Title: Assessment of Utility of Total Kidney Volume for Trial Enrichment*

Authors:
John Lawrence
US Food and Drug Administration
10903 New Hampshire Ave.
Silver Spring, MD 20993-002
John.Lawrence@fda.hhs.gov

Abstract:
A consortium recently submitted an application to qualify Total Kidney Volume (TKV) as a biomarker for enriching trials of patients with Autosomal Dominant Polycystic Kidney Disease (ADPKD). TKV, in combination with patient age and baseline estimated glomerular filtration rate (eGFR), was qualified as a prognostic enrichment biomarker for ADPKD subjects at high risk for a progressive decline in renal function, defined as a confirmed 30% decline in eGFR. Risk scores can be derived from datasets to select patients to enrich future trials. The utility of this approach is discussed. We used a model from the historical datasets to define a risk score and applied it to a hypothetical trial. In the hypothetical trial, we simulated data according to a model that was fit to an independent dataset. We found the risk score may not select the higher risk patients very well. The R package SurvDisc can be used to try out different assumptions and enrichment criteria. This will be illustrated in the presentation.

Keywords: chronic kidney disease, mixed effects models, renal disease, measurement errors, counting process

---

* This article reflects the views of the author and should not be construed to represent FDA's views or policies.

Introduction. Consider a clinical trial where a measurement of function is taken on a patient at the start of the trial and at fixed times after randomization. The primary endpoint is the time to a pre-defined value of the measurement. This value can be a fixed value that is the same for all patients, a percent decline of the patient's baseline value, or an absolute amount of decline from the patient's baseline measurement. For example, in trials of Chronic Kidney Disease (CKD), the estimated Glomerular Filtration Rate (eGFR) is used as the measure of kidney function. This is measured at baseline and at pre-defined times (for example, every 3 months) after baseline. An event occurs when the patient's eGFR declines by a fixed amount, which could be a 30% decline, for example [1,2]. The event can be defined either as a single measurement that crosses the defined threshold or the definition can require a second measurement to confirm the decline. The latter definition (requiring confirmation) is more common in CKD trials. In CKD trials, there are several reasons for having a confirmation measurement. From the clinical standpoint, it can be required to rule out cases of Acute Kidney Injury (AKI), which are events where the kidney function has not decreased although temporarily the eGFR is markedly lower than the true kidney function. The confirmatory measurement can either occur at the next scheduled visit or there can be at an unscheduled visit at a short time after the initial qualifying event (e.g. 1 month after). The interval should be long enough so that the potential AKI can be resolved. Another example is long term maintenance trials for Major Depressive Disorder. There, the primary endpoint is the time to relapse, which can be defined as having a pre-defined score on the Hamilton Depression Rating Scale, the Montgomery-Asberg Depression Rating Scale, or the Clinical Global Impression Scale; the exact choice and the threshold varies from trial to trial [3]. The purpose of this article is to discuss survival analysis for events that are defined in any of these ways.

Let the time points where measurements are taken after baseline be $t_1, t_2, \dots, t_J$. Of note, the measurement times are assumed to be the same for all subjects, measurements are observed at all scheduled time points and there is no missing data. In practice, there is a time window for each visit. Small differences in the event time are not considered clinically relevant and do not reflect differences in outcome since the precise timing of the measurement at each visit is considered independent of the response. For example, if the first visit is scheduled for Day 90 and if one subject has the visit on Day 88 while a second subject has the visit on Day 93 and both subjects have an event, those two subjects are considered to have an event at the same time in the analysis. In addition, in many trials the primary endpoint is the time to the first event. After a patient has an event, there are several things that could happen and any are possible depending on the protocol for the trial: the patient could continue to be followed and stay on the randomized study drug, the patient could discontinue from the trial completely and have no further follow-up, the patient could be unblinded and go onto any open-label treatment strategy but continue to be followed and have measurements taken. Also, depending on the protocol, subjects in a clinical trial could have variable number of visits by design. For example, if the total duration of the trial is 5 years with visits every 3 months and 2 years of accrual, then by design the first patient could have a maximum of 20 visits where the last patient randomized could have a maximum of 12 visits. Other trials could have 12 visits (3 years) for every patient regardless of when they were randomized.

For subject i, we will observed J Bernoulli random variables: $B_{i,1}, B_{i,2}, \dots, B_{i,J}$ with $P[B_{i,j} = 1] = \pi_{i,j}$ and all of these Bernoulli random variables are independent (within a subject and between subjects). These Bernoulli random variables are defined as the indicator of the event that the measurement has declined by the fixed amount from baseline at the respective visit. In the following, we will assume the fixed amount is a 30% decline for illustration.

The rates $\pi_{i,j}$ can depend on subject-specific fixed and random effects. For example, the individual eGFR measurements can be modeled using a longitudinal mixed effects model. Suppose the true (unobserved) baseline GFR for patient i is denoted by $Y_{i,0}^*$ and the observed value of eGFR at baseline is denoted by

$Y_{i,0} = Y_{i,0}^* + \varepsilon_{i,0}$. Furthermore, suppose trt_i is the treatment assignment indicator (0 or 1) for subject i and there is a set of covariates measured at baseline, $\vec{X}_i'$, that modify the rate of change of GFR over time. This vector of covariates could include the latent variable $Y_{i,0}^*$. This model also includes a random slope and intercept. One possible longitudinal model for the measured eGFR at time t_j is defined by $Y_{i,j} = Y_{i,0}^* + \alpha_i + \left(\beta_i + \delta\ \text{trt}_i + \vec{X}_i'\vec{\theta}\right)t_j + \varepsilon_{i,j}$. Here, $\vec{\theta}$ is a fixed effect parameter vector. All measurement errors $\varepsilon_{i,j}$ are assumed normally distributed with mean 0 and variance σ^2; $(\alpha_i, \beta_i)'$ are bivariate normal random effects with mean $(0,0)'$ with some covariance matrix; all measurement errors and random effect vectors are independent within and between subjects. Alternatively, a linear mixed effects model can be used for the natural logarithm of the response variable with the obvious changes. Importantly, the actual data need not come from this model at all and the treatment effect can in reality be much more complicated than expressed in this simple model [4,5,6]. Indeed, part of the attractiveness of analyzing time to event is to avoid the reliance on the model assumptions needed for this longitudinal analysis. However, if the data actually do follow this model then

$$\begin{aligned}\pi_{i,j} = P\left[B_{i,j} = 1\right] &= P\left[Y_{i,j} < 0.7Y_{i,0}\right] \\ &= P\left[Y_{i,0}^* + \alpha_i + \left(\beta_i + \delta\ \text{trt}_i + \vec{\theta}\vec{X}_i'\right)t_j + \varepsilon_{i,j} < 0.7\left\{Y_{i,0}^* + \varepsilon_{i,0}\right\}\right] \\ &= P\left[\varepsilon_{i,j} < 0.7\varepsilon_{i,0} - 0.3Y_{i,0}^* - \alpha_i - \left(\beta_i + \delta\ \text{trt}_i + \vec{\theta}\vec{X}_i'\right)t_j\right] \\ &= \Phi\left(\frac{0.7\varepsilon_{i,0} - 0.3Y_{i,0}^* - \alpha_i - \left(\beta_i + \delta\ \text{trt}_i + \vec{\theta}\vec{X}_i'\right)t_j}{\sigma}\right)\end{aligned}$$

where $\Phi(x)$ denotes the standard normal distribution function. So, conditional on $Y_{i,0}^*$, $\varepsilon_{i,0}, \alpha_i$, and β_i, these $B_{i,j}$ are independent Bernoulli random variables with rates $\pi_{i,j}$ that can be calculated from the above formula. Table 1 shows the $\pi_{i,j}$ for other ways of defining the event based on this same longitudinal model.

Table 1. Formula for $\pi_{i,j} = P\left[B_{i,j} = 1\right]$ based on the longitudinal model used in the text for different definitions of events.

Definition of event	$\pi_{i,j}$
Decline of $\Delta \times 100\%$ from baseline	$\Phi\left(\frac{(1-\Delta)\varepsilon_{i,0} - \Delta Y_{i,0}^* - \alpha_i - \left(\beta_i + \delta\ \text{trt}_i + \vec{\theta}\vec{X}_i'\right)t_j}{\sigma}\right)$
Decline of absolute amount Δ from baseline	$\Phi\left(\frac{\varepsilon_{i,0} - \Delta - \alpha_i - \left(\beta_i + \delta\ \text{trt}_i + \vec{\theta}\vec{X}_i'\right)t_j}{\sigma}\right)$
Decline below an absolute constant M	$\Phi\left(\frac{\text{M} - Y_{i,0}^* - \alpha_i - \left(\beta_i + \delta\ \text{trt}_i + \vec{\theta}\vec{X}_i'\right)t_j}{\sigma}\right)$

Unconfirmed events. The survival function for subject i is

$$S_i(t_j) = P\left[\text{No event for subject } i \text{ up to and including time } t_j\right] = \prod_{k=1}^{j}\left\{1 - \pi_{i,k}\right\}$$

and the hazard rate is

$$h_i(t_j) = h_{i,j} = P\left[\text{Event at time } t_j \middle| \text{No event before time } t_j\right] = \pi_{i,j}$$

The probability mass function is

$$P\left[\text{First event for subject } i \text{ occurs at time } t_j\right] = \pi_{i,j}\prod_{k=1}^{j-1}\left\{1 - \pi_{i,k}\right\}$$

Take the longitudinal model described previously and assume the patient's true baseline function is 100, the mean response without measurement error is declining by 8 units per year so that function will decline by 30% from baseline (30 units) in 3.75 years. In the model, we could assume $Y_{i,0}^{*} = 100$, $\varepsilon_{i,0} = \alpha_i = \beta_i = \mathrm{trt}_i = 0$, and $\vec{\theta}\vec{X}_i' = -8$ so that $0.7\varepsilon_{i,0} - 0.3Y_{i,0}^{*} - \alpha_i - \left(\beta_i + \delta\,\mathrm{trt}_i + \vec{\theta}\vec{X}_i'\right)t_j = -30 + 8t_j$. Now, suppose the measurement error has standard deviation $\sigma = 5$ and measurements are taken every 3 months (indefinitely until the patient has an event for the sake of the following calculations). For instance, at time $t_{16} = 4$ years, we have $h_{i,16} = \pi_{i,16} = \Phi\left(\frac{-30+8(4)}{5}\right) \approx 0.6554$. But, under the same assumptions except changing the treatment assignment to $\mathrm{trt}_i = 1$ and assuming $\delta = 2$, we would have $h_{i,16} = \pi_{i,16} = \Phi\left(\frac{-30+6(4)}{5}\right) \approx 0.1151$. The hazard ratio for treatment is approximately 0.1756 at time $t_{16} = 4$, but would be approximately 0.0623 at time $t_8 = 2$. The hazard ratio is not constant under this model with these parameters. Furthermore, the expected time when the event will be observed (reverting back to the case $\mathrm{trt}_i = 0$) can be calculated numerically from the probability mass function as 3.506. In other words, the measurement error creates a negative bias in the estimated time to event even when there is no measurement error at baseline.

Confirmation of events at unscheduled visit. Assume that by the trial design, whenever a patient has a qualifying measurement below the threshold at a scheduled visit at time t_j, they will return for an unscheduled visit at time $t_j + \varepsilon$. If the scheduled visits are every 3 months, then ε could be approximately 2 to 4 weeks. The calculations are similar to those for unconfirmed events with the exception that the probability of a confirmed event at time t_j is

$$\pi_{i,j} P\left[\text{measurement is below threshold at uncheduled visit at time } t_j + \varepsilon\right]$$

With $\varepsilon = 0.1$ and the same longitudinal model assumptions used in the example for unconfirmed events together with an assumption that the measurement error at time $t_j + \varepsilon$ is independent and identically distributed to the other measurement errors, the hazard rate at time $t_{16} = 4$ years in the control group subject would be $h_{i,16} = \Phi\left(\frac{-30+8(4)}{5}\right)\Phi\left(\frac{-30+8(4.1)}{5}\right) \approx 0.4668$. The expected value of the time to the first confirmed event is approximately 3.92. So, the observed time to confirmed event is now positively biased in this case. The hazard ratio for the treatment effect is again not constant: approximately 0.0038 at year 2 and approximately 0.0345 at year 4. A rough approximation to the hazard ratio for this type of confirmation is the square of the hazard ratio for unconfirmed events (better for smaller values of ε).

Confirmation of events at scheduled visit. Next, we consider the problem of calculating the discrete time hazard function for confirmed events at scheduled visits. The definition of the survival function for subject i, as usual, is $S_i(t) = P[\text{no confirmed event for subject } i \text{ up to and including time } t]$. The survival function for the first two time points can be computed easily from the rates $\pi_{i,j}$ as follows: $S_i(t_1) = 1 - \pi_{i,1}\pi_{i,2}$ and $S_i(t_2) = 1 - \pi_{i,1}\pi_{i,2} - \pi_{i,2}\pi_{i,3} + \pi_{i,1}\pi_{i,2}\pi_{i,3}$. For $2 < j < J$, the survival function can be computed recursively. First, note that

$$1 - S_i(t_j) = P\left[\bigcup_{k=1}^{j}\left\{B_{i,k} = B_{i,k+1} = 1\right\}\right]$$

$$= P\left[\left\{B_{i,j} = B_{i,j+1} = 1\right\} \text{ or } \bigcup_{k=1}^{j-1}\left\{B_{i,k} = B_{i,k+1} = 1\right\}\right]$$

$$= \pi_{i,j}\pi_{i,j+1} + \left\{1 - S_i(t_{j-1})\right\} - P\left[\left\{B_{i,j} = B_{i,j+1} = 1\right\} \text{ and } \bigcup_{k=1}^{j-1}\left\{B_{i,k} = B_{i,k+1} = 1\right\}\right]$$

then observe that

$$1 - S_i(t_j) = P\left[\bigcup_{k=1}^{j}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$

$$= P\left[\{B_{i,j-1} = B_{i,j} = 1\} \text{ or } \{B_{i,j} = B_{i,j+1} = 1\} \text{ or } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$

$$= P[B_{i,j-1} = B_{i,j} = 1] + P[B_{i,j} = B_{i,j+1} = 1] + P\left[\bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$
$$- P[\{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \{B_{i,j} = B_{i,j+1} = 1\}]$$
$$- P\left[\{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$
$$- P\left[\{B_{i,j} = B_{i,j+1} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$
$$+ P\left[\{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \{B_{i,j} = B_{i,j+1} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$

$$= \pi_{i,j-1}\pi_{i,j} + \pi_{i,j}\pi_{i,j+1} + \{1 - S_i(t_{j-2})\} - \pi_{i,j-1}\pi_{i,j}\pi_{i,j+1}$$
$$- P\left[\{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right] - \pi_{i,j}\pi_{i,j+1}\{1 - S_i(t_{j-2})\}$$
$$+ P\left[\{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \{B_{i,j} = B_{i,j+1} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$

Now, the prior equation implies

$$P\left[\{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right] = \pi_{i,j-1}\pi_{i,j} + S_i(t_{j-1}) - S_i(t_{j-2})$$

Also, we can see that

$$P\left[\{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \{B_{i,j} = B_{i,j+1} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$

$$= P\left[\{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$

$$\times P\left[\{B_{i,j} = B_{i,j+1} = 1\} \;\middle|\; \{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]$$

$$= P\left[\{B_{i,j-1} = B_{i,j} = 1\} \text{ and } \bigcup_{k=1}^{j-2}\{B_{i,k} = B_{i,k+1} = 1\}\right]\pi_{i,j+1}$$

Hence,

$$1 - S_i(t_j)$$
$$= \pi_{i,j-1}\pi_{i,j} + \pi_{i,j}\pi_{i,j+1} + \{1 - S_i(t_{j-2})\} - \pi_{i,j-1}\pi_{i,j}\pi_{i,j+1} - \{\pi_{i,j-1}\pi_{i,j} + S_i(t_{j-1}) - S_i(t_{j-2})\}$$
$$- \pi_{i,j}\pi_{i,j+1}\{1 - S_i(t_{j-2})\} + \{\pi_{i,j-1}\pi_{i,j} + S_i(t_{j-1}) - S_i(t_{j-2})\}\pi_{i,j+1}$$
$$= 1 + S_i(t_{j-1})\{\pi_{i,j+1} - 1\} + S_i(t_{j-2})\{\pi_{i,j}\pi_{i,j+1} - \pi_{i,j+1}\}$$

and finally,

$$S_i(t_j) = S_i(t_{j-1})\{1 - \pi_{i,j+1}\} + S_i(t_{j-2})\pi_{i,j+1}\{1 - \pi_{i,j}\}$$

Since there are no measurements after time t_J that could confirm a possible qualifying measurement at this time, $S_i(t_J) = S_i(t_{J-1})$.

The hazard function at time t_j, denoted by $h_i(t_j)$ or simply $h_{i,j}$, is the probability of having a confirmed event at time t_j given that there was no event before that time. Hence, $S_i(t_j) = \{1 - h_{i,j}\}S_i(t_{j-1})$ for $j > 1$. Thus, $h_{i,1} = 1 - S_i(t_1)$ and the hazard function can be calculated from $h_{i,j} = 1 - \frac{S_i(t_j)}{S_i(t_{j-1})}$ for $j > 1$.

With the same assumptions used in the example for unconfirmed events, the hazard rate at time $t_{16} = 4$ years in the control group subject would be $h_{i,16} \approx 0.4159$. The expected value of the time to the first confirmed event is approximately 3.93. The hazard ratio for the treatment effect is approximately 0.0037 at year 2 and approximately 0.048 at year 4.

$\mathrm{N}_i(t_k) = \sum_{j=1}^{k} B_{i,j}B_{i,j+1}$ is a discrete time counting process that counts the number of events up to and including time t_k. We can see that this is a non-Markov process by considering two scenarios where $\mathrm{N}_i(t_2) = 1$. In one case, $P[\mathrm{N}_i(t_3) = 2 | B_{i,1} = B_{i,2} = 1 \text{ and } B_{i,3} = 0] = 0$ but in a second case $P[\mathrm{N}_i(t_3) = 2 | B_{i,2} = B_{i,3} = 1 \text{ and } B_{i,1} = 0] = \pi_{i,4}$. In the case of unconfirmed events or confirmation at unscheduled visits, the corresponding counting processes have the Markov property.

Summary and conclusions. There are some notable issues in the analysis of trials where the endpoint is the time to a pre-defined value of some measured endpoint and the measurements are at fixed times. First, time is discrete and special attention should be made to how the ties are handled. Small differences in the random timing of visits should be ignored in many cases and all events at the same visit should be treated as occurring at the same time in those cases. The frequency of ties could be larger than usual. There are several common methods for handling ties used in commercial statistical software packages [7, 8, 9]. The method for grouped survival data described in [10] is the preferred method but is not available in many of those packages. Efron's method [8] tends to be the best among the three generally available methods [11]. The proportional hazards assumption is often not correct, so attention should be paid to the way the treatment effect is described. The estimated hazard ratio may still be a reasonable summary of the treatment effect [12]. We provided formulas in this article for the survival function, the hazard function, and the probability mass function for events with confirmation (in two different definitions of confirmation) and without confirmation. We then compared the hazard functions and hazard ratios under one example model for the longitudinal data. The expected value of the time to event is negatively biased with measurement error and no confirmation, but was positively biased with confirmation.

References

1. Levey AS, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *American Journal of Kidney Diseases* **2014; 64(6)**, pp. 821-835.

2. Thompson A, Lawrence J, Stockbridge N. GFR decline as an end point in trials of CKD: a viewpoint from the FDA. *American Journal of Kidney Diseases* **2014; 64(6)**, pp. 836-837.

3. Borges S, et al. Review of maintenance trials for major depressive disorder: a 25-year perspective from the US Food and Drug Administration. *The Journal of Clinical Psychiatry* **2014; 75(3)**, pp. 205-214.

4. Greene T, et al. Utility and validity of estimated GFR–based surrogate time-to-event end points in CKD: a simulation study. *American Journal of Kidney Diseases* **2014; 64(6)**, pp. 867-879.

5. Greene T. A model for a proportional treatment effect on disease progression. *Biometrics* **2001; 57(2)**, pp. 354-360.

6. Levey AS, Greene T, Beck GJ, et al. Dietary protein restriction and the progression of chronic renal disease: what have all of the results of the MDRD Study shown? *Journal of the American Society of Nephrology* **1999; 10(11)**, pp. 2426-2439.

7. Breslow NE. Analysis of Survival Data under the Proportional Hazards Model. *International Statistical Review / Revue Internationale de Statistique* **1975; 43(1)**, pp. 45–57. doi:10.2307/1402659.

8. Efron B. The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association* **1974; 72(359)**, pp. 557–565. doi:10.1080/01621459.1977.10480613.

9. Kalbfleisch JD and Prentice RL. *The statistical analysis of failure time data.* John Wiley & Sons: New Jersey, 2011.

10. Prentice RL and Gloeckle LA. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 1978; 57-67.

11. Hertz-Picciotto I and Rockhill B. Validity and efficiency of approximation methods for tied survival times in Cox regression. *Biometrics* 1997; 1151-1156.

12. Yang P and Fleming TR. Simultaneous Use of Weighted Logrank and Standardized Kaplan-Meier Statistics. *Journal of Biopharmaceutical Statistics* **2006; 16(2)**, 241-252.