Power and sample size analysis for the Royston-Parmar combined test in clinical trials with a time-to-event outcome: Correction and program update

Patrick Royston
MRC Clinical Trials Unit at UCL
90 High Holborn
London WC1V 6LJ
UK.
Phone: +44 (0) 20 7670 4736
Email: j.royston@ucl.ac.uk

28 August 2018

Abstract

The changes made to Royston (2018) and to `power_ct` are: (i) in section 2.4 ('Sample-size calculation for the combined test'), to replace ordinary least squares regression using `regress` with grouped probit regression using `glm`; (ii) in section 4 ('Examples'), to revisit the worked examples of sample size estimation in light of the revised estimation procedure; (iii) to update the help file entry for option `n(`*numlist*`)`. The updated software is version 1.2.0.

Example	n()	n_{sim}	OLS		Probit	
			n_{est}	95% CI	n_{est}	95% CI
1	600, 650, 700	5000	643	640, 646	643	631, 654
2 (initial)	200, 500, 1000	500	405	405, 405	401	370, 433
2 (refined)	350, 400, 450	5000	383	381, 384	383	376, 389
3	874, 971, 1117	5000	1048	1021, 1074	1049	1027, 1070

Table 1: Original (OLS) and revised (probit) estimates of sample size and 95% CI for examples of three different time-dependent HR patterns.

The method of estimating a confidence interval (CI) for the required sample size, n, described in section 2.4 of Royston (2018), is incorrect. It does not account properly for uncertainty in the estimated power of the combined test at each of the candidate sample sizes specified in the `n`(*numlist*) option. The result is that CIs for the estimated sample size may be misleadingly narrow.

As before, the revised version of `power_ct` documented here uses simulation to estimate the power, ω, of the combined test at the suggested sample sizes, n. The relation

$$\Phi^{-1}(\omega) = b_0 + b_1\sqrt{n} \tag{1}$$

between probit transformed power, $\Phi^{-1}(\omega)$, and square root transformed sample size, $\sqrt{n}$, is still assumed.

Let n_{sim} be the number of simulations specified in `simulate()` and $r \in [0, n_{sim}]$ be the number of simulation samples in which the combined test rejects the null hypothesis with a given n. Previously, parameters b_0 and b_1 and their variance-covariance matrix were estimated by ordinary least squares regression of the inverse probit of the estimated power, $\omega = r/n_{sim}$, on $\sqrt{n}$. The required sample size, n_{est}, for the target power, ω_0, was determined by inversion and back-transformation of (1), giving $n_{est} = \left\{\left[\Phi^{-1}(\omega_0) - b_0\right]/b_1\right\}^2$. A delta-method, normal-based confidence interval for n_{est} was found by using `nlcom`, for example `nlcom ((invnormal('omega0') - _b[_cons])/_b[sqrtn])^2`.

Rather than using OLS regresssion, an appropriate way to estimate b_0 and b_1 in (1) and their covariance matrix is by probit regression for grouped data (`bprobit`) of r on $\sqrt{n}$, with binomial denominator n_{sim}. As of Stata 14, `bprobit` is no longer developed or supported by StataCorp. The recommended method of fitting standard or grouped probit models is `glm`. Here, I would code something like `glm r sqrtn, family(binomial 'simulate') link(probit)`. After model estimation, `nlcom` can be run exactly as before to get n_{est} and its CI.

I now revisit the three examples given in sections 4.2, 4.3 and 4.4 of Royston (2018). Table 1 compares the original and revised values of n_{est} and its 95% CI. [TABLE 1 NEAR HERE]

While n_{est} is little affected by the method of estimation, its CI may change considerably. Writing informally, while I would accept $n_{est} \simeq 643$ with a CI of $(631, 654)$ in example 1, it

is clear that $(370, 433)$ is too wide in example 2; in other words, the correct, probit-based CI shows that sample size is not yet sufficiently precisely estimated for n_{est} to be acceptable. Given the probit results in table 1, it makes sense to refine the values initially supplied in `n()`, for example by using the current point estimate of n and its CI.

The changes made to `power_ct` are (i) to replace OLS regression with probit regression using `glm`, and (ii) to update the help file entry for option `n(`*numlist*`)`. The update is version 1.2.0.

Acknowledgement

I thank Ian White for pointing out the problem with estimation of the sample size, particularly its CI.

References

Royston, P. (2018). Power and sample size analysis for the Royston-Parmar combined test in clinical trials with a time-to-event outcome, *Stata Journal* **18**: 3–21.