

Design Considerations for a Two-stage Study with a Continuous Outcome and a Rare Exposure

Hung-Mo Lin^{1*} and John M. Williamson²

¹Department of Health Evidence and Policy, Mount Sinai School of Medicine, New York, NY 10029, USA

²Center for Global Health Research, Centers for Disease Control and Prevention/Kenya Medical Research Institute, 1578 Kisumu-Busia Road, Kisian Kisumu, Kenya

Abstract

We consider the scenario in which the severity of a disease is characterized by a normally distributed response and the chance of being exposed (yes/no) to the risk factor is extremely rare. A screening test is employed to oversample subjects who may be at risk for the disease because expensive laboratory tests are needed to measure the outcome of interest and to confirm the true exposure status. Considerations of sample size and cost are discussed for this type of two-stage design with the objectives of 1) minimizing the number of subjects in the Stage II, and 2) overcoming the problem of a rare exposure. In particular, with an imperfect screening tool, one must take into account the sensitivity and specificity of the screening test, and the uncertainty of the estimates of the exposure prevalence in the survey population. The Penn State Children Sleep Disorder Study (PSCSDS) is used for illustration.

Keywords: Cost; Negative predictive value; Positive predictive value; Rare Exposure; Sample size; Screening; sensitivity; Specificity; *t*-test; Two-stage design

Introduction

In large-scale epidemiologic studies, it is often easy and relatively inexpensive to employ an imperfect screening test (e.g. a questionnaire survey or telephone interview) to determine whether a subject is at increased risk of developing a disease. However, expensive laboratory tests are needed to confirm the true exposure status. Here we consider the situation in which the severity of a disease is characterized by a normally distributed response (e.g. blood pressure) rather than a dichotomized yes/no variable, and the likelihood of being exposed to the risk factor is extremely rare. To investigate the association between exposure and disease, it may only be possible to enroll a subset of a survey sample into a second stage study to ascertain the true exposure status and measure the response of interest because of cost and time considerations.

A valid strategy to follow the above scenario is to select random samples from the survey sample and then analyze the second stage sample using the 2-sample *t*-test. This will result in unbiased estimates of exposure effects, but is highly inefficient. A more powerful approach may be achieved by oversampling the potentially exposed subjects into the second stage study. With an imperfect screening test, it is not easy to derive an explicit sample size formula; one must take into account the sensitivity (SE) and specificity (SP) of the screening test, and the uncertainty of the estimate of the exposure prevalence in the survey population. Another important consideration for optimal study design is the cost for both Stage I and Stage II studies given a fixed amount of the total budget.

Covariate misclassification problems for two-stage design are abundant. Researchers in this area typically concentrate on a binary disease outcome and odds ratio (OR) or relative risk modeling. Almost all work has focused on the framework where disease status is ascertained in the first stage of study, and a subset of sample is collected in the second stage for the purpose of validation of the covariate or collection of additional information on potential confounding factors. White [1], Cain and Breslow [2], and Breslow and Cain [3] laid the foundation for the two-stage design. Zhao and Lipsitz [4] unified and compared twelve types of sampling - three subtypes in the first stage

and four subtypes in the second stage. An overview of cost-efficient designs was presented by Spiegelman [5] and Hanley et al. [6] and Schaubel et al. [7] consider power and sample size issues for two-stage studies with a binary outcome.

The focus of the study design here is different from previous work in several aspects. First, the outcome of interest is a continuous variable and is available only for the second stage sample. Second, the screening test is imperfect and the exposure status cannot be ascertained after the Stage II sample is collected. Therefore, it is important to address this uncertainty of exposure status in the sample size calculation. Third, given a fixed sample size for the Stage II study, it is critical to derive a scheme that would efficiently allocate the Stage I sample to the sample II study.

Our current work is motivated by the Penn State Children Sleep Disorder Study (PSCSDS), designed as a population-based study of the prevalence and correlates of SDB among young children. Many studies have shown that sleep-disordered breathing (SDB) is significantly associated with hypertension and obesity [8,9]. The PSCSD study indicated that children with SDB have increased blood pressure at the borderline of being classified as having hypertension [10]. The study estimated that less than 2% of the children have a mild to moderate form of SDB, defined as an apnea/hypopnea index (AHI) ≥ 5 .

In the first stage of the PSCSDS, questionnaires were sent to the parents to identify some of the signs and symptoms of their children's sleep disorder, such as snoring, breath cessation or difficulty breathing, restless sleep, daytime sleepiness, and school or behavior problems. However, it was not possible to confirm the presence of clinically diagnosed SDB based solely upon parental reports. Therefore the study

***Corresponding author:** Hung-Mo Lin, Department of Anesthesiology, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1010, New York City, New York 10029, USA, Email: Hung-Mo.Lin@mountsinai.org

Received April 16, 2012; Accepted May 25, 2012; Published May 26, 2012

Citation: Lin HM, Williamson JM (2012) Design Considerations for a Two-stage Study with a Continuous Outcome and a Rare Exposure. J Biomet Biostat 3:144. doi:10.4172/2155-6180.1000144

Copyright: © 2012 Lin HM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

required a second stage which involved the selection of a subset of children from those parents who returned the questionnaire. These children then spent one night in the sleep laboratory and only then could SDB status and blood pressure be ascertained. The objective of the sample size calculation is to determine the number of children needed to participate in the Stage II study, and to derive an optimal allocation scheme for sampling subjects from the Stage I study to the Stage II study.

Methods

Sample size calculation for the stage II study

Notation for the Stage I and II studies can be referred to in Table 1. Let w denote the sampling weight for the exposure group; that is the proportion of subjects in the exposure group of the total sample size N in the Stage II study. Assume that w is determined *a priori* by design. For example, if w is 0.5, equal numbers of subjects will be recruited in the exposure and the non-exposure groups. Further assume that subjects will be selected from the Stage I study population, with the goal to have N_1 and N_0 individuals in the exposed and non-exposed groups, respectively. Assume that the outcome of interest for each group follows a normal distribution $N(\mu_i, \sigma_i^2)$, for $i = 0$ (unexposed) and 1 (exposed). The hypothesis of interest is: $H_0: \mu_1 - \mu_0 = 0$ versus $H_1: \mu_1 - \mu_0 \neq 0$. Standard statistical software packages can be used to calculate equal or unequal sample sizes for comparing two group means in the typical 2-sided 2-sample t -test setting (Table 1).

Adjusting for uncertainty of the exposure group membership

Investigators are usually able to assume that exposure status is known when conducting sample size calculations. However, in the current design setting, adjustment for sample size calculation may be needed because membership in the exposed or non-exposed groups cannot be fully determined until the subjects are enrolled in the Stage II study. Unfortunately, calculating N_1 and N_0 during the design stage requires an assumption concerning the proportion of the population classified in each group. This leads to a potential problem in the sample size calculation; after the study is conducted and the group membership is ascertained for each individual, the actual size of each group may not be the same as what was prespecified because of binomial variability. As a result, the actual power based on the calculated sample size is different from that targeted.

Stage I Notation

E	True exposure status with $E = 1$ if exposed and $E = 0$ if non-exposed
X	Screening test of the exposure status with $X = 1$ if positive and $X = 0$ if negative
τ	Probability of a positive screening test = $\Pr[X = 1]$
M	Total sample size of the Stage I study
PPV	Positive predictive value of the screening test = $\Pr[E = 1 X = 1]$
NPV	Negative predictive value of the screening test = $\Pr[E = 0 X = 0]$
C_1	Cost per person in Stage I study

Stage II Notation

f_{x1}	Sampling fraction of those testing positive at Stage I
f_{x0}	Sampling fraction of those testing negative at Stage I
r_{SF}	Sampling fraction ratio $f_{x1} : f_{x0}$
N	Total sample size of Stage II study with N_0 and N_1 in the two exposure groups
w	Sampling weight for the exposure group in Stage II study
C_2	Cost per person in Stage II study
f_{x1}	Sampling fraction of those testing positive at Stage I

Table 1: Notation for Stage I and Stage II Studies.

Lin et al. [11] have previously investigated this issue for comparing two independent means when group membership is subject to uncertainty. They have concluded that the need to increase the sample size to retain targeted power depends only slightly on the values of standardized effect size, but more noticeably on the group weight - the probability of belonging to the exposure group. However, the difference should be negligible unless the group weights are fairly dissimilar between the exposed and non-exposed groups. In general, only 2 to 4 extra subjects are needed when $w = 0.2$, and no more than two extra subjects are needed for $w = 0.3$ across all standardized effect sizes (from 0.2 to 3) to achieve at least 80% study power. If $w = 0.1$, the additional sample size required is 7 to 8 (10 to 11) for targeted power of 80% (90%). However, in the case of a very rare exposure, it is most likely that the proportion of subjects from the exposure group will be less than 10%. Therefore, the impact of very low sampling fraction for the exposure group on sample size calculation deserves special attention.

Briefly, Lin et al. [11] proposed to perform the usual sample size calculation for comparing two groups and then calculate the expected power, which is the weighted average of power estimates across the range of possible N_1 and N_0 values. If the expected power is less than the targeted power, one should increase the total sample size by 1 and repeat until the expected power achieves the targeted power. Here we show in Table 2 the unadjusted sample size and the additional samples required as a function of targeted power (90% and 80%) and standardized effect size (0.2 to 3) for different sampling weights (0.01 to 0.10) of the exposure group in the Stage II study. In the situation where the sampling fraction of the exposure group is extremely low, the additional sample size required is considerably more appreciable when $w \leq 0.1$ (Table 2).

Design considerations for the stage I study

After the sample size for the Stage II study is determined with targeted N_1 and N_0 individuals in the exposed and non-exposed groups, respectively, the next question is how to design a cost-efficient Stage I study. In the case of a very rare exposure and a screening test that produces positive results with a high probability (i.e., specific but not sensitive), the negative predicted value (NPV) will be very high (e.g. > 95%) and the positive predicted value (PPV) will be very low (e.g. < 10%). The important factors influencing the efficiency and cost of the entire study include 1) the sampling fractions of the high and low exposure risk groups, 2) the accuracy of the screening test, and 3) the cost of the screening and laboratory tests in the Stage II study (Table 3).

Table 3 shows the expected cell counts of the exposure (E) cross-classified by screening test (X). Note that in order to acquire at least N_1 subjects from the exposure group to enter the Stage II study, the screening sample, M , needs to be at least $N_1/\Pr(E=1)$. Note that the most efficient design with maximum power is to have equal sample sizes between the two exposure groups (i.e., $M(1-\tau)f_{x0}$ (NPV)+ $M\tau f_{x1}$ (1-PPV)= $M(1-\tau)f_{x0}$ (1-NPV)+ $M\tau f_{x1}$ (PPV) in Table 3). It follows that the sampling fractions for f_{x1} and f_{x0} should satisfy

$$r_{SF} = \frac{f_{x1}}{f_{x0}} = \frac{(1-\tau)(1-2NPV)}{\tau(1-2PPV)}, \text{ with } 0 < r_{SF}.$$

When the exposure is rare, it might not be feasible to have an equal sample size study. In such a case, the sampling fractions f_{x1} and f_{x0} should satisfy

$$r_{SF} = \frac{f_{x1}}{f_{x0}} = \frac{(1-\tau)(1-w-NPV)}{\tau(w-NPV)}, \text{ with } 0 < r_{SF}, \quad (1)$$

Targeted power = 90%

Std ES	w = 0.010		w = 0.025		w = 0.050		w = 0.075		w = 0.100	
	N_{unadj}	Extra	N_{unadj}	Extra	N_{unadj}	Extra	N_{unadj}	Extra	N_{unadj}	Extra
0.2	26536	126	10779	49	5533	23	3789	14	2921	10
0.4	6636	126	2697	48	1385	23	949	15	732	10
0.6	2951	126	1200	49	617	23	423	15	327	10
0.8	1661	127	676	49	348	23	239	15	185	10
1.0	1064	128	434	49	224	23	154	15	119	11
1.2	739	129	302	50	156	24	108	14	84	10
1.4	544	130	222	51	115	24	80	15	62	11
1.6	417	131	171	51	89	24	62	15	48	11
1.8	330	133	135	52	71	24	49	15	39	10
2.0	268	134	110	52	58	24	40	16	32	11
2.2	222	136	92	52	48	25	34	15	27	11
2.4	187	138	77	54	41	25	29	16	23	11
2.6	159	142	66	55	35	26	25	16	20	11
2.8	138	144	57	56	31	26	22	16	18	11
3.0	120	147	50	57	27	27	19	17	16	11

Targeted power = 80%

Std ES	w = 0.010		w = 0.025		w = 0.050		w = 0.075		w = 0.100	
	N_{unadj}	Extra	N_{unadj}	Extra	N_{unadj}	Extra	N_{unadj}	Extra	N_{unadj}	Extra
0.2	19823	82	8053	32	4133	16	2831	10	2183	7
0.4	4958	82	2015	32	1035	15	710	9	547	7
0.6	2205	82	897	32	461	16	317	9	245	7
0.8	1241	83	506	32	261	15	179	10	139	7
1.0	795	83	324	33	168	15	116	9	90	7
1.2	553	84	226	33	117	16	81	10	63	7
1.4	407	84	167	32	87	15	60	10	47	7
1.6	312	86	128	33	67	16	47	10	37	7
1.8	247	87	102	33	53	16	37	10	29	7
2.0	201	88	83	34	44	16	31	10	24	7
2.2	166	90	69	35	37	16	26	10	21	7
2.4	140	91	58	36	31	17	22	11	18	7
2.6	120	93	50	36	27	17	19	11	16	7
2.8	104	95	44	36	24	17	17	11	14	7
3.0	91	97	38	38	21	17	15	11	12	8

Table 2: The unadjusted sample size (N_{unadj}) and the additional sample required (extra) as a function of targeted power, standardized effect size (Std ES) and sampling weight for the exposure group (w). A two-sided two-sample t - test was used for hypothesis testing at the $\alpha = 0.05$ level.

to obtain a pre-specified w . However, because $0 < r_{SP}$, the specification of w is subject to the following constraint:

$$PPV < w < 1 - NPV \quad \text{or} \quad 1 - NPV < w < PPV. \quad (2)$$

Suppose the estimated sample size is N for the Stage II study, then the sampling fraction for those testing negative at Stage I is

$$f_{x0} = \min\left(\frac{N}{(1 - \tau + \tau \cdot r_{SF})M}, 1\right). \quad (3)$$

Suppose the total cost of the study needs to be constrained to the amount of $C_T = C_1 \times M + C_2 \times M [(1 - \tau) f_{x0} + \tau f_{x1}]$. The cost of Stage II can be also deemed as fixed because N subjects are required to attain the pre-specified power in Stage II. Therefore, the number of subjects in Stage I is constrained to be $M = (C_T - C_2 \cdot N) / C_1$, and f_{x0} in (3) is

$$f_{x0} = \min\left(\frac{N \cdot C_1}{(1 - \tau + \tau \cdot r_{SF})(C_T - C_2 \cdot N)}, 1\right). \quad (4)$$

The penn state children sleep disorder study

The main hypothesis of the Penn State Children Cohort was that children with SDB have increased blood pressure at the borderline of

being classified as having hypertension [10]. The study estimated that less than 2% of the children have a mild to moderate form of SDB (AHI ≥ 5).

Suppose we wish to design a study with 90% power to detect an increase in systolic blood pressure in children with SDB by at least 10 mm Hg compared to the normal group, assuming a common SD=12 using a two sample t-test with a 0.05 two-sided significance level. Questionnaires will be sent to parents to identify some of the signs and symptoms of their children’s sleep disorder, such as snoring; breathe cessation or difficulty breathing, restless sleep, daytime sleepiness, and school or behavior problems. A child will be considered at risk of SDB if more than two signs or symptoms are present. Many of the children are expected to fall into the at-risk category without actually having the disorder. Therefore, the screening tool based on the survey is expected to be specific but not sensitive.

For illustration, we assume that $NPV = 0.99$, $PPV = 0.06$ and $\tau = 0.20$, which is consistent with $Pr[E=1] = 0.02$. Based on equation (2), the sampling weight, w , needs to be between 0.01 and 0.06. Clearly, a study of equal sample sizes for the two exposure groups, although is most powerful, is not feasible.

Here we first consider the scenario in which the investigators plan to collect 3000 screening samples, which is the estimated number of eligible children from the five neighboring schools. The expected number of SDB cases is 60 ($3000 \times 0.02 = 60$) from the screening sample. Assume first that the investigators are unaware of the restriction placed on w . To acknowledge that SDB is rare among young children, they choose a *priori* a small w of 0.05 for the sample size calculation in the Stage II study, which result in needing a total of 340 subjects to achieve 90% power using the two sample t -test. The next question is to determine who is to be invited to the sleep laboratory based on the screening results. Substituting $w = 0.05$, NPV = 0.99, PPV = 0.06 and $\tau = 0.20$ in equation (1) will result in $r_{SF} = 16$ and then $f_{x0} = 2.83\%$ because $3000 \times (1-0.2) \times f_{x0} + 3000 \times 0.2 \times 16 \times f_{x0} = 340$ (see Table 3). Subsequently, our method suggests that the investigators should randomly sample 45.33% ($N = 272$) of the subjects from those who had a positive screening test, and 2.83% ($N = 68$) of the subjects who had a negative screening test.

Using the given information on the NPV and PPV, one can show that the expected numbers of SDB and non-SDB cases from the selected Stage I samples should reach the targeted sample sizes as desired for the Stage II study. However, the true SDB status cannot be confirmed until much later. To account for the uncertainty of the SDB status during the design stage, we need approximately 24 additional subjects in the Stage II study to guarantee 90% targeted power (Std ES = $10/12 = 0.83$; see Table 2). Correspondingly, the investigators need to randomly sample 48.5% ($N = 291$) of the subjects from those who had a positive screening test, and 3.0% ($N = 73$) of the subjects who had a negative screening test.

Cost Considerations

Suppose that the cost for each subject in the Stage II study is \$300 and \$10 for each survey questionnaire. The total cost for aforementioned scenario is $\$300 \times 364 + \$10 \times 3000 = \$1,39,200$. Can the cost be reduced to approximately \$100 000 by perhaps increasing the survey sample and decreasing the laboratory sample, or by decreasing both the survey and laboratory samples? The two aforementioned constraints require that

	$E = 0$	$E = 1$	
$X = 0$	$M(1-\tau) f_{x0}$ (NPV)	$M(1-\tau) f_{x0}$ (1-NPV)	$M(1-\tau) f_{x0}$
$X = 1$	$M\tau f_{x1}$ (1-PPV)	$M\tau f_{x1}$ (PPV)	$M\tau f_{x1}$
	N_0	N_1	N

- M Total sample size of the Stage I study
- τ Probability of a positive screening test = $\Pr[X=1]$
- f_{x0} Sampling fraction of those testing negative from the screening test at Stage I
- f_{x1} Sampling fraction of those testing positive from the screening test at Stage I
- N Total sample size of the Stage II study with N_0 and N_1 in the non-exposure group and the exposure group, respectively

Table 3: Expected Cell Counts of the Exposure (E) cross-classified by Screening Test (X)

Total Cost	M	f_{x0}	f_{x1}	$N_{x=0}$	$N_{x=1}$	N	N_0	N_1	Targeted Power
\$100 000	1330	.005	1.00	6	266	272	256	16	89%
\$105 000	1830	.004	.774	6	283	289	272	17	90%
\$110 000	2330	.003	.608	6	283	289	272	17	90%
\$115 000	2830	.003	.500	6	283	289	272	17	90%
\$120 000	3330	.002	.425	6	283	289	272	17	90%
\$125 000	3830	.002	.370	6	283	289	272	17	90%
\$130 000	4330	.002	.327	6	283	289	272	17	90%
\$135 000	4830	.001	.293	6	283	289	272	17	90%
\$140 000	5330	.001	.266	6	283	289	272	17	90%

Table 4: Sample Size and Cost Considerations for the PSCSDS.

$M > 19/0.02 = 950$, and that $0.01 < w < 0.06$. To maximize the power for the Stage II study so that w could be as close to 0.06 as possible, we use $N_1: N_0 = 1:16$ ($w = 0.059$). The sample sizes of $N_1 = 16$ and $N_0 = 256$ will result in 89% power, or $N_1 = 17$ and $N_0 = 272$ to guarantee the targeted power of at least 90%. Table 4 displays several designs under different cost considerations. Compared to the first scenario where $w = 0.05$, a slight increase of w to 0.059 and a decrease in number of screening samples could greatly reduce the cost of the study (Table 4).

Discussion

We considered the design of a two-stage study in which disease severity is characterized by a normally distributed variable and the chance of exposure to the risk factor is extremely rare. Although power and sample size calculation for the two-sample t -test with unequal group sizes has been studied, a unique two-stage study design is implemented to increase power and decrease cost. First, a screening test is employed to a large population to oversample subjects who may be at risk for the disease because expensive laboratory tests are needed to confirm the true exposure status. However, with an imperfect screening tool one must take into account the SE and SP of the screening test, and the uncertainty of the estimates of the exposure prevalence in the survey population. We first focused on sample size calculation using the typical two-sample t -test for the Stage II study. We corrected for uncertainty of the group sizes. Based on the estimated sample size for each group, we then consider the design issues for the Stage I study to increase power and decrease cost.

In order to achieve a more optimal Stage I design, information concerning the accuracy of the screening test has been assumed to be known. We recommend that investigators conduct a pilot study with subjects drawn from the same study population to estimate the SE, SP, NPV, and PPV of the screening test. Alternatively, different assumptions for the misclassification parameters can be varied in sensitivity analyses to determine how they might impact the study design.

The proposed two-stage design is motivated with the objectives of 1) minimizing the number of subjects in Stage II stage due to expensive cost, and 2) overcoming the problem of a rare exposure. In the situation where the outcome measure is inexpensive to measure, it is plausible to obtain both screening and outcome information simultaneously in a large sample without the concern of inadequate study power. However, adjustment for misclassification of the exposure status is still needed at the analysis stage and investigators will still need to conduct either an internal or external validation study. Recently, researchers have applied designs combining both types of external and internal validation data. Detailed discussion can be found in Greenland [12], Thurston et al. [13], and Lyles et al. [14].

References

1. White JE (1982) A two stage design for the study of the relationship between a rare exposure and a rare disease. Am J Epidemiol 115: 119-128.
2. Cain KC, Breslow NE (1988) Logistic regression analysis and efficient design for two-stage studies. Am J Epidemiol 128: 1198-1206.
3. Breslow NE, Cain KC (1988) Logistic regression for two-stage case-control data. Biometrika 75: 11-20.
4. Zhao LP, Lipsitz S (1992) Designs and analysis of two-stage studies. Stat Med 11: 769-782.
5. Spiegelman D (1994) Cost-efficient study designs for relative risk modeling with covariate measurement error. J Stat Plan Inference 42: 187-208.
6. Hanley JA, Csizmadia I, Collet JP (2005) Two-stage case-control studies: Precision of parameter estimates and considerations in selecting sample size. Am J Epidemiol 162: 1225-1234.

7. Schaubel D, Hanley J, Collet JP, Boivin JF, Sharpe C, et al. (1997) Two-stage sampling for etiologic studies. Sample size and Power. *Am J Epidemiol* 146: 450-458.
8. Tauman R, Gozal D (2006) Obesity and obstructive sleep apnea in children. *Paediatr Respir Rev* 7: 247-259.
9. Arens R, Muzumdar H (2010) Childhood obesity and obstructive sleep apnea syndrome. *J Appl Physiol* 108: 436-444.
10. Bixler EO, Vgontzas AN, Lin HM, Liao D, Calhoun S, et al. (2009) Sleep disordered breathing in children in a general population sample: Prevalence and risk factors. *Sleep* 32: 731-736.
11. Lin HM, McClintock SK, Williamson JM (2011) Correction for two-group sample size calculation with uncertain group membership. *J Data Sci* 9: 155-170.
12. Greenland S (1988) Variance estimation for epidemiologic effect estimates under misclassification. *Stat Med* 7: 745-757.
13. Thurston SW, Williams PL, Hauser R, Hu H, Hernandez-Avila M, et al. (2003) A comparison of regression calibration approaches for designs with internal validation data. *J Stat Plan Inference* 131: 175-190.
14. Lyles R, Zhang F, Drews-Botsch C (2007) Combining internal and external validation data to correct for exposure misclassification: a case study. *Epidemiology* 18: 321-328.