(s) needs to be built into the OV design.

5.4.1 Historic Rate Calculation

During OV, adjusted rates for a historical target should be calculated and the diagnostics for this single model examined. The historic analysis is performed in order to make the best representation of the historic cohort, not to select between potential COSs.

At this stage, diagnostics should be used to inform the rate modeling and help identify deficiencies in the adjusted rate model (eg, categorical bins with too few counts may lead to poor diagnostics and the bins may need to be resized[27]). That is, the definition of $\mathbf{X}$ may be amended in light of the historical rate model diagnostics, provided the original analytic intent of $\mathbf{X}$ is preserved.

The results of this model building stage will be a finalized definition of $\mathbf{X}$ that will be used for all subsequent simulation analysis including the following OV steps as well as the final analysis report. The set of diagnostic measures used in this stage of the analysis will also be preserved in subsequent steps in the analysis.

5.4.2 Submodel Scenario Rate Calculations

Many of the same considerations around iterations and aggregation from the "Diagnostics Across Simulation Iterations" subsection are applicable in the OV situation, and in the same way the assessment of model quality needs to be an important part of the OV design prior to starting modeling.

In OV, adjusted rate diagnostics across iterations for a single COS should perform favorably in comparison to the historic rate model's diagnostics and considered when selecting the best COS. A COS whose averaged rate (over iterations) performed better than that of another COS may not be preferable if there is wide variability in model quality across iterations. In particular during OV, nonconvergence[28] of rates during a single iteration should likely be a disqualifying result for a given COS. Diagnostic issues at this

[26] Under this OV design, each COS is treated just like a scenario in the example figures above, and the same figures used for the final analysis report are used for the OV steps.

[27] Of note, the "binning" of continuous covariates into subgroups that we have historically done in simulation analysis may translate into risk-adjustment models that are more flexible and therefore fit better, due to not imposing linearity on all continuous covariates.

[28] "Nonconvergence" is used a broad sense here, and is intended to include any model deficiencies that would be disqualifying.

stage of the analysis only indicate deficiencies in a given COS not deficiencies in the overall adjusted rate model, because the form, as defined by $\mathbf{X}$, was finalized in the "Historic Rate Calculation" stage.

5.5 Diagnostics of Simulated Scenarios

At this stage in the analysis, when a baseline and proposed scenarios are being simulated and the results analyzed, diagnostics of the rate calculations again need to be examined. All considerations related to iterations from the "Diagnostics Across Simulation Iterations" section apply here.

The diagnostic results of the simulated scenarios now inform the interpretation of the rate estimates, and are not used for any sort of "selection." That is, the diagnostics are used as usual to quanitify the quality of the rate estimates. A simulated scenario may need to be considered as "noninterpretable" if the rates are not able to be estimated with consistent quality across iterations.

Some potential ideas for formalizing diagnostic analysis:

- determine thresholds for whether or not a simulation scenario can be interpreted for analysis based on standard regression results
  - thresholds may apply across iterations where the overall quality needs to be above some level
  - thresholds may also apply at the iteration level, where a single iteration that has poor diagnostics may lead to the scenario overall being noninterpretable
- aggregate *P* values or overall fit metrics like Akaike information criterion across iterations to help quantify a simulation scenario's rate estimates quality
- do not aggregate and examine all goodness-of-fit statistics and diagnostic plots, perhaps with faceting by each iteration

6 Extensions of a Calculated Rate

6.1 Cumulative Incidence

From "Survival and Event History Analysis" by Aalen et al, a cumulative incidence function (probability) for each event is derived directly from the rate function definitions and relationships:

$$F_z(t) = \int_0^t h_z(t) e^{-\int_0^t \sum_k h_k(t) dt} dt \tag{10}$$

Where $F_z(t)$ is a regular probability function (cumulative distribution function) for each event type.

Note that this is not an estimated quantity (yet). This is purely a mathematical derivation from the definitions of the rates, h_z. The probability associated with each event, F_k, is a function not only of that event's rate function, h_z, but all other event rates via $\sum_k h_k(t)$.

6.1.1 Cumulative Incidence with a Constant Rate

Starting from equation (10), what is a formula for cumulative incidence at time t if all rates are constant, $h_z(t) = h_z, \forall t, z$?

$$\begin{aligned} F_z(t) &= h_z \int_0^t e^{-t\sum_k h_k} dt \\ &= \frac{-h_z}{\sum_k h_k} e^{-t\sum_k h_k} - \frac{-h_z}{\sum_k h_k} e^{-0\sum_k h_k} \\ &= \frac{-h_z}{\sum_k h_k} e^{-t\sum_k h_k} + \frac{h_z}{\sum_k h_k} \\ &= \frac{h_z}{\sum_k h_k} (1 - e^{-t\sum_k h_k}) \end{aligned} \tag{11}$$

At this point, a cumulative incidence estimate can be calculated by simply plugging in a rate estimate as described above for the unadjusted case, equation (5), or the adjusted case, equation (9).

$$\hat{F}_z(t) = \frac{\hat{h}_z}{\sum_k \hat{h}_k} (1 - e^{-t\sum_k \hat{h}_k}) \tag{12}$$

This is a simple formula for a cumulative incidence in a competing risk framing under constant (hazard) rates. This is internally consistent with the earlier rate estimates, and interpretable as "given these average rates, we would see these cumulative incidence functions". Or perhaps, the rates are treated as the driver that leads to the resulting cumulative incidence.

7 References

- Aalen OO, Borgan Ø, Gjessing HK. *Survival and Event History Analysis: A Process Point of View*. Springer; 2008:114-115.