Planning Studies accounting for competing risks

Aurélien Latouche, Raphaël Porcher

PLANNING STUDIES ACCOUNTING FOR COMPETING RISKS

A. LATOUCHE AND R. PORCHER

Abstract. Recently, with the growth of statistical developments for competing risks analysis, some methods have been proposed to compute sample size and plan a study in such a framework. These methods differ from a modelling approach: some are based on the Cox regression model for the cause-specific hazard accounting or not for accrual and follow–up period while another deals with the Fine and Gray regression model for the subdistribution hazard of a competing risk. Nonetheless, these formulas rely on similar key parameters, namely the difference one wishes to detect and the proportion of patients who will actually experience the event of interest. Choices for these parameters have a drastic effect on the sample size or equivalently the power. In this work, we compare these approaches and give practical advices for planning studies in a competing risks framework.

1. Introduction

Competing risks analysis is now a standard approach in the analysis of clinical trials, particularly in the cardiovascular and the oncology field. Regression modelling and test for equality of cumulative incidence tend to be used routinely and are now also available in standard statistical softwares. It appears that the advance in analyses accounting for competing risks do not translate in the crucial phase of planning the trial. Some recent works proposed sample size formula that rely on two proportional hazards models, namely the Cox model for the cause–specific hazard function [1, 2] and the Fine and Gray model for the hazard function associated to the cumulative incidence function [3]. These formulas, based on Schoenfeld's formula for the Cox model [4], rely on similar key parameters, namely the difference one wishes to detect and the proportion of patients who will actually experience the event of interest. Of note, the previous works pointed out that planning a study ignoring competing risks may lead to an underpowered study. Thus from a pratical point of view, the practionner has different options to plan a study accounting for competing risks.

In this article we compare these approach and supplement the sample size formula for the Fine and Gray model with the determination of the proportion of event of interest which was lacking. A modified version of this sample size based on Renyi–type test is also proposed.

2. Sample size calculation

Let us consider a clinical trial where n patients will be randomly assigned to a control group C or an experimental treatment group E, in respective proportions p_C and p_E. Suppose that the patients are exposed to K disctinct and exclusive failure causes, defining a competing risks setting. In such a case, data typically consist in both a failure time $T \geq 0$ and a failure cause $\epsilon \in \{1, \ldots, K\}$. Among these K failure causes, some may represent transition to transient (non-fatal) states, but we ignore future transitions and focus on time to the first failure (event). We also consider that the main endpoint of the trial is the occurrence of one of these K failure causes. In the sequel we will consider that 2 competing events act on the population without any loss of generality. In addition, we assume the time to event data are subject to administrative only *i.e.* there are no losses to follow–up during the trial. Patient entry are staggered over an accrual period and all patients are followed until the end of the trial. Under this condition, censoring times are independent of the failure times.

Two main strategies for analysis of the trial can be identified. Most common analyses focus on comparing the cause-specific hazard under the control and the experimental treatment, where the hazard of failure from cause 1 is defined as: $\lambda_{1.}(t) = dF_{1.}(t)/S(t)$, where the subscript "$\cdot$" denotes treatment arm (with C for control arm and E for experimental arm). In previous expression, $F_{1.}$ is the cumulative incidence function of failure from the cause of interest, *i.e.* $F_{1.}(t) = \Pr(T \leq t, \epsilon = 1)$ and $S_{.}(t) = 1 - (F_{1.}(t) + F_{2.}(t))$ is the event free survival function. In such a case, comparisons are often performed against proportional hazards alternatives, using a Cox model [5]. The other strategy consists in comparing the corresponding event probabilities $F_{1.}(t)$, either directly with the Gray's test [6] or using a Cox-like model for the associated hazard [7]: $\alpha_{1.}(t) = dF_{1.}(t)/(1 - F_{1.}(t))$. Detailed discussion of the differences between both approaches can be found in [8].

Assuming proportional hazards leads to following models:

$$\lambda_{1E}(t) = \lambda_{1C}(t)\theta$$

and

$$\alpha_{1E}(t) = \alpha_{1C}(t)\gamma,$$

where θ and γ are the hazard ratio and the subdistribution hazard ratio, respectively. Each is the measure of treatment effect under the considered approach. The hypotheses tested are :

$$H_0 : \log\theta = 0 \quad vs \quad H_1 : \log\theta \neq 0$$

or

$$H_0 : \log\gamma = 0 \quad vs \quad H_1 : \log\gamma \neq 0,$$

according to the modelisation choice for quantifying treatment benefit.

In this setting, sample size formulas have been developed for both the Cox model [1, 2] and the Fine and Gray model [3]. In each case, a required number of events e is computed according to the type I and II error rates (α and β, respectively) and the assumed hazard ratio under the alternative hypothesis. Indeed, both formulas are akin to the "classical" formula obtained for the Cox model (in a survival analysis setting) [4], and differ by the hazard ratio considered. This leads to

$$e = \frac{\left(z_{\alpha/2} + z_{\beta}\right)^2}{\left(\log\theta\right)^2 \times p_C \times p_E}$$

for a cause–specific analysis and

$$e = \frac{\left(z_{\alpha/2} + z_{\beta}\right)^2}{\left(\log\gamma\right)^2 \times p_C \times p_E}$$

for a subdistribution analysis, where z_α denotes the $(1-\alpha)$-quantile of the standard normal distribution.

To plan the trial, one has to compute the total number of subjects n to be recruited to obtain the required number of events e, as $n = e/\Psi$ where Ψ is the probability of observing the event of interest during the trial. The probability $\Psi_{\cdot}$ in each treatment arm depends on

$F_{1.}$, the duration of the accrual period τ_1, the total duration of the trial τ_2 and the distribution of random entry times with density ϕ through:

$$\Psi_{.} = \int_0^{\tau_1} F_{1.}(\tau_2 - s)\phi(s)ds. \tag{1}$$

$\Psi_{.}$ is computed for both groups, and subsequently combined as:

$$\Psi = p_C\Psi_C + p_E\Psi_E.$$

Several works have been devoted to derivation of Ψ for survival studies under different acrual patterns [9, 10]. For a cause–specific analysis, assuming uniform entry time and time-homogeneous Markov model for each transition, the expected proportion of events expresses as [2]:

$$\Psi_{.} = \frac{\lambda_{1.}}{\lambda_{.}} \times \left[1 - \frac{\exp(-\lambda_{.}(\tau_2 - \tau_1) - \exp(-\lambda_{.}\tau_2)}{\lambda_{.}\tau_1}\right], \tag{2}$$

where $\lambda_{1.}, \ldots, \lambda_{K.}$ represent the K constant hazard rates, and $\lambda_{.} = \sum_{k=1}^{K} \lambda_{k.}$.

When using a subdistribution-based analysis [3], one also has to determine a value for Ψ. Unfortunately, as both a Cox model and a Fine and Gray model cannot hold simultaneously, except in the degenerate case where no other failure cause than the cause of interest occurs, the expression (2) for Ψ should not be used for both treatment arms if proportional cause-specific hazards are not assumed. However, it is still possible to assume a parametric form for F_{1C} (for instance that given above), and use the straightforward relation between cumulative incidences in both groups, i.e.

$$F_{1E}(t) = 1 - [1 - F_{1C}(t)]^{\gamma}.$$

Even under simple parametric models, evaluating equation (1) with such an expression of $F_{1E}(t)$ is not trivial, but Simpson's rule provides the following approximation

$$\Psi_E = \frac{1}{6}\left[F_{1E}(\tau_2) + 4F_{1E}\left(\tau_2 - \frac{\tau_1}{2}\right) + F_{1E}(\tau_2 - \tau_1)\right]. \tag{3}$$

To illustrate the implication in terms of number of event to be observed e, we use the example taken from Schulgen et al. as a working example [2] . The 4D trial is a randomized, double–bind, placebo–controlled trial to evaluate the efficacy of antihyperlipidemic treatment,

in reducing cardiovascular mortality and frequency of non–fatal myocardial infarction. In this trial more than one primary event of interest and competing risks are present. Notably, it can be summarized by a multi–state model with initial state "alive" and the outcome states for the primary endpoint "non–fatal myocardial infarction" (state 1) and "death of cardiovascular cause" (state 2) and the additional outcome state death of other causes (state 3).

In the sequel, we considered the composite endpoint combining states 1 and 2 (non–fatal myocardial infarction and death of cardiovascular cause), death of other causes being the competing event. Reparameterizing the model accordingly leads to $\lambda_{1C} = 0.26$, $\lambda_{1E} = 0.19$ and $\lambda_{2C} = \lambda_{2E} = 0.14$. Thus $\theta = \lambda_{1E}/\lambda_{1C} = 0.73$ and $\gamma = \log(1 - F_{1E}(4))/\log(1 - F_{1C}(4)) = 0.75$. With $\alpha = 5\%$ and $\beta = 10\%$, the required number of events is 424 and 504, respectively.

Under uniform accrual assumption with an accrual duration $\tau_1 = 1.5$ years and a total duration $\tau_2 = 4$ years, the proportion Ψ obtained for a cause–specific model are $\Psi_C = 0.470$ and $\Psi_E = 0.377$. For a subdistribution analysis, very similar results are obtained with $\Psi_C = 0.470$ and $\Psi_E = 0.379$. This would lead to a trial of 1002 patients for a cause–specific analysis and 1190 patients for a subdistribution analysis.

In this example, both approaches lead to very similar anticipated proportions of events. Still using the hazard rates of the 4D trial, the differences between both proportions were very limited for trial durations varying from 3 to 9 years and accrual durations varying from 6 months to 3.5 years. In all these cases, the relative difference between the anticipated proportions of events was always less than 2%. For shorter trials, the anticipated proportion of events was found to be slightly larger when using equation (3), while the converse was found for longer durations. Reducing the accrual period led to closer values for Ψ obtained with both methods, whereas longer accrual periods yielded larger differences. In the example, both hazard ratios θ and γ were quite similar. Nonetheless, as the required number of events is inversely proportional to the logarithm of the hazard ratio, the influence of this apparently small difference was quite important in terms of sample size. Actually, as the relative difference between both hazard ratios was less than 3%, the relative difference between the numbers of events is only slightly inferior to 20%.

[Figure 1 about here.]

This is also illustrated on Figure (1), that displays the influence of a relative variation in parameter on the relative variation in sample size. It appears that differences in the proportion of events of interest, which may arise from the choice of one model *vs.* the other or from a misspecification of the true accrual or survival pattern, have little influence on the sample size, or equivalently the power. On the contrary, even a reasonable error on the targeted hazard ratio may lead to dramatic modifications of the sample size. This stresses the importance to select the model that will be used with care. In particular, if it is the planned analysis to analyze the trial by comparison of the subdistribution functions, the sample size should not be computed on the basis of a cause-specific analysis, and conversely.

3. Sample size formulas for Renyi–type tests

In previous section, both sample size formulas depend on the model chosen for inference, irrespective to the verification of this model assumptions. However, it is well known that the power of both tests is sensible to proportional hazard. As trial planning precedes the check of model assumptions it may be useful to anticipate possible departures from the proportional hazards by using a test statistic less sensitive to this proportionality assumption. This is precisely the case of Renyi–type tests *aka* supremum log–rank tests in the classical survival framework [11]. Moreover, recent work on sample size has shown that this test is nearly as efficient as the logrank test when hazards are proportional, and can accomodate with broader range of alternatives where the log–rank has no power to distinguish between groups [12]. Additionally, Renyi–type test statistics have already been extended to the comparison of subdistribution functions in the unpublished Ph.D. thesis of Bajorunaite [13]. We thus derive the sample size formula for this test.

We briefly present the sample size for the supremum log–rank statistic. Standardized weighted logrank are based on the counting process integral

$$V_n(t) = n^{-1/2} \int_0^t \hat{W}_n(s) \frac{R_C(u)R_E(u)}{R_C(u) + R_E(u)} \left\{ \frac{dN_C(u)}{R_C(u)} - \frac{dN_E(u)}{R_E(u)} \right\},$$

with $N_\cdot$ and $R_\cdot$ are the counting and at–risk processes in treatment arm "$\cdot$" and where $\hat{W}_n$ is an estimated weight function. Let $\sigma_n^2(t)$ be the variance of $V_n(t)$ and τ be the total duration

of the trial, then $T_n(\tau)$ is the standardized weighted logrank where $T_n(t) = V_n(t)/\sigma_n(t)$. The corresponding supremum version is $\sup_{t\in[0,\tau]} |T_n(t)|$.

The sample size formula is based on the limiting distribution of $\sup_{t\in[0,\tau]} |T_n(t)|$ with proportional hazards contiguous alternative. Statistical derivations can be found in [12] and the methodology to compute sample size can be summarize as follow: a sample size $\tilde{n}$ is first computed via Schoenfeld's formula; then an adujstment for the supremum version is computed as $n = R \times \tilde{n}$. This adjustment factor R is surprisingly only a function of α and β anf thus does not depend neither on the weight functions nor the true underlying model for the hazard function. An `R` function is available to compute this factor and can be found at `http://www.bios.unc.edu/~kosorok/renyi.html`.

Similarly, Gray's test is based on the scores of the form:

$$Z_\cdot = \int_0^\tau W_\cdot(u) \left\{ d\hat{\Gamma}_{1\cdot}(u) - d\hat{\Gamma}_1^0(u) \right\},$$

with $W_\cdot$ is a positive weight function, $\hat{\Gamma}_{1\cdot}(t)$ is the estimated cumulative subdistribution hazard for the cause of interest in treatment arm "$\cdot$" and $\hat{\Gamma}_1^0(t)$ is the cumulative subdistribution hazard estimates pooling treatment arms C and E.

In the two-sample problem and taking weight equal to 1, Renyi-type analogous for Gray's test can be based on:

$$Z_1(t) = \int_0^t \frac{R_C(u)R_E(u)}{R_C(u) + R_E(u)} \left\{ \frac{dN_{1C}(u)}{R_C(u)} - \frac{dN_{1E}(u)}{R_E(u)} \right\}, \tag{4}$$

where $N_{1\cdot}(t)$ is the number of events of interest in the treatment arm "$\cdot$" by t and $R_\cdot(t) = I(\tau \geq t)Y_\cdot(t)[1 - \hat{F}_{1\cdot}(t-)]/\hat{S}_\cdot(t-)$.

The Gray's test variance is complicated to evaluate but Klein and Bajorunaite [14] have proposed a simplified estimator for the variance of $Z_1(t)$, that reduces to:

$$\hat{\sigma}^2(t) = \int_0^t \frac{R_C(u)R_E(u)}{R_C(u) + R_E(u)} \times \frac{R_E(u)\left(1 - \hat{F}_{1E}(u-)\right) + R_C(u)\left(1 - \hat{F}_{1C}(u-)\right)}{R_C(u)\left(1 - \hat{F}_{1E}(u-)\right) + R_E(u)\left(1 - \hat{F}_{1C}(u-)\right)} \times \left\{ \frac{dN_{1C}(u) + dN_{1E}(u)}{R_C(u) + R_E(u)} \right\}. \tag{5}$$

In this two-sample case, Gray's test correspond to the test statistic $T_G = Z_1(\tau)/\hat{\sigma}(\tau)$, whereas the corresponding Renyi-type supremum version is $T_R = \sup_{t\in(0,\tau]} |Z_1(t)/\hat{\sigma}(\tau)|$ [13].

The work of Latouche et al. [3] showed that Schoenfeld's formula can be transposed to the Fine and Gray model with a similar expression, where the subdistribution hazard ratio γ stands for the hazard ratio θ. Thus when analysing data based on subdistribution hazard the first step is to compute the Schoenfled's like formula from Latouche et al. [3] then to correct it with the factor R .

4. A Simulation Study

A limited simulation study was conducted to illustrate the performance of the proposed test statistic T_R, comparatively to the others. Parameters of the simulations were determined to grossly mimic the 4D trial with different situations. Two scenarios were considered: cause-specific simulations and subdistribution simulations. In the first case, independent latent failure times for each failure cause (1 and 2) were generated for each treatment arm. Then, the time to first event and the corresponding failure cause were taken for (T, ε). As latent failure times were independent, the cause-specific hazards were equal to the marginal hazards of each failure cause. In the second case, time to failure from each cause were directly generated from improper Gompertz models, such as the one recently proposed by Jeong and Fine [15] for direct modelling of subdistribution functions . Briefly the subdistribution of failure from cause k in each arm is expressed as a function of two parameters as:

$$F_{k.}(t) = 1 - \exp\left[\beta_{k.}\left\{1 - \exp(\nu_{k.}t)\right\}/\nu_{k.}\right],$$

with $\nu_{k.} < 0$ and $|\beta_{k.}| < \infty$. Parameters were chosen to ensure that the limits of the subditributions of both failure causes summed up to unity. With such a model, the subdistribution hazard for failure from cause 1 can be expressed as $\alpha_{1.}(t) = \beta_{1.}\exp(\nu_{1.}t)$. In particular, this enables to generate data arizing from the model of Fine and Gray as long as $\nu_{1C} = \nu_{1E}$. For each scenario (cause–specific or subdistribtion),time to event were generated for the following hypothesis:

(a): under H_0

(b): under a proportional hazards alternative

(c): under a nonproportional hazards alternative

In cause-specific scenarios (a) and (b) independent exponential failure time were considered with respective hazards $\lambda_{1C} = 0.26$ and $\lambda_{2C} = 0.14$ in the control group while $\lambda_{1E} = 0.19$ and $\lambda_{2E} = 0.14$ in the experimental group. For the nonproportional alternative (c) λ_{1E} was piecewise constant, keeping $\lambda_{2C} = 0.14$ as shown on Figure (2).

[Figure 2 about here.]

In all cases we compared the performance of the logrank test, supremum logrank, Gray test and our proposed Renyi–type test. The sample size and the parameter were taken from our working example.

Under the null hypotesis, the values of $\nu_{k\cdot}$ and $\beta_{k\cdot}$ were taken to reach a prevalence for the event of interest $P(\varepsilon = 1) = 0.65$, with $\nu_{1\cdot} = \nu_{2\cdot} = -0.3$ and $\beta_{k\cdot} = \nu_{k\cdot} \log(1 - 0.65)$ (=0.315). The different settings for the alternatives hypothesis are displayed on Figure (2). For the simulation study we considered the two sample sizes respectively based on a log–rank that is 1002 patients and 1190 patients for a subdistribution analysis .

To evaluate the sensitivity of the test statistics namely log–rank, Supremum log–rank, Gray , and Renyi type test T_R, 10,000 data sets where generated in each case and the observed power and size were computed. The table 1 summarizes the simulation study.

[Table 1 about here.]

Under proportional cause–specific hazards (*i.e.* exponential model) and a sample size of 1002 patients, the log-rank expectidly performed better than the other tests. The proposed Renyi test was the farthest from the nominal 90% power with a 85.1% observed power. If we consider nonproportional hazards, the observed power diminished for all tests but the supremum log–rank test achieves the greater observed power (79.5%). This exhibits one advantages over standard log-rank. The proposed test performed better than Gray test with a 76.9% observed power. Increasing the sample size to 1190, the proposed test performed well with an observed power of 91%. The other test statistics led to over powered trials. Theses two settings clearly favored the log–rank test and the supremum log-rank test.

In the setting of proportional subdistribution hazards (*i.e.* Gompertz model) with a trial of 1002 patients, both Gray and Renyi tests achieved an observer power over 80% while the

observed power of ohter tests collapsed. Indeed, the Gray test is expected to have an optimal power when subdistibution hazards are proportional. When subdistributions were nonproportional, none of the tests statistics appeared to be clearly favored. Both Gray and Renyi tests observed power decreased to 73.9% and 75.5% respectively. Log–rank and supremum log–rank observed power were 66.1% and 68.2%. Gray's test was found to be more sensitive to nonproportionality that the Renyi type test. Increasing sample size to 1190, with proportional subdistribution hazards, only Gray's and Renyi tests observed powers were close to 90% , the log-rank and supremum tests observed power being 15% under the targeted power. Such a scenario favored subdistribution-based inference.

The Renyi type test is designed to accomodate departure form the proportional cause–specific hypothesis. This is particularly the case for nonproportional subdistribution hazards (indeed proportional cause specific hazards implies nonproportional subdistributions hazards) where our proposed test observed power was the highest with 82.6%. The proposed test statistics T_R can accomodate either proportional hazards or nonproportional hazards with better observed power when sample size formula accounting for competing risks (meaning more patients).

Moreover, the observed size was close to the 5% targeted size in all settings for all test statistics.

5. Discussion

When analysing data from a clinical trial accounting for the occurence of competing events, some complementary analysis can be conducted. As illustrated in Figure 1, imprecisions in parameter determination can lead to recruit much more patients than necessary or, on the contrary, to a dramatic decrease of the power of the trial. Parameter determination is thus a key issue when planning a trial, especially in survival trials with competing risks, where these parameters will heavily depend on assumptions on the distributions of failure time that are are almost unverifiable. The choice of the sample size formula is thus motivated by the corresponding quantity of interest. The small simulation study we conducted covered a combination of sample size and proportionality assumptions. It should be noted that in practice

only the method for computing sample size is part of the choice, whereas model assumptions can only be verified retrospectively. Proportionality assumptions are constraints toward test statitics employed for inference. Thus, we recommend the use of the proposed Renyi type test when planning studies accounting for competing risks because it can accomodate with broader range of alternatives such as nonproportional hazards.

References

[1] Pintilie M. Dealing with competing risks: testing covariates and calculating sample size. *Statistics in Medicine* 2002; **21**:3317–3324.

[2] Schulgen G, Olschewski M, Krane V, Wanner C, Ruf G, Schumacher M. Sample sizes for clinical trials with time-to-event endpoints and competing risks. *Contemporay Clinical Trials* 2005; **26**:386–396.

[3] Latouche A, Porcher R, Chevret S. Sample size formula for proportional hazards modelling of competing risks. *Statistics in Medicine* 2004; **23**:3263–3274.

[4] Schoenfeld DA. Sample-size formula for the proportionnal-hazards regression model. *Biometrics* 1983; **39**:499–503.

[5] Cox DR. Regression models and life tables. *Journal of the Royal Statistical Society, Series B* 1972; **34**:187–220.

[6] Gray RJ. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics* 1988; **116**:1141–1154.

[7] Fine JP, Gray RJ. A proportional hazards model for subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**:496–509.

[8] Beyersmann J, Dettenkofer M, Bertz H, Schumacher M. A competing risks analysis of bloodstream infection after stem-cell transplantation using subdistribution hazards and cause-specific hazards. *submitted* 2006; .

[9] Lachin JM, Foulkes MA. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics* 1986; **42**:507–519.

[10] Lakatos E. Sample size determination in clinical trials with time-dependent rates of losses and noncompliance. *Controlled Clinical Trials* 1986; **7**:189–199.

[11] Fleming TR, Harrington DP, O'Sullivan M. Supremum versions of the log-rank and generalized wilcoxon statistics. *Journal of the American Statistical Association* 1987; **82**:312–320.

[12] Eng K, Kosorok MR. A sample size formula for the supremum log-rank statistic. *Biometrics* 2005; **61**:86–91.

[13] Bajorunaite R. Comparison of failure probabilites in the presence of competing risks. Ph.D. thesis. Medical College of Wisconsin 2004.

[14] Klein JP, Bajorunaite R. Inference for competing risks. In *Advances in Survival Analysis*, Balakrishnan N, Rao C (eds). number 23 in Hanbook of statistics. Elsevier: Amsterdam, 2004; .

[15] Jeong JH, Fine J. Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2006; **55**:187–200.

Département de Biostatistique et Informatique Médicale, Hopital Saint–Louis, Paris, France

E-mail address: aurelien.latouche@paris7.jussieu.fr

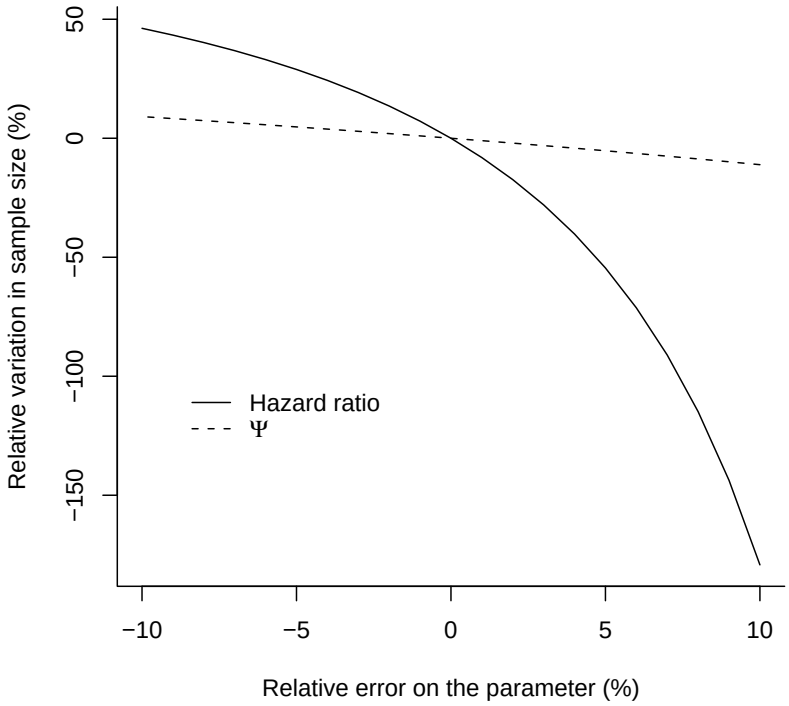

FIGURE 1. Influence of misspecification of the hazard ratio or the proportion of events of interest (Ψ) on the number of subjects

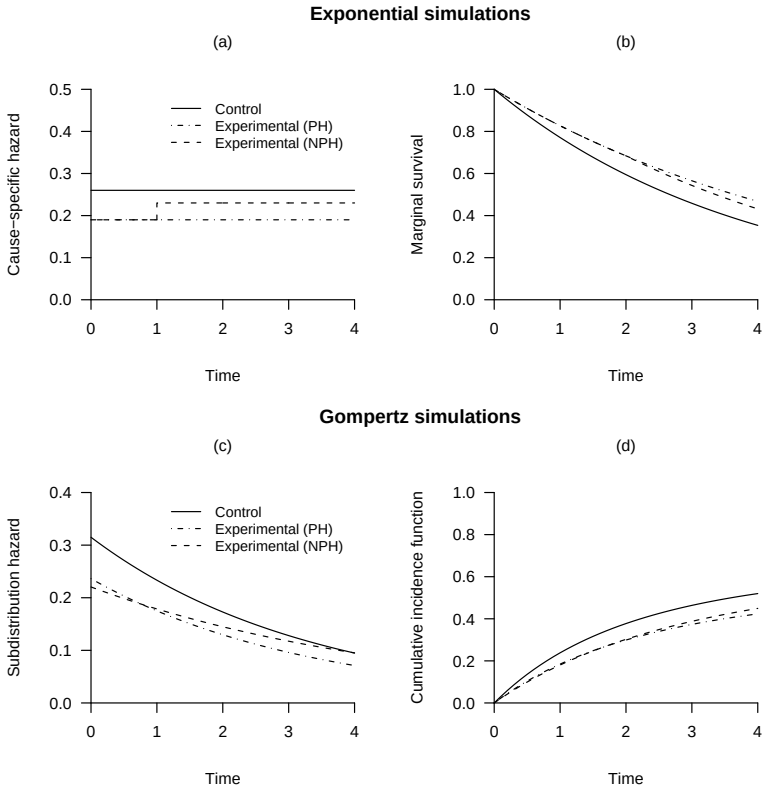

FIGURE 2. Simulated hazard (left) (cause–specific (a) subdistribution (c)) being proportional (PH) or not (NPH) and corresponding marginal survival function (b) and cumulative incidence function (d)

TABLE 1. Size and Power Comparison, Relative to Modelisation and Test Statistics Considering Proportional (PH) or Non-Proportional Hazards (NPH)

	$N = 1002$			$N = 1190$		
Test	Size	Power (PH)	Power (NPH)	Size	Power (PH)	Power (NPH)
Exponential model						
Log-rank	0.053	0.908	0.796	0.051	0.942	0.847
Supremum log-rank	0.049	0.890	0.808	0.048	0.928	0.861
Gray	0.053	0.880	0.755	0.050	0.923	0.813
Adapted Renyi	0.050	0.863	0.783	0.046	0.910	0.835
Gompertz model						
Log-rank	0.050	0.695	0.674	0.050	0.765	0.742
Supremum log-rank	0.046	0.686	0.697	0.049	0.749	0.763
Gray	0.050	0.860	0.743	0.051	0.912	0.812
Adapted Renyi	0.047	0.836	0.767	0.049	0.888	0.826