Sample Size Formulas for Different Study Designs

Supplement Document for *Wang, X. and Ji, X., 2020. Sample size estimation in clinical research: from randomized controlled trials to observational studies. Chest, 158(1), pp.S12-S20.*

Xiaofeng Wang, PhD[1,*] and Xinge Ji, MS [1]

[1]Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

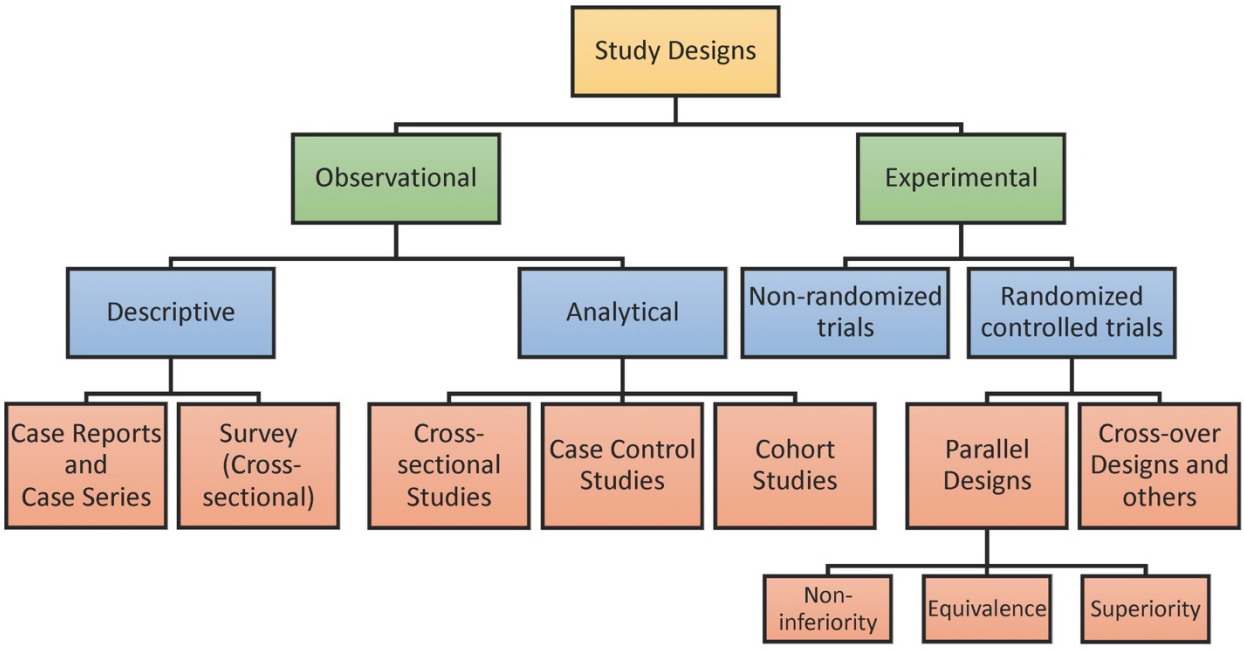
[*]Correspondence to:

Xiaofeng Wang, PhD

Department of Quantitative Health Sciences, Lerner Research Institute

Cleveland Clinic, 9500 Euclid Ave./ JJN3-01, Cleveland, OH 44195

Email: wangx6@ccf.org



We present the hypothesis tests and the corresponding sample size estimation formulas by the type of study design and type of outcome.

1. Randomized controlled trials

The majority of RCTs in clinical research are parallel-group trials. An analysis of the RCTs indexed in PubMed between 2000 and 2006 found that 78% were parallel, and 16% were crossover [1]. For brevity, we restrict discussion to the sample size estimation of the parallel design. We refer to the book by Chow et al. [2] for other subtypes of RCT designs and refer to [3,4] for non-randomized interventional studies.

In a parallel RCT, the outcome of interest could be a continuous, dichotomous, or a time-to-event variable. There are three commonly-used types of trials using a parallel RCT design: non-inferiority, equivalence, and superiority trials [5]. A non-inferiority trial aims to demonstrate that a new treatment is not worse than an active control treatment already in use by a small pre-specified amount. This amount is known as the non-inferiority margin. An equivalence trial is to show that the true treatment difference lies between a lower and an upper equivalence margin of clinically acceptable differences. When an RCT aims to show that one treatment is superior to another, the trial (test) is called a superiority trial (test).

In many RCT designs, more participants are randomized to the treated group than to the control group. This imbalance may encourage people to join in a trial because their chance of being randomized to the treated group is greater than to the control group. When we present the formulas for RCTs below, we denote k be the ratio of the sample size of treatment group n_T to the sample size of control group n_C, so that $n_T = kn_C$.

1.1 Continuous outcomes

- Non-inferiority design

The testing hypotheses are:

$$H_0: \mu_T - \mu_C \leq -\delta \text{ vs. } H_1: \mu_T - \mu_C > -\delta$$

where $\delta > 0$ denotes the non-inferiority margin, which is a (clinically meaningful) minimal detectable difference. The sample sizes are [2]

$$n_C = \left(1 + \frac{1}{k}\right)\sigma^2\left(\frac{z_{1-\alpha} + z_{1-\beta}}{d + \delta}\right)^2; n_T = kn_C$$

where σ^2 is the variance, and $d = \mu_T - \mu_C$ is known as the *allowable difference*, which is the true mean difference between the new treatment group (μ_T) and the control group (μ_C). In many applications, d is set to be zero. The z_γ denotes the standard normal deviate, i.e. $P(Z < z_\gamma) = 1 - \gamma$. A standard normal deviate is a realization of a standard normal random variable. For example, the $z_{1-\beta}$ is 0.84 at 80% power and 1.28 at 90% power.

- Equivalence design

The testing hypotheses are:

$$H_0: |\mu_T - \mu_C| \geq \delta \text{ vs. } H_1: |\mu_T - \mu_C| < \delta$$

where $\delta > 0$ denote the equivalence margin. We have

$$n_C = \left(1 + \frac{1}{k}\right)\sigma^2\left(\frac{z_{1-\alpha} + z_{1-\beta/2}}{\delta - |d|}\right)^2; n_T = kn_C.$$

- Superiority design

The testing hypotheses are:

$$H_0: \mu_T - \mu_C \leq \delta \text{ vs. } H_1: \mu_T - \mu_C > \delta$$

where $\delta > 0$ denote the superiority margin. We have

$$n_c = \left(1 + \frac{1}{k}\right)\sigma^2\left(\frac{z_{1-\alpha} + z_{1-\beta}}{d - \delta}\right)^2; n_T = kn_C.$$

1.2 Dichotomous outcomes – based on proportion difference

- Non-inferiority design

The testing hypotheses are

$$H_0: p_T - p_C \leq -\delta \text{ vs. } H_1: p_T - p_C > -\delta$$

where $\delta > 0$ denote the non-inferiority margin. We have [2]

$$n_c = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{d + \delta}\right)^2 \left[\frac{p_T(1-p_T)}{k} + p_C(1-p_C)\right]; n_T = kn_C$$

where $d = p_T - p_C$ is the difference between the true response rates of the new treatment group (p_T) and the control group (p_C).

- Equivalence design

The testing hypotheses are:

$$H_0: |p_T - p_C| \geq \delta \text{ vs. } H_1: |p_T - p_C| < \delta$$

where $\delta > 0$ denote the equivalence margin. We have

$$n_c = \left(\frac{z_{1-\alpha} + z_{1-\beta/2}}{\delta - |d|}\right)^2 \left[\frac{p_T(1-p_T)}{k} + p_C(1-p_C)\right]; n_T = kn_C.$$

- Superiority design

The testing hypotheses are:

$$H_0: p_T - p_C \leq \delta \text{ vs. } H_1: p_T - p_C > \delta$$

where $\delta > 0$ denote the superiority margin. We have

$$n_c = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{d - \delta}\right)^2 \left[\frac{p_T(1-p_T)}{k} + p_C(1-p_C)\right]; n_T = kn_C.$$

1.3 Dichotomous outcomes – based on odds ratio

Odds ratio has been frequently used to assess the association between a binary exposure variable and a binary disease outcome. The odds ratio between the treatment and the control is defined as

$$OR = \frac{p_T(1-p_C)}{p_C(1-p_T)}.$$

In RCTs, it is often of interest to investigate the odds ratio of a treatment for the disease under study.

- Non-inferiority design

The testing hypotheses are:

$$H_0: OR \leq \exp(-\delta) \quad \text{vs.} \quad H_1: OR > \exp(-\delta).$$

Note that here $\delta > 0$ denote the non-inferiority margin in log-scale. We have [2]

$$n_c = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\log(OR) + \delta}\right)^2 \left[\frac{1}{kp_T(1-p_T)} + \frac{1}{p_C(1-p_C)}\right]; n_T = kn_C.$$

- Equivalence design

The testing hypotheses are:

$$H_0: |\log(OR)| \geq \delta \quad \text{vs.} \quad H_1: |\log(OR)| < \delta.$$

where $\delta > 0$ denote the equivalence margin in log-scale. We have

$$n_C = \left(\frac{z_{1-\alpha} + z_{1-\beta/2}}{\delta - |\log(OR)|}\right)^2 \left[\frac{1}{kp_T(1-p_T)} + \frac{1}{p_C(1-p_C)}\right]; n_T = kn_C.$$

- Superiority design

The testing hypotheses are:

$$H_0: OR \leq \exp(\delta) \quad \text{vs.} \quad H_1: OR > \exp(\delta)$$

where $\delta > 0$ denote the superiority margin in log-scale. We have

$$n_C = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\log(OR) - \delta}\right)^2 \left[\frac{1}{kp_T(1-p_T)} + \frac{1}{p_C(1-p_C)}\right]; n_T = kn_C.$$

1.4 Time-to-event outcomes – based on hazard ratio

In clinical research, investigators may be interested in evaluating the effect of the test drug on the time to event. The analysis of time-to-event data is often referred to as survival

analysis. Basic concepts regarding survival and hazard functions in the analysis of time-to-event data can be found from Clark [6]. Assuming that the proportional hazards assumption holds in a study, the hazard ratio is defined as

$$HR = \lambda_T(t)/\lambda_C(t), \qquad for\ t \geq 0,$$

where $\lambda_T(t)$ is the hazard for the treatment group and $\lambda_C(t)$ is the hazard for the control group.

- Non-inferiority design

The testing hypotheses are:

$$H_0: HR \leq \exp(-\delta) \quad \text{vs.} \quad H_1: HR > \exp(-\delta)$$

where $\delta > 0$ denote the non-inferiority margin in log-scale. Following the theoretical results by [7,8], the total number of events (deaths) required in the two groups is

$$\frac{(k+1)^2}{k}\left(\frac{z_{1-\alpha} + z_{1-\beta}}{\log(HR) + \delta}\right)^2.$$

Let us assume that the probabilities that a person experiences an event in the control and treatment groups during the trial are π_c and π_T, respectively. The combined probability of the event is then $\pi = (\pi_C + \pi_T)/2$. The sample sizes are given by [7]:

$$n_C = \frac{k+1}{\pi k}\left(\frac{z_{1-\alpha} + z_{1-\beta}}{\log(HR) + \delta}\right)^2; n_T = kn_C.$$

Investigators could have a reasonable guess for π_c and π_T from previous studies. If there is no prior knowledge, one may assume an exponential survival model and estimate π_c and π_T using explicit formulas (see Formula (3) in cohort studies below).

- Equivalence design

The testing hypotheses are:

$$H_0: |\log(HR)| \geq \delta \quad \text{vs.} \quad H_1: |\log(HR)| < \delta$$

where $\delta > 0$ denote the equivalence margin in log-scale. We have

$$n_C = \frac{k+1}{\pi k}\left(\frac{z_{1-\alpha} + z_{1-\beta/2}}{\delta - |\log(HR)|}\right)^2 ; n_T = kn_C.$$

- Superiority design

The testing hypotheses are:

$$H_0: HR \leq \exp(\delta) \quad \text{vs.} \quad H_1: HR > \exp(\delta)$$

where $\delta > 0$ denote the superiority margin in log-scale. We have

$$n_C = \frac{k+1}{\pi k}\left(\frac{z_{1-\alpha} + z_{1-\beta}}{\log(HR) - \delta}\right)^2 ; n_T = kn_C.$$

2. Observational Studies

Here we only discuss the sample size estimation for two-sided tests. One-sided tests for all cases below are dealt with by changing $(1-\alpha/2)$ to $(1-\alpha)$ in all equations. In observational studies, investigators often can obtain more samples in the control group than in the case group (in case-control studies) or in the unexposed group than in the exposed group (in cohort studies). This imbalance may encourage investigators to collect more data in a study (See our discussion in Section 7: Strategies for reducing sample size). Let n_0 be the sample size of the control/unexposed group and n_1 be the sample size of the case/exposed group. We set k to be the allocation ratio of the sizes of the two groups; that means $n_0 = kn_1$.

2.1 Case-control study – Unmatched

Case-control study is a study that compares patients who have a disease or outcome of interest (cases) with patients who do not have the disease or outcome (controls). It looks back retrospectively to compare how frequently the exposure to a risk factor is present in each group to determine the relationship between the risk factor and the disease. Denote p_0 the probability of exposure in the control group, and p_1 the probability of exposure in the case group. We test

$$H_0: p_0 = p_1 \quad \text{vs.} \quad H_1: p_0 \neq p_1$$

The above hypotheses are equivalent to

$$H_0: OR = 1 \quad \text{vs.} \quad H_1: OR \neq 1,$$

where $OR = p_1(1-p_0)/(p_0(1-p_1))$ is the odds ratio between the case and control groups.

The required sample sizes are [9]

$$n_1 = \frac{\left(z_{1-\alpha/2}\sqrt{(k+1)\bar{p}(1-\bar{p})} + z_{1-\beta}\sqrt{p_0(1-p_0)+kp_1(1-p_1)}\right)^2}{k(p_1-p_0)^2}, n_0 = kn_1 \quad (1)$$

where $\bar{p} = (kp_0 + p_1)/(k+1)$.

If one employs a correction for continuity (an adjustment that is made when a discrete distribution is approximated by a continuous distribution) in statistical analysis, one should use the modified formula [10]:

$$n_{1,cc} = \frac{n_1}{4}\left[1 + \sqrt{1 + \frac{2(k+1)}{kn_1|p_1-p_0|}}\right]^2, n_{0,cc} = kn_{1,cc}. \quad (2)$$

In general situations, equation (2) is preferable to equation (1).

2.2 Case-Control study – Matched

The matched case-control study design has been commonly applied in public health research. Matching of cases and controls is employed to control the effects of known potential confounding variables. The sample size formula was developed by Dupont [11]. To compute the sample size, we need to provide α, β, p_0, p_1, and the correlation coefficient r for exposure in matched pairs of case-control patients. Note that due to the correlation of the paired samples, the original definition of odds ratio in the unmatched case-control study is not valid any more. The odds ratio for a matched case-control study is given by

$$OR_M = \frac{p_1(1-p_0) - r\sqrt{p_1(1-p_1)p_0(1-p_0)}}{p_0(1-p_1) - r\sqrt{p_1(1-p_1)p_0(1-p_0)}}.$$

We test the following hypotheses:

$$H_0: OR_M = 1 \quad \text{vs.} \quad H_1: OR_M \neq 1.$$

The sample sizes are calculated by

$$n_1 = \frac{\left[\frac{1}{\sigma} z_{1-\alpha/2} + z_{1-\beta}\right]^2}{\delta^2}, n_0 = kn_1,$$

where

$$\delta = \frac{1}{\sigma}\left[\sum_{j=1}^{k} \frac{jt_j \cdot OR}{j \cdot OR + k - j + 1} - 1\right], \quad \sigma = \left(\sum_{j=1}^{k} \frac{jt_j(k-j+1) \cdot OR}{j \cdot OR + k - j + 1}\right)^{1/2},$$

$$t_j = p_1 \binom{k}{j-1} p_{0+}^{j-1}(1-p_{0+})^{k-j+1} + (1-p_1)\binom{k}{j} p_{0-}^{j}(1-p_{0-})^{k-j}, \quad j = 1, \ldots, k,$$

$$p_{0+} = p_0 + r\sqrt{(1-p_1)p_0(1-p_0)/p_1}, \qquad p_{0-} = p_0 - r\sqrt{p_1 p_0(1-p_0)/(1-p_1)}.$$

Note that r can be estimated from previous studies. When r is not known, Dupont [11] suggested that it is better to use a small arbitrary value, say 0.2, than it is to assume independence (a value of 0).

2.3 Cohort study – Independent

The sample size formula for an independent cohort study uses the same formula as in an unmatched case-control study [9,10]. Now we assume p_0 is the probability of event in the unexposed group, and p_1 is the probability of event in the exposed group. We shall test

$$H_0: p_0 = p_1 \quad \text{vs.} \quad H_1: p_0 \neq p_1$$

The above hypotheses are the same as

$$H_0: RR = 1 \quad \text{vs.} \quad H_1: RR \neq 1,$$

where $RR = p_1/p_0$ is the relative risk between the exposed and the unexposed groups. The sample size equations are the same as equations (1) and (2) in the independent case-control

study. If an investigator prefers to calculate the sample size based on the relative risk, we refer to Woodward's formula [12], which is another representation of equation (1).

2.4 Cohort study – Paired

For paired cohort studies with dichotomous response variables, our primary interest may be the relative risk of an event between exposed and unexposed patients. It is possible to derive the sample size formula based on the relative risk in an analogous fashion to that of matched case-control study by Dupont [11]. However, here we present a simpler formula by Breslow and Day [13]. Consider the paired sample for the exposed and unexposed groups, where $n_0 = n_1$. We test

$$H_0: RR = 1 \quad \text{vs.} \quad H_1: RR \neq 1,$$

The sample size equation is given by

$$n_0 = n_1 = \frac{\left(\frac{1}{2} z_{1-\alpha/2} + z_\beta \sqrt{p^*(1-p^*)}\right)^2}{p_a p_b (p^* - 0.5)^2},$$

where $p^* = p_b/(p_a + p_b)$, $p_a = p_0(1-p_1) - r\sqrt{p_1(1-p_1)p_0(1-p_0)}$, and $p_b = p_1(1-p_0) - r\sqrt{p_1(1-p_1)p_0(1-p_0)}$. The correlation coefficient r can be estimated from previous studies. When r is not known, we may use a small arbitrary value, for example 0.2.

2.5 Cohort study – Time to event outcomes based on exponential survival model

For the time-to-event outcome in an RCT, we presented a simple formula derived by Schoenfeld [7,8]. We can obtain the expected number of events (deaths) in a trail by specifying the hazard odds under the alternative hypothesis. Investigators may follow a sufficient number of patients (by assuming the average probability that a person experiences an event during the trial) long enough so that the requisite number of events is attained.

More often, we want to consider the following in the proper design of a cohort study with a time-to-event endpoint. Consider a two-arm prospective cohort study with accrual time period

T_a and the follow-up time period T_b. Here the accrual period is the period of time that patients are entering the study, and the follow-up period denotes the period of time after accrual has ended before the final analysis is conducted.

We assume that the proportional hazards assumption holds in the study.

$$HR = \lambda_1(t)/\lambda_0(t), \qquad for\ t \geq 0,$$

where $\lambda_0(t)$ is the hazard for the unexposed group and $\lambda_1(t)$ is the hazard for the exposed group. The testing hypotheses are

$$H_0: HR = 1 \quad \text{vs.} \quad H_1: HR \neq 1.$$

Let D_0 and D_1 be the expected number of events in the unexposed and exposed groups. Rubinstein et al. [14] showed the following relationship for the two-sided test with a significant level of α and the power β:

$$D_0^{-1} + D_1^{-1} = \left(\frac{\log(HR)}{z_{1-\alpha/2} + z_{1-\beta}}\right)^2.$$

If we further assume the survival model is an exponential survival model, the probability π_0 or π_1 that a subject in the unexposed or exposed group experiences an event is given by

$$\pi_i = 1 - \frac{e^{-\lambda_i T_b}\left(1 - e^{-\lambda_i T_a}\right)}{\lambda_i T_a}, \qquad i = 0,1. \tag{3}$$

With some algebra, the sample sizes are given by [15]

$$n_1 = \left(\frac{1}{k\pi_0} + \frac{1}{\pi_1}\right)\left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\log HR}\right)^2; n_0 = kn_1.$$

2.6 Cross-sectional study

Here we describe the sample size calculations where the problem is to compare the means of two independent samples in a cross-sectional study. To test

$$H_0: \mu_0 = \mu_1 \quad \text{vs.} \quad H_1: \mu_0 \neq \mu_1$$

where μ_0 and μ_1 are the means of the end-points in group one and group two. Let n_0 be the sample size in group one, n_1 be the sample size in group two, $k = n_1/n_2$ be the ratio of two sample sizes, and σ^2 be the variance of the two samples (assumed common). We have [12]

$$n_1 = \left(1 + \frac{1}{k}\right)\sigma^2 \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\mu_0 - \mu_1}\right)^2 ; n_0 = kn_1. \quad (4)$$

This formula (4) is also appropriate to the continuous outcomes in cohort studies and case-control studies, although the major end-point of interest of them is usually a proportion rather than a mean. For the dichotomous outcomes in cross-sectional studies, the sample size formulas (1) and (2) in the unmatched case-control study are also suitable here.

2.7 Descriptive survey – Cross-sectional

In descriptive studies, the purpose is to describe one or more characteristics in one particular group using means or proportions. Since the studies are not involved in hypothesis testing, we need to know the *margin of error* in order to compute sample size. The margin of error is defined as half the width (or "radius") of a confidence interval for a particular statistic from a survey. It reflects how precise the statistic, such as mean or proportion, is expected to be.

In studies designed to estimate a mean, the sample size equation [16] is given by

$$n = \frac{z_{1-\alpha/2}^2 \sigma^2}{e^2}.$$

where σ^2 is the variance of the population, and e is the margin of error of the mean.

In studies designed to estimate a proportion, the sample size equation [16] is

$$n = \frac{z_{1-\alpha/2}^2 p(1-p)}{e^2}.$$

where p is the estimate of the proportion to be measured, and e is the margin of error of the proportion.

References

1. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ.* 2010;340:c723.
2. Chow S-C, Shao J, Wang H, Lokhnygina Y. *Sample Size Calculations in Clinical Research.* Third ed: Chapman and Hall/CRC; 2017.
3. Bernardo MVP, Lipsitz SR, Harrington DP, Catalano PJ. Sample size calculations for failure time random variables in non‐randomized studies. *Journal of the Royal Statistical Society: Series D (The Statistician).* 2000;49(1):31-40.
4. Lipsitz SR, Parzen M. Sample size calculations for non‐randomized studies. *Journal of the Royal Statistical Society: Series D (The Statistician).* 1995;44(1):81-90.
5. Lesaffre E. Superiority, equivalence, and non-inferiority trials. *Bull NYU Hosp Jt Dis.* 2008;66(2):150-154.
6. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *British journal of cancer.* 2003;89(2):232.
7. Schoenfeld D. The Asymptotic Properties of Nonparametric-Tests for Comparing Survival Distributions. *Biometrika.* 1981;68(1):316-319.
8. Schoenfeld D. Sample-Size Formula for the Proportional-Hazards Regression-Model. *Biometrics.* 1983;39(2):499-503.
9. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions.* Third ed: John Wiley & Sons; 2013.
10. Fleiss JL, Tytun A, Ury HK. A simple approximation for calculating sample sizes for comparing independent proportions. *Biometrics.* 1980;36(2):343-346.
11. Dupont WD. Power calculations for matched case-control studies. *Biometrics.* 1988;44(4):1157-1168.
12. Woodward M. Formulae for sample size, power and minimum detectable relative risk in medical studies. *Journal of the Royal Statistical Society: Series D (The Statistician).* 1992;41(2):185-196.
13. Breslow NE, Day NE, Heseltine E, Breslow NE. *Statistical Methods in Cancer Research: The Design and Analysis of Cohort Studies.* International Agency for Research on Cancer; 1987.
14. 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(9-10):469-479.
15. Faries DE, Obenchain R, Haro JM, Leon AC. *Analysis of observational health care data using SAS.* SAS Institute; 2014.
16. Cochran WG. *Sampling Techniques.* John Wiley & Sons; 1977.