Sample Size Estimation In Cohort Studies For Testing Of Relative Risk

Subramanian Chandrasekaran, Jayadevan Sreedharan, Aji Gopakumar

Abstract: Cohort study is one of the epidemiological research designs to determine incidence rate and risk associated with the health outcome. The study identifies the risk factors of the outcome variable by estimating relative risk (RR) and test of hypothesis about RR. Determining an adequate sample size (n) with sufficient subjects is an important step in any research for generalization of the sample results. This paper discusses a method for calculating adequate sample size in a cohort study, specifically to test the 'significance of RR in a regression model'. The formula proposed in this paper includes significance level (α), power (1-β), degree of effect to be detected (RR) and desired precision. Simulation is carried out to evaluate the performance of the proposed sample size formula in deriving adequate 'n'.

Index Terms: Adequate sample size, epidemiological design, cohort study, relative risk, incidence rate, precision, logistic regression model.

1. INTRODUCTION

Determination of sample size is an important step in any research. Calculation of adequate sample size leads to inclusion of sufficient number of study subjects to detect significant effect size and facilitate valid generalization of sample results to the whole population. There is different sample size formulas derived according to the design of the study and statistical techniques that expected to apply [1], [2]. Cohort study is one of the observational studies in Medical research, aims to estimate the incidence rate and test the significance of risk ratio. In this design, a cohort is selected on the basis of exposure status as exposed/unexposed. Incidence is measured by regular follow up of the cohort for a period of time. The study involves large number of subjects for a long period of time according to the exposure and outcome of interest. A cohort study is generally carried out after exploring a hypothesis which is strengthened from quicker studies such as cross sectional or case-control studies. The choice of the cohort for a research specifically depends upon the investigated hypothesis and the population has been initially surveyed to explore the presence of exposure. Cohort studies are more suitable if the research is related to an exposure which is rare. In case of rare exposure, lesser number of people may be exposed to the factor and therefore study population consists of deliberate selection of larger exposed group (/highly exposed). As an involvement of large number of subjects in the research, variety of common exposures in relation to several health outcomes can be studied in cohort studies. A similar kind of comparison group is required to be chosen including the participants who are 'unexposed' or 'exposed at low level' to an exposure understudy [3], [4], [5]. Cohort studies are mainly of two types, prospective and retrospective studies.

In prospective cohort studies, information on exposure will be collected at present and hence large number of people should be followed up for longer period of time from exposure to disease onset. In retrospective cohort studies, information on exposure will be available from the existing authentic records and consequently time to wait for effect (risk of disease) may be reduced / eliminated. As the focus of cohort studies is follow up till the development of the disease (/multiple outcomes), researchers ensure absence of outcome at entry level. But it may not be ensured all the time, especially in conditions with long latent period and those cases get excluded from the research. Similarly, subjects are often removed from the study for various reasons and hence cohort size should be adequately collected taking all these points into consideration. Exposure effect is measured using Relative Risk (RR) or Risk Difference considering equal/unequal follow-up duration. In order to calculate the exposure effect (RR), incidence needs to be measured among exposed and unexposed; then comparison to be done across the groups. In order to derive valid conclusion, recruitment of sufficient and representative sample is a vital step prior to initiation of any kind of research. Results of the study may mislead if size of the sample is inadequate, however each steps in research are appropriate. Both very large and small samples fail to identify significance of real effect, hence calculation of adequate number of subjects needs to be calculated according to the type of statistical test and study design [6], [7], [8]. Cohort studies generally aim the comparison of exposed and unexposed groups in terms of incidence of health outcome. Objective of this paper is to derive a new formula for adequate sample size in a cohort study. Studies reported that sample size formula (n) for testing of RR in a cohort study includes the factors such as significance level (α), power (1-β), number of control subjects per experimental subjects (m), incidence rates (p_1 & p_2) among exposed and unexposed group [9], [10]. A study with estimation of RR doesn't need to include power term (Z_β) in its sample size formula, but it is necessarily to be included if the purpose of the study is to test a hypothesis related to RR.

- *Subramanian Chandrasekaran is a PhD holder (Statistics) and Associate Professor of Dept. of Statistics, Annamalai University, India. PH- +919715263275. E-mail: manistat@yahoo.co.in5*
- *Jayadevan Sreedharan has double PhD in Biostatistics and Epidemiology, Professor of Dept. of Community Medicine, Gulf Medical University, UAE. PH-+9710796101. E-mail: drjayadevans@gmail.com*
- *Aji Gopakumar is a Research Scholar, currently pursuing PhD under the Dept. of Statistics, Annamalai University, India. PH- +9710992363. E-mail: ajitha_gopan@yahoo.co.in*

2 EXISTING METHOD OF SAMPLE SIZE CALCULATION IN COHORT STUDIES

Since selection of sample in a cohort study is based on exposure status, study will be initiated with two groups such as exposed and unexposed; RR will be the parameter of interest

[11], [12]. The null hypothesis for the test of RR can be expressed as H0: RR=1. Null hypothesis indicates that event rate is same among exposed and unexposed group. In other words, the proportion of disease is same across exposed and unexposed groups. Therefore the null hypothesis can also be written as H0: P1=P2 against the two tailed alternative hypothesis H1: P1≠P2. Here P1 and P2 are the incidence rates among exposed and unexposed group respectively. Implies, the test of hypothesis about RR is equivalent to test of two sample proportions [13]. Since RR is the risk ratio to be detected and P2 the disease rate among unexposed, both are generally known in most cases. Therefore P1 can be expressed as P1= (RR) P2. The existing formula for determining sample size in two-sided test of RR is given as follows [14], [15]

$$n = \frac{[Z_{1-\alpha/2}\sqrt{2p(1-p)} \ + \ Z_{1-\beta}\sqrt{[p_1(1-p_1)+p_2(1-p_2)]}]^2}{[p_1-p_2]^2}$$

Where p= $(p_1+p_2)/2$ and the constraint $0 < RR < 1/P_2$

$Z_{(1-\alpha/2)}$ is the standard normal variate at the chosen level of significance (z=1.96 and 2.58 at 5% and 1% level of significance respectively) and $Z_{(1-\beta)}$ is the power term which commonly fixed at 80%.

3 PROPOSED SAMPLE SIZE FOR COHORT STUDY

Theoretically, sample size (n) is the ratio of standard deviation (SD) to the standard error (SE),

$$n = \frac{SD^2}{SE^2} \quad \text{-------------------------------- (A)}$$

The current study interested to focus on the simple regression model and the sample size required for finding association between the response variable and one predictor variable which are dichotomous. The corresponding logistic model can be expressed in the following function,

$$y = \frac{e^{x\beta}}{1+e^{x\beta}}$$

where e^{β} is the relative risk (risk ratio).

In reference to Confidence Interval (CI) of the parameter of interest, CI can be expressed as

CI = (estimate) ± (Critical value) (SE of the estimate)

where (Critical value x SE of the estimate) is considered as the margin of error (L).

In a cohort study, investigator's interest is to test the significance of Relative Risk (risk ratio). Hence the confidence interval for the relative risk can be expressed in the following notation,

$e^{\beta} \pm Z(\alpha/2)SE\,(e^{\beta})$ ………..……(1)

Since we have relative risk and its standard error as follows,

Relative risk, $e^{\beta}=p_1/p_2$ = (incidence rate among exposed group)(incidence rate among unexposed group)

Implies, Log e^{β}=Log p_1/p_2

$$\text{Standared Error of } (Log\ \ e^{\beta}) = \sqrt{\frac{1-p_1}{n_1p_1}+\frac{1-p_2}{n_2p_2}}$$

$$\text{SE of } (\beta) = \sqrt{\frac{1-p_1}{n_1p_1}+\frac{1-p_2}{n_2p_2}}$$

From equation(1),

$$L = Z_{\frac{\alpha}{2}}\ \text{SE}\ (\widehat{\beta})$$

$$= Z_{\frac{\alpha}{2}}\sqrt{\frac{1-p_1}{n_1p_1}+\frac{1-p_2}{n_2p_2}} = Z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n}(\frac{1-p_1}{p_1}+\frac{1-p_2}{p_2})}$$ with an assumption of $n=n_1=n_2$

$$\sqrt{n} = \frac{Z\alpha_{/2}\sqrt{(\frac{1-p_1}{p_1}+\frac{1-p_2}{p_2})}}{L} \quad \text{……………..….. (2)}$$

Since margin of errror represents shift from the true estimate, denominator of equation (2) can be represented as difference in the disease rate (p_1-p_2)

$$\sqrt{n} = \frac{Z\alpha_{/2}\sqrt{(\frac{1-p_1}{p_1}+\frac{1-p_2}{p_2})}}{p_1-p_2} \quad \text{………………...(3)}$$

Since the null hypothesis tested in a cohort study is H_0: RR=1 (or H_0: $p_1=p_2$), an equivalent hypothesis can be expressed as H_0: $p_1-p_2=0$ against H_1: $p_1-p_2\neq0$

If the null hypothesis is true (under H_0), the upper 100(α/2)th percent point of the distribution of p which centered at 0 can be espressed as

$a = 0 + Z_{\frac{\alpha}{2}}\sqrt{2p(1-p)/n}$ …………..(4)

the distribution if the alternative hypothesis is true (at H_1 where distribution centered at p_1-p_2),

$a = (p_1-p_2) - Z_{\beta}\sqrt{p_1(1-p_1)/n+p_2(1-p_2)/n}$

$a + Z_{1-\beta}$ SE(p_1-p_2) = p_1-p_2 ……………….(5)

when expression of equation (4) substituted in equation (5), we will get

$$Z_{\frac{\alpha}{2}}\sqrt{2p(1-p)/n} \ + \ Z_{\beta}\sqrt{p_1(1-p_1)/n+p_2(1-p_2)/n} = p_1-p_2$$ ……………..(6)

From equation 3, we have $p_1-p_2 = \frac{Z\alpha_{/2}\sqrt{(\frac{1-p_1}{p_1}+\frac{1-p_0}{p_0})}}{\sqrt{n}}$

while substituting for p_1-p_2 in equation (6), we get

$$\frac{Z_{\alpha}\sqrt{2P'(1-P')/n} \ + \ Z_{\beta}\sqrt{\frac{1}{n}[p_1(1-p_1)+p_2(1-p_2)]}}{Z\alpha_{/2}\sqrt{(\frac{1-p_1}{p_1}+\frac{1-p_2}{p_2})}} = \frac{1}{\sqrt{n}}$$

Devide both sides by $\frac{1}{\sqrt{n}}$

$$\frac{1}{\frac{1}{\sqrt{n}}\sqrt{n}} = \frac{\frac{1}{\sqrt{n}}[Z_{\alpha}\sqrt{2P'(1-P')} \ + \ Z_{\beta}\sqrt{[p_1(1-p_1)+p_2(1-p_2)]}]}{\frac{1}{\sqrt{n}}\ \ Z\alpha_{/2}\sqrt{(\frac{1-p_1}{p_1}+\frac{1-p_2}{p_2})}}$$

$$\sqrt{n} = \frac{Z_{\alpha}\sqrt{2P'(1-P')} \ + \ Z_{\beta}\sqrt{[p_1(1-p_1)+p_2(1-p_2)]}}{Z\alpha_{/2}\sqrt{\frac{1}{n}(\frac{1-p_1}{p_1}+\frac{1-p_2}{p_2})}}$$

$$n = \frac{[Z_{\alpha}\sqrt{2P'(1-P')} \ + \ Z_{\beta}\sqrt{[p_1(1-p_1)+p_2(1-p_2)]}]^2}{[Z\alpha_{/2}\sqrt{\frac{1}{n}(\frac{1-p_1}{p_1}+\frac{1-p_2}{p_2})}]^2}$$

$$n = \frac{[Z_{\alpha}\sqrt{2P'(1-P')} \ + \ Z_{\beta}\sqrt{[p_1(1-p_1)+p_2(1-p_2)]}]^2}{(Precision)^2}$$

is the adequate sample size required to test the significance of RR in a cohort study. Formula includes p_1 and p_2 which are the incidence rates among exposed and unexposed group. Z_{α} and Z_{β} are the standard normal variates for level of significance and

power respectively. RR is the required relative risk to be detected.
Precision can be defined as,
Precision = [Standard Normal Variate] x [Standard Error]
p_1 = (RR) p_2 and p = $\frac{p_0+p_1}{2}$
Hypothesis about the relationship or causation can be proven with the help of a cohort study, which is considered as the best method of causation (Grimes & Schulz, 2002). Current research paper introduced a sample size formula which is derived by taking both 'inferential techniques and study design' into consideration.

4 SIMULATION RESULTS

4.1 Simulation Results with Existing Sample Size Formula

In order to observe the performance of proposed sample size formula, simulation is carried out and various trails were performed for different values of significance level and power. For comparison, minimum sample size for the test of RR is calculated by the existing method in Table 1. 'n' is calculated for fixed level of significance (α=0.05), power (80%) and different values of P2. Calculated sample size is presented in the table 1 according to the RR to be detected.

4.2 Simulation Results by the Proposed Method of Sample Size

To verify the accuracy of proposed sample size formula, sample size is calculated at different values of precision as well as various disease rates among the unexposed group (p2). Level of significance and power has fixed for varied disease rates and precision. In order to find 'n' for different levels of precision, SE can be identified from previous studies which is then multiplied with the Z score. Repeated trials were performed for different power values and significance levels in accordance with the RR to be detected. Simulation results revealed that lesser samples are required to detect larger RR and for higher incidence rates. The findings agree with the theoretical concept of requirement of larger samples to detect a subtle exposure effect. Adequate sample size is presented in table 2 for α level 0.05, 1-β at 0.80 and various values of p2 (0.01-0.2). Precision is calculated at various values of Standard Error (SE) which is multiplied with Z score (table 2-4). From tables 1-4, proposed sample size gives an adequate large 'n' in a cohort study to follow up for longer period of time. Since the calculation consists of precision that required for valid inferential statistics, comparatively a better large 'n' is reflected as adequate sample size for different values of standard error.

5 CONCLUSION

Sample size proposed in this research is calculated for the desired precision of the estimate; it is considered as the allowable or acceptable error in the estimate. Adequate sample size depends on how precise an investigator needs his descriptive and inferential statistics. Precision has an important role for the estimation of statistics; accordingly marginal of error is fixed for determination of sample size. Hence major components which are affecting the precision are margin of error and confidence level. The proposed sample size formula calculates adequate sample size considering these two major components. An adequate large sample reduces the margin of error and subsequently increase the precision; Confidence level ensures the results found are accurate within this margin of error. Compared to the size obtained from existing method, the sample size proposed in this study gives an adequate larger sample considering the required precision.

ACKNOWLEDGMENT

The authors would like to acknowledge Annamalai University, India and Gulf Medical University, UAE for the support provided to conduct this research as a part of PhD completion.

REFERENCES

[1] M.A. Pourhoseingholi, M. Vahedi and M. Rahimzadeh, "Sample size calculation in medical studies," Gastroenterol Hepatol Bed Bench, vol. 6, no. 1, pp. 14-7, 2013.
[2] P. Dattalo, Determining sample size: Balancing power, precision and practicality. Oxford University Press Inc, New York, 2008.
[3] S.I. Dos Santos, Cancer Epidemiology: Principles and methods. International Agency for Research on Cancer, World Health Organization, ISBN: 978-92-832-0405-3, 1999.
[4] D. Barrett and H. Noble, "What are cohort studies?" Evidence-Based Nursing, vol. 22, no. 4, pp. 95-96, 2019.
[5] H. Liu, Y. Shen, J. Ning, and J. Qin. "Sample size calculations for prevalent cohort designs," Stat Methods Med Res., vol. 26, no. 1, 280–291, 2017.
[6] C.J. Mann, "Observational research methods. Research design II: cohort, cross sectional, and case-control studies," Emergency Medicine Journal, vol. 20, no. 1, pp. 54-60, 2003.
[7] C.J. Mann, "Observational research methods—Cohort studies, cross sectional studies, and case–control studies," African Journal of Emergency Medicine, vol. 2, no. 1, pp. 38-46, 2012.
[8] Centre for Evidence-Based Medicine (CEBM). "Study Designs". 2014. Retrieved from https://www.cebm.net/2014/04/study-designs/
[9] V. Kasulevičius, V. Sapoka, R. and Filipavičiūtė, "Sample size calculation in epidemiological studies". Gerontologija; vol 7, no. 4, pp. 225–231, 2016.
[10] J. Charan and T. Biswas, "How to Calculate Sample Size for Different Study Designs in Medical Research?" Indian Journal of Psychological Medicine, vol. 35, no. 2, pp. 121-126, 2013.
[11] P.M. Bernard, C. Lapointe, "Calculation of sample size for the detection of relative risk with a predetermined confidence level and power". Rev Epidemiol Sante Publique, vol. 38, no. 3, pp. 255-60, 1990.
[12] K.J. Lui, "Exact equivalence test for risk ratio and its sample size determination under inverse sampling," Stat Med., vol. 16, no. 15, pp. 1777-86, 1997.
[13] H. Wang and S.C. Chow. "Sample size calculation for comparing proportions," 2nd Ed., Chapman and Hall/CRC, New York, 2007.
[14] S. Lemeshow, D.W. Hosmer and J. Klar, "Sample size requirements for studies estimating odds ratios or relative risks". Stat Med., vol. 7, no. 7, pp. 759-64, 1988.
[15] S. Lemeshow, D.W. Hosmer, J. Klar and S.K. Lwanga. "Adequacy of Sample Size in Health Studies," World Health Organization, 1990.

TABLE 1: *SAMPLE SIZE BY EXISTING FORMULA FOR FIXED VALUES OF α=0.05, 1-β=0.80 AND INCIDENCE RATE AMONG UNEXPOSED GROUP (P_2)*

RR	P_2 - INCIDENCE RATE AMONG UNEXPOSED GROUP																
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15	0.18	0.20
0.1	962	478	316	236	187	155	132	115	101	90	82	74	68	63	58	47	42
0.2	1339	665	440	328	261	216	184	160	141	126	114	103	95	87	81	66	58
0.3	1907	947	627	467	371	307	261	227	200	179	161	147	135	124	115	94	83
0.4	2811	1396	924	688	546	452	384	334	295	263	237	216	198	182	169	137	122
0.5	4358	2163	1431	1065	846	699	595	516	455	407	367	333	305	281	260	211	187
0.6	7293	3618	2393	1781	1413	1168	993	862	760	678	611	556	508	468	433	351	311
0.7	13824	6855	4532	3371	2674	2209	1878	1629	1435	1280	1153	1048	959	882	816	661	583
0.8	33034	16374	10821	8044	6378	5267	4474	3879	3416	3046	2743	2491	2277	2094	1935	1565	1380
0.9	139845	69285	45765	34005	26949	22245	18885	16365	14405	12837	11554	10485	9581	8805	8133	6565	5781
1.1	155232	76832	50699	37632	29792	24565	20832	18032	15854	14112	12687	11499	10494	9632	8885	7143	6272
1.2	40728	20148	13288	9858	7800	6428	5448	4713	4141	3684	3309	2998	2734	2508	2312	1854	1626
1.3	18953	9371	6177	4580	3622	2983	2526	2184	1918	1705	1531	1386	1263	1158	1066	853	747
1.4	11140	5505	3626	2687	2124	1748	1480	1278	1122	997	894	809	737	675	621	496	433
1.5	7435	3672	2418	1791	1414	1163	984	850	745	662	593	536	488	447	411	327	285
1.6	5376	2653	1746	1292	1020	839	709	612	536	476	426	385	350	320	294	234	203
1.7	4105	2025	1332	985	777	638	539	465	407	361	323	292	265	242	222	176	153
1.8	3262	1608	1057	781	616	506	427	368	322	285	255	230	209	191	175	138	120
1.9	2671	1316	865	639	503	413	348	300	262	232	208	187	170	155	142	112	97
2	2240	1103	724	534	421	345	291	250	219	193	173	156	141	128	118	92	80
2.5	1164	571	374	275	216	176	148	127	110	97	87	78	70	63	58	45	38
3	748	366	239	175	137	111	93	80	69	60	54	48	43	39	35	26	22
3.5	539	263	171	125	97	79	66	56	48	42	37	33	29	26	24	17	14
4	415	202	131	95	74	60	49	42	36	31	27	24	21	19	17	12	10
4.5	335	162	105	76	59	47	39	33	28	24	21	18	16	14	13	9	7

TABLE 2: *ADEQUATE SAMPLE SIZE FOR FIXED VALUES α=0.05, 1-β=0.80, AND VARYING PRECISION & P_2 (PROPOSED METHOD)*

RR	P_2 - INCIDENCE RATE AMONG UNEXPOSED GROUP																
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.15	0.18	0.2
0.1	1058	525	348	259	206	170	145	126	111	99	90	82	75	69	64	52	46
0.2	1460	725	480	358	284	235	200	174	153	137	124	112	103	95	88	72	63
0.3	2065	1025	679	505	401	332	282	245	216	193	174	159	145	134	124	101	89
0.4	3026	1502	994	740	587	486	413	359	316	282	255	232	212	195	181	147	130
0.5	4668	2316	1532	1140	904	748	636	552	486	434	391	356	325	300	277	225	199
0.6	7776	3856	2549	1896	1504	1243	1056	916	807	720	649	589	539	496	459	372	328
0.7	14682	7277	4809	3575	2835	2341	1989	1724	1518	1354	1219	1107	1012	931	860	696	613
0.8	34961	17321	11441	8501	6737	5561	4721	4091	3601	3209	2889	2621	2395	2201	2033	1641	1445
0.9	147544	73064	48237	35824	28376	23410	19864	17204	15135	13480	12126	10997	10042	9224	8514	6859	6032
1.1	162910	80590	53150	39430	31198	25710	21790	18850	16563	14734	13237	11990	10935	10030	9246	7417	6502
1.2	42645	21085	13898	10305	8149	6711	5685	4915	4316	3837	3445	3118	2841	2605	2399	1920	1681
1.3	19804	9786	6447	4777	3776	3108	2631	2273	1995	1772	1590	1438	1310	1200	1104	882	770
1.4	11618	5738	3778	2798	2210	1818	1538	1328	1164	1034	927	838	762	698	642	511	446
1.5	7741	3821	2514	1861	1469	1207	1021	881	772	685	614	554	504	461	423	336	293
1.6	5587	2756	1813	1341	1058	869	734	633	554	491	440	397	361	330	303	240	208
1.7	4261	2101	1381	1021	805	661	558	481	421	373	333	301	273	249	229	180	156
1.8	3381	1666	1094	808	637	522	441	380	332	294	263	237	215	196	179	141	122
1.9	2765	1362	894	660	520	426	359	309	270	239	213	192	174	159	145	114	98
2	2316	1140	748	552	434	356	300	258	225	199	177	160	144	132	120	94	81
2.5	1197	587	384	282	221	181	152	130	113	99	88	79	71	65	59	45	38
3	767	375	244	179	140	114	95	81	70	62	54	48	43	39	35	27	22
3.5	551	268	174	127	99	80	67	57	49	43	37	33	30	26	24	17	14
4	423	206	133	97	75	61	50	42	36	31	28	24	21	19	17	12	10
4.5	341	165	106	77	60	48	39	33	28	24	21	18	16	14	13	9	7

TABLE 3: *ADEQUATE SAMPLE SIZE FOR FIXED VALUES OF A=0.05, 1-B=0.80, AND VARYING P_2 & SE (PRECISION=Z SCORE X SE)*

RR	P_2 - INCIDENCE RATE AMONG UNEXPOSED GROUP															
	0.01		0.02		0.03		0.04		0.05		0.06		0.07		0.08	
	SE	N	SE	N	SE	N	SE	N	SE	N	SE	N	SE	N	SE	N
0.1	0.005	1058	0.009	525	0.014	348	0.018	259	0.023	206	0.028	170	0.032	145	0.037	126
0.2	0.004	1460	0.008	725	0.012	480	0.016	358	0.020	284	0.024	235	0.029	200	0.033	174
0.3	0.004	2065	0.007	1025	0.011	679	0.014	505	0.018	401	0.021	332	0.025	282	0.029	245
0.4	0.003	3026	0.006	1502	0.009	994	0.012	740	0.015	587	0.018	486	0.021	413	0.024	359
0.5	0.003	4668	0.005	2316	0.008	1532	0.010	1140	0.013	904	0.015	748	0.018	636	0.020	552
0.6	0.002	7776	0.004	3856	0.006	2549	0.008	1896	0.010	1504	0.012	1243	0.014	1056	0.016	916
0.7	0.002	14682	0.003	7277	0.005	4809	0.006	3575	0.008	2835	0.009	2341	0.011	1989	0.012	1724
0.8	0.001	34961	0.002	17321	0.003	11441	0.004	8501	0.005	6737	0.006	5561	0.007	4721	0.008	4091
0.9	0.001	147544	0.001	73064	0.002	48237	0.002	35824	0.003	28376	0.003	23410	0.004	19864	0.004	17204
1.1	0.001	162910	0.001	80590	0.002	53150	0.002	39430	0.003	31198	0.003	25710	0.004	21790	0.004	18850
1.2	0.001	42645	0.002	21085	0.003	13898	0.004	10305	0.005	8149	0.006	6711	0.007	5685	0.008	4915
1.3	0.002	19804	0.003	9786	0.005	6447	0.006	4777	0.008	3776	0.009	3108	0.011	2631	0.012	2273
1.4	0.002	11618	0.004	5738	0.006	3778	0.008	2798	0.010	2210	0.012	1818	0.014	1538	0.016	1328
1.5	0.003	7741	0.005	3821	0.008	2514	0.010	1861	0.013	1469	0.015	1207	0.018	1021	0.020	881
1.6	0.003	5587	0.006	2756	0.009	1813	0.012	1341	0.015	1058	0.018	869	0.021	734	0.024	633
1.7	0.004	4261	0.007	2101	0.011	1381	0.014	1021	0.018	805	0.021	661	0.025	558	0.029	481
1.8	0.004	3381	0.008	1666	0.012	1094	0.016	808	0.020	637	0.024	522	0.029	441	0.033	380
1.9	0.005	2765	0.009	1362	0.014	894	0.018	660	0.023	520	0.028	426	0.032	359	0.037	309
2	0.005	2316	0.010	1140	0.015	748	0.020	552	0.026	434	0.031	356	0.036	300	0.041	258
2.5	0.008	1197	0.015	587	0.023	384	0.031	282	0.038	221	0.046	181	0.054	152	0.061	130
3	0.010	767	0.020	375	0.031	244	0.041	179	0.051	140	0.061	114	0.071	95	0.082	81
3.5	0.013	551	0.026	268	0.038	174	0.051	127	0.064	99	0.077	80	0.089	67	0.102	57
4	0.015	423	0.031	206	0.046	133	0.061	97	0.077	75	0.092	61	0.107	50	0.122	42
4.5	0.018	341	0.036	165	0.054	106	0.071	77	0.089	60	0.107	48	0.125	39	0.143	33

TABLE 4: ADEQUATE SAMPLE SIZE FOR FIXED VALUES OF α=0.05, 1-β=0.80, AND VARYING P_2 & SE (PRECISION=Z SCORE X SE)

	P_2 - INCIDENCE RATE AMONG UNEXPOSED GROUP																	
	0.09		0.1		0.11		0.12		0.13		0.14		0.15		0.18		0.2	
RR	SE	N	SE	N	SE	N	SE	N	SE	N	SE	N	SE	N	SE	N	SE	N
0.1	0.041	111	0.046	99	0.051	90	0.055	82	0.060	75	0.064	69	0.069	64	0.083	52	0.092	46
0.2	0.037	153	0.041	137	0.045	124	0.049	112	0.053	103	0.057	95	0.061	88	0.073	72	0.082	63
0.3	0.032	216	0.036	193	0.039	174	0.043	159	0.046	145	0.050	134	0.054	124	0.064	101	0.071	89
0.4	0.028	316	0.031	282	0.034	255	0.037	232	0.040	212	0.043	195	0.046	181	0.055	147	0.061	130
0.5	0.023	486	0.026	434	0.028	391	0.031	356	0.033	325	0.036	300	0.038	277	0.046	225	0.051	199
0.6	0.018	807	0.020	720	0.022	649	0.024	589	0.027	539	0.029	496	0.031	459	0.037	372	0.041	328
0.7	0.014	1518	0.015	1354	0.017	1219	0.018	1107	0.020	1012	0.021	931	0.023	860	0.028	696	0.031	613
0.8	0.009	3601	0.010	3209	0.011	2889	0.012	2621	0.013	2395	0.014	2201	0.015	2033	0.018	1641	0.020	1445
0.9	0.005	15135	0.005	13480	0.006	12126	0.006	10997	0.007	10042	0.007	9224	0.008	8514	0.009	6859	0.010	6032
1.1	0.005	16563	0.005	14734	0.006	13237	0.006	11990	0.007	10935	0.007	10030	0.008	9246	0.009	7417	0.010	6502
1.2	0.009	4316	0.010	3837	0.011	3445	0.012	3118	0.013	2841	0.014	2605	0.015	2399	0.018	1920	0.020	1681
1.3	0.014	1995	0.015	1772	0.017	1590	0.018	1438	0.020	1310	0.021	1200	0.023	1104	0.028	882	0.031	770
1.4	0.018	1164	0.020	1034	0.022	927	0.024	838	0.027	762	0.029	698	0.031	642	0.037	511	0.041	446
1.5	0.023	772	0.026	685	0.028	614	0.031	554	0.033	504	0.036	461	0.038	423	0.046	336	0.051	293
1.6	0.028	554	0.031	491	0.034	440	0.037	397	0.040	361	0.043	330	0.046	303	0.055	240	0.061	208
1.7	0.032	421	0.036	373	0.039	333	0.043	301	0.046	273	0.050	249	0.054	229	0.064	180	0.071	156
1.8	0.037	332	0.041	294	0.045	263	0.049	237	0.053	215	0.057	196	0.061	179	0.073	141	0.082	122
1.9	0.041	270	0.046	239	0.051	213	0.055	192	0.060	174	0.064	159	0.069	145	0.083	114	0.092	98
2	0.046	225	0.051	199	0.056	177	0.061	160	0.066	144	0.071	132	0.077	120	0.092	94	0.102	81
2.5	0.069	113	0.077	99	0.084	88	0.092	79	0.099	71	0.107	65	0.115	59	0.138	45	0.153	38
3	0.092	70	0.102	62	0.112	54	0.122	48	0.133	43	0.143	39	0.153	35	0.184	27	0.204	22
3.5	0.115	49	0.128	43	0.140	37	0.153	33	0.166	30	0.179	26	0.191	24	0.230	17	0.255	14
4	0.138	36	0.153	31	0.168	28	0.184	24	0.199	21	0.214	19	0.230	17	0.276	12	0.306	10
4.5	0.161	28	0.179	24	0.196	21	0.214	18	0.232	16	0.250	14	0.268	13	0.321	9	0.357	7