Statistica Applicata Vol. 12, n. 1, 2000

THE IMPACT OF DIFFERENT EFFECT SIZE DEFINITIONS ON SAMPLE SIZE FOR COMPARING TWO PROPORTIONS

Giuseppe Reina*, Bruno M. Cesana and Ettore Marubini***

* *Istituto di Statistica Medica e Biometria, Università di Milano*
** *Ospedale Maggiore di Milano, IRCCS Laboratorio Epidemiologico*

Abstract

The paper focuses on the fact that the well–known definition of Effect Size can be considered the starting point to derive the most common approximate formulae for sample sizes required for the comparison of two proportions both in non inferiority and in equivalence clinical trials. It is shown how the comparison among different mathematical expressions may be performed by examining the properties of the specific Effect Size definition adopted, and that the latter is the pivot of the expression under consideration. As a consequence, it is possible to achieve a formal justification of the differences found in sample sizes either among the approximate formulae or between them and the referent one calculated for the Fisher's exact test. The Effect Size definition through the arcsin metameter results to be the most satisfactory since it allows to obtain sample sizes very similar to those required by the unconditional test and, with an appropriate correction, to the sample sizes calculated for the conditional Fisher's exact test.

1. INTRODUCTION

Determining an adequate sample size for testing a hypothesis is a crucial step in planning biomedical researches. When the attention of the researcher is focused on the comparison of two population means of Gaussian variables, it is well known that the *Effect Size (ES)* is the core of the expressions suitable for sample size computation. This parameter is the postulated difference between the two population means standardised by dividing it by the common standard deviation of the measures in their respective population. Being a number, free of the original measurement unit, it enables the researcher to index what can be defined the degree of departure from the null hypothesis. In the case of the comparison of two proportions, the expression for sample size computation is usually based on approximating the distribution of the corresponding test statistic by the Gaussian distribution; however the definition of ES cannot be univocal as the standard deviation of the binomial distribution is a function of the mean.

In debating the clinical relevance of statistical decisions, Burnand et al. (1990) adopted the following ES:

$$f(p_1, p_0) = \frac{p_1 - p_0}{\sqrt{\bar{p}(1-\bar{p})}} \tag{1}$$

where: p_1 = true proportion of positive responses in the treated group (active)
p_0 = true proportion of positive responses in the control group ($p_1 \geq p_0$)
$\bar{p} = (p_1 + p_0)/2$

In discussing the problem of calculating the number of observations required by some common nonparametric tests, Noether (1987) gives an expression (see Appendix 1) from which the following ES can be attained:

$$h(p_1, p_0) = \frac{p_1 - p_0}{\sqrt{p_0(1-p_0)}} \tag{2}$$

Finally, following Cohen (1969), if the arcsin metameter is adopted, ES becomes:

$$g(p_1, p_0) = 2\left(\arcsin\sqrt{p_1} - \arcsin\sqrt{p_0}\right) \tag{3}$$

This note aims at studying the relationships between the previously defined ES and at making explicit the effect of different definitions of ES on sample size calculation for the comparison of two proportions both in non inferiority and in equivalence clinical studies.

2. METHODOLOGICAL CONSIDERATIONS

Figure 1 shows the behaviour of (1) (continuous line), (2) (dashed line) and (3) (dotted line) as functions of p_1 for three fixed values of p_0 = 0.1, 0.5, 0.9. Symmetrically, Figure 2 gives the behaviour of (1), (2) and (3) as functions of p_0 for p_1= *0.1, 0.5, 0.9*.

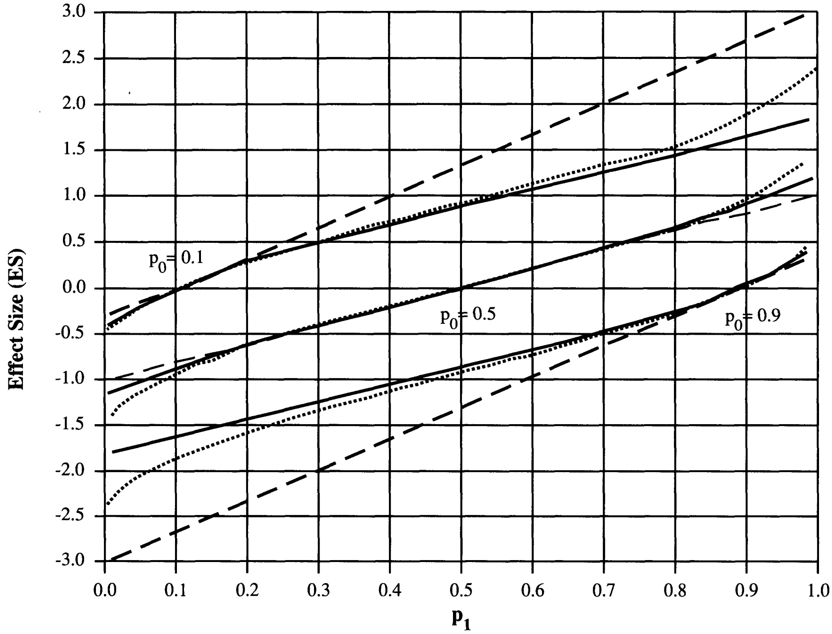

Fig. 1: Patterns of different ES definitions: $f(p_1,p_0)$ (continuous line), $g(p_1,p_0)$ (dottet line) and $h(p_1,p_0)$ (dashed line) as functions of p_1 for p_0= 0.1, 0.5, 0.9.

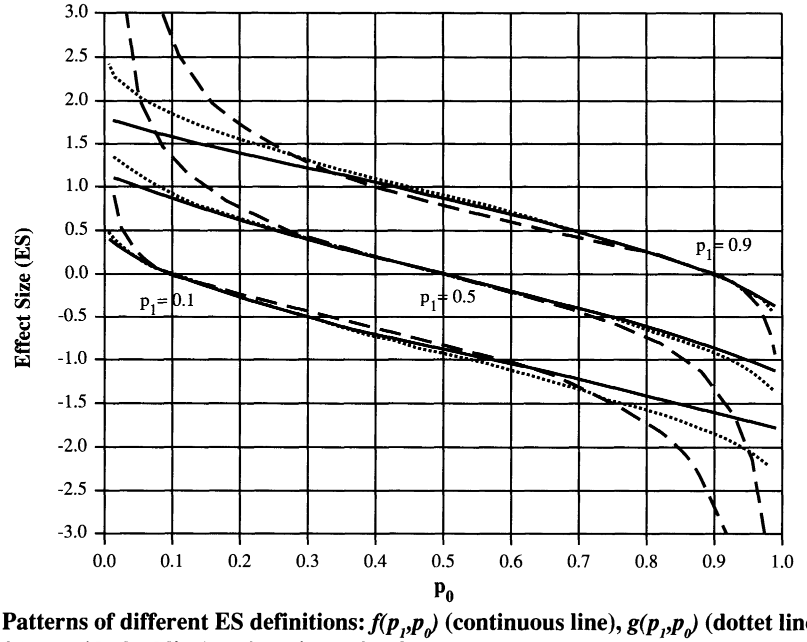

Fig. 2: Patterns of different ES definitions: $f(p_1,p_0)$ (continuous line), $g(p_1,p_0)$ (dottet line) and $h(p_1,p_0)$ (dashed line) as functions of p_0 for p_1= 0.1, 0.5, 0.9.

At a glance it appears that $f(.)$ and $g(.)$ tend to overlap as $\delta = p_1 - p_0$ tends to zero, i.e. to the value of δ under the null hypothesis. This suggests studying the behaviour of the two functions in the neighbourhood of $\delta = 0$ by means of series expansion.

In Appendix 2 it is shown that $h(.)$ is the common series expansion of both $f(.)$ and $g(.)$.

3. NON INFERIORITY STUDIES

In order to investigate the role of the previous definitions of ES on sample size computed by assuming that the distribution of the test statistic T is asymptotically Gaussian, the already mentioned Noether's formulation appears to be particularly convenient. As shown in Appendix 1, the core of this formulation is the noncentrality factor, $Q(T)$, for the test statistic T.

With reference to $f(p_1, p_0)$, it turns out:

$$Q(T) = \left[\frac{p_1 - p_0}{\sqrt{\bar{p}(1-\bar{p})}} \sqrt{\frac{n}{2}} \right]^2$$

and

$$n(f) = \left\{ \frac{z_\alpha \sqrt{2\bar{p}(1-\bar{p})} + z_\beta \sqrt{p_1(1-p_1) + p_0(1-p_0)}}{(p_1 - p_0)} \right\}^2 \tag{4}$$

where z_α and z_β are the two standardised Gaussian deviates corresponding to the Type I error probability, α, and to the Type II error probability, β, respectively. In this note sample size computation for a one-tailed test was considered; for a two-tailed test $z_{\alpha/2}$ must replace z_α. In both cases n is the sample size for each of the two groups to be compared.

This equation corresponds to (3.14) of Fleiss (1973) and to (3.2) of Machin and Campbell (1987).

With reference to $h(p_1, p_0)$, it turns out:

$$Q(T) = \left[\frac{p_1 - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{\frac{n}{2}} \right]^2$$

and

$$n(h)=\left\{\frac{z_{\alpha}\sqrt{2p_0(1-p_0)}+z_{\beta}\sqrt{p_1(1-p_1)+p_0(1-p_0)}}{(p_1-p_0)}\right\}^2. \tag{5}$$

Finally, with reference to $g(p_1, p_0)$, it turns out:

$$Q(T)=\left[2\left(\arcsin\sqrt{p_1}-\arcsin\sqrt{p_0}\right)\sqrt{\frac{n}{2}}\right]^2$$

and

$$n(g)=\frac{(z_{\alpha}+z_{\beta})^2}{2\left(\arcsin\sqrt{p_1}-\arcsin\sqrt{p_0}\right)^2}. \tag{6}$$

This equation was utilised by many Authors to calculate sample size tables and graphs: among the others, Cochran and Cox (1957), Mace (1974), Cohen (1969), Gail and Gart (1973) for $n > 30$, and Feigl (1978).

4. EQUIVALENCE STUDIES

So far, we have considered studies suitable for testing treatment difference in which the standard formulation typically involves a null hypothesis of no difference between the two groups. On the contrary the null hypothesis that the two treatment groups do not differ by more than a pre-specified biologically or clinically immaterial amount η, is formulated in equivalence studies. Let now p_1 and p_0 indicate the true response proportions to the new and to the standard treatment respectively.

If one defines:

$$\begin{cases} H_0 : p_1 - p_0 \geq \eta \\ H_a : p_1 - p_0 < \eta \end{cases} \tag{7}$$

rejecting the null hypothesis implies equivalence of the two treatments (Blackwelder, 1982). The choice of α and β values must be carefully addressed remembering that in these studies the definition of Type I and Type II errors are an exact switch of those applied to non inferiority testing. The statistic is now:

$$T=\hat{p}_1-\hat{p}_0-\eta$$

where $\hat{p}_1$ and $\hat{p}_0$ are the estimates of p_1 and p_0 respectively. The variance of this statistic cannot be written straightforwardly as a function of p_1 and p_0 owing to the null hypothesis constraint (7). However, as shown by Farrington and Manning (1990), an estimate can be obtained by means of the maximum likelihood estimates $\tilde{p}_1$ and $\tilde{p}_0$ calculated by solving a cubic equation in $\tilde{p}_1$ as $\tilde{p}_0 = \tilde{p}_1 - \eta$.

Accordingly, $Q(T)$ becomes:

$$Q(T) = \left[\frac{p_1 - p_0 - \eta}{\sqrt{\tilde{p}_1(1-\tilde{p}_1) + \tilde{p}_0(1-\tilde{p}_0)}} \sqrt{n} \right]^2 \tag{8}$$

and

$$n(fm) = \left\{ \frac{z_\alpha \sqrt{\tilde{p}_1(1-\tilde{p}_1) + \tilde{p}_0(1-\tilde{p}_0)} + z_\beta \sqrt{p_1(1-p_1) + p_0(1-p_0)}}{p_1 - p_0 - \eta} \right\}^2 . \tag{9}$$

By replacing in (8) $\tilde{p}_1$ and $\tilde{p}_0$ with p_1 and p_0 respectively, one obtains:

$$Q(T) = \left[\frac{p_1 - p_0 - \eta}{\sqrt{p_1(1-p_1) + p_0(1-p_0)}} \sqrt{n} \right]^2 \tag{10}$$

and:

$$n(ms) = \left[p_1(1-p_1) + p_0(1-p_0) \right] \left[\frac{z_\alpha + z_\beta}{p_1 - p_0 - \eta} \right]^2 \tag{11}$$

which coincides with the equation (6) of Makuch and Simon (1978) and the one given by Blackwelder (1982, Table I). It is easy to see that for $\eta = 0$, expression (10) suggests a fourth definition of ES; the corresponding n is equal to the approximation mentioned by Machin and Campbell (1987) in their equation (3.3).

5. COMPARISON OF RESULTS OBTAINED WITH DIFFERENT APPROACHES

5.1 NON INFERIORITY STUDIES

In biomedicine, one of the most widely used tests to assess the significance of differences between two proportions is the Fisher's (1935) *conditional* exact test for a 2x2 contingency table together with the Mantel-Haenszel (1959) extension to combine several 2x2 tables. Therefore, Table 1 reports sample sizes for different values of $\delta = p_1 - p_0$ with $\alpha = 0.05$ (one tailed test) and $\beta = 0.2$, obtained by using expressions: (4) $n(f)$, (5) $n(h)$ and (6) $n(g)$ together with the exact sample sizes $n(e)$ (see Haseman, 1978) for the Fisher's *conditional* test. It appears that in the explored range of δ, sample sizes $n(f)$ and $n(g)$ are similar, if not equal, whereas sample sizes $n(h)$ differ substantially from the other two. In comparison to $n(f)$ and $n(g)$, $n(h)$ sample sizes are undersized for $p_0 < 0.5$ with a difference progressively decreasing as p_0 approaches 0.5. On the contrary, for $p_0 > 0.5$, $n(h)$ oversize $n(f)$ and $n(g)$ with a difference progressively increasing as p_0 approaches 1.0. This reflects the behaviour of $h(p_1, p_0)$, in relation to that of $f(p_1, p_0)$ and of $g(p_1, p_0)$, shown by Figures 1 and 2. However it is worth noticing that both $n(f)$ and $n(g)$ underestimate the exact $n(e)$ values whereas $n(h)$, for p_0 greater than some 0.70, overestimate $n(e)$.

Tab. 1: Sample size for demonstrating the difference of two proportions in non inferiority studies ($\alpha = 0.05$, $\beta = 0.2$, one tailed test).

n(h) from (5); ***n(f)*** from (4); ***n(g)*** from (6); ***n(e)*** = Fisher's exact test.

p_0	p_1	$n(h)$	$n(f)$	$n(g)$	$n(e)$
.05	.10	269	343	334	372
	.15	74	111	105	126
	.20	36	60	55	67
	.25	21	39	35	45
	.30	14	28	25	34
	.35	10	21	19	25
	.40	8	17	15	20
	.45	6	14	12	17
.10	.15	476	541	536	575
	.20	126	157	154	173
	.25	58	79	76	89
	.30	34	49	47	56
	.35	22	34	32	39
	.40	16	25	24	30

%

continued

p_0	p_1	n(h)	n(f)	n(g)	n(e)
	.45	12	20	19	24
	.50	9	16	15	19
	.55	7	13	12	16
	.60	6	11	10	13
.15	.20	658	714	711	750
	.25	170	197	196	215
	.30	78	95	94	106
	.35	45	57	56	65
	.40	29	39	38	46
	.45	21	28	28	34
	.50	15	22	21	26
	.55	12	17	17	22
	.60	9	14	13	17
	.65	8	11	11	15
.20	.25	815	862	861	901
	.30	209	231	230	249
	.35	95	109	108	121
	.40	54	64	64	73
	.45	35	43	42	49
	.50	24	31	30	36
	.55	18	23	23	27
	.60	14	18	18	23
	.65	11	14	14	17
	.70	9	12	12	15
.25	.30	947	986	985	1025
	.35	241	259	259	277
	.40	108	120	120	132
	.45	62	70	69	78
	.50	40	46	46	54
	.55	28	32	32	37
	.60	20	24	24	30
	.65	16	19	19	23
	.70	12	15	15	18
	.75	10	12	12	15
.30	.35	1054	1085	1084	1123
	.40	266	281	281	302
	.45	119	128	128	142
	.50	67	74	74	84
	.55	43	48	48	55
	.60	30	33	33	41
	.65	22	25	25	31
	.70	17	19	19	23
	.75	13	15	15	18

%

continued

p_0	p_1	n(h)	n(f)	n(g)	n(e)
.35	.40	1136	1159	1159	1196
	.45	286	296	296	318
	.50	128	134	134	143
	.55	72	76	76	85
	.60	46	49	49	56
	.65	32	34	34	41
	.70	23	25	25	31
	.75	18	19	19	23
.40	.45	1194	1208	1208	1245
	.50	299	306	306	321
	.55	133	137	137	144
	.60	75	77	77	85
	.65	48	49	49	56
	.70	33	33	33	41
	.75	24	24	24	30
	.80	18	18	18	23
.50	.55	1235	1233	1233	1270
	.60	308	306	306	321
	.65	136	134	134	143
	.70	76	74	74	84
	.75	48	46	46	54
	.80	33	31	30	36
	.85	24	22	21	26
.65	.70	1111	1085	1084	1123
	.75	273	259	259	277
	.80	119	109	108	121
	.85	65	57	56	65
.70	.75	1020	9862	9852	1025
	.80	250	231	230	249
	.85	108	95	94	106
	.90	59	49	47	56
.75	.80	905	862	861	901
	.85	219	197	196	215
	.90	94	79	76	89
	.95	51	39	35	45
.80	.85	764	714	711	750
	.90	183	157	154	173
	.95	77	60	55	67

5.2 EQUIVALENCE STUDIES

Table 2 reports sample sizes computed from (9) *n(fm)* and from (11) *n(ms)* for $\alpha = 0.05$ (one-tailed test) and three values of $\beta = 0.20, 0.10, 0.05$. It appears that, in the explored range of p_1, p_0 and η, the use of p_1 and p_0 as large sample approximations to $\tilde{p}_1$ and $\tilde{p}_0$ respectively, enables calculating sample sizes in good accord with those obtained after estimating $\tilde{p}_1$ and $\tilde{p}_0$ under the constraint $\tilde{p}_1 - \tilde{p}_0 = \eta$.

Tab. 2: Sample size for demonstrating the difference of two proportions in equivalence studies (α = 0.05, β = 0.05, 0.10, 0.20, one tailed test)

(η = maximum acceptable difference for equivalence, *n(fm)* = Farrington and Manning (1990), *n(ms)* = Makuch and Simon (1978))

			β = 0.20		β = 0.10		β = 0.05	
p_0	p_1	η	n(fm)	n(ms)	n(fm)	n(ms)	n(fm)	n(ms)
0.10	0.10	0.05	457	446	630	617	795	780
0.15	0.15	0.05	636	631	880	874	1111	1104
	0.10	0.10	551	538	761	746	959	942
	0.15	0.10	163	158	224	219	283	276
0.20	0.20	0.05	794	792	1099	1097	1389	1386
	0.15	0.10	716	711	991	985	1252	1245
	0.20	0.10	200	198	277	275	349	347
	0.15	0.15	182	178	251	247	316	312
	0.20	0.15	90	88	124	122	157	154
0.30	0.30	0.05	1038	1039	1438	1439	1817	1819
	0.25	0.10	981	984	1359	1362	1718	1721
	0.30	0.10	259	260	359	360	454	455
	0.25	0.15	245	246	339	341	429	431
	0.30	0.15	115	116	159	160	201	203
	0.25	0.20	108	110	150	152	190	192
	0.30	0.20	64	65	89	90	113	114
0.40	0.40	0.05	1185	1188	1642	1645	2076	2078
	0.35	0.10	1151	1157	1595	1602	2017	2024
	0.40	0.10	295	297	409	412	517	520
	0.35	0.15	286	290	396	401	502	506
	0.40	0.15	130	132	181	183	229	231
	0.35	0.20	126	129	175	178	221	225
	0.40	0.20	72	75	101	103	128	130

%

continued

p_0	p_1	η	$\beta = 0.20$ n(fm)	$\beta = 0.20$ n(ms)	$\beta = 0.10$ n(fm)	$\beta = 0.10$ n(ms)	$\beta = 0.05$ n(fm)	$\beta = 0.05$ n(ms)
0.50	0.50	0.05	1234	1237	1710	1713	2162	2165
	0.45	0.10	1224	1231	1697	1705	2146	2154
	0.50	0.10	307	310	426	429	538	542
	0.45	0.15	304	308	421	427	533	539
	0.50	0.15	135	138	188	191	238	241
	0.45	0.20	133	137	185	190	235	240
	0.50	0.20	75	78	105	108	133	136
0.60	0.60	0.05	1185	1188	1642	1645	2076	2078
	0.55	0.10	1200	1206	1663	1670	2103	2111
	0.60	0.10	295	297	409	412	517	520
	0.55	0.15	298	302	413	418	523	528
	0.60	0.15	130	132	181	183	229	231
	0.55	0.20	131	134	182	186	230	235
	0.60	0.20	72	75	101	103	128	130
0.70	0.70	0.05	1038	1039	1438	1439	1817	1819
	0.65	0.10	1078	1082	1494	1499	1889	1894
	0.70	0.10	259	260	359	360	454	455
	0.65	0.15	268	271	372	375	470	474
	0.70	0.15	115	116	159	160	201	203
	0.65	0.20	118	121	164	167	208	211
	0.70	0.20	64	65	89	90	113	114
0.80	0.80	0.05	794	792	1099	1097	1389	1386
	0.75	0.10	860	860	1191	1191	1505	1505
	0.80	0.10	200	198	277	275	349	347
	0.75	0.15	216	215	299	298	377	377
	0.80	0.15	90	88	124	122	157	154
	0.75	0.20	96	96	133	133	168	168
	0.80	0.20	51	50	70	69	89	87
0.90	0.90	0.05	457	446	630	617	795	780
	0.85	0.10	551	538	761	746	959	942
	0.90	0.10	121	112	166	155	208	195
	0.85	0.15	143	135	197	187	247	236
	0.90	0.15	57	50	78	69	97	87
	0.85	0.20	66	60	91	83	113	105
	0.90	0.20	34	28	46	39	57	49

6. CONCLUDING REMARKS

As regards non inferiority studies, the ES definition given by (2) should be avoided; in fact, depending on p_0 and δ, it leads to underestimate or overestimate the sample size obtained by the exact *conditional* approach.

Otherwise, the sample size underestimation given by ES as defined by (1) and (3), can be satisfactorily adjusted as shown by Casagrande Pike and Smith (1978) and Walter (1979), respectively; in addition, Gordon and Watson (1994) reported that this latter is very appropriate even for testing hypotheses regarding low probability levels.

Furthermore, it is worth noticing that the ES as defined by (3) leads to sample sizes which agree quite well with those given by Suissa and Shuster (1985) for the exact *unconditional* approach.

In conclusion, the ES definition in terms of the arcsin metameter appears to be the most satisfactory as it can be easily used for proper sample size computation both for the *conditional* and the *unconditional* test.

As regards equivalence studies, it appears that the ES definition of Makuch and Simon (1978) can be satisfactorily used in practice.

7. APPENDIX 1

This appendix aims at aiding the reader to obtain expressions (4), (5), (6), (9) and (11) by starting from the general Noether's (1987) formulation.

Assume that the distribution of a statistic T is asymptotically Gaussian with mean $\mu(T)$ and standard error $\sigma(T)$; in particular assume that, under the null hypothesis, the two parameters are $\mu_0(T)$ and $\sigma_0(T)$. Noether (1987) defines:

- the ratio of the two standard errors under the alternative and the null hypothesis: $\phi = \sigma(T) / \sigma_0(T)$
- the noncentrality factor for the statistic T:

$$Q(T) = \left[\frac{\mu(T) - \mu_0(T)}{\sigma_0(T)} \right]^2$$

Approximate sample sizes are obtained from the expression: $Q(T) = \left(z_\alpha + \phi z_\beta\right)^2$.

Assume the equality of sample sizes for the two groups to be compared: $n_1 = n_0 = n$.

With reference to $ES = f(p_1, p_0)$:

$\mu_0(T) = 0 \qquad \mu(T) = p_1 - p_0$

$$Q(T) = \left[f(p_1, p_0)\sqrt{\frac{n}{2}} \right]^2 = \left[\frac{p_1 - p_0}{\sqrt{\bar{p}(1-\bar{p})}} \sqrt{\frac{n}{2}} \right]^2.$$

Therefore $\sigma_0^2(T) = \dfrac{2\bar{p}(1-\bar{p})}{n}$, consequently $\phi = \dfrac{\sqrt{p_1(1-p_1) + p_0(1-p_0)}}{\sqrt{2\bar{p}(1-\bar{p})}}$

and (4) follows after some algebra.

With reference to $ES = h\,(p_1, p_0)$:

$\mu_0(T) = 0 \qquad \mu(T) = p_1 - p_0$

$$Q(T) = \left[h(p_1, p_0)\sqrt{\frac{n}{2}} \right]^2 = \left[\frac{p_1 - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{\frac{n}{2}} \right]^2.$$

Therefore $\sigma_0^2(T) = \dfrac{2p_0(1-p_0)}{n}$, consequently $\phi = \dfrac{\sqrt{p_1(1-p_1) + p_0(1-p_0)}}{\sqrt{2p_0(1-p_0)}}$

and (5) follows after some algebra.

With reference to $ES = g\,(p_1, p_0)$:

$\mu_0(T) = 0 \qquad \mu(T) \cong 2\left(\arcsin\sqrt{p_1} - \arcsin\sqrt{p_0}\right)$

$$Q(T) = \left[g(p_1, p_0)\sqrt{\frac{n}{2}} \right]^2 = \left[2\left(\arcsin\sqrt{p_1} - \arcsin\sqrt{p_0}\right)\sqrt{\frac{n}{2}} \right]^2.$$

Therefore, as $Var\left[2\left(\arcsin\sqrt{\hat{p}_1} - \arcsin\sqrt{\hat{p}_0}\right)\right] \simeq \dfrac{2}{n}$ (see Cox, 1970) both under the null and the alternative hypothesis, it results $\phi = 1$ and (6) follows after some algebra.

8. APPENDIX 2

After writing $p_1 = p_0 + \delta$ and $\bar{p} = p_0 + \frac{\delta}{2}$, $f(.)$ can be rewritten:

$$f(p_1, p_0) = \frac{\delta}{\sqrt{\left(p_0 + \frac{\delta}{2}\right)\left(1 - p_0 - \frac{\delta}{2}\right)}}$$

which in the neighbourhood of $\delta = 0$ becomes:

$$f(p_1, p_0) \cong \frac{p_1 - p_0}{\sqrt{p_0(1 - p_0)}} = h(p_1, p_0).$$

With regard to $g(.)$, by expanding $arcsin\sqrt{p_0 + \delta}$ in the neighbourhood of $\delta = 0$ one obtains:

$$arcsin\sqrt{p_0 + \delta} \cong arcsin\sqrt{p_0} + \frac{\delta}{2\sqrt{p_0(1 - p_0)}}.$$

Hence, in the neighbourhood of $\delta = 0$, $g(.)$ becomes:

$$g(p_1, p_0) = 2\left\{\arcsin\sqrt{p_0} + \frac{\delta}{2\sqrt{p_0(1 - p_0)}} - \arcsin\sqrt{p_0}\right\} \cong \frac{p_1 - p_0}{\sqrt{p_0(1 - p_0)}} = h(p_1, p_0).$$

Therefore $f(.)$ and $g(.)$ have the same series expansion, $h(p_1, p_0)$, in the neighbourhood of $\delta = 0$.

ACKNOWLEDGEMENTS

We thank the anonymous referees for the useful suggestions to improve this paper.

REFERENCES

BLACKWELDER W.C., 1982, "Proving the null hypothesis" in clinical trials, *Controlled Clinical Trials,* 3, 345–353.

BURNAND B., KERNAN W.N., FEINSTEIN A.R., 1990, Indexes and boundaries for "quantitative significance" in statistical decisions, *Journal of Clinical Epidemiology, 43, 1273–1284.*

CASAGRANDE J.T., PIKE M.C. and SMITH P.G., 1978, An improved approximate formula for calculating sample sizes for comparing two binomial distributions, *Biometrics*, 34, 483–486.

COCHRAN W.G. and COX G., 1957, *Experimental designs*, John Wiley and Sons, New York.

COHEN J., 1969, *Statistical power analysis for the behavioral sciences*, Academic Press, New York.

COX D.R., 1970, *The analysis of binary data*, Methuen and Co. Ltd, London.

FARRINGTON C.P. and MANNING G., 1990, Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine*, 9, 1447-1454.

FEIGL P., 1978, A graphical aid for determining sample size when comparing two independent proportions, *Biometrics*, 34, 111–122.

FISHER R.A., 1935, The logic of inductive inference, *Journal of the Royal Statistical Society, Series A*, 98, 39–54.

FLEISS J.L., 1973, *Statistical methods for rates and proportions*, John Wiley and Sons, New York.

GAIL M. and GART J.J., 1973, The determination of sample sizes for use with the exact conditional test in 2x2 comparative trials, *Biometrics*, 29, 441–448.

GORDON I. and WATSON R., 1994, A Note on Sample Size Determination for Comparison of Small Probabilities, *Controlled Clinical Trials*, 15, 77–79.

HASEMAN J.K., 1978, Exact sample sizes for use with the Fisher-Irwin test for 2x2 tables, *Biometrics*, 34, 106–109.

MACE A.E., 1974, *Sample-size determination*, Reinhold Publishing Corporation, Huntington, New York.

MACHIN D. and CAMPBELL M.J., 1987, *Statistical tables for the design of clinical trials*, Blackwell Scientific Publications, Oxford.

MAKUCH R. and SIMON R., 1978, Sample size requirements for evaluating a conservative therapy, *Cancer Treatment Reports*, 62, 1037–1040.

MANTEL N. and HAENSZEL W., 1959, Statistical aspects of the analysis of data from retrospective studies of disease, *Journal of fhe National Cancer Institute*, 22, 719–748.

NOETHER G.E., 1987, Sample size determination for some common nonparametric tests, *Journal of the American Statistical Association*, 82, 645–647.

SUISSA S. and SHUSTER J. J., 1985, Exact unconditional sample sizes for the 2 x 2 binomial trial, *Journal of the Royal Statistical Society A*, 148, 317–327.

WALTERS D.E., 1979, In Defence of the Arc Sine Approximation, *The Statistician,* 28, 219–222.

L'INFLUENZA DI DIFFERENTI DEFINIZIONI DI "DIMENSIONE DELL'EFFETTO" SUL CALCOLO DELLA DIMENSIONE CAMPIONARIA PER IL CONFRONTO DI DUE PROPORZIONI

Riassunto

In questo lavoro si evidenzia come la ben nota definizione di "Dimensione dell'Effetto" (Effect Size) possa essere considerata il punto di partenza per ricavare le formule approssimate di uso più comune per la determinazione del numero di osservazioni richieste per il confronto di due proporzioni sia in studi clinici di differenza che di equivalenza in efficacia. Inoltre, è messo in evidenza come il confronto fra differenti espressioni di calcolo della dimensione campionaria possa essere eseguito tramite lo studio della specifica definizione di "Dimensione dell'Effetto" utilizzata e che sta alla base dell'espressione considerata. Ne consegue la formale giustificazione delle differenze sia tra le dimensioni campionarie ottenute dalle formule approssimate che tra quelle riscontrate rispetto alla dimensione campionaria di riferimento del test esatto di Fisher. La definizione di "Dimensione dell'Effetto" ottenuta mediante il metametro arcoseno è particolarmente soddisfacente per il fatto che essa permette di ottenere numerosità campionarie sovrapponibili sia a quelle calcolate per il confronto tra due proporzioni mediante l'approccio non condizionato che, mediante un'appropriata correzione, a quelle richieste per il test esatto condizionato di Fisher.