Published in final edited form as:

Stat Med. 2014 March 15; 33(6): 958–970. doi:10.1002/sim.6005.

Sample size estimation for time-dependent ROC

H. Li^{a,*} and C. Gatsonis^b

^aDepartment of Preventive Medicine, Rush University Medical Center, Chicago, IL 60612

^bDepartment of Biostatistics, Brown University, Providence, RI 02912

Abstract

In contrast to the usual ROC analysis with a contemporaneous reference standard, the time-dependent setting introduces the possibility that the reference standard refers to an event at a future time and may not be known for every patient due to censoring. The goal of this research is to determine the sample size required for a study design to address the question of the accuracy of a diagnostic test using the area under the curve in time-dependent ROC analysis. We adapt a previously published estimator of the time-dependent AUC which is a function of the expected conditional survival functions. This estimator accommodates censored data. The estimation of the required sample size is based on approximations of the expected conditional survival functions and their variances, derived under parametric assumptions of an exponential failure time and an exponential censoring time. We also consider different patient enrollment strategies. The proposed method can provide an adequate sample size to ensure that the test's accuracy is estimated to a prespecified precision. We present results of a simulation study to assess the accuracy of the method and its robustness to departures from the parametric assumptions. We apply the proposed method to design of a study of PET as predictor of disease free survival in women undergoing therapy for cervical cancer.

Keywords

Censoring; Parametric survival model; Sample size estimation; Time-dependent AUC; Time-dependent ROC; Uniform accrual period

1. Introduction

The methodology of ROC curves to evaluate diagnostic accuracy of tests has been well established in the literature (Pepe [1] and Zhou et al. [2]). The area under the ROC curve (AUC) is a commonly used summary measure of diagnostic accuracy and it can be interpreted as the probability that the test result from a randomly chosen individual with the target condition exceeds that from a randomly chosen individual without the target condition. In the usual ROC formulation, presence or absence of the target condition is a characteristic of a study subject that is fixed in time.

ROC analysis has also been proposed for situations in which the target condition refers to a future event, such as onset of disease, disease progression, or death. This "time-dependent" ROC analysis in effect derives ROC curves for the prediction of events at future time points, i.e. subjects are initially nondiseased but can become diseased during the course of the study. For the time-dependent ROC curve, the test result is the value of the baseline biomarker and the binary reference standard is an indicator of whether event (e.g. death) occurred by the time of interest. Thus the time-dependent ROC analysis permits an assessment of how the predictive ability of a marker may vary over time [20].

The usual measures of diagnostic performance, such as sensitivity, specificity, and AUC can be defined and estimated for each time point. The average AUC over all time points has also been proposed as an appropriate measure of performance over a time range. Time-dependent ROC methods were discussed by Heagerty et al. [3], Heagerty and Zheng [4], and several other authors in recent years. Chambless et al. [5] presented an estimator of $AUC(t)$ (AUC for prediction at time t) which is based on conditional survival functions estimated by survival analysis methods. These authors showed via simulation studies that the estimator of $AUC(t)$ is unbiased, efficient, and performs well in practical situations. The estimation of time-dependent ROC curves and their functionals needs to account for the fact that censoring may have occurred in the study. In addition, several other researchers studied time-dependent ROC for several different scenarios, for instance, Zheng and Heagerty ([6] and [7]) introduced approaches for constructing time-dependent ROC for longitudinal data. Uno et al. [8] proposed a modified C-statistic which could consistently estimate a conventional concordance measure free of censoring. Saha and Heagerty [9] developed the time-dependent ROC in the presence of competing risks. Chiang et al. [10] developed a nonparametric estimator for the time-dependent AUC.

In this paper, we discuss the estimation of sample size for a study designed to address the question of how well a diagnostic marker measured at a particular point in time (e.g. baseline) can distinguish between subjects who experience a particular outcome and subjects who do not in a follow-up interval $[0, t]$. For example, a common goal in studies of Positron Emission Tomography (PET) in cancer is to investigate the accuracy of the change in standardized uptake value from baseline to an early point in the course of therapy to predict overall survival. We consider two possible enrollment strategies to the study: a) all patients enter the study at the same time (single entry) and b) patients enter the study over a period of time (staggered entry). In the latter case, there is an accrual period, during which patients enter the study at a certain rate and then they are followed for an additional time period. We consider both patient enrollment strategies in this paper. As is well known in survival analysis, failure times may be censored. We consider two basic types of censorship in this paper: a) during both the accrual and follow-up period, patients may be lost to follow-up, and b) when the end of follow-up is reached, patients who have no events are said to be "administratively censored".

In Section 2, we describe how to estimate the variance of the estimator of the time-dependent AUC of Chambless et al. [5] and how to estimate the sample size using this estimator. We assume an exponential failure time and an exponential censoring time and consider two scenarios for patient enrollment: single entry and staggered entry. In Section 3,

we test the performance and assess the robustness of the proposed method using simulated data. In Section 4, we apply the proposed method to design of a study of PET as predictor of disease free survival in women undergoing therapy for cervical cancer. We discuss our findings in Section 5.

2. Methods

In this section we first detail the specific notation, assumptions, and basic models and then we present the derivation to obtain the required sample size. To simplify the presentation, we consider death as the event of interest.

2.1. Assumptions and Notation

We make the following assumptions:

1. For staggered entry, during the accrual period of r years, patients enter the study according to a uniform distribution and are followed for additional f years. The time point of interest in the time-dependent ROC analysis is denoted by τ which is equal to $r + f$.
2. For the derivation of the sample size estimate, we use a fully parametric model. Specifically:
 - i. All measurements are assumed independent for different patients.
 - ii. The times from entry to death are assumed to be exponential distribution with rate λ .
 - iii. The intervals between entry and loss to follow-up are assumed to be independent of the entry and death times.
 - iv. The loss to follow-up times are assumed to follow an exponential distribution with rate γ .
3. The accrual time is assumed to be independent of the failure time.

Let Y_i denote the baseline biomarker measurement for patient i , for a total of N patients. The baseline measurements (Y_i, Y_j) for a pair of patients are assumed to be independent. We define $t_i = \min(T_i, C_i)$ where T_i and C_i are the death time and censoring time for patient i , respectively, and are assumed to be independent random variables.

The indicator of the true disease status for patient i at time τ is defined as

$$D_i(\tau) = \begin{cases} 1 & \text{if patient } i \text{ is dead within } [0, \tau] \\ 0 & \text{if patient } i \text{ is alive at time } \tau \\ \text{missing} & \text{if patient } i \text{ is censored before time } \tau \end{cases}$$

Patients whose disease status are missing could be either dead or alive at time τ . The censoring indicator of the failure time is defined as

$$\delta_i = \begin{cases} 1 & \text{if } T_i \leq C_i \\ 0 & \text{otherwise} \end{cases}$$

and the censoring indicator for patient i at time τ is defined as $\delta_i(\tau) = I(t_i \leq \tau)(1 - \delta_i)$, where

$$\Delta_i(\tau) = \begin{cases} 1 & \text{if patient } i \text{ is censored within } [0, \tau] \\ 0 & \text{otherwise} \end{cases}$$

We denote the probability that patients are alive, dead, or censored at time τ as follows:

$$\begin{aligned} p_a(\tau) &= P(D(\tau) = 0) \\ p_d(\tau) &= P(D(\tau) = 1) \\ p_c(\tau) &= P(D(\tau) = \text{missing}) \end{aligned}$$

We suppress the time τ in the notation for $p_a(\tau)$, $p_d(\tau)$, and $p_c(\tau)$ in the rest of this paper.

2.2. Estimation of the Area Under a Time-dependent ROC Curve

Chambless et al. [5] defined the time-dependent AUC at time τ as

$$\begin{aligned} AUC(\tau) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u|D(\tau)=1) g(v|D(\tau)=0) du dv \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1-s(\tau|u))s(\tau|v)g(u)g(v)dudv \\ &= \frac{E[1-s(\tau|Y)]E[s(\tau|Y)]}{E[(1-s(\tau|U))s(\tau|V)I(U>V)]} \\ &= \frac{E[1-s(\tau|Y)]E[s(\tau|Y)]}{E[(1-s(\tau|U))s(\tau|V)I(U>V)]} \end{aligned}$$

where g represents a density function, s represents a survival function, U and V are independent observations of Y , and $I(U > V)$ is the indicator function for $U > V$ (1 when true, 0 otherwise). They also proposed the following estimator for $AUC(\tau)$

$$\hat{AUC}(\tau) = \frac{\hat{E}[(1-s(\tau|U))s(\tau|V)I(U>V)]}{\hat{E}[1-s(\tau|Y)]\hat{E}[s(\tau|Y)]} \quad (1)$$

and suggested that each expectation is estimated as a sample mean.

2.3. Joint Model of Markers and Failure Time

In order to derive the estimate of the required sample size, we assume a fully parametric model. In particular, the failure time follows an exponential distribution with parameter λ . The hazard function for patient i is defined as

$$\lambda_i = \exp(\beta_0 + \beta_1 Y_i)$$

which is a function of this patient's baseline biomarker measurement Y_i . The survival function for patient i given the baseline biomarker measurement Y_i is

$$s_i(\tau|Y_i) = \exp(-\exp(\beta_0 + \beta_1 Y_i) \tau) \quad (2)$$

The estimated variances of the regression coefficients can be approximated as

$$\begin{aligned} \text{var}(\hat{\beta}_0^*) &\approx \frac{1}{N(1-\hat{p}_c)} \\ \text{var}(\hat{\beta}_1) &\approx \frac{1}{M_1 N(1-\hat{p}_c)} \end{aligned}$$

where $M_1 = \text{var}Y$ and $\hat{p}_c$ is the estimated value of p_c . In order to simplify the variance estimation for regression coefficients, we center the biomarker value and $\beta_0^* = \beta_0 + \beta_1 \bar{Y}$, where $\bar{Y}$ is the mean of the biomarker measurements. The detailed derivation of the variance terms (Appendix A) proceeds similarly to the approach of Bernardo et al. [11].

For our study setting, every patient's biomarker measurement is measured at baseline but the "disease" status (vital status in our case) is measured at a future time τ . According to equation (2), survival is a function of baseline biomarker measurements. For example, patients with larger baseline biomarker values may have higher risk to death and may die within the interval $[0, \tau]$ and patients with smaller baseline biomarker values may have lower risk to death and may survive past time τ . Therefore, it is reasonable to assume different baseline biomarker distributions for patients with different disease status at time τ . Given the true disease status at time τ , we assume the baseline biomarker follows a normal distribution, specified conditionally on failure status at time τ .

1. for patient i who is dead within $[0, \tau]$

$$Y_i|D_i(\tau) = \text{INormal}(\mu_d, \sigma_d^2)$$

2. for patient i who is alive at time τ

$$Y_i|D_i(\tau) = \text{ONormal}(\mu_a, \sigma_a^2)$$

3. for patient i who is censored before time τ

$$Y_i|D_i(\tau) = \text{missingONormal}(\mu_c, \sigma_c^2)$$

where

$$\begin{aligned} \mu_c &= p_0 \mu_d + (1 - p_0) \mu_a \\ \sigma_c^2 &= p_0^2 \sigma_d^2 + (1 - p_0)^2 \sigma_a^2 \\ p_0 &= \frac{P(\Delta(\tau)=1, D(\tau)=1)}{P(\Delta(\tau)=1)} \end{aligned}$$

Note that p_0 is the probability that a patient who was censored before time τ would actually be dead at time τ . We assume that $p_0 = p_d$. The detailed derivation of the biomarker distribution for patients censored before τ is given in Appendix B.

2.4. Sample Size Estimation

Depending on the primary goal of a diagnostic accuracy study, the derivation of the sample size may be based on hypothesis testing or estimation via confidence intervals ([1] and [2]). In this paper we derive the sample size required to ensure that the test accuracy is estimated to a prespecified precision (half-length of the corresponding confidence interval). In this section, we discuss the derivation of sample size based on the estimator of $AUC(\tau)$ shown in equation (1). The asymptotic $(1 - \alpha)\%$ confidence interval for $\hat{AUC}(\tau)$ which is $\left[\hat{AUC}(\tau) - Z_{1-\alpha/2} * sd(\hat{AUC}(\tau)), \hat{AUC}(\tau) + Z_{1-\alpha/2} * sd(\hat{AUC}(\tau)) \right]$ where $sd(\hat{AUC}(\tau))$ is the standard error of $\hat{AUC}(\tau)$. If we assume the length of the confidence interval for $\hat{AUC}(\tau)$ is $2l$, then

$$Z_{1-\alpha/2}^2 var(\hat{AUC}(\tau)) = l^2 \quad (3)$$

In order to estimate the sample size, we first need an approximation for $var(\hat{AUC}(\tau))$, which is a function of sample size N and other parameters. Taking log transformations on both sides of equation (1), we obtain

$$\log(\hat{AUC}(\tau)) = \log \hat{E}_0 - \log(\hat{E}[1 - s(\tau|Y)]) - \log(\hat{E}[s(\tau|Y)]) \quad (4)$$

where

$$\hat{E}_0 = \hat{E}[(1 - s(\tau|U)) s(\tau|V) I(U > V)]$$

Using the delta method (chapter 7 in Casella et al. [12]), we obtain

$$var(\hat{AUC}(\tau)) \approx (\hat{AUC}(\tau))^2 var(\log(\hat{AUC}(\tau))) \quad (5)$$

From equation (4), we have

$$\begin{aligned} var(\log(\hat{AUC}(\tau))) = & var(\log \hat{E}_0) + var(\log(\hat{E}[1 - s(\tau|Y)])) + var(\log(\hat{E}[s(\tau|Y)])) \\ & - 2cov(\log \hat{E}_0, \log(\hat{E}[1 - s(\tau|Y)])) - 2cov(\log \hat{E}_0, \log(\hat{E}[s(\tau|Y)])) \\ & + 2cov(\log(\hat{E}[1 - s(\tau|Y)]), \log(\hat{E}[s(\tau|Y)])) \end{aligned} \quad (6)$$

Equation (6) includes 6 variance and covariance terms. By repeated applications of the delta method, all variance and covariance terms are functions of

$\hat{E}[s(\tau|Y)]$, $\hat{E}[1 - s(\tau|Y)]$, $\hat{E}[(1 - s(\tau|U)) s(\tau|V) I(U > V)]$, the variance of each estimated expectation, and the covariance between them (detailed derivation for each of variance and covariance terms is given in Appendix E).

2.4.1 Approximation and Estimation of Expected Survival Functions—In this section, we discuss how to estimate and approximate the expected conditional survival

functions according to different patient enrollment strategies (detailed derivations of all estimations and approximations are given in Appendix D).

2.4.1.1 Estimation of Expected Survival Functions: The expectation of the conditional survival function can be estimated by the sample mean, i.e.

$$\hat{E}[s(\tau|Y)] = \frac{\sum_i s(\tau|Y_i)}{N} \quad (7)$$

In order to estimate the expectation, we group patients into three groups: dead, alive, and censored patients at time τ and equation (7) can be written as

$$\hat{E}[s(\tau|Y)] = \hat{p}_d \hat{E}s_d + \hat{p}_a \hat{E}s_a + \hat{p}_c \hat{E}s_c \quad (8)$$

where s_d , s_a , and s_c are the survival functions of dead, alive, and censored patients at time τ . The probabilities p_d , p_a , and p_c can be estimated using the assumed distributions for failure time and censoring time, and they are functions of the accrual period r , the follow-up period f , the loss to follow-up rate γ , the estimated regression coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$, and the biomarker values (detailed derivations are given in Appendix C). Similarly,

$$\hat{E}[1 - s(\tau|Y)] = \hat{p}_d \hat{E}[1 - s_d] + \hat{p}_a \hat{E}[1 - s_a] + \hat{p}_c \hat{E}[1 - s_c] \quad (9)$$

and

$$\begin{aligned} \hat{E}_0 &= \hat{E}[(1 - s(\tau|U))s(\tau|V)I(U > V)] \\ &= \frac{\sum_{i,j} (1 - s(\tau|U_i))s(\tau|V_j)I(U_i > V_j)}{N_1} \end{aligned} \quad (10)$$

where N_1 is the total number of comparisons for all (U, V) pairs of distinct biomarker observations. We group the summation in equation (10) into the following nine groups: dead patients vs. alive patients (da), alive patients vs. dead patients (ad), dead patients vs. censored patients (dc), censored patients vs. dead patients (cd), censored patients vs. alive patients (ca), alive patients vs. censored patients (ac), censored patients vs. censored patients (cc'), dead patients vs. dead patients (dd'), and alive patients vs. alive patients (aa'). Then equation (10) can be rewritten as

$$\begin{aligned} \hat{E}_0 &= \hat{p}_a \hat{p}_d \hat{q}_1 \hat{E}[(1 - s_d)s_a] + \hat{p}_a \hat{p}_d \hat{q}_2 \hat{E}[(1 - s_a)s_d] + \hat{p}_d \hat{p}_c \hat{q}_3 \hat{E}[(1 - s_d)s_c] + \hat{p}_c \hat{p}_d \hat{q}_4 \hat{E}[(1 - s_c)s_d] \\ &+ \hat{p}_c \hat{p}_a \hat{q}_5 \hat{E}[(1 - s_c)s_a] + \hat{p}_c \hat{p}_a \hat{q}_6 \hat{E}[(1 - s_a)s_c] + \hat{p}_d \hat{p}_d \hat{q}_7 \hat{E}[(1 - s_d)s_d] + \hat{p}_d \hat{p}_d \hat{q}_8 \hat{E}[(1 - s_d)s_d] \\ &+ \hat{p}_c \hat{p}_c \hat{q}_9 \hat{E}[(1 - s_c)s_c] \end{aligned} \quad (11)$$

The detailed definitions and estimations of proportions q_1 to q_9 are given in Appendix F. Since we need to form distinct biomarker pairs when we compare dead with dead, alive with alive, and censored with censored patients in equation (10), we use s_d to denote the survival function for one dead patient who forms a pair with another dead patient whose survival function is denoted by $s_{d'}$. The definition and estimation of s_d and $s_{d'}$ are the same. The similar idea applies to s_a and $s_{a'}$, and s_c and $s_{c'}$.

2.4.1.2 Approximation of Estimated Expectations of Survival Functions: In this section, we discuss how to approximate the estimated expectations (8), (9), and (11) for single and staggered entries.

1) Single Entry: When all patients enter the study simultaneously, say at time zero, r is equal to zero and we evaluate the survival at time τ . Therefore, after centering the biomarker value, the survival function (2) can be estimated as

$$\hat{s}_i(\tau|Y_i) = \exp\left(-\exp\left(\hat{\beta}_0^* + \hat{\beta}_1 Y_i^*\right)\tau\right)$$

where Y_i^* is the centered biomarker value. In order to approximate the survival function, we need to specify the biomarker values. We use the assumed mean biomarker value of dead, alive, and censored patients to replace those of individual dead, alive, and censored patients (see also Elandt-Johnson et al. [13]). Therefore, the expected survival functions for dead patients can be approximated as

$$\begin{aligned}\hat{E}s_h &\approx \hat{E}\hat{s}_h \\ &= \exp\left(-\exp\left(\hat{\beta}_0^* + \hat{\beta}_1 \mu_h^*\right)\tau\right)\end{aligned}\quad (12)$$

where $h = d, a, c$ and μ_h^* is the mean of the centered biomarker values for h . We can derive the approximation of $\hat{E}[1 - s_h]$ and $\hat{E}[(1 - s_h)s_k]$ using the similar method (see Appendix D).

2) Staggered Entry: The estimator of $AUC(\tau)$ for staggered entry is still represented by equation (1), but we need to estimate the survival functions differently. Again by using the assumed mean biomarker value of dead, alive, and censored patients to replace those of individual dead, alive, and censored patients, the survival function of dead patients can be approximated as

$$\begin{aligned}\hat{E}s_h &\approx \hat{E}\hat{s}_h \\ &= \frac{\exp(-\hat{\lambda}_h f) - \exp(-\hat{\lambda}_h \tau)}{\hat{\lambda}_h r}\end{aligned}\quad (13)$$

where

$$\hat{\lambda}_h = \exp\left(\hat{\beta}_0^* + \hat{\beta}_1 \mu_h^*\right)$$

and $h = d, a, c$ and μ_h^* is the mean of the centered biomarker values for h . We can derive the approximation of $\hat{E}[1 - s_h]$ and $\hat{E}[(1 - s_h)s_k]$ using the similar method (see Appendix D).

Hence, for single entry or staggered entry, equation (8) can be approximated as

$$\hat{E}[s(\tau|Y)] \approx \hat{p}_d \hat{E}\hat{s}_d + \hat{p}_a \hat{E}\hat{s}_a + \hat{p}_c \hat{E}s_c \quad (14)$$

Similarly equation (9) and (11) can be approximated as

$$\hat{E}[1 - s(\tau|Y)] \approx \hat{p}_d \hat{E}[1 - \hat{s}_d] + \hat{p}_a \hat{E}[1 - \hat{s}_a] + \hat{p}_c \hat{E}[1 - \hat{s}_c] \quad (15)$$

and

$$\begin{aligned} \hat{E}_0 \approx & \hat{p}_d \hat{p}_a \hat{q}_1 \hat{E}[(1 - \hat{s}_d) \hat{s}_a] + \hat{p}_d \hat{p}_a \hat{q}_2 \hat{E}[(1 - \hat{s}_a) \hat{s}_d] + \hat{p}_d \hat{p}_c \hat{q}_3 \hat{E}[(1 - \hat{s}_d) \hat{s}_c] \\ & + \hat{p}_d \hat{p}_c \hat{q}_4 \hat{E}[(1 - \hat{s}_c) \hat{s}_d] + \hat{p}_c \hat{p}_a \hat{q}_5 \hat{E}[(1 - \hat{s}_c) \hat{s}_a] + \hat{p}_c \hat{p}_a \hat{q}_6 \hat{E}[(1 - \hat{s}_a) \hat{s}_c] \\ & + \hat{p}_d \hat{p}_d \hat{q}_7 \hat{E}[(1 - \hat{s}_d) \hat{s}_d] + \hat{p}_a \hat{p}_a \hat{q}_8 \hat{E}[(1 - \hat{s}_a) \hat{s}_a] + \hat{p}_c \hat{p}_c \hat{q}_9 \hat{E}[(1 - \hat{s}_c) \hat{s}_c] \end{aligned} \quad (16)$$

Equations (14)-(16) are functions of $f, r, \hat{\beta}_0, \hat{\beta}_1, \gamma, \mu_a$, and μ_d . According to different patient enrollment strategies, p_a, p_d and p_c are estimated using the corresponding equations in Appendix C, and $q_1 - q_9$ are estimated using the corresponding equations in Appendix F.

2.4.2 Approximation of Variance and Covariance of the Estimated

Expectations—Using equations (14)-(16), we approximate the estimates of the following 6 variances and covariances:

$$\text{var}(\hat{E}_0), \text{var}(\hat{E}[s(\tau|Y)]), \text{var}(\hat{E}[1 - s(\tau|Y)]), \text{cov}(\hat{E}_0, \hat{E}[1 - s(\tau|Y)]), \text{cov}(\hat{E}_0, \hat{E}[s(\tau|Y)]) \quad \text{and} \quad . \text{ For example,}$$

$$\begin{aligned} \text{var} \hat{E}[s(\tau|Y)] & \approx \text{var} \hat{E}[\hat{s}(\tau|Y)] \\ & = \text{var}(\hat{p}_d \hat{E} \hat{s}_d) + \text{var}(\hat{p}_a \hat{E} \hat{s}_a) + \text{var}(\hat{p}_c \hat{E} \hat{s}_c) \end{aligned}$$

We can derive the variance for $\hat{p}_d \hat{E} \hat{s}_d, \hat{p}_a \hat{E} \hat{s}_a$, and $\hat{p}_c \hat{E} \hat{s}_c$ using the delta method. Similarly, we can derive the other 5 variance and covariance terms. These variances and covariances are functions of $\text{var} \hat{\lambda}_d, \text{var} \hat{\lambda}_a$, and $\text{var} \hat{\lambda}_c$, where

$$\text{var} \hat{\lambda}_h = \hat{\lambda}_h^2 \left(\text{var} \hat{\beta}_0^* + (\mu_h^*)^2 \text{var} \hat{\beta}_1 \right)$$

and $h = d, a, c$. Since the approximation and estimation of the variance and covariance are complicated, we present the detailed derivations in Appendix E.

2.4.3 Approximation of the Variance of $A\hat{U}C(\tau)$ —In the previous sections, we describe how to approximate

$\hat{E}[s(\tau|Y)], \hat{E}[1 - s(\tau|Y)], \hat{E}[(1 - s(\tau|U)) s(\tau|V) I(U > V)]$, the variance of each, and the covariance between them according to different patient enrollment strategies. In this section, we describe how to approximate the variance of $A\hat{U}C(\tau)$.

As we mentioned in Section 2.4, equation (6) is a function of

$\hat{E}[s(\tau|Y)], \hat{E}[1 - s(\tau|Y)], \hat{E}[(1 - s(\tau|U)) s(\tau|V) I(U > V)]$, the variance of each estimated expectation, and the covariance between them. After substituting the approximate variances and covariances derived in Section 2.4.2 into equation (6), we obtain

$$\begin{aligned} \text{var} \left(\log \hat{AUC}(\tau) \right) &\approx S_1 \text{var} \hat{\beta}_0^* + S_2 \text{var} \hat{\beta}_1 \\ &= \frac{S_1}{N(1-\hat{p}_c)} + \frac{S_2}{M_1 N(1-\hat{p}_c)} \end{aligned} \quad (17)$$

where S_1 and S_2 are functions of $f, r, \hat{\beta}_0, \hat{\beta}_1, \gamma, \mu_a, \mu_d$ and three approximate expectations $\hat{E}[\hat{s}(\tau|Y)]$, $\hat{E}[1 - \hat{s}(\tau|Y)]$ and $\hat{E}[(1 - \hat{s}(\tau|U)) \hat{s}(\tau|V) I(U > V)]$. The detailed expressions of S_1 and S_2 are given in Appendix E.

Following equations (5) and (17), we obtain the approximation of the variance of $\hat{AUC}(\tau)$, which is

$$\begin{aligned} \text{var} \hat{AUC}(\tau) &\approx \left(\hat{AUC}(\tau) \right)^2 \text{var} \left(\log \hat{AUC}(\tau) \right) \\ &= \frac{\left(\hat{AUC}(\tau) \right)^2 S_1}{N(1-\hat{p}_c)} + \frac{\left(\hat{AUC}(\tau) \right)^2 S_2}{M_1 N(1-\hat{p}_c)} \end{aligned}$$

2.4.4 Sample Size Estimation

2.4.4.1 Sample Size Estimation Equation: Using equation (3), we have

$$\frac{\left(\hat{AUC}(\tau) \right)^2 S_1}{N(1-\hat{p}_c)} + \frac{\left(\hat{AUC}(\tau) \right)^2 S_2}{M_1 N(1-\hat{p}_c)} = \frac{l^2}{Z_{1-\alpha/2}^2}$$

Hence the total sample size N is estimated as

$$N = \frac{Z_{1-\alpha/2}^2 \left(\hat{AUC}(\tau) \right)^2 (S_1 M_1 + S_2)}{l^2 (1 - \hat{p}_c) M_1} \quad (18)$$

where $\hat{AUC}(\tau)$ is a pre-specified value. In applying (18) we will assume $\hat{AUC}(\tau)$ is equal to its asymptotic mean which would be the true AUC. The sample size N is a function of a pre-specified $AUC(\tau)$, $l, \alpha, \mu_d, \sigma_d, \mu_a, \sigma_a, \gamma, r, f, \hat{\beta}_0$, and $\hat{\beta}_1$. As a summary, Table 1 lists the required input information for the sample size estimation.

2.4.4.2 Sample Size Estimation Steps, Constrains, and Parameter Selections: The estimation of the required sample size can proceed in the following steps according to the specified patient enrollment strategy:

1. Make assumptions for the required inputs listed in Table 1.
2. Estimate p_a, p_d, p_c , and q_1 to q_9 .
3. Obtain the approximations for $\hat{E}[s(\tau|Y)]$, $\hat{E}[1 - s(\tau|Y)]$, and $\hat{E}[(1 - s(\tau|U)) s(\tau|V) I(U > V)]$ using equations (14)-(16).

4. Calculate the approximations of the estimated variances and covariances:

$$\text{var}(\hat{E}_0), \text{var}(\hat{E}[\hat{s}(\tau|Y)]), \text{var}(\hat{E}[1 - \hat{s}(\tau|Y)]), \text{cov}(\hat{E}_0, \hat{E}[1 - \hat{s}(\tau|Y)]), \text{cov}(\hat{E}_0, \hat{E}[\hat{s}(\tau|Y)]),$$

and get the estimated values for S_1 and S_2 .

5. Substitute the values of S_1 , S_2 , and other parameters into equation (18) to obtain the estimated sample size N .

The parameters listed in Table 1 include three types of information: 1) biology information which specifies the baseline biomarker distribution and survival:

$\mu_a, \mu_d, \sigma_a, \sigma_d, \hat{\beta}_0$, and $\hat{\beta}_1$, 2) study design information: accrual period r , follow-up period f , and loss to follow-up rate γ , and 3) estimation information: prespecified $AUC(\tau)$, l , and α which specify how accurate of the sample size estimate. We could specify r , f , and γ according to how we want to design the study and specify l and α according to how accurate of the sample size estimate we want. However, the prespecified value of AUC most likely is from literature, historical data, or previous studies. The accepted value of AUC depends on the nature of the study marker. Some markers may provide better distinctions between diseased and nondiseased patients. Some other markers, which are only available markers we can use based on current discovery and development, may not provide very good distinctions. For the biology related parameters, μ_a , σ_a , μ_d , and σ_d are baseline biomarker distributions for nondiseased patients and diseased patients which can be obtained through literature review, historical data, or previous studies. For β_0 and β_1 , it might be possible to find the estimates through literature review or previous studies (which is what we did for the simulation study and the example); otherwise, 1) if we can find the hazard estimate from the literature, we might be possible to estimate β_0 and β_1 , or 2) if we can not find any information about β_0 and β_1 we can use the following two constrains to obtain the possible estimates or ranges for β_0 and β_1 : a) $s_d \approx 0.5$, and b) $\lambda_d * \tau \approx p_d * (1 + \lambda_d * \tau)$ which are derived based on information required for the sample size estimation, or 3) if we can find partial information about β_0 and β_1 , we may combine the partial information and constrains from 2) to obtain estimates of β_0 and β_1 .

3. Results

3.1. Sample Size Estimation Results

We use the following assumptions to illustrate the approach for estimating the sample size: $AUC(\tau) = 0.6$, $l = 0.1$, $\alpha = 0.05$, $\gamma = 0, 0.05, 0.1$, and 0.15 , $\tau = 3$ with three (r, f) pairs (1, 2), (1.5, 1.5), and (2, 1), $\mu_a = 2$, $\mu_d = 2.5$, $\sigma_a = 1$, $\sigma_d = 1$, $\hat{\beta}_0 = -3$, and $\hat{\beta}_1 = 0.4$. The assumed parameter values for $\mu_a, \mu_d, \sigma_a, \sigma_d, \hat{\beta}_0$, and $\hat{\beta}_1$ are based on results from Matin [14], Gao et al. [15], and Lilja et al. [16]. We also estimate sample sizes for $\tau = 4$ with four (r, f) pairs (1, 3), (1.5, 2.5), (2, 2), and (3, 1) and keeping all other parameter assumptions the same. Table 2 presents the sample size estimation results for single entry and staggered entry by varying input values with and without loss to follow-up. The value $\widehat{AUC}(\tau)$ is the estimated $AUC(\tau)$ given the parameters using equation (1) and the sample size N is estimated based on equation (18). From Table 2 we can see that a) the sample size is larger when the loss to follow-up rate is higher, b) when τ is the same, the sample size is larger when the accrual period r is longer, c) when the accrual period r is the same, the sample size

is smaller if the follow-up period f is longer, d) when the follow-up period f is the same, the sample size is smaller if the accrual period r is longer, and e) when τ is the same, the study with staggered entry requires larger sample size than the study with single entry.

3.2. Simulation Results

In the previous sections, we discussed how to estimate sample size N based on parametric models linking the biomarkers to failure times. In this section, we examine the performance of the proposed approach using simulated data.

We used the following assumptions for the simulation:

1. Sample size is N with 500 replicates.
2. Conditionally on the true disease status at τ , the baseline biomarker distribution is ([14], [15], and [16])

$$Y \sim \text{Normal}(\mu_d, \sigma_d^2) \text{ for } D(\tau)=1, \text{ where } \mu_d=2.5 \text{ and } \sigma_d=1,$$

$$Y \sim \text{Normal}(\mu_a, \sigma_a^2) \text{ for } D(\tau)=0, \text{ where } \mu_a=2 \text{ and } \sigma_a=1,$$

$$Y \sim \text{Normal}(\mu_c, \sigma_c^2) \text{ for } D(\tau)=\text{missing}, \text{ where we assume } p_0 = p_d.$$

3. The true disease status at time τ , $D(\tau)$, follows a multinomial distribution with the cell probabilities $\hat{p}_d, \hat{p}_a$, and $\hat{p}_c$ estimated using equations in Appendix C.
4. Coefficient estimates $\hat{\beta}_0 = -3$ and $\hat{\beta}_1 = 0.4$ in the hazard function (Martin [14]).
5. Failure time T follows an exponential distribution with rate $\hat{\lambda}$.
6. Censoring time follows an exponential distribution with rate $\gamma = 0, 0.05, 0.1$, and 0.15 .

We simulated datasets of size N , obtained in Section 3.1, with 500 replicates. After estimating $AUC(\tau)$ for each of 500 replicates, we computed the 95% confidence interval for $AUC(\tau)$ and compared the half length of the confidence interval with l , which was the pre-specified half length of the 95% confidence interval for $AUC(\tau)$ in the sample size estimation. Since the assumption that failure time follows an exponential distribution may not always hold in reality, we also assessed the robustness of the method when the underlying distribution of the survival time was a Weibull distribution with various values of the shape parameter θ .

Tables 3 presents the sample size estimation results for single entry and staggered entry by comparing $\hat{l}$, which is the estimated half length of the 95% confidence interval of $AUC(\tau)$ for the sample size N with 500 replicates, with l when the failure time was simulated to follow an exponential distribution. Comparison between l and $\hat{l}$ indicates good concordance between them, with $\hat{l}$ having no clear systematic differences from l . For illustration purpose, we only present the comparison results for the following scenarios: $r=1, f=2, \gamma=0.05, 0.1$, and 0.15 for staggered entry and $\tau=3, \gamma=0.05, 0.1$, and 0.15 for single entry. We have similar results for all other scenarios.

When the failure time follows a Weibull distribution, some of the estimates involved in deriving the sample size do not have closed-form solutions. We use simulated datasets to assess the robustness of the proposed method when the failure time follows a Weibull distribution. For each shape parameter θ we simulated another 500 datasets in which the failure time was simulated to follow a Weibull distribution with sample size N estimated using equation (18) by assuming an exponential survival model. When θ is close to 1 (for example between 0.95 and 1.05 in Table 4), there is a good concordance between l and $\hat{l}$, with $\hat{l}$ having no clear systematic differences from l . When the failure time distribution departs from an exponential, i.e. the shape parameter is far away from 1 in both sides, our method tends to underestimate the sample size. In this case, simulation will be one way to estimate the sample size. The starting point of the sample size for the simulation could be the N estimated through equation (18) with the assumptions of the exponential distribution of the failure time and all other parameters.

4. Example

In this section we illustrate the use of our method in deriving the required sample size for estimating the diagnostic accuracy of a biomarker used for predictive progression free survival. The example is the phase II trial of ^{64}Cu -ATSM PET/CT assessment of tumor hypoxia in cervical cancer from the American College of Radiology Imaging Network (ACRIN Protocol 6682). This study is designed to examine whether ^{64}Cu -ATSM uptake assessed by PET/CT is a biomarker predictive of the behavior of cervical cancer and of patient outcome. In this study accrual is expected to be completed within 18 months and patients will be followed for up to 3 years to assess progression of disease and overall survival. Overall tumor ^{64}Cu -ATSM uptake is measured by the tumor-to-muscle activity ratio (T/M) from pre-therapy ^{64}Cu -ATSM-PET /CT. Previous studies showed that for ^{60}Cu -ATSM a T/M threshold of 3.5 best discriminated those likely or unlikely to have cancer recurrence (Dehdashti et al. [17]). Studies also showed that the semiquantitative measure of tumor uptake is similar between ^{64}Cu -ATSM and ^{60}Cu -ATSM, and a good correlation (0.95) was observed between the uptake of these two agents.

The sample size consideration for the primary endpoint in the study protocol was addressed by Cox-regression model. One of the secondary endpoints for this study is to estimate the diagnostic accuracy of ^{64}Cu -ATSM uptake as a predictor of several important events, such as progression free survival and overall survival. For our illustrative example we will determine the sample size required for estimating the diagnostic accuracy of the baseline T/M ratio as a predictor of 3-year progression free survival. The required parameters of biology information were selected based on previous clinical studies of Cu-ATSM, studies of Cu-ATSM-PET in patients with cervical cancer, and the research results in the paper by Dehdashti et al. [17]. The parameters required for the study design follow the design for the primary endpoints. We make the following assumptions to estimate the sample size:

1. The mean and standard deviation of the T/M ratio given the failure status at year 3:
 $\mu_a = 2.5, \mu_d = 4.5, \sigma_a = 2.5, \sigma_d = 2.5$.
2. $\hat{\beta}_0 = -3, \hat{\beta}_1$ is between 0.25 and 0.3.

3. $r=1.5$ years, $f=1.5$ years, $l=0.1$, and $\alpha=0.05$.
4. The possible value for γ is 0, 0.05, and 0.1.

Table 5 presents the sample size estimates. It confirms the findings obtained using simulated datasets.

5. Discussion

In this paper, we present an approach to estimating the required sample size for a study designed to estimate the time-dependent ROC at a specified time point. The area under the curve provides an overall estimate of the discriminating ability of the current value of a marker against a reference standard defined by the occurrence of an event at a future time. The sample size estimation method can be tailored to different patient enrollment strategies and is based on parametric assumptions of an exponential failure time and an exponential censoring time. In particular, we adapted an estimator of the time-dependent AUC which accommodates censoring [5]. The simulation results presented in Tables 2 - 4 indicate that our proposed method produces reliable results when the failure time follows an exponential distribution or does not depart too far from it. Because of the complexity of the model and the multitude of inter-dependent parameters involved, practical guidelines of how to adjust the sample size estimates in case of departures from exponentiality do not seem feasible. However, our approach guides the analyst through the variety of required inputs, uses analytic formulas when feasible, and can be implemented with software which is programmed in R.

Time-dependent ROC analysis provides an assessment of the discriminating performance of a marker that is complementary but distinct from the assessment via a hazard ratio. Whereas the hazard ratio characterizes the likelihood of the occurrence of a future event conditionally on the value of a marker, the time-dependent ROC characterizes the separation of marker values conditionally on the future event. Thus, the hazard ratio is a measure of association between marker and outcome which does not characterize the discrimination between diseased and nondiseased patients. As pointed in the literature ([18] and [19]), a marker displaying an independent strong association may not necessarily contribute meaningfully to improved classification. The time-dependent ROC, which is a statistical method developed to evaluate the performance of prognostic markers for risk of an event where the event time is subject to censoring, is an appropriate approach instead of hazard ratio when discrimination for a future event is deemed important. We focused our method on sample size for the estimation of the area under the curve because the AUC provides a measure of performance that is averaged across all possible threshold values of the marker.

Our method for sample size determination makes maximum use of analytic equations for several of the involved quantities. However, the final computation requires the specification of a number of quantities, some of which would be difficult to assess without extensive prior information. When such inputs are possible to specify, the approach of this paper is fast and easy to implement. The use of simulation to estimate sample size is an alternative approach that, in principle at least, can be widely applicable and is flexible to allow the analyst to use parameter values and distribution information suitable to the particular study. A simulation

approach might also be complementary to analytic based approach, for example if the analytic formulas are unreliable for small samples. In this case, additional simulation might be considered based on the initial guess from the analytic calculation to obtain more realistic values. However, simulation methods would also require extensive prior information and would also rely on detailed parametric assumptions. In addition a simulation-based method would be computationally intensive, especially if in the exploration of sensitivity to modeling assumptions.

Our approach relies on parametric assumptions beyond those discussed already about the survival distribution. For example, we assume that baseline biomarker values or a transformation of them follow a normal distribution. Although such a parametric assumption is common in determining sample size for ROC studies, its impact on the results of the computation needs further study. Finally, in this paper we address sample size determination for only one of the metrics and settings addressed by time-dependent ROC analysis. A number of other questions and metrics have been proposed in literature that require the development of appropriate sample size computations. For example, methods are needed to determine the required sample size for estimating the average AUC over a range of time points.

Acknowledgments

Work partially supported by NCI grant U01 CA079778 (PI: C Gatsonis).

References

1. Pepe, MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press; 2003.
2. Zhou, X.; Obuchowski, NA.; McMLISH, DK. Statistical Methods in Diagnostic Medicine. Wiley-Interscience; 2002.
3. Heagerty PJ, Lumley T, Pepe MS. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics*. 2000; 56:337–344. [PubMed: 10877287]
4. Heagerty PJ, Zheng Y. Survival Model Predictive Accuracy and ROC Curves. *Biometrics*. 2005; 61:92–105. [PubMed: 15737082]
5. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine*. 2006; 25:3437–3486.
6. Zheng Y, Heagerty PJ. Semiparametric Estimation of Time-Dependent ROC Curves for Longitudinal Marker Data. *Biostatistics*. 2004; 5:615–632. [PubMed: 15475423]
7. Zheng Y, Heagerty PJ. Prospective Accuracy for Longitudinal Markers. *Biometrics*. 2007; 63:332–341. [PubMed: 17688486]
8. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for Evaluating Overall Adequacy of Risk Prediction Procedures with Censored Survival Data. *Statistics in Medicine*. 2010; 29:1105–1117. [PubMed: 21484848]
9. Saha P, Heagerty PJ. Time-Dependent Predictive Accuracy in the Presence of Competing Risks. *Biometrics*. 2010; 66:999–1011. [PubMed: 20070296]
10. Chiang CT, Wang SH, Hung H. Random weighting and Edgeworth expansion for the nonparametric time-dependent AUC estimator. *Statistic Sinica*. 2009; 19:969–979.
11. Bernardo MVP, Lipsitz SR, Harrington DP, Catalano PJ. Sample size calculations for failure time random variables in non-randomized studies. *The Statistician*. 2000; 49:31–40.
12. Casella, G.; Berger, RL. Wadsworth and Brooks/Cole. Addison-Wesley; 1990. Statistical Inference.

13. Elandt-Johnson, RC.; Johnson, NL. Survival Models and Data Analysis. John Wiley and Sons; 1980.
14. Matin MA. Effect of Using Bias-Corrected Estimators in Logistic Regression Model in Small Samples: Prostate-Specific Antigen (PSA) Data. *Data Science Journal*. 2006; 5:100–107.
15. Gao F, Manatunga AK, Chen S. Developing Multivariate Survival Trees with a Proportional Hazards Structure. *Journal of Data Science*. 2006; 4:343–356.
16. Lilja H, Ulmert D, Bjork T, Becker C, Serio AM, Nilsson J, Abrahamsson P, Vickers AJ, Berglund G. Long-term Prediction of Prostate Cancer up to 25 Years before Diagnosis of Prostate Cancer Using Prostate Kallikreins Measured at Age 44 to 50 Years. *Journal of Clinical Oncology*. 2007; 25:431–436. [PubMed: 17264339]
17. Dehdashti F, Grigsby PW, Lewis JS, Laforest R, Siegel BA, Welch MJ. Assessing Tumor Hypoxia in Cervical Cancer by PET with Cu-Labeled Diacetyl-Bis (N-Methylthiosemicarbazone). *J Nucl Med*. 2008; 49:201–205. [PubMed: 18199612]
18. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker. *American Journal of Epidemiology*. 2004; 159:882–890. [PubMed: 15105181]
19. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattane MW. Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. *Epidemiology*. 2009; 21:128–138. [PubMed: 20010215]
20. Gatsonis C. Receiver Operating Characteristic Analysis for the Evaluation of Diagnosis and Prediction. *Radiology*. 2009; 252:593–596. [PubMed: 19952022]

Table 1

Required Input in Sample Size Estimation

Input	Interpretation
μ_a	mean baseline biomarker value for alive patients
μ_d	mean baseline biomarker value for dead patients
σ_a	standard deviation of the baseline biomarker distribution for alive patients
σ_d	standard deviation of the baseline biomarker distribution for dead patients
$\hat{\beta}_0, \hat{\beta}_1$	estimated regression coefficients in the hazard function
l	pre-specified half length of the $(1 - \alpha)\%$ confidence interval for $AUC(\tau)$
α	type I error
$AUC(\tau)$	pre-specified value for the area under ROC curve at time τ
r	accrual period
f	follow-up period
γ	loss to follow-up rate

Table 2

Sample Size Estimates

Staggered Entry	τ	r	f	γ	$\widetilde{AUC}(\tau)$	N	a	d	c
	3	2	1	0	0.61	161	0.77	0.23	0
				0.05	0.60	179	0.69	0.22	0.08
				0.10	0.60	200	0.63	0.21	0.16
				0.15	0.59	221	0.57	0.20	0.23
	3	1.5	1.5	0	0.61	143	0.74	0.26	0
				0.05	0.60	163	0.66	0.25	0.09
				0.10	0.60	186	0.59	0.23	0.18
				0.15	0.59	208	0.52	0.22	0.25
	3	1	2	0	0.61	128	0.71	0.29	0
				0.05	0.60	149	0.63	0.27	0.10
				0.10	0.60	173	0.55	0.26	0.19
				0.15	0.59	196	0.49	0.24	0.27
4	1.5	2.5	0	0.61	93	0.65	0.35	0	
			0.05	0.60	113	0.55	0.33	0.12	
			0.10	0.59	136	0.46	0.31	0.23	
			0.15	0.59	157	0.39	0.29	0.32	
4	2	2	0	0.61	103	0.67	0.33	0	
			0.05	0.60	122	0.57	0.31	0.12	
			0.10	0.59	147	0.49	0.29	0.22	
			0.15	0.59	168	0.42	0.27	0.31	
4	3	1	0	0.61	125	0.72	0.28	0	
			0.05	0.60	144	0.64	0.26	0.10	
			0.10	0.60	166	0.56	0.24	0.19	
			0.15	0.59	187	0.50	0.23	0.27	
4	1	3	0	0.61	85	0.62	0.38	0	
			0.05	0.60	103	0.52	0.34	0.13	
			0.10	0.59	126	0.43	0.32	0.24	
			0.15	0.59	145	0.36	0.30	0.34	
Single Entry	τ	γ	$\widetilde{AUC}(\tau)$		N	a	d	c	
	3	0	0.61	103	0.67	0.33	0		
		0.05	0.60	124	0.57	0.31	0.12		
		0.10	0.60	149	0.49	0.29	0.22		

		0.15	0.59	171	0.42	0.27	0.31
4	0	0.61		71	0.58	0.42	0
	0.05	0.59		88	0.47	0.38	0.15
	0.10	0.59		108	0.38	0.35	0.27
	0.15	0.59		123	0.31	0.32	0.37

Table 3

Results from a Simulation Study Comparing the Estimated Half Length of the 95% Confidence Interval of $AUC(\tau)$ with a Pre-specified Value l

Staggered Entry	τ	r	f	γ	l	N	$\hat{l}$	$l - \hat{l}$	$\hat{l}/l$
	3	1	2	0.05	0.1	149	0.099	0.001	0.99
				0.10	0.1	173	0.102	-0.002	1.02
				0.15	0.1	196	0.099	0.001	0.99
Single Entry	τ			γ	l	N	$\hat{l}$	$l - \hat{l}$	$\hat{l}/l$
	3			0.05	0.1	124	0.105	-0.005	1.05
				0.10	0.1	149	0.099	0.001	0.99
				0.15	0.1	171	0.097	0.003	0.97

Table 4

Assessing the Robustness of the Method

θ	τ	γ	l	N	$\hat{l}$	$l - \hat{l}$
0.33	3	0.05	0.1	124	0.08	0.02
		0.10	0.1	149	0.078	0.024
		0.15	0.1	171	0.08	0.02
0.67	3	0.05	0.1	124	0.079	0.021
		0.10	0.1	149	0.076	0.024
		0.15	0.1	171	0.08	0.02
0.95	3	0.05	0.1	124	0.103	-0.003
		0.10	0.1	149	0.098	0.002
		0.15	0.1	171	0.096	0.004
1	3	0.05	0.1	124	0.105	-0.005
		0.10	0.1	149	0.099	0.001
		0.15	0.1	171	0.097	0.003
1.05	3	0.05	0.1	124	0.101	-0.001
		0.10	0.1	149	0.097	0.003
		0.15	0.1	171	0.102	-0.002
1.5	3	0.05	0.1	124	0.11	-0.01
		0.10	0.1	149	0.13	-0.03
		0.15	0.1	171	0.16	-0.06
2	3	0.05	0.1	124	0.14	-0.04
		0.10	0.1	149	0.15	-0.05
		0.15	0.1	171	0.17	-0.07

Table 5

Sample Size for Example

Staggered Entry	<i>AUC</i>	<i>r</i>	<i>f</i>	γ	β_1	<i>N</i>	<i>a</i>	<i>d</i>	<i>c</i>
	0.71	1.5	1.5	0	0.25	329	0.71	0.29	0
				0.05		391	0.63	0.28	0.09
				0.10		461	0.56	0.26	0.18
	0.71	1.5	1.5	0	0.3	262	0.65	0.35	0
				0.05		313	0.58	0.33	0.09
				0.10		371	0.51	0.31	0.18
Single Entry	<i>AUC</i>	τ		γ	β_1	<i>N</i>	<i>a</i>	<i>d</i>	<i>c</i>
	0.71	3		0	0.25	213	0.63	0.37	0
				0.05		272	0.54	0.34	0.12
				0.10		337	0.45	0.32	0.22
	0.71	3		0	0.3	156	0.56	0.44	0
				0.05		196	0.48	0.41	0.11
				0.10		244	0.40	0.38	0.22