



*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

The Stata Journal (2018)  
18, Number 4, pp. 995–996

## 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  
University College London  
London, UK  
j.royston@ucl.ac.uk

**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 (OLS) 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; and iii) to update the help file entry for the option `n(numlist)`. The updated software is version 1.2.0.

**Keywords:** `st0510_1`, `power_ct`, update, randomized controlled trial, time-to-event outcome, restricted mean survival time, log-rank test, Cox test, combined test, treatment effect, hypothesis testing, flexible parametric model

### Introduction

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 properly account for uncertainty in the estimated power of the combined test at each candidate sample size specified in the `n(numlist)` option. Thus, 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,  $\omega$ , of the combined test at the suggested sample sizes,  $n$ . The relation

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

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

Let  $n_{\text{sim}}$  be the number of simulations specified in `simulate()`, and let  $r \in [0, n_{\text{sim}}]$  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 OLS regression of the inverse probit of the estimated power,  $\omega = r/n_{\text{sim}}$ , on  $\sqrt{n}$ . The required sample size,  $n_{\text{est}}$ , for the target power,  $\omega_0$ , was determined by inversion and back-transformation of (1), giving  $n_{\text{est}} = [\{\Phi^{-1}(\omega_0) - b_0\}/b_1]^2$ . A delta-method, normal-based confidence interval for  $n_{\text{est}}$  was found using `nlcom`, for example, `nlcom ((invnormal('omega0') - _b[_cons])/_b[sqrtn])^2`.

Rather than using OLS regression, 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  $n_{\text{sim}}$ . As of Stata 14, StataCorp no longer develops or supports `bprobit`. The recommended method of fitting standard or grouped probit models is `glm`. Here we would code something like `glm r sqrt(n), family(binomial 'simulate') link(probit)`. After model estimation, `nlcom` can be run exactly as before to get  $n_{\text{est}}$  and its CI.

We 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  $n_{\text{est}}$  and its 95% CI.

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

Example	$n()$	$n_{\text{sim}}$	OLS		Probit	
			$n_{\text{est}}$	95% CI	$n_{\text{est}}$	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

While  $n_{\text{est}}$  is little affected by the method of estimation, its CI may change considerably. Writing informally, while I would accept  $n_{\text{est}} \simeq 643$  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  $n_{\text{est}}$  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 the option `n(numlist)`. The update is version 1.2.0.

## Acknowledgment

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

## Reference

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.