Jia and Lynn *Trials* (2015) 16:213
DOI 10.1186/s13063-015-0727-9

TRIALS

METHODOLOGY Open Access

A sample size planning approach that considers both statistical significance and clinical significance

Bin Jia and Henry S Lynn*

Abstract

Background: The CONSORT statement requires clinical trials to report confidence intervals, which help to assess the precision and clinical importance of the treatment effect. Conventional sample size calculations for clinical trials, however, only consider issues of statistical significance (that is, significance level and power).

Method: A more consistent approach is proposed whereby sample size planning also incorporates information on clinical significance as indicated by the boundaries of the confidence limits of the treatment effect.

Results: The probabilities of declaring a "definitive-positive" or "definitive-negative" result (as defined by Guyatt *et al.*, CMAJ 152(2):169-173, 1995) are controlled by calculating the sample size such that the lower confidence limit under H_1 and the upper confidence limit under H_0 are bounded by relevant cut-offs. Adjustments to the traditional sample size can be directly derived for the comparison of two normally distributed means in a test of nonequality, while simulations are used to estimate the sample size for evaluating the hazards ratio in a proportional-hazards model.

Conclusions: This sample size planning approach allows for an assessment of the potential clinical importance and precision of the treatment effect in a clinical trial in addition to considerations of statistical power and type I error.

Keywords: clinical significance, confidence interval, sample size

Background

The importance of confidence intervals is clearly attested by journal guidelines [1-3] as they "convey information about magnitude and precision of effect simultaneously, and keep these two aspects of measurements closely linked" [4]. For clinical trials, the CONSORT statement [5] stipulates the reporting of the "estimated effect size and its precision (such as 95% confidence interval)" and "how sample size was determined," but traditional sample size calculations for testing scientific hypotheses consider only statistical significance and power. The precision and clinical importance of the effect that can be depicted by confidence intervals is ignored. Under the usual practice, one calculates the sample size needed to declare some "clinically important difference" statistically significant at the α-level with 1 - β probability. The problem is that there is substantial subjectivity in quantifying this difference, and this can turn the sample size calculation into a moot exercise for choosing a difference to justify the number of patients the study can afford [6]. Frequently, the selected difference ends up larger than what is usual, and thus many studies may display large differences but lack the precision to make them statistically significant. Such shortcomings have led some to argue for reform of current sample size conventions in order to avoid misinterpretation of completed studies and harm to scientific research [7].

What would be helpful is a sample size estimation procedure that provides information on the confidence interval to supply users with information on the clinical significance and precision of the treatment effect in addition to power and statistical significance. Beal [8] suggested selecting sample size such that there is a high probability of the half-width of the confidence interval being less than some prescribed length, conditional on the interval containing the parameter of interest. Similarly, Liu [9] chose the sample size to yield a short confidence

* Correspondence: hslynn@shmu.edu.cn
Department of Biostatistics, School of Public Health, Key Laboratory on Public Health Safety of the Ministry of Education, Fudan University, 130 Dong'an Road, Shanghai 200032, China

interval width but conditional on the rejection of the null hypothesis H_0. Jiroutek *et al.* [10] combined the two by considering the probability of attaining a certain interval width conditional on both rejection of H_0 and inclusion of the true parameter. Cesana *et al.* [11,12] introduced a two-step procedure by first obtaining the sample size according to power and then iteratively increasing the sample size until the probability of obtaining confidence intervals with widths less than the expected interval width under H_1 exceeds a specified level.

In the above methods, the user either has to designate an interval length as reference or rely on the expected interval width, which may not be clinically relevant. A more straightforward alternative is to calculate a sample size such that the confidence limits of the parameter will be bounded by designated cut-offs. Specifically, the sample size is chosen such that according to the confidence limits the result can be deemed "definitive-positive" if there is indeed an effect or deemed "definitive-negative" if there is none. According to Guyatt *et al.* [13], a "definitive-positive" result implies that the lower confidence limit (*LCL*) of the parameter is not only larger than zero, implying a "positive" and statistically significant study, but above a relevant nonzero threshold. Conversely, a "definitive-negative" result implies that the upper confidence limit (*UCL*) is below some nonzero threshold. In hypothesis testing, one does not know whether H_1 or H_0 is true and can only control the probabilities of making a false positive or false negative error. Likewise, in this approach, we control the probabilities of declaring a "definitive-positive" or "definitive-negative" result by calculating the sample size such that *LCL* under H_1 and *UCL* under H_0 are bounded by fixed cut-offs. The following section demonstrates these concepts first for continuous normally distributed data and then for time-to-event data.

Methods

Normally distributed data

Consider a randomized 1:1 clinical trial comparing the mean responses between the treatment and control groups. When the response (or appropriately transformed response) can be regarded as normally distributed, the assessment of the treatment effect can be formulated as a hypothesis test of H_0: $\mu_1 - \mu_0 = 0$ versus H_1: $\mu_1 - \mu_0 \neq 0$. The sample size is then given by

$$n = \frac{\sigma^2\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{\delta^2}, \quad (1)$$

where Z_γ is the γth quantile of the standard normal distribution, (μ_0,σ_0) and (μ_1,σ_1) are the means and standard deviations of the control and treatment groups, respectively, $\sigma^2 = \sigma_0^2 + \sigma_1^2$, and $\delta = \mu_1 - \mu_0$ is the clinically important difference to be detected at level α with power 1 - β.

We first examine how likely the above sample size will yield a "definitive-negative" or "definitive-positive" result by calculating, respectively, the probabilities $\Pr(UCL < k_0\delta \mid H_0)$ and $\Pr(LCL > k_1\delta \mid H_1)$ for k_0, $k_1 \in [0,1]$. Without loss of generality, assume $\delta > 0$ and let $\bar{D}$ be the sample estimate of the treatment difference. If σ is known, then

$$\begin{aligned}\Pr(UCL < k_0\delta|H_0) &= \Pr\left(\bar{D} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < k_0\delta|H_0\right)\\ &= \Pr\left(Z < k_0\delta\frac{\sqrt{n}}{\sigma} - Z_{1-\alpha/2}\right)\\ &= \Pr\left(Z < (k_0-1)Z_{1-\alpha/2} + k_0Z_{1-\beta}\right), \text{and}\end{aligned} \quad (2)$$

$$\begin{aligned}\Pr(LCL > k_1\delta|H_1) &= \Pr\left(\bar{D} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} > k_1\delta|H_1\right)\\ &= \Pr\left(Z > (k_1\delta-\delta)\frac{\sqrt{n}}{\sigma} + Z_{1-\alpha/2}\right)\\ &= \Pr\left(Z > (k_1-1)Z_{1-\beta} + k_1Z_{1-\alpha/2}\right),\end{aligned} \quad (3)$$

where Z is the standard normal variable. As k_0, k_1 vary from 0 to 1, these two probability functions are mirror images about 1/2, with $\Pr(LCL > \delta/2 \mid H_1) = \Pr(UCL < \delta/2 \mid H_0)$. At the boundaries of 0 and 1, $\Pr(LCL > 0 \mid H_1) = \Pr(UCL < \delta \mid H_0) = 1 - \beta$.

Based on the derivations of equations (2) and (3), it can be shown that if the sample size is increased to $n_0 = n/k_0^2$ then $\Pr(UCL < k_0\delta \mid H_0) = 1 - \beta$ for $k_0 \in (0,1)$ and if it is increased to $n_1 = n/(1-k_1)^2$ then $\Pr(LCL > k_1\delta \mid H_1) = 1 - \beta$ for $k_1 \in (0,1)$. For example, with $k_0 = k_1 = 1/2$ and sample size $n_0 = n_1 = 4n$ both $\Pr(LCL > \delta/2 \mid H_1) = \Pr(UCL < \delta/2 \mid H_0) = 1 - \beta$. Note that if $k_0 = k_1 < 1/2$ then $n_0 > n_1$ and a larger sample size is required to establish a "definitive-negative" compared to a "definitive-positive" result. Conversely, if $k_0 = k_1 > 1/2$, then $n_0 < n_1$, and a larger sample size is needed to establish a "definitive-positive" result. In general, if

$$k_0 = 1-k_1 \text{ and } n_0 = n_1 = n/k_0^2, \quad (4)$$

then $\Pr(UCL < k_0\delta \mid H_0) = \Pr(LCL > k_1\delta \mid H_1) = 1 - \beta$. For example, if $k_0 = 2/3$, $k_1 = 1/3$ and $n_0 = n_1 = 9n/4$ then $\Pr(LCL > \delta/3 \mid H_1) = \Pr(UCL < 2\delta/3 \mid H_0) = 1 - \beta$.

Time-to-event data

We extend our proposed method to include time-to-event data, and use this case to show how a simulation-based approach can be used to estimate the sample size when the validity of normal approximation may be in doubt. In situations where a closed-form sample size formula is not readily available or difficult to derive, simulation provides an alternative and offers greater flexibility for adapting to

more complicated analyses. Briefly, the initial sample size required to detect the clinically important difference δ at power 1 - β is first calculated and then iteratively increased until Pr($LCL > k_1\delta \mid H_1$) and Pr($UCL < k_0\delta \mid H_0$) reach desired levels. The hazard ratio Δ is chosen as the parameter of interest with its corresponding confidence limits *LCL* and *UCL* being estimated using Cox regression. In the following description, we select for simplicity and convenience a single common cut-off by letting $k_0 = k_1 = 1/2$.

Under the proportional hazards assumption, the initial total sample size N_0 for detecting $\delta = \log_e\Delta$ at level α and power 1 - β can be estimated using Schoenfeld's [14] formula,

$$N_0 = \frac{\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{P_0 P_1 \left(\log_e \Delta\right)^2} \frac{1}{1-\pi_c}, \quad (5)$$

where π_c is the overall censoring proportion, and P_0 and P_1 are the proportion of subjects in the treatment and control groups, respectively. (Another choice is to use Freedman's [15] formula, which gives a slightly smaller sample size.)

Time-to-event data are simulated from the exponential distribution since it is most widely used to model time-to-event data under the proportional hazards assumption. Specifically, we simulate exponential survival times T_i and exponential censoring times L_i for subjects $i = 1, \ldots, N_0/2$ in each group, and consider a subject censored whenever $T_i < L_i$. According to Halabi and Bahadur [16], the parameters for the survival and censoring time distributions are given by

$$2\pi_c = \frac{\lambda_c}{(\lambda_0 + \lambda_c)} + \frac{\lambda_c}{(\lambda_1 + \lambda_c)}, \quad (6)$$

where λ_0, λ_1 are the hazard rates of the exponential survival times for the control and treatment groups, respectively, and λ_c is the hazard rate for the exponential censoring time. When $\pi_c = 0.5$, equation (6) reduces to the simple relationship

$$\lambda_c = \sqrt{\lambda_0 \lambda_1}. \quad (7)$$

We set $\lambda_0 = 1$ and select four values, (1.25, 1.5, 1.75, 2.0), for the hazard ratio $\Delta \equiv \lambda_1/\lambda_0 = \lambda_1$. For each value of Δ, the procedure goes through the following steps:

1. With α = 0.05, β = 0.2, $P_0 = P_1 = 0.5$, $\pi_c = 0.5$, and $\delta = \log_e(\Delta)$, calculate the initial total sample size N_0 using (5);
2. Simulate $N_0/2$ independent samples of exponential survival and censoring times for the treatment and control groups with corresponding parameters $\lambda_0 = 1$, λ_1, and $\lambda_c = \sqrt{\lambda_1}$;
3. Compare the survival times between the treatment and control groups using Cox regression and compute the 95% confidence interval for $\log_e(\Delta)$;
4. Repeat steps (2) and (3) for 10,000 iterations and estimate Pr($LCL > \delta/2 \mid H_1$) using the proportion of iterations where $LCL > \delta/2$;
5. Set Δ = 1 and repeat steps (2) and (3) 10,000 times to estimate Pr($UCL < \delta/2 \mid H_0$) using the proportion of times when $UCL < \delta/2$;
6. Replace N_0 with a larger sample size and repeat steps (2) through (5) until the estimates for both Pr($LCL > \delta/2 \mid H_1$) and Pr($UCL < \delta/2 \mid H_0$) are greater than some desired level (for example, 0.8).

The above procedure was programmed using SAS 9.2, and a sample SAS program is provided in the Appendix as reference.

Results

For comparing the means of normally distributed outcomes, Figure 1 shows that when α = 0.05 and power = 0.8, Pr($LCL > k\delta \mid H_1$) decreases steadily from 0.8 to 0.025 while Pr($UCL < k\delta \mid H_0$) increases steadily from 0.025 to 0.80 as *k* varies from 0 to 1. In fact, these two probability functions are mirror images about $k = 1/2$, where they both equal 0.288. This implies that a trial designed to detect a clinically important difference δ at the 5% significance level with 80% power will be "definitive-positive" about 29% of the time if one wants to say with 95% confidence that the treatment effect must be at least δ /2.

For time-to-event data, the initial total sample size ($N_0 = 1264$) for detecting a hazard ratio Δ = 1.25 is almost $5/(1 - \pi_c)$ or ten times larger than that ($N_0 = 132$) for detecting Δ = 2.00 according to Schoenfeld's [14] formula. At these initial sample sizes, the estimates of Pr($LCL > 0 \mid H_1$) ranged from 0.79 to 0.81 as expected, while Pr($UCL < \delta \mid H_0$) ranged from 0.70 to 0.77, slightly less than 0.8. Similarly, estimates for Pr($LCL > \delta/2 \mid H_1$) ranged from 0.27 to 0.29, close to what is expected for normally distributed data, while estimates of Pr($UCL < \delta/2 \mid H_0$) are slightly lower than expected, ranging from 0.23 to 0.27. For a specific example, say Δ = 1.75, then

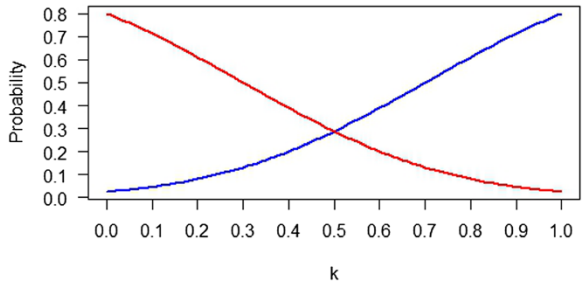

Figure 1 Plot of Pr($LCL > k\delta \mid H_1$) (red curve) and Pr($UCL < k\delta \mid H_0$) (blue curve) for $k \in [0,1]$, α = 0.05, β = 0.80 in a comparison of normally distributed mean responses with known σ between treatment and control groups for a 1:1 randomized clinical trial.

N_0 = 204 according to (5) and the estimates of α and β are 0.0485 and 0.2044, respectively. The β estimate implies that 79.6% of the samples have $LCL > 0$ under H_1. But the mean LCL is 0.16, thus as shown in Table 1 only 27.7% of the samples have $LCL > \delta/2 = \log_e(1.75)/2 = 0.28$. Correspondingly, 95.2% of the samples under H_0 have confidence intervals that include zero, but since the mean UCL is 0.42 only 25.4% of the samples have $UCL < 0.28$.

Table 1 suggests that sample sizes need to be larger by four to five times the initial sample size before estimates of both $\Pr(LCL > \delta/2 \mid H_1)$ and $\Pr(UCL < \delta/2 \mid H_0)$ are above 0.8. For example, with Δ = 1.75, the mean LCL for samples under H_1 equals 0.38 when the sample size reaches 938 (4.6 times N_0), and 85.0% of the samples then have $LCL > \delta/2 = 0.28$. In addition, at this sample size, the mean UCL for samples under H_0 equals 0.19, and 80.2% of the samples have $UCL < 0.28$. In terms of confidence interval width, the final sample sizes yield confidence interval widths that are between 0.4 to 0.5 times narrower than those at the initial sample sizes. For example, with Δ = 1.75 and a final sample size of 938, the mean confidence interval widths are 0.37 and 0.39 under H_0 and H_1, respectively, and 0.46 times narrower than the corresponding mean confidence interval widths at the initial sample size of 204.

Discussion

Many researchers realize that a traditional sample size calculation for testing H_0: $\mu_1 - \mu_0 = 0$ versus H_1: $\mu_1 - \mu_0 \neq 0$ with α = 0.05 and 80% power to detect a clinically important difference δ implies that: 1) 95% of its 95% confidence intervals for $\mu_1 - \mu_0$ will include zero when H_0 is true, and 2) 80% of the 95% confidence intervals will exclude zero when H_1 (that is, $\mu_1 - \mu_0 = \delta$) is true. However, a confidence interval with a LCL that is barely larger than zero may indicate a statistically significant treatment effect but be unconvincing to investigators who desire a "definitive-positive" result [13]. In contrast, a confidence interval that includes zero and demonstrates a "statistically nonsignificant" effect may be more convincing as a "definitive-negative" result when its UCL is small. Therefore, we propose that information on $\Pr(LCL >$ cut-off $\mid H_1)$ and $\Pr(UCL <$ cut-off $\mid H_0)$ be available to assist investigators in gauging the clinical significance of the treatment effect. For example, a plot similar to Figure 1 can be provided as a supplement to the usual sample size calculation or the investigator can directly estimate the sample size required such that LCL and UCL are bounded by relevant cut-offs with high probability. This offers a more consistent approach since the confidence interval becomes an important component in the design of clinical trials and not solely for analysis.

One question for this method concerns how a clinically relevant cut-off can be selected. Since δ, the clinically important difference, is already defined in the original sample size calculation, a convenient choice is to specify the cut-off with respect to δ. Given the uncertainty involved in quantifying δ and the tendency to inflate it [6], we set the cut-off equal to $k\delta$ for $k \in (0,1)$. This bypasses the need to additionally specify a confidence interval reference width [8-10] or calculate an expected confidence interval width [11,12]. For example, δ /2 can be used as the cut-off since it gives equal consideration to the expected precision of symmetrical intervals under H_0 and H_1. However, it should be stressed that there is no requirement for intervals under H_0 and H_1 to be given equal emphasis or for the boundaries of LCL and UCL to be the same. A researcher may well choose different cut-offs corresponding to a "definitive-positive" and a "definitive-negative" result; for example, $LCL > 3\delta/4$ and $UCL < \delta/4$ or $LCL > \delta/3$ and $UCL < 2\delta/3$.

Previous considerations of sample size estimation by controlling statistical power and precision often involve complex calculations even for normally distributed or binary outcomes. The current proposal is pedagogically straightforward as it simply focuses on the position of the confidence limits in relation to clinically relevant boundaries. Greenland [17] designed a method that provides high power to discriminate between the parameter values under H_0 and H_1. A sample size was chosen such that the discriminatory power, $\min\{\Pr(LCL > 0 \mid H_1), \Pr(UCL < \delta \mid H_0)\}$, equals a specified level. Our method also focuses on the probabilities of the lower and upper confidence limits being bounded, but the boundaries are different as Greenland was not thinking of clinically important effect sizes but the original parameter values under H_0 and H_1.

Table 1 Clinical significance and precision of the log-hazard ratio according to the initial and final sample sizes

Δ	$\log_e(\Delta)$	$^b\lambda_c$	*N*		$\Pr(LCL > \delta/2 \mid H_1)$	$^e\text{CIW}_1$	$\Pr(UCL < \delta/2 \mid H_0)$	$^d\text{CIW}_0$
1.25	0.22	1.12	[a]Initial	1264	0.2925	0.322	0.2651	0.314
			[c]Final	5402	0.8241	0.155	0.8016	0.151
1.50	0.41	1.22	[a]Initial	384	0.2759	0.602	0.2658	0.577
			[c]Final	1694	0.8349	0.285	0.8039	0.273
1.75	0.56	1.32	[a]Initial	204	0.2766	0.850	0.2536	0.804
			[c]Final	938	0.8496	0.392	0.8021	0.371
2.00	0.69	1.41	[a]Initial	132	0.2700	1.087	0.2344	1.018
			[c]Final	632	0.8503	0.487	0.8052	0.457

The [a]initial *N* calculated using equation (5), Schoenfeld's [14] formula, is the total sample size required to detect a hazard ratio Δ at the 5% level with 80% power, assuming equal subject allocation and a 0.5 overall censoring proportion. [b] λ_c is the hazard rate for the exponential censoring time given by equation (7), and $\delta = \log_e(\Delta)$. The [c]final *N* is the total sample size such that both $\Pr(LCL > \delta/2 \mid H_1)$ and $\Pr(UCL < \delta/2 \mid H_0)$ are at least 0.8 as estimated by the proportion of times LCL and UCL are bounded by δ /2 in 10,000 iterations. [d]CIW_0 and [e]CIW_1 are the mean width of the 95% confidence intervals under H_0 and H_1, respectively.

The condition $LCL > k_1\delta$ corresponds to the alternative hypothesis for a superiority test of H_0: $\mu_1 - \mu_0 \le k_1\delta$ versus H_1: $\mu_1 - \mu_0 > k_1\delta$. However, the sample size n_1 to attain a "definitive-positive" result is different from the sample size for the superiority test since the former is two-sided while the latter is one-sided. For example, with $\alpha = 0.05$, $\beta = 0.2$, $\sigma^2 = 2$, $\delta = 1$, and $k_1 = 1/2$, equations (1) and (4) imply that $n_1 = 4\times16 = 64$, while the sample size for the superiority test, as given by

$$\frac{\sigma^2\left(Z_{1-\alpha}+Z_{1-\beta}\right)^2}{\left(\delta-k_1\delta\right)^2},$$

equals 50. More importantly, our method calculates not only the sample size involving $LCL > k_1\delta$ but also that for $UCL < k_0\delta$.

Conclusions

In summary, our proposed method allows the researcher to calculate the sample size for a clinical trial not only according to the specifications of statistical significance (that is, α and β) but also in terms of clinical significance as judged by the boundaries of the confidence limits. For normally distributed data, simple formulae are available and their results serve as a reference for sample size planning when analyzing other types of data. For example, to ensure that *LCL* and *UCL* are both bounded by δ /2 the sample size needs to be increased 4-fold when comparing normally distributed means. Likewise, when evaluating the hazard ratio for time-to-event data, simulation results also suggest that sample sizes need to be 4 to 5 times larger. The results of our method indicate that sample size needs to be increased but our intention is not to mandate larger sample sizes per se. Such an effort may be futile since in practice cost constraints force clinical trials to aim for the smallest possible sample size What is important is that researchers be informed, for example by a graph similar to Figure 1, as to how their sample size will affect judgments of clinical significance using confidence intervals. In this respect, our proposal directs attention back to the importance of gauging effect sizes using confidence intervals, and is consistent with the predicted confidence intervals Goodman and Berlin [6] advocated to help investigators better understand the idea of statistical power when calculating sample size.

Appendix

Sample SAS program to estimate the total sample size for testing H_0: $\Delta = 1$ versus H_1: $\Delta \ne 1$ such that $\Pr(LCL > \delta/2 \mid H_1) = \Pr(UCL < \delta/2 \mid H_0) = 1 - \beta$. Survival and censoring times are assumed to be exponentially distributed, and the overall censoring proportion equals 0.5. The initial sample size is estimated using Schoenfeld's [14] formula for detecting $\delta = \log_e(\Delta)$ with 80% power at the 5% significance level.

```
%macro Calc_N(lambda0=, lambda1=);
data base;
  p_cen = 0.5; ** Overall censoring proportion **;

  lambdac = sqrt(&lambda0*&lambda1); ** lambdac = hazard rate of censoring distribution **;
  hr = &lambda1/&lambda0; ** hazard ratio **;

  E = 4*(probit(0.975)+probit(0.8))**2/log(hr)**2;  ** Schoenfeld: total expected events **;
  E = ceil(E);
  if mod(E,2) = 0 then ntot = E;  else ntot = E+1;   ** round to nearest even integer **;
  n0 = ceil(ntot/(1 - p_cen));  ** total sample size accounting for censoring **;

  %global l0 l1 lc;
  %let l0=&lambda0;
  %let l1=&lambda1;
  call symput('lc',lambdac);
  %let lc=&lc;
run;
proc print data=base noobs split="*";
  var p_cen lambdac hr e n0;
  label p_cen="Censoring*Proportion" lambdac="Hazard Rate of*Censoring Variable"
        hr="Hazard Ratio" e="Estimated No.*of Events" n0="Total Sample*Size";
  title3 "Initial Parameters and Sample Size (Alpha=0.05, Beta=20%)";
run;
%mend;

%macro sim_n(sampsize=, nrep=, seed=);
data simout;
  call streaminit(&seed);
  do rep=1 to &nrep;
    do trt=0 to 1;
      do i=1 to &sampsize/2;
        if trt=0 then do;
           t1 = rand('exponential')/&l0;  * simulate survival time for control group *;
           t0 = t1;
         end;
        if trt=1 then do;
           t1 = rand('exponential')/&l1;  * simulate survival time for trt group under H1 *;
           t0 = rand('exponential')/&l0;  * simulate survival time for trt group under H0 *;
         end;

         c=rand('exponential')/&lc;  ** exponential censoring **;
         if t1 < c  then censor1=0;
         if t1 >= c then do; t1 = c; censor1=1; end;

         if t0 < c  then censor0=0;
         if t0 >= c then do; t0 = c; censor0=1; end;

        output;
      end;
    end;
  end;
run;
%survreg(simout);
%mend;

%macro survreg(indat);
ods output ParameterEstimates=out1(where=(parameter='trt') keep=parameter hazardratio hrlowercl hruppercl
probchisq);
ods listing close;
options nonotes;
proc phreg data=&indat;
  model t1*censor1(1) = trt / rl;
  by rep;
run;
ods output ParameterEstimates=out0(where=(parameter='trt') keep=parameter hazardratio hrlowercl hruppercl
probchisq);
proc phreg data=&indat;
  model t0*censor0(1) = trt / rl;
  by rep;
run;
options notes;
ods listing;

data out1; set out1;
  if probchisq >= 0.05 then type2error=1; else type2error=0;
  lcl = log(hrlowercl);
  ucl = log(hruppercl);
  ciw = ucl - lcl;
  cutoff = log(&l0) + log(&l1/&l0)/2;
  if lcl > cutoff then gtcutoff=1; else gtcutoff=0;

  label lcl="Lower 95% CL for ln(HR)"
        ucl="Upper 95% CL for ln(HR)"
        ciw="Width of 95% CI for ln(HR)"
        gtcutoff="Pr(LCL > delta/2 | H1)";
run;
proc means data=out1 maxdec=4 mean lclm uclm;
  var type2error lcl ucl ciw gtcutoff ;
  title3 "Simulation Results (&nrep Repetitions) Under H1 When N=&sampsize";
run;

data out0; set out0;
  if probchisq < 0.05 then type1error=1; else type1error=0;
  lcl = log(hrlowercl);
  ucl = log(hruppercl);
  ciw = ucl - lcl;
  cutoff = log(&l0) + log(&l1/&l0)/2;
  if ucl < cutoff then ltcutoff=1; else ltcutoff=0;

  label lcl="Lower 95% CL for ln(HR)"
        ucl="Upper 95% CL for ln(HR)"
        ciw="Width of 95% CI for ln(HR)"
        ltcutoff="Pr(UCL < delta/2 | H0)";
run;
proc means data=out0 maxdec=4 mean lclm uclm;
  var type1error lcl ucl ciw ltcutoff;
  title3 "Simulation Results (&nrep Repetitions) Under H0 When N=&sampsize";
run;
%mend;

** lambda0 = hazard rate for control group   **
** lambda1 = hazard rate for treatment group **;
%calc_n(lambda0=1,lambda1=1.5);

%Sim_N(sampsize=384, nrep=10000, seed=12345);
```

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials; LCL: lower confidence limit; UCL: upper confidence limit.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
HSL conceived the study, performed the analyses, and drafted the manuscript. BJ participated in the analyses and drafted the manuscript. Both authors have read and approved the final manuscript.

Acknowledgements
None. This research was not supported by any external funding resources.

Received: 9 April 2014 Accepted: 22 April 2015
Published online: 12 May 2015

References

1. Simon R, Wittes RE. Methodologic guidelines for reports of clinical trials. Cancer Treat Rep. 1985;69:1–3.
2. Bailar JC, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Ann Intern Med. 1988;108:266–73.
3. Lang T. Documenting research in scientific articles: guidelines for authors. 1. Reporting research designs and activities. Chest. 2006;130:1263–8.
4. Rothman K. Modern epidemiology. Boston: Little Brown; 1986.
5. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134:663–94.
6. Goodman S, Berlin J. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Ann Intern Med. 1994;121:200–6.
7. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. BMC Med. 2010;8, e17.
8. Beal SL. Sample size determination for confidence intervals on the population mean and on the difference between two population means. Biometrics. 1989;45:969–77.
9. Liu XS. Implications of statistical power for confidence intervals. Br J Math Stat Psychol. 2012;65:427–37.
10. Jiroutek MR, Muller KE, Kupper LL, Stewart PW. A new method for choosing sample size for confidence interval-based inferences. Biometrics. 2003;59:580–90.
11. Cesana BM, Reina G, Marubini E. Sample size for testing a proportion in clinical trials: a 'two-step' procedure combining power and confidence interval expected width. Am Stat. 2001;55:288–92.
12. Cesana BM. Sample size for testing and estimating the difference between two paired and unpaired proportions: a 'two-step' procedure combining power and the probability of obtaining a precise estimate. Stat Med. 2004;23:2359–73.
13. Guyatt G, Jaeschke R, Heddle N, Cook D, Shannon H, Walter S. Basic statistics for clinicians: 2. Interpreting study results: confidence intervals. Can Med Assoc J. 1995;152:169–73.
14. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. Biometrics. 1983;39:499–503.
15. Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. Stat Med. 1982;1:121–9.
16. Halabi S, Bahadur S. Sample size determination for comparing several survival curves with unequal allocations. Stat Med. 2004;23:1793–815.
17. Greenland S. On sample-size and power calculations for studies using confidence intervals. Am J Epidemiol. 1988;128:231–7.