QÜESTIIÓ, vol. 26, 1-2, p. 259-271, 2002

INVERSE SAMPLING AND TRIANGULAR SEQUENTIAL DESIGNS TO COMPARE A SMALL PROPORTION WITH A REFERENCE VALUE*

VÍCTOR MORENO[1,2]
ISAAC MARTÍN[3]
FERRAN TORRES[2]
MANUEL HORAS[2]
JOSÉ RIOS[2]
JUAN R. GONZÁLEZ[1]

Inverse sampling and formal sequential designs may prove useful in reducing the sample size in studies where a small population proportion p is compared with a hypothesized reference proportion p_0. These methods are applied to the design of a cytogenetic study about chromosomal abnormalities in men with a daughter affected by Turner's syndrome. First it is shown how the calculated sample size for a classical design depends on the parameterization used. Later this sample size is compared with the required sample size in an inverse sampling design and a triangular sequential design using four different parameterizations (absolute differences, log-odds ratio, angular transform and Sprott's transform). The expected savings in sample size, when the alternative hypothesis is true, are 20% of the fixed sample size for the inverse sampling design and 40% for the triangular sequential design.

Keywords: Sample size, inverse sampling, sequential methods, triangular sequential test

AMS Classification (MSC 2000): 62L05 Statistics, Sequential methods

* Please address all correspondence to: Victor Moreno. Servicio de Epidemiología y Registro del Cáncer. Instituto Catalán de Oncología. Gran Vía km 2.7.Hospitalet, 08907 Barcelona. Spain.
E-mail: V.Moreno@ico.scs.es. Tel: +34-93 260 78 12. Fax: +34-93 260 77 87.
[1] Instituto Catalán de Oncología. Gran Vía km 2.7, Hospitalet, 08907 Barcelona, Spain.
[2] Laboratorio de Bioestadística y Epidemiología. Facultad de Medicina, Universidad Autónoma de Barcelona. Barcelona, Spain.
[3] Departamento de Estadística y Econometría. Universidad Carlos III, Madrid, Spain.

– Received February 2000.
– Accepted June 2001.

1. INTRODUCTION

The sample size needed in many biological experiments is so large when the characteristic of interest is a rare event that it is appealing to explore different sampling schemes oriented to reduce the number of observations needed. In this paper it is shown how inverse sampling and formal sequential designs may prove useful in reducing the sample size in some specific situations.

The methods described were motivated by the design of a cytogenetic study where the aim was to compare the proportion of chromosomal abnormalities observed in an affected individual with a known reference value for healthy individuals. Similar situations occur in preclinical studies of toxicity or while monitoring rare adverse events in clinical trials, especially in phase IV studies.

The hypothesis of interest was that the proportion of spermatozoa carrying chromosomal abnormalities in men with a daughter affected by Turner's syndrome was higher than that observed in control men (without daughters affected by Turner's syndrome). Turner's syndrome appears when one sexual chromosome is lost and the karyotype results in 22 pairs of somatic chromosomes but only one sexual chromosome X. The affected individual is a female with some specific phenotypic characteristics. The chromosomal studies to determine abnormalities involve complex experiments in which hamster oocytes are fused with human spermatozoa.

The proportion of spermatozoa with the sexual chromosome missing in normal men has been accurately estimated in several studies to be around 0.003. In the present study, due to the technical difficulties in determining the abnormalities, it was thought sensible that the results from a sample of cases (men with a daughter affected by Turner's syndrome) could be compared to the already known proportion in control men.

The aim was to determine whether a population proportion p differs from a hypothesized reference proportion p_0. In our study, doubling the reference proportion was considered an increase interesting to detect. Thus, the interest was to reject the null hypothesis that $p = p_0 = 0.003$ when $p \geqslant 0.006$ with power 0.80. Only the one sided alternative that the proportion was higher in men with an affected daughter than controls was considered and a conservative significance level of 0.025 was adopted.

In the classical fixed sample size design, a sample size should be calculated to satisfy the power requirements. Data should be collected but not examined and a decision taken until the complete sample size had been achieved. In our situation, because the proportion of abnormal spermatozoa was very small, the sample size required was large and further, as will be shown later, the parameterization used to compare the proportions affected the required sample size in a significant amount.

Table 1. Formulas for Z and V and sample size according to the parameterization of θ, the difference between p and p_0.

θ	Z	V	n
Log-odds ratio $\log(p(1-p_0)/(p_0(1-p)))$	$r-np_0$	$np_0(1-p_0)$	$\dfrac{(z_\alpha+z_\beta)^2}{\log^2\left(\dfrac{p(1-p_0)}{p_0(1-p)}\right)p_0(1-p_0)}$
Probability difference $p-p_0$	$\dfrac{r-np_0}{p_0(1-p_0)}$	$\dfrac{n}{p_0(1-p_0)}$	$\dfrac{(z_\alpha+z_\beta)^2\,p_0(1-p_0)}{(p-p_0)^2}$
Angular transformation $\arcsin\sqrt{p}-\arcsin\sqrt{p_0}$	$2\dfrac{r-np_0}{\sqrt{p_0(1-p_0)}}$	$4n$	$\dfrac{(z_\alpha+z_\beta)^2}{4\left(\arcsin\sqrt{p}-\arcsin\sqrt{p_0}\right)^2}$

Alternatives to the fixed sample size design may prove interesting in reducing the sample size needed while preserving the statistical errors at the predefined levels. Two such alternatives will be revised in this paper. First, inverse sampling, that consists on sampling until the first r occurrences of an event are seen in a sample. Secondly, a formal sequential analysis based in successive examinations of accumulating data with a prespecified stopping rule.

2. SAMPLING PROCEDURES

2.1. Fixed sample size design

Several formulas may be used to calculate the sample size needed to compare a proportion with a reference value. The notation used by Whitehead (1992), that allows deriving a variety of these formulas from a unified theory, will be followed. This notation will also be used to design and analyze formal sequential tests.

A sequence $x_1, x_2, \ldots, x_n$ of independent binary observations with event probability p is observed and each one is coded 1 if the event appears or 0 otherwise. The null hypothesis to test is $H_0: \ p = p_0$. This is equivalent to testing whether a parameter $\theta = g(p, p_0)$ is equal to zero, where g is a function that parameterizes the difference between p and p_0 in an appropriate measurement scale.

As a Bernoulli experiment, the likelihood of p based on n observations is

$$L(p) = p^r (1-p)^{n-r}$$

where $r = x_1 + x_2 + \cdots + x_n$ is the sum of responses. The log-likelihood is

$$l(p) = r \log(p/(1-p)) + n \log(1-p).$$

Following Whitehead's notation (Whitehead, 1992), in this log-likelihood p can be substituted by θ from $g(p, p_0)$. The log-likelihood can be approximated by a Taylor's series expansion of second order and two statistics derived, namely Z and V.

$$l(\theta) = \text{const} + \theta Z - {}^1\!/_2 \theta^2 V + O(\theta^3)$$

From the series expansion we obtain that $Z = l_\theta(0)$ and $V = -l_{\theta\theta}(0)$, where $l_\theta(0)$ and $l_{\theta\theta}(0)$ denote respectively the first and second derivatives of $l(\theta)$ with respect to θ, evaluated at $\theta = 0$. Z is the efficient score for θ, a cumulative measure of the difference between p and p_0, and V is Fisher's information about θ contained in Z. The actual formula for Z and V are different depending on the choice of the function g. Table 1 shows three possible forms, the first is based on log-odds ratio scale, which measures relative differences, the second is based on absolute differences, and the third uses the angular transformation, which stabilizes the variance of the proportion. Note that Z, the cumulative difference, is always a function of $r - np_0$, that is, the observed minus expected number of responses. Also V, the information, is always proportional to the sample size, n.

For large sample sizes and small values of θ, the distribution of Z is approximately normal with mean θV and variance V. This normal approximation can be used to calculate the required sample size for a fixed sample design. Z, as an efficient score, may be used as the test statistic with working significance level α, and power $1 - \beta$. If Z is greater than some value $k = k(\alpha, \beta)$, then the null hypothesis is rejected at the level of significance α and it is concluded that the proportion in experimental group p is superior to the hypothesized p_0. The requirements for the one sided test are

$$\begin{aligned} P(Z \geqslant k/\theta = 0) &= \alpha \\ P(Z \geqslant k/\theta = \theta_R) &= 1 - \beta \end{aligned}$$

where θ_R is the difference which, if present, should be detected. A fixed sample study will satisfy these requirements if the information V and k are given by

$$\begin{aligned} V &= \left\{ \left(z_\alpha + z_\beta\right) / \theta_R \right\}^2 \\ k &= \left(z_\alpha + z_\beta\right) z_\alpha / \theta_R \end{aligned}$$

where z_γ denotes the upper $100(1-\gamma)$ percentage point of the normal distribution. Formulas for V are used to translate a requirement value for V into a required total sample

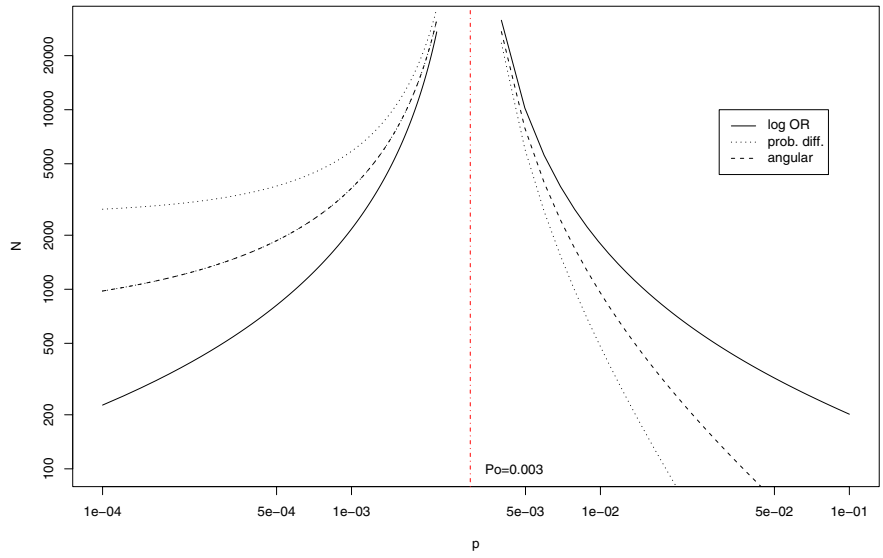

Figure 1. Sample size needed for various parameterizations.

size n. The formulas for Z, V and n are shown in table 1 for each parameterization studied.

The three parameterizations give different values for the sample size n, depending on the values of p and p_0. We can see in figure 1 that

$$\begin{aligned} &\text{If } p > p_0 \qquad n_{\log OR} > n_{\text{angular}} > n_{\text{prob. diff.}} \\ &\text{If } p < p_0 \qquad n_{\log OR} < n_{\text{angular}} < n_{\text{prob. diff.}} \end{aligned}$$

For the final statistical analysis of the difference between p and p_0 the chosen test also affects the results. An exact test will be used since the proportions compared are small, though exact tests are more conservative due to the discreteness of the response variable. An equivalent form to the one sided exact test, will be the calculation of the exact lower limit of the $100(1-2\alpha)\%$ confidence interval for the estimated proportion p assuming a binomial distribution. This limit can be calculated easily using the relation between the F and the binomial distributions (Jowett, 1963):

$$p_{\text{lower}}(r,t,\alpha) = r/\left\{r + (t+1)\, f_\alpha\left(2(t+1), 2r\right)\right\},$$

Table 2. Sample size and performance of fixed sample design with different parameterizations and tests for $\alpha = 0.025$ and power $= 0.80$.

	Sample	Exact test		χ^2test	
Method	Size	α	Power	α	Power
Log-odds ratio	5415	0.0159	0.8939	0.0280	0.9272
Probability difference	2608	0.0154	0.6010	0.0309	0.7003
Angular transformation	3795	0.0249	0.8157	0.0452	0.8704

where r is the number of cases with the characteristic, t is the number of cases without it and $f_\alpha(a,b)$ is the upper $100(\alpha)$ percentile of the F distribution with a and b degrees of freedom. If a bilateral test was used, the upper limit of the confidence interval could be calculated with the formula:

$$p_{\text{upper}}(r,t,\alpha) = r\, f_\alpha(2r,2t) / \{r\, f_\alpha(2r,2t) + t\}.$$

In our study, p_0 was 0.003 and the value of p we wished to detect was 0.006. We can see in table 2 the sample size calculated using the different definitions of θ with significance level $\alpha = 0.025$ and power 0.80. The parameterization of θ results in differences in the sample size needed, especially when the proportion is small. The statistical test used for the analysis is also important. In table 2 we can see the observed type I error rate and power after 50000 simulations for the fixed sample size design with a classical chi-square test without continuity correction, and for the exact test. As expected, for each parameterization of θ, the exact test gives more conservative results. The most accurate results, with respect to the predefined error rates, correspond to the angular transformation, which is in concordance with known results about comparison of two binomial proportions (Haseman, 1978).

2.2. Inverse sampling

The inverse sampling design is an old method (Haldane, 1945; Finney, 1949) to estimate a proportion p. The method consists on sampling until exactly r occurrences of an event appear in a study and counting the needed sample size n. In a classical fixed sample design sampling continues until the complete sample size n is attained and the number of occurrences r are counted. For both designs, the proportion p is estimated as r/n. The differences appear in the way the variance of this proportion is calculated. Methods to calculate the confidence interval for p when the inverse sampling design is used have been described by George and Elston (1993), for the special case when sampling continues until the first occurrence of an event of interest. They use the geometric

Table 3. Sample size needed with an inverse sampling design to detect $p = 0.006$ for given r and $\alpha = 0.025$.

r	$T_{\max}$	Mean n ($p = 0.006$)	Observed α	Expected power	Observed power
10	1591	1426	0.0259	0.4921	0.4967
11	1822	1612	0.0271	0.5402	0.5416
12	2058	1803	0.0256	0.5859	0.5814
13	2297	1985	0.0257	0.6281	0.6263
14	2540	2167	0.0266	0.6677	0.6673
15	2787	2355	0.0228	0.7045	0.7137
16	3036	2533	0.0258	0.7379	0.7384
17	3288	2721	0.0244	0.7685	0.7671
18	3542	2899	0.0264	0.7960	0.8005
19	3798	3084	0.0278	0.8208	0.8142
20	4056	3249	0.0283	0.8431	0.8499
21	4316	3432	0.0282	0.8630	0.8658
22	4578	3595	0.0292	0.8807	0.8858
23	4841	3780	0.0237	0.8963	0.8979
24	5106	3954	0.0253	0.9102	0.9122
25	5372	4112	0.0269	0.9223	0.9265

$T_{\max}$: number of cases without the event needed to observe before r events so that the lower 95% confidence interval around p doesn't include $p_0 = 0.003$.

distribution to calculate the confidence interval and demonstrate that the length is shorter than the one calculated by use of direct binomial sampling under certain situations. This is because of the fact that no occurrences in the first $t = n - 1$ trials is more informative than 1 occurrence in n trials. Nevertheless, the length of the confidence interval calculated on the basis of the first single case ($r = 1$) may be too wide for general utility. Lui (1995*a*; 1995*b*) describes the extension of this procedure to accommodate any finite number of cases ($r > 1$), and calculates the confidence interval using the exact method based on the relations between the negative binomial, the binomial and the F distributions as previously described.

For the comparison between p and p_0, an r large enough should be chosen so that, with probability $1 - \beta$, the lower bound of the lower $100(1 - 2\alpha)\%$ confidence limit around p will exceed the hypothesized proportion p_0. Classical inverse sampling should continue including subjects until r events appear. However, in the one sided design a

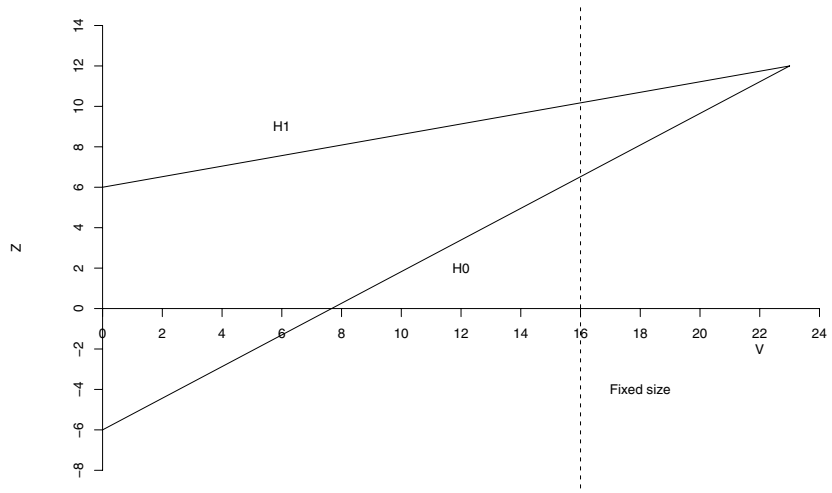

Figure 2. Continuation region for the triangular sequential test.

maximum value of $t\,(t_{\max})$ exists so that if the r events have not been observed before $t_{\max}$, the confidence interval will always include p_0. The design can be modified to stop sampling either when the r cases have been found or when $t_{\max}$ have been reached. We have checked, through simulations, that this truncation of the inverse sampling design did not alter the theoretic operating characteristic of test, and calculated the average sample size needed to finish the study, which was significantly reduced when the null hypothesis was not true.

Table 3 summarizes the results of simulations done to evaluate the power to detect a significant difference when $p = 0.006$ with the inverse sampling design using different values of r. $T_{\max}$, the maximum number of cases required for each value of r, can be calculated searching the quintile of the negative binomial distribution with p_0. The power to detect a given difference $p_R > p - p_0$ can be calculated from the negative binomial distribution function for those r and $t_{\max}$. In our case, with $r = 18$, the power to detect $p > p_0$ when $p = 0.006$ is 0.80 and the maximum number on cases without event needed to monitor, $t_{\max}$, is 3542. Note that, in this situation the average sample size is 2899, which corresponds to an 18.8% reduction of the maximum sample size for this number of cases and a 24% reduction relative to the fixed sample size design using the angular transform formula. In conclusion, if we were to use the inverse sampling method, we would continue sampling until either 18 cases had been found or until we reached 3542 subjects without the event.

2.3. Sequential methods

Formal sequential methods have proven useful in reducing the sample size needed to test hypotheses in some situations. For the current problem, among the different types of sequential methods available, the triangular test as defined by Whitehead (1992) has been chosen for comparison with inverse sampling. This sequential design is simple to implement and is attractive for practical situations.

In a sequential method the sample size needed is a random variable. The implementation of the method has two phases, the design and the analysis. In the design phase the sequential rule is defined given the following values: the difference of interest to be detected (θ_R), the test characteristics in terms of power ($1-\beta$) and significance level (α) and the shape of the boundaries chosen for the sequential rule. In the analysis phase, as data accumulates, repeated evaluations of the sequential rule are made. The values of Z, the cumulative difference between p and p_0, and V, the information that depends on n, are calculated at each inspection. These two values are plotted against each other, that is, Z against V, in a graph where the sequential rule has been drawn. For the triangular sequential test used here, the sequential rule consists on two straight line boundaries making the shape of a triangle as it is illustrated in Figure 2. The area inside these boundaries is called the continuation region. As data accumulates, the path from (Z,V) is drawn and sampling continues while the path is whithin the continuation region. Whenever this path crosses one of the boundaries, a decision is taken. If the upper boundary is crossed, the null hypothesis is rejected. Otherwise, if the lower boundary is crossed, the null hypothesis is accepted. The position of the boundaries are computed to assure a given power and significance level. Approximate p-values can be calculated depending on the position of the point where the boundaries were crossed in the last inspection. The computer program PEST (Brunier & Whitehead, 1993) performs all necessary calculations for the design and analysis of this sequential method.

As with the fixed sample design, the parameterization used for θ is important in our study because the proportions compared are small and then the normal approximation used is not as good as desired. We have compared the results of inverse sampling with all three parameterizations described in the fixed sample size paragraph and a new one, proposed by Sprott (1973) which was also explored by Whitehead (1981). The Sprott's method has the property that the expected value of the third derivative of the likelihood function is zero and gives a very good normal approximation even for extreme situations. For this parameterization,

$$\theta = \{\eta(p) - \eta(p_0)\}\{p_0(1-p_0)\}^{-1/3},$$

where

$$\eta(p) = \int_0^p \frac{dt}{\{t(1-t)\}^{2/3}}$$

Table 4. Theoretic and simulated sample sizes for the situation $p_0 = 0.003$, $p = 0.006$, $\alpha = 0.025$ and power $= 0.80$.

	Theoretic values			Simulated values		
Method	$E(n)$	Median (n)	Max (n)	α	Power	Mean (n)
Log-odds ratio	3608	3470	8365	0.0354	0.9127	2553
Probability difference	1738	1672	4029	0.0406	0.6758	1477
Angular transformation	2530	2433	5890	0.0376	0.8122	2018
Sprott's transformation	2147	1977	3795	0.0244	0.7905	2301
Inverse Sampling $r = 18$			3559	0.0264	0.8005	2899

All sample sizes calculated when $p = 0.006$, equivalent fixed sample size $n = 3795$.

This integral can be calculated numerically. The comparison of sequential methods and inverse sampling is shown in table 4. Although the sequential trials were designed with a type I error equal to 0.025 and a power of 0.80, simulations show that the triangular sequential test results in a type I error slightly greater than the specified for all the parameterizations studied, except for the Sprott's transformation. This parameterization adjusts very finely to the design characteristics; the angular transformation maintains the desired power of 0.8, while the log-odds ratio parameterization results in an excess power reaching 0.9 and the probability difference only 0.7. The average final sample size parallels the results of the attained power. The log-odds ratio parameterization needs more patients than the angular transformation and the lower number is seen for the probability difference, though with this parameterization the desired power is not attained. The Sprott's transformation needs a sample size intermediate between the log-odds ratio and the angular transformation. Note that, even for the log-odds ratio parameterization, the average sample size is smaller than the inverse sampling design. The Sprott's transformation, that gives best results in power and type I error, needs on average about 20% less sample size than the equivalent inverse sampling design and about 40% less sample size than the equivalent fixed sample design.

3. OTHER ALTERNATIVES

Alternative hypothesis different from $p = 0.006$ have been explored. We have chosen smaller values for the difference of interest ($p = 0.0035$, that corresponds to a 17% relative increase, $p = 0.004$ that corresponds to a 25% relative increase) and greater values ($p = 0.009$ that corresponds to a three-fold relative increase). Table 5 shows the summary results for the simulations using these alternatives. When the difference of interest is very small ($p = 0.0035$ or $p = 0.004$), the test characteristics are preserved better for

Table 5. Sample size and performance for triangular sequential tests with different parameterizations and inverse sampling for other alternative hypothesis p ($\alpha = 0.025$ and power $= 0.80$).

Method	p	α	Power	Average sample size (n)
Log-odds ratio	0.0035	0.0271	0.8364	67587
	0.004	0.0290	0.8545	18136
	0.009	0.0403	0.9524	853
Probability difference	0.0035	0.0280	0.7744	60274
	0.004	0.0300	0.7513	14553
	0.009	0.0600	0.6743	394
Angular transformation	0.0035	0.0266	0.8044	63984
	0.004	0.0307	0.8067	16287
	0.009	0.0451	0.8079	581
Sprott's transformation	0.0035	0.0239	0.7992	65992
	0.004	0.0250	0.7961	17316
	0.009	0.0219	0.7728	718
Inverse sampling				
r t_{max}				
337 100339	0.0035	0.0217	0.7819	95619
99 26731	0.004	0.0223	0.7883	24432
8 1146	0.009	0.0242	0.8076	846
Fixed sample size for angular transform formula:				
$p = 0.0035$ 101542				
$p = 0.004$ 27232				
$p = 0.009$ 1213				

all parameterizations. The triangular sequential test using the Sprott's parameterization needs, on average, a 35% smaller sample than the fixed size design. However, the sample sizes needed to detect these small differences are prohibitive for practical purposes. For alternatives of greater magnitude ($p = 0.009$), the sample size is reduced but with low performance of the tests characteristics.

Inverse sampling design keeps the test characteristics in all cases, but average sample size needed is greater than sequential tests.

4. DISCUSSION

We have explored alternative designs to reduce the sample size needed when the interest is to compare a small proportion with a reference value. Inverse sampling design, which stops sampling when a predefined number of events have been observed is an easy procedure to implement, and in our study could save 24% of the fixed sample size design in optimal situations, when the real proportion doubled the reference value of 0.003. The maximum sample size needed with this design is always inferior (6%) to the equivalent with fixed sample.

Formal sequential designs, based on continuous boundaries as the triangular sequential test, can reduce even more the sample size needed in optimal situations, up to 40%. However this method also has some limitations. As Whitehead (1992) stresses and we have checked for the design of our experiment, it is important to choose an adequate parameterization for the difference to be tested. The transformation proposed by Sprott performs well with respect to error rates, but other parameterizations explored have type I error rates greater than the specified and should be used with caution in situations similar to our case, where the reference proportion is small. Exact sequential methods have been developed for special designs, including this one (Stallard & Todd, 2000), but are not easy to implement.

The boundaries of these sequential methods impose the risk of needing more observations than the fixed sample case in some situations. For the triangular test, the maximum sample size in our experiment would be up to 30% more in the extreme case, which might occur if the real proportion is about half the difference between the reference proportion and the population proportion p_0. Though the situation where the true proportion is not as high as expected might not be so rare in practice, the median sample size of the triangular test is always smaller than the equivalent fixed design. For example, in our study the reference proportion p_R was 0.006 and the population proportion p_0 was 0.003. In the case that the real proportion was about 0.0045, the expected sample size would be 2147 and the 90^{th} percentile 3340, still 12% less than the 3795 needed with a fixed design as calculated by the angular transform formula.

The observed gain in sample size for the triangular test can be compared with the expected ones shown in table 4, derived from sequential theory (Whitehead, 1992). These theoretical sample sizes show some disagreement with the simulated values that parallel the observed and expected type I error rate. For the log-odds ratio, probability difference and angular transformation parameterizations, the observed mean sample size is smaller than the expected. These parameterizations show a lower type I error coverage than expected. For the Sprott's parameterization the observed mean sample size is slightly greater than the expected and type I error coverage is correct.

In conclusion, to compare a small proportion with a reference value, the inverse sampling design and formal sequential methods like the triangular test may prove useful to save sample size. The use of the triangular sequential method, that is based on approximations to the likelihood function should be cautious since the type I error rate is slightly increased and power varies unless the Sprott's parameterization is used. With these sequential methods the researcher must accept a small risk of needing to exceed the sample size that would be used with a classical design.

5. REFERENCES

Brunier, H. & Whitehead, J. (1993). *PEST Version 3, Operating Manual*. Reading: The University of Reading.

Finney, D. J. (1949). «On a method of estimating frequencies». *Biometrika*, 36, 233-234.

George, V. T. & Elston R. C. (1993). «Confidence limits based on first occurrence of an event». *Statist. Med.*, 12, 686-690.

Haldane, J. B. S. (1945). «On a method of estimating frequencies». *Biometrika*, 33, 222-225.

Haseman, J. K. (1978). «Exact sample sizes for use with Fisher-Irwin test for 2×2 tables». *Biometrics*, 34, 106-109.

Jowett, G. H. (1963). «The relationship between the binomial and F distributions». *The Statistician*, 13, 55-57.

Lui, K.-J. (1995). «Confidence limits for the population prevalence rate based on the negative binomial distribution». *Statist. Med.*, 14, 1471-1477.

Lui, K.-J. (1995). «Notes on conditional confidence limits under inverse sampling». *Statist. Med.*, 14, 2051-2056.

Sprott, D. A. (1973). «Normal likelihoods and their relation to large sample theory of estimation». *Biometrika*, 60, 457-465.

Stallard, N. & Todd, S. (2000). «Exact sequential test for single samples of discrete responses using spending functions». *Statist. Med.*, 19, 3051-64.

Whitehead, J. (1981). «Use of the sequential probability ratio test for monitoring the percentage germination of accessions in seed banks». *Biometrics*, 37, 129-136.

Whitehead, J. (1992). *The design and analysis of sequential clinical trials*. London: Ellis Horwood.