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[1]. 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.

[1]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 $\hat{\mathbf{h}_z} = exp(\mathbf{W}\hat{\theta_z})$[2].

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.

[2]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 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.

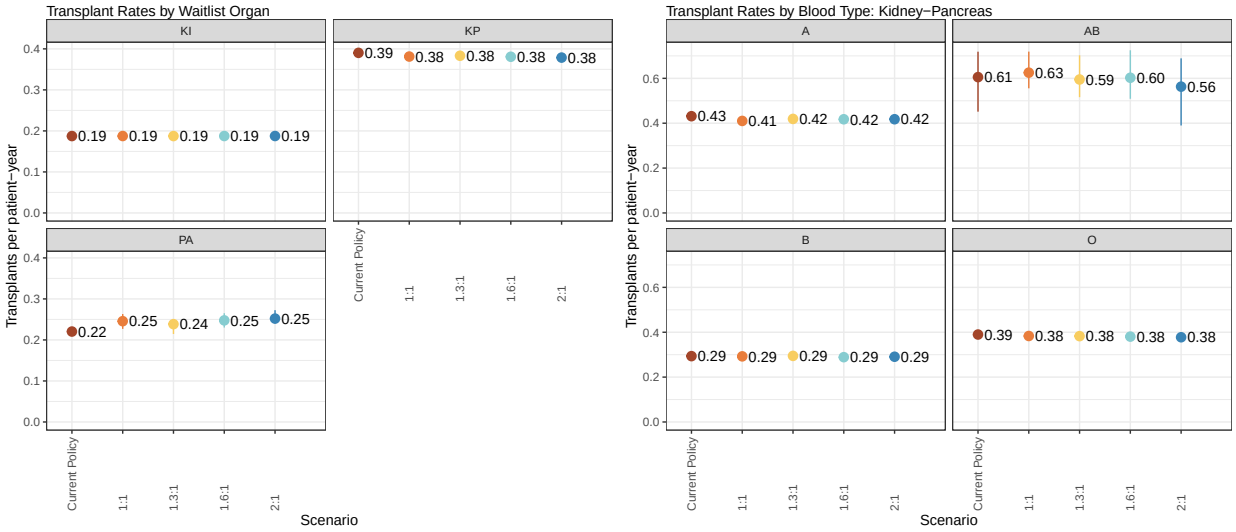

Figure 1: Example Rate Figures from KIPA2023_01

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[3]. Furthermore, there is a nice correspondence between the figures in the SRTR Annual Data Report.

[3]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[4]. 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[5]). 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[6] of rates during a single iteration should likely be a disqualifying result for a given COS. Diagnostic issues at this

[4]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.

[5]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.

[6]"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.