
Zhenyu Jia


Xiao Xu


James Koziol


Wei-De Zhong

# Sample Size Calculation for Single-Arm Phase 2 Prostate Cancer Trials

Zhenyu Jia
The University of Akron, Akron, OH
Northeast Ohio Medical University, Rootstown, OH
Guizhou Normal College, Guiyang, China

Xiao Xu
The University of Akron, Akron, OH

James Koziol
The Scripps Research Institute, La Jolla, CA

Wei-De Zhong
Guangzhou Medical University, Guangzhou, China

Prostate cancer remains one of the major health issues for men worldwide and is the second leading cause of cancer-related death in males in the United States. Each year in the United States, there are approximately 195,000 radical prostatectomies.[1] Prostatectomy provides excellent disease control for the majority of patients with clinically localized prostate cancer. However, for patients at a high risk of relapse, additional (adjuvant) therapy may be needed to prevent disease recurrence. Kattan's nomograms, a set of alignment charts or 2-dimensional diagrams, have been widely used in disease management for identifying patients who may need adjuvant therapy.[2,3] These nomograms were developed based on the retrospective data of thousands of prostate cancer patients in order to calculate the probability of survival (overall survival or progression-free survival [PFS]) using patients' clinical data, such as age, Gleason score, lymph nodes, size of tumor, metastasis status, etc.

Oncologists often design early phase 2 trials as single-arm studies, with dichotomous clinical outcomes as primary efficacy end points; for example, biochemical recurrence in prostate cancer.[4,5] There are hypothesized population values for the target end points of interest, and comparison of observed outcomes from the trial with these population values are then utilized to justify further clinical testing.[6] Nevertheless, the hypothesized population values may vary due to the inclusion of different patient cohorts with distinct clinical characteristics. Here, nomograms are particularly useful, as they provide individual target probabilities for the dichotomous outcome of interest in the early phase 2 setting. This, in turn, leads to the seminal question of how many patients are needed to power such a single-arm study. Here, we present a simple approach to calculating sample size for early single-arm prostate cancer trial studies with a dichotomous outcome and heterogeneous response probabilities. The underlying concept can be applied to early phase 2 trials for other types of diseases.

## Methods

In a typical phase 2 trial, let $p$ be the proportion of the population that does not undergo disease recurrence (PFS) when treated with adjuvant therapy on trial. Our goal is to test $H_0: p = p_0 \; vs. H_A: p > p_0$, where $p_0$ is the proportion of PFS (hypothesized value) for the control population. Let $X_i$ be independent, dichotomous (eg, success/failure or relapse/nonrelapse), but with individual probabilities of success $p_i$ from the hypothesized control population, ie, $X_i|p_i \sim \text{Bin}(1, p_i)$. For prostate cancer trials, the $p_i$'s may be the priori probabilities of PFS determined from Kattan's nomograms.[2,3] For exam-

*Dr Jia is Assistant Professor of Biostatistics at The University of Akron and Northeast Ohio Medical University. He is extensively engaged in the design and analysis of both preclinical studies and clinical trials.*

*Ms Xu was a graduate student of The University of Akron in Ohio, having recently graduated with a master's degree in statistics. Her research interests include biostatistics and sample size calculation for clinical trials.*

*Dr Koziol is Head of the Division of Biomathematics at The Scripps Research Institute. He has been actively involved in the design and analysis of laboratory and clinical studies over the past 25 years, first at the Cancer Center of the University of California at San Diego, then as statistician for the Scripps General Clinical Research Center.*

*Dr Zhong is Professor of Medicine and Pathology at Guangzhou First People's Hospital, Guangzhou Medical University. His research interests include prostate cancer and benign prostate hyperplasia.*

ple, a web application (www.mskcc.org/cancer-care/adult/prostate/prediction-tools) based on the 1999 Kattan nomogram[3] can be used to calculate the PFS probabilities at various years after prostatectomy based on clinical variables including age, margin status, tumor stage, Gleason primary score, Gleason secondary score, preoperative prostate-specific antigen level, seminal vesicle status, lymph node status, and year of prostatectomy. We assume that the $p_i$'s are *i.i.d.* with $\mathrm{E}(p_i) = p_0$; for example, $p_i$ can follow a beta distribution. Let $T_n$ denote the number of "successes" in the trial, namely, $T_n = \sum_{i=1}^{n} X_i$, and $n$ is the sample size. Suitably large values of $T_n$ will lead to rejection of the null hypothesis. Our key observation here is that, following the conditional expectation formula and the law of total variance, we have

$$\mathrm{E}(T_n) = np_0, \quad (1)$$

and

$$\mathrm{Var}(T_n) = np_0(1-p_0), \quad (2)$$

regardless of the underlying distribution of the $p_i$, and in particular, the variance of the $p_i$. Elementary proofs of equations (1) and (2) are given in the Appendix.

In planning a phase 2 trial, we hope to determine the sample size $n$ based on the values of $p_i$'s (calculated using the nomograms) and the desirable effect size $d = p - p_0$. Note that $p$ is not available in the planning stage. The effect size is the size of difference between the observed survival proportion and the hypothesized survival proportion, beyond which experimenters would confirm the efficacy of the therapy strategy. Suppose we take a random sample of size $n$ from the hypothesized population, then $\hat{p}_0 = \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{\sum_{i=1}^{n} p_i}{n}$. The 95% CI for the $p_0$ is

$$\hat{p}_0 \pm 1.96 \times \sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n}}, \quad (3)$$

where the margin of error is $1.96 \times \sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n}}$. If the observed $p$ from the trial study is greater than the upper bound of the CI, then the classic proportion test would be significant, ie, the effect size $d$ is not less than the margin of error. Therefore, the required minimum sample size $n$ to power the trial study can be expressed as

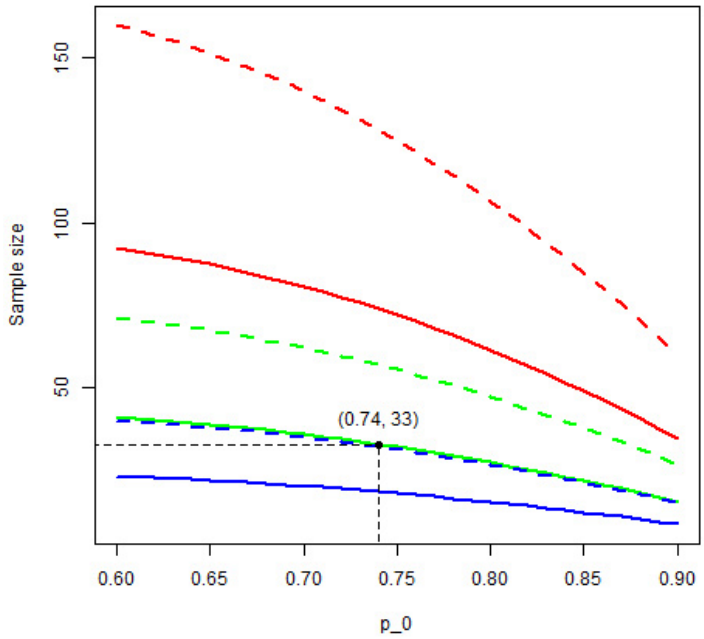
$$n = \frac{1.96^2 \times \hat{p}_0(1-\hat{p}_0)}{d^2}, \quad (4)$$

In a clinical test setting, we may already have recruited $m$ patients for the trial, and we would like to determine if these $m$ patients have enough power to test the efficacy of the therapy, and, if not, how many more patients we need to achieve a significant result. There will be 3 steps for determining the required sample size.

Step 1: Calculate the individual $p_i$'s for these $m$ patients using Kattan's nomogram,[2,3] and calculate $\hat{p}_0 = \frac{\sum_{i=1}^{m} p_i}{m}$;

Step 2: Determine the target effect size $d$, (eg, 0.1, 0.15, or 0.2), beyond which we believe the treatment is significantly effective in reducing disease recurrence.

**Figure** **Required Sample Size When Different CIs and Various Effect Sizes Are Considered**




Solid and dashed lines indicate 95% and 99% CIs, respectively. Red, green, and blue lines indicate margin of error equal to 0.1, 0.15, and 0.2, respectively.

Prostate cancer remains one of the major health issues for men worldwide and is the second leading cause of cancer-related death in males in the United States.

Step 3: Use equation (4) to calculate the required sample size $n$. If $m > n$, then the trial can be started promptly; otherwise, if $m < n$, we need $n - m$ additional patients who are similar to the previously recruited patients with regard to the clinical features.

**Table 1** Observed Outcomes and Nomogram Values of 24 Patients in a Clinical Trial

| Recurrence at 2 Years | Probability of Nonrecurrence at 2 Years Estimated by Nomogram |
|---|---|
| No | 0.99 |
| No | 0.98 |
| No | 0.96 |
| No | 0.95 |
| No | 0.94 |
| No | 0.92 |
| No | 0.91 |
| No | 0.91 |
| No | 0.90 |
| No | 0.87 |
| No | 0.85 |
| No | 0.82 |
| No | 0.81 |
| No | 0.75 |
| No | 0.69 |
| No | 0.69 |
| No | 0.63 |
| No | 0.55 |
| No | 0.52 |
| No | 0.52 |
| No | 0.40 |
| No | 0.21 |
| Yes | 0.76 |
| Yes | 0.31 |

**Table 2** Sample Size Calculation for the Clinical Trial

| | Effect Size | | |
|---|---|---|---|
| CI | $d = 0.1$ | $d = 0.15$ | $d = 0.2$ |
| 95% | 73 | 33 | 18 |
| 99% | 127 | 56 | 32 |

## Results

We investigated how the required sample size changes with the change of $p_0$ (**Figure**). We let $p_0$ take value between 0.6 and 0.9 with an increment of 0.01, and then calculate the required sample size when 95% or 99% CI and various effect sizes (0.1, 0.15, and 0.2) are considered. Generally, the required sample size decreases with the increase of $p_0$. Moreover, the wider the CI, the more patients are needed. On the other hand, the larger the effect size, the fewer patients are needed. The graph itself can serve as a "sample size nomogram" for calculating sample size for a single-arm trial. We will demonstrate the use of this sample size nomogram in a real trial study for prostate cancer adjuvant therapy.

Between 2001 and 2006, 24 subjects with high-risk prostate cancer were treated with open radical prostatectomy followed by adjuvant multimodality therapy.[4] **Table 1** shows the clinical outcomes of these patients in the trial. The first column represents the observed outcomes at 2 years after being treated with the therapy; the second column represents the probabilities of nonrecurrence (chance of "survival" without therapy) at 2 years estimated by the nomogram. Using the method proposed here, we are able to evaluate if these 24 patients are adequate to achieve a significant result. First, we calculate $\hat{p}_0 = \frac{\sum_{i=1}^{m} p_i}{m} = 0.74$. The required samples for various combinations of CIs and effect sizes are listed in **Table 2**. The required sample size for 95% CI and effect size of 0.15 is 33, which has been marked in the sample size nomogram (Figure). Note that the observed nonrecurrence rate is 22/24 = 0.92. Therefore, the difference between the observed nonrecurrence proportion $p$ and the hypothesized proportion $p_0$ is 0.18. If we treat 0.18 as the desirable effect size, the required sample size would become 23, which is less than 24, the number of patients enrolled in the clinical trial. Thus, we conclude that 24 patients in this clinical trial are adequate to power the test.

## Discussion

In this study, we focus on early phase 2 trials where only 1 arm (treatment arm) is available due to ethical issues and patient acceptance. In this setting, comparison of observed outcomes from the trial with the hypothesized population values or end points such as response rate are utilized to demonstrate clinical impact or justify further clinical testing in the absence of direct comparison data.[7-11] However, heterogeneity in the trial cohort is not often taken into account.

With the recent explosion in genomic information relating to patient responses to therapeutic regimens, treatment of choice is becoming more personalized and precise, with treatment plans and protocols individually tailored to patients' expected responses.[12-17] As Singer noted,[18] personalized medicine is therefore prompting redesigns of clinical trials. Our approach to the design of phase 2 trials is in this spirit: we postulate that expected responses are heterogeneous and explicitly plan the trial to satisfy appropriate statistical criteria while accounting for this expected heterogeneity in outcomes. Our mathematical development in this setting allows rapid calcu-

lation of sample sizes analogous to more conventional phase 2 trial planning. Our method applies regardless of the underlying distribution of the postulated response probabilities, leading to robust calculation of sample sizes for these single-arm phase 2 trials.

## Recognition of patient heterogeneity relating to expected responses to particular therapeutic regimens is becoming increasingly common.

The underlying goal of a conventional phase 2 design typically is to estimate the response probability of an interventional agent in a cohort of patients, and, if the response probability in a cohort of patients is sufficiently high, progression to a phase 3 trial may be warranted. This design ignores the fact that patients are quite heterogeneous in terms of clinical characteristics and by extension in terms of their "expected" outcomes (eg, response probabilities). There have been previous attempts to modify phase 2 designs to address patient heterogeneity and have motivated our current work. In particular, Simon and Maitournam[19] have proposed a "targeted" design for phase 2 trials, in which patients are initially stratified into 2 cohorts: those who are predicted to be responsive to the investigative treatment, and a remaining proportion. Though an improvement over the conventional design, it too may be insufficiently sensitive to patient heterogeneity, as response probabilities are again assumed to be constant within each subcohort. Our proposed design explicitly accommodates clinical heterogeneity through individual postulated response probabilities. Heller and colleagues[20] provide further justification for individualized outcome probabilities; their logistic regression model for analysis of phase 2 trials expressly incorporates patient covariates in assigning expected outcomes.

## Conclusion

We have developed a simple but robust approach to sample size estimation for single-arm phase 2 clinical trials with heterogeneous outcome probabilities. Recognition of patient heterogeneity relating to expected responses to particular therapeutic regimens is becoming increasingly common in clinical practice, and our approach formalizes incorporation of this heterogeneity into the design of early phase 2 clinical trials. ◆

## Acknowledgments

This work is supported by the Natural Science Foundation of Guizhou Province QKH-J [2013]2238; International Scientific and Technological Cooperation Projects of Guizhou Province QKHW-G [2014]7004; National Natural Science Foundation of China (81170699, 81272813), and National Institutes of Health grant 1 P01 HL119165-01A1.

## References

1. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin*. 2011;61:69-90.
2. Kattan MW, Eastham JA, Stapleton AM, et al. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst*. 1998;90:766-771.
3. Kattan MW, Wheeler TM, Scardino PT. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. *J Clin Oncol*. 1999;17:1499-1507.
4. Ebrahimi K, Ruckle H, Harper JD, et al. Post prostatectomy multimodality adjuvant therapy for patients at high risk for prostate cancer relapse. *J Urol*. 2008;179:558-559.
5. Jia Z, Lilly MB, Koziol JA, et al. Generation of "virtual" control groups for single arm prostate cancer adjuvant trials. *PLoS One*. 2014;9:e85010.
6. Koziol JA. New tricks with old dogs: personalised medicine and clinical trials. *Br J Cancer*. 2011;105:599-601.
7. Rubinstein L. Phase II design: history and evolution. *Chin Clin Oncol*. 2014;3:48.
8. Brookmeyer R, Crowley J. A confidence interval for the median survival time. *Biometrics*. 1982;38:29-41.
9. Korn EL, Freidlin B. Conditional power calculations for clinical trials with historical controls. *Stat Med*. 2006;25:2922-2931.
10. Rubinstein LV, Gail MH, Santner TJ. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *J Chronic Dis*. 1981;34:469-479.
11. Korn EL, Liu PY, Lee SJ, et al. Meta-analysis of phase II cooperative group trials in metastatic stage IV melanoma to determine progression-free and overall survival benchmarks for future phase II trials. *J Clin Oncol*. 2008;26:527-534.
12. Antonarakis ES, Lu C, Wang H, et al. AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med*. 2014;371:1028-1038.
13. Armstrong AJ, Li J, Beaver J, et al. Genomic analysis of circulating tumor cells (CTCs) from men with metastatic castration resistant prostate cancer (mCRPC) in the context of enzalutamide therapy. *J Clin Oncol*. 2014;32(suppl 4). Abstract 65.
14. Santoni M, Santini D, Massari F, et al. Heterogeneous drug target expression as possible basis for different clinical radiological response to the treatment of primary and metastatic renal cell carcinoma: suggestions from bench to bedside. *Cancer Metastasis Rev*. 2014;33:321-331.
15. Kompier LC, van der Aa MN, Lurkin I, et al. The development of multiple bladder tumour recurrences in relation to the FGFR3 mutation status of the primary tumour. *J Pathol*. 2009;218:104-112.
16. Cheng L, Zhang S, MacLennan GT, et al. Bladder cancer: translating molecular genetic insights into clinical practice. *Hum Pathol*. 2011;42:455-481.
17. Cutcutache I, Suzuki Y, Tan IB, et al. Exome-wide sequencing shows low mutation rates and identifies novel mutated genes in seminomas. *Eur Urol*. 2015;68:77-83.
18. Singer E. Personalized medicine prompts push to redesign clinical trials. *Nat Med*. 2005;11:462.
19. Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clin Cancer Res*. 2004;10:6759-6763.
20. Heller G, Kattan MW, Scher HI. Improving the decision to pursue a phase 3 clinical trial by adjusting for patient-specific factors in evaluating phase 2 treatment efficacy data. *Med Decis Making*. 2007;27:380-386.

## Appendix

1. Proof of the conditional expectation of equation (1).

   We assume that the $p_i$'s are *i.i.d.* with $\mathrm{E}(p_i) = p_0$. Let $T_n$ denote the number of "successes" in the trial, namely, $T_n = \sum_{i=1}^{n} X_i$, where $n$ is the sample size and $X_i \sim Bernoulli(p_i)$.

   $$\mathrm{E}(T_n) = \mathrm{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} \mathrm{E}\left(\mathrm{E}(X_i|p_i)\right) = \sum_{i=1}^{n} \mathrm{E}(p_i) = \sum_{i=1}^{n} p_0 = np_0.$$

2. Proof of the law of total variance of equation (2).

   $$\mathrm{Var}(T_n) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \sum_{i=1}^{n} \left(\mathrm{E}\left(\mathrm{Var}(X_i|p_i)\right) + \mathrm{Var}\left(\mathrm{E}(X_i|p_i)\right)\right) =$$

   $$\sum_{i=1}^{n} \left(\mathrm{E}\left(p_i(1-p_i)\right) + \mathrm{Var}(p_i)\right) = \sum_{i=1}^{n} \left(\mathrm{E}(p_i) - \mathrm{E}(p_i^2) + \mathrm{Var}(p_i)\right) = \sum_{i=1}^{n} \left(\mathrm{E}(p_i) - \mathrm{E}(p_i^2) +$$

   $$\mathrm{E}(p_i^2) - \left(\mathrm{E}(p_i)\right)^2\right) = \sum_{i=1}^{n} (p_0 - p_0^2) = np_0(1-p_0).$$