Ann. Inst. Statist. Math.
Vol. 49, No. 4, 681-692 (1997)

CONTROLLING TYPE II ERROR WHILE CONSTRUCTING TRIPLE SAMPLING FIXED PRECISION CONFIDENCE INTERVALS FOR THE NORMAL MEAN

M. S. SON[1], L. D. HAUGH[1], H. I. HAMDY[1,2] AND M. C. COSTANZA[1,2]

[1]*Statistics Program, College of Engineering and Mathematics, The University of Vermont, 16 Colchester Avenue, Burlington, VT 05401, U.S.A.*
[2]*Medical Biostatistics, College of Medicine, The University of Vermont, 24C Hills Science Building, Burlington, VT 05401, U.S.A.*

(Received January 16, 1996; revised November 1, 1996)

Abstract. The rationale and methodology for estimating a mean with a fixed width confidence interval through sampling in three stages are extended to cover the additional problem of testing hypotheses concerning shifts in the mean with controlled Type II error. The coverage probability and operating characteristic function of the confidence interval based on the integrated approach are derived and compared with those of the usual triple sampling confidence interval. The extended methodology leads to better coverage probability and uniformly better Type II error probabilities. Achieving the additional objective of controlling Type II error inevitably implies a two- to threefold increase in the required optimal sample size. Some suggestions for dealing with this apparent limitation are discussed from a practical viewpoint. It is recommended that an integrated approach to estimation and testing based on confidence intervals be incorporated in the design stage for credible inferences.

Key words and phrases: Coverage probability, operating characteristic function, optimal design.

1. Introduction

Statistical hypothesis tests and confidence intervals can be used to achieve similar inferential objectives. But confidence intervals, in general, provide more information concerning the reliability of the inference. It is well known that a confidence interval shows by its length the precision of estimation, as well as which parameter values would not be rejected if they were hypothesized as (point) null values. This classical view of the duality between the union of all non-rejectable point null hypotheses and a single confidence interval that is taught in most introductory statistical methods courses is what Tukey ((1991), p. 102) was referring to in passing when he wrote:

> "Many of us are familiar with *deriving* a confidence interval from an infinite array of tests of significance, one for each potential null hypothesis".

His main point, however, was made in the two subsequent sentences, to wit:

> "Fewer of us, perhaps, have thought of the *use* of a confidence interval as the reverse process. This is the most important reason for a confidence interval...".

In this 'reverse' view, the intersection of an infinite array of confidence intervals is utilized to determine a single set of plausible point null hypotheses.

Regardless of the method used to derive a hypothesis test, careful consideration of the issue of Type II error should be made to ensure that a credible inference will be obtained. And although it does not appear to be standard practice, the operating characteristics of confidence intervals should be addressed. In the context of the classical use of confidence intervals to test hypotheses, Lehmann ((1986), pp. 89–96) studied the relationship between uniformly most powerful one-sided tests and corresponding lower or upper confidence bounds. In medical applications there appears to be some awareness of the need to consider the power of tests based on confidence intervals. For example, Bristol (1989) computed the length of the confidence interval obtained using the sample size required to control the power of the corresponding test, as well as the power obtained using the sample size required to control the length of the confidence interval.

In designing and performing tests of hypotheses based on confidence intervals using the 'reverse process' that Tukey suggested, one would be concerned with controlling Type II errors in advance. This is essentially the approach employed when statistical quality control charts are designed and used to detect shifts in a process mean (see e.g. Montgomery (1982) and Rahim (1993)). Apart from studies such as the latter, however, the relationship between confidence intervals and the power of tests derived therefrom has received relatively little attention in the statistical literature.

In order to achieve a targeted coverage probability in constructing a fixed width confidence interval for a parameter in the presence of nuisance parameters, a variety of multistage sampling techniques have been developed since the 1940's by many authors. Of these techniques, Stein's (1945) two stage group sampling provides at least the desired coverage, but it can lead to substantial oversampling if the initial sample size is much less than the optimal fixed sample size, as demonstrated by Ghosh and Mukhopadhyay (1981). Anscombe's (1953) and Chow-Robbins' (1965) one-by-one sequential sampling improves upon this drawback, but the desired coverage is attained asymptotically, and it is impractical to implement when decisions need to be reached quickly. A major breakthrough was achieved when Hall (1981) proposed an elegant three stage group sampling technique (triple sampling) as a reasonable compromise between two stage and purely sequential sampling for estimating the normal mean. However, since the triple sampling coverage probability was also attained asymptotically due to the nature of the optimal sample size ordinarily used, he suggested modifying the procedure through increasing the sample size by a small known number of extra observations after termination of the final sampling stage to improve the coverage.

To the best of our knowledge, no one else has studied the use of triple sampling fixed width confidence intervals to test hypotheses in Tukey's (1991) sense, either without or with imposing an additional requirement of some kind of Type II error control. Recently, Costanza *et al.* (1995) evaluated the sensitivity of fixed width confidence intervals to detecting shifts in the normal mean based on Hall's (1981) triple and modified triple sampling *versus* the corresponding fixed size sampling procedure. They found the (unmodified) triple sampling fixed width confidence intervals were more sensitive to shifts occurring within the intervals than their fixed sample size counterparts. However, the corresponding Type II error probabilities were still large. Although the use of Hall's (1981) modified triple sampling improved the coverage probability, it also led to increases in the Type II error probabilities for shifts occurring both inside and far outside the confidence intervals. This occurred because the usual "optimal" fixed sample sizes used to establish the triple sampling estimation procedures did *not* reflect any requirements regarding the control of Type II error. They conjectured that the use of another "optimal" fixed sample size that *did* reflect some form of Type II error control would probably improve the coverage probability as well. In this paper we investigate this latter approach in depth, as well as examine other relevant issues involved in controlling the Type II error probabilities of triple sampling fixed length confidence intervals.

Specifically, in Section 2 we describe the hypotheses to be tested when controlling Type II error, and we derive the corresponding (approximate) optimal fixed sample size. In Section 3 a triple sampling procedure designed to estimate the optimal sample size is proposed and its asymptotic properties are examined and compared to those of the usual triple sampling procedure without control for Type II error. In Section 4 we derive the asymptotic operating characteristic function of our proposed triple sampling procedure. In Section 5 some numerical computations for small to moderate size samples are presented to supplement the asymptotic findings. Discussion of some practical and sampling design issues appears in Section 6.

2. Constructing fixed precision confidence intervals with controlled Type II error

Let $X_1, X_2, \dots$ be a sequence of independent and normally distributed random variables with mean μ and variance σ^2, both unknown. Suppose it is required to estimate μ by a confidence interval such that the precision, $\pm d$, and the coverage probability, $100(1-\alpha)\%$, of the interval are predetermined. Suppose also that a random sample $X_1, X_2, \dots, X_n$ of size n (≥ 2) is available, from which we compute the usual sample measures $\bar{X}_n$ and S_n^2 for μ and σ^2, respectively. We propose the confidence interval $I_n = (\bar{X}_n - d, \bar{X}_n + d)$ for μ.

Moreover, suppose that for all $k \geq 0$ the constructed interval is to be used to test the null hypothesis

$$H_0: \mu = \mu_0 \in I_n$$

versus the alternative hypothesis

$$H_1: \mu = \mu_1 = \mu_0 \pm d(1+k) \notin I_n,$$

such that the probability of Type II error when the true mean is, in fact, μ_1 for a prespecified value of k is required to be at most $100\beta\%$. We emphasize that under H_0, *all* of the values in I_n would be considered "acceptable" (or "non-rejectable") values (or equally represent "*the*" true value) of μ. Consequently, H_1 only makes sense as such for shifts of magnitude $d(1+k)$ occurring *outside* I_n. (For clarity, we mention that the hypotheses and the representation of shifts in μ studied in this paper differ slightly from those considered by Costanza *et al.* (1995).)

As far as we are aware, in all previous work related to constructing fixed width confidence intervals of length $2d$ with prespecified coverage probability $100(1-\alpha)\%$, but *without explicit* control of Type II error, the required optimal fixed sample size has been taken to be

$$n^{\circ} = a^2\sigma^2/d^2, \tag{2.1}$$

where a is the upper $\alpha/2$ critical point of the $N(0,1)$ distribution. However, if, as in this paper, we impose an *additional* requirement that Type II error probability be explicitly controlled in some form, the required optimal sample size must accordingly be modified (i.e., *increased*).

More specifically, we first consider the usual (approximate (see e.g. Brownlee (1965), pp. 117–118)) optimal fixed sample size required to control the Type II error probabilities of detecting shifts in μ of magnitude $\pm d(1+k)$ units asay from $\mu = \mu_0$ *outside* the interval for a prespecified value of k at a prespecified value β, given by

$$n^{\bullet} = (a+b)^2\sigma^2/d^2(1+k)^2, \tag{2.2}$$

where b is the upper β critical point of the $N(0,1)$ distribution. (In this and subsequent derivations we assume that the value of $\mu = \mu_0$ is located in the center of the interval in order to provide equal Type II error probabilities for equidistant shifts to $\mu = \mu_1$ outside the interval in either direction.) For k close to 0 in (2.2), we expect a coverage probability greater than the targeted value, since (2.2) will become greater than (2.1). However, for larger k we expect a coverage probability less than the targeted value, since the effect of increasing k will dominate the effect of b, and (2.2) will become less than (2.1).

Now it is certainly reasonable to keep the coverage probability at least at the targeted value independent of the value of k. It is also natural to be more concerned about small values of k, since potential values of $\mu = \mu_1$ close to the endpoints of the interval from outside are less likely to be detected. With $k = 0$ in (2.2), we obtain an upper bound for the optimal sample size,

$$n^* = (a+b)^2\sigma^2/d^2, \tag{2.3}$$

which maintains at least the targeted coverage through focusing on the value of k of *most* concern. It also follows that the Type II error probabilities, $\beta(k)$ for arbitrary $k > 0$, corresponding to the use of (2.3) will be less than those corresponding to the use of (2.2) for *all* $k \geq 0$.

In summary, utilizing n^* in (2.3) as the "required" optimal sample size has the following desirable features: (1) it ensures *at least* the targeted coverage,

$100(1-\alpha)\%$, independent of k; (2) it has Type II error probabilities which are *less* than the prespecified $100\beta\%$ at the particular shift(s) of interest indexed by the prespecified value of k; (3) it actually has Type II error probabilities $\beta(k)$ that are *uniformly less* than those corresponding to the use of (2.2) for *all* $k \geq 0$. Henceforth we utilize n^* in (2.3) as the optimal sample size required to achieve our objectives in defining our proposed procedures.

Clearly, (2.1) is the special case of (2.3) with $b = 0$, which implies that the corresponding $\beta(0) \cong 0.50$. In other words, the usual classical (i.e., "uncontrolled" for Type II error) application of the confidence interval to test H_0 *versus* H_1 has the property that its Type II error probability is *only implicitly* controlled at 50% for shifts occurring at the endpoints of the interval. Moreover, from the classical standpoint, the apparent "fact" that the corresponding "Type II error probabilities" are even higher for values of μ = "μ_1" lying within the interval might be viewed as a disadvantage. However, the use of the confidence interval to test H_0 *versus* H_1 that is being considered in this paper implies that this apparent drawback is actually an advantage in the sense that all potential values of μ lying *within* the interval are legitimate values only under H_0. Thus, the designation "Type II error" does not even apply for these values of μ (instead, "sensitivity to departures from the center of the interval" is a more appropriate descriptor).

If σ^2 was known, a fixed sample size procedure based on n^* in (2.3) to construct I_{n^*} would be employed. Since σ^2 is unknown, a triple sampling procedure could be used to simultaneously estimate n^* via estimation of σ^2 in constructing the required interval for μ.

3. Type II error controlled triple sampling confidence intervals for the mean

The triple sampling procedure for estimating n^* in (2.3) begins with a pilot sample $X_1, X_2, \ldots, X_m$, $m \geq 2$, from which $\bar{X}_m$ and S_m^2 are computed as initial estimates of μ and σ^2. The second stage sample size is determined according to the decision rule

$$N_1 = \max\{m, [\gamma(a+b)^2 S_m^2/d^2] + 1\}, \tag{3.1}$$

where $0 < \gamma < 1$ is a design factor which represents the fraction of n^* to be estimated in this stage and $[\cdot]$ denotes the largest integer function. (Empirical results support the choice of $\gamma = 0.50$ and $m = 5$ to 15 in practice.) If the decision is to continue sampling, we combine the pilot sample with $N_1 - m$ additional observations to bring the final (third stage) sample size to

$$N = \max\{N_1, [(a+b)^2 S_{N_1}^2/d^2] + 1\}. \tag{3.2}$$

If necessary, we observe $N - N_1$ further observations and terminate the sampling process. We then compute $\bar{X}_N$. Consequently we construct the Type II error controlled confidence interval

$$I_N = (\bar{X}_N - d, \bar{X}_N + d) \quad \text{for} \quad \mu.$$

The main asymptotic features of the triple sample size N are given in Theorem 3.1.

THEOREM 3.1. *For the controlled triple sampling procedure* (3.1)–(3.2) *and optimal fixed* n^* *in* (2.3), *as* $d \to 0$ *we have*

(i) $E(N) = n^* - 2\gamma^{-1} + 1/2 + o(1)$

(ii) $\text{Var}(N) = 2\gamma^{-1}n^* + o(d^{-2})$

(iii) $E|N - n^*|^3 = o(d^{-4})$

(iv) $E(h(N)) = h(n^*) + (2\gamma^{-1})\{(\gamma - 4)h'(n^*) + 2n^*h''(n^*)\} + o(d^{-4}|h'''(n^*)|)$,

where $h(> 0)$ *is any real valued, continuously differentiable function in a neighborhood of* n^* *such that* $\sup_{n \geq m} |h'''(n)| = O(|h'''(n^*)|)$, *and primes mean derivatives.*

The proofs of Parts (i)–(iii) of Theorem 3.1 are given in Hall (1981) and Hamdy (1988), so they are omitted. The proof of Part (iv) of Theorem 3.1 involves a straightforward application of the expectation of a Taylor series expansion of $h(N)$ around n^*, Parts (i)–(iii), and the assumption that h''' is bounded.

Remark 1a. Part (i) of Theorem 3.1 shows that, on the average, the triple sample size is approximately the optimal fixed sample size in (2.3). Conditioning on N, it is easy to show that $E(\bar{X}_n) = \mu$. Parts (ii) and (iii) of Theorem 3.1 indicate that the triple sampling technique is in fact as efficient as the Anscombe (1953)–Chow and Robbins (1965) purely sequential sampling technique. Again conditioning on N and applying Part (iv) of Theorem 3.1 to evaluate $E(N^{-1})$ provide $\text{Var}(\bar{X}_N) = (\sigma^2/n^*) - (\gamma - 8)(\sigma^2/2\gamma n^{*2}) + o(d^4)$. Moreover, the special case of $h(N) = e^{tN}$ in Part (iv) can be used to establish the asymptotic normality of N. Part (iv) of Theorem 3.1 is also used subsequently to obtain both the coverage probability and the operating characteristic function of I_N. Our approach to the latter based on considering a general continuously differentiable function of N is more direct than an approach based on employing the integer moments of N, which has been utilized in previous work (see Hamdy *et al.* (1988)).

Remark 1b. Theorem 3.1 holds for the triple sampling procedure suitably based on *any* optimality criteria and the corresponding optimal fixed sample size. In particular, it holds for Hall's (1981) (uncontrolled) triple sampling procedure with optimal sample size n° in (2.1).

The following Theorem 3.2 extends the results in Theorem 3.1 to obtain the coverage probability of I_N.

THEOREM 3.2. *For the controlled triple sampling procedure* (3.1)–(3.2) *and optimal fixed* n^* *in* (2.3), *as* $d \to 0$ *the coverage probability of* I_N *is given by*

$$\begin{aligned} P(\mu \in I_N) &= \{2\Phi(a+b) - 1\} - Q_0(n^*, \gamma) + o(d^2) \\ &> (1 - \alpha) - Q_0(n^*, \gamma) + o(d^2), \end{aligned}$$

where $Q_0(n^*, \gamma) = (a+b)\phi(a+b)\{(a+b)^2 - \gamma + 5\}(2\gamma n^*)^{-1}$ *and* $\phi(\cdot)$ *is the probability density function of* $N(0,1)$.

To prove Theorem 3.2 we note that $P(\mu \in I_N) = 2E\Phi(d\sqrt{N}/\sigma) - 1$. To evaluate $E\Phi(d\sqrt{N}/\sigma)$, we make use of Part (iv) of Theorem 3.1 to complete the proof.

Remark 2. We note that for most reasonable choices of α and β and even for small values of n^*, $Q_0(n^*, \gamma)$ is negligible. The coverage probability of Hall's (1981) (uncontrolled) triple sampling procedure with optimal sample size n° in (2.1) follows from Theorem 3.2 with equality instead of the inequality for $b = 0$. Moreover, the corresponding $Q_0(n^\circ, \gamma)$ term is larger than $Q_0(n^*, \gamma)$ in Theorem 3.2. Thus, Hall's procedure attains the coverage probability only asymptotically, whereas our procedure exceeds the targeted coverage for any value of n^*. This latter result *could* be construed as indicating that our procedure is too liberal in terms of Type II error probabilities. However, we also note that since the n^* required by our procedure is *necessarily* much larger than (e.g.) the "corresponding" n° required by Hall's procedure, the impact of this apparent drawback would seem to be of less concern. The use of $n^\bullet$ in (2.2) would attain the targeted coverage, but the control of Type II error would be achieved *only* at the particular prespecified value of k. We opt to use n^* in our subsequent developments for the reasons stated previously during the course of its derivation in Section 2.

4. Operating characteristic function of Type II error controlled confidence intervals

In this section we investigate the capability of the controlled triple sampling fixed width confidence interval I_N to signify potential shifts in the true mean of "distance" k (measured in units of the precision d) occurring *outside* the interval when it is erroneously thought that such shifts never took place. For all $k \geq 0$, the relevant hypotheses to be tested are

$$H_0 : \mu = \mu_0 \in I_N \qquad \textit{versus} \qquad H_1 : \mu = \mu_1 = \mu_0 \pm d(1+k) \notin I_N.$$

The Type II error controlled conditional probability, $\beta_{TC}(k)$, of *not* detecting such a shift when, in fact, such a shift actually took place can be written as

$$\begin{aligned}
\beta_{TC}(k) &= P(\mu \in I_N \mid H_1) \\
&= P(\bar{X}_N - d \leq \mu \leq \bar{X}_N + d \mid H_1) \\
&= \sum_{n=m}^{\infty} P(-(2+k)d \leq \bar{X}_N - \mu_1 \leq -kd, N = n) \\
&= \sum_{n=m}^{\infty} P(-(2+k)d \leq \bar{X}_N - \mu_1 \leq -kd \mid N = n) P(N = n).
\end{aligned}$$

The event $(N = n)$ depends on S_n^2 from the rule (3.2) for $n = m, m+1, \ldots$. On the other hand, the normality implies that S_n^2 and $\bar{X}_n$ are independent for all n. Therefore,

$$\beta_{TC}(k) = \sum_{n=m}^{\infty} P(-(2+k)d\sqrt{n}\sigma^{-1} \leq (\bar{X}_n - \mu_1)\sqrt{n}\sigma^{-1} \leq -kd\sqrt{n}\sigma^{-1}) P(N = n)$$

$$= \sum_{n=m}^{\infty} \{\Phi(-kd\sqrt{n}\sigma^{-1}) - \Phi(-(2+k)d\sqrt{n}\sigma^{-1})\}P(N=n)$$
$$= E\{\Phi(-kd\sqrt{N}\sigma^{-1})\} - E\{\Phi(-(2+k)d\sqrt{N}\sigma^{-1})\}.$$

We then apply Part (iv) of Theorem 3.1 to obtain

$$E\{\Phi(-kd\sqrt{N}\sigma^{-1})\} = \Phi(-k(a+b)) + Q_1(n^*, \gamma, k) + o(d^2) \quad \text{and}$$
$$E\{\Phi(-(2+k)d\sqrt{N}\sigma^{-1})\} = \Phi(-(2+k)(a+b)) + Q_2(n^*, \gamma, k) + o(d^2),$$

where

$$Q_1(n^*, \gamma, k) = (4\gamma n^*)^{-1}k(a+b)\phi(-k(a+b))\{k^2(a+b)^2 - \gamma + 5\} \quad \text{and}$$
$$Q_2(n^*, \gamma, k) = (4\gamma n^*)^{-1}(2+k)(a+b)\phi(-(2+k)(a+b))\{(2+k)^2(a+b)^2 - \gamma + 5\}.$$

Finally, we obtain the following Theorem.

THEOREM 4.1. *For the controlled triple sampling procedure* (3.1)–(3.2) *and optimal fixed n^* in* (2.3), *as $d \to 0$ the operating characteristic function is given by*

$$\beta_{TC}(k) = \Phi\{-k(a+b)\} - \Phi\{-(2+k)(a+b)\} + Q_1(n^*, \gamma, k) - Q_2(n^*, \gamma, k) + o(d^2).$$

Remark 3. The operating characteristic function, $\beta_T(k)$, of Hall's (1981) (uncontrolled) triple sampling procedure with optimal sample size n° in (2.1) follows from Theorem 4.1 with $b = 0$. Since for all $k \geq 0$ we have

$$\Phi\{-k(a+b)\} - \Phi\{-(2+k)(a+b)\} > \Phi\{-k(a)\} - \Phi\{-(2+k)(a)\}$$

and

$$Q_1(n^\circ, \gamma, k) - Q_2(n^\circ, \gamma, k) > Q_1(n^*, \gamma, k) - Q_2(n^*, \gamma, k),$$

we have $\beta_T(k) > \beta_{TC}(k)$ uniformly in k.

5. Numerical comparisons of controlled vs. uncontrolled confidence intervals

In this section we supplement the asymptotic results for the uncontrolled and controlled triple sampling procedures with some numerical computations for small to moderate samples. The optimal sample sizes were n° in (2.1) for Hall's (1981) triple sampling procedure and n^* in (2.3) for our proposed controlled procedure in (3.1)–(3.2). For consistency, we employed the same values of the optimal n° with the corresponding values of the fixed precision d, as well as level of significance $\alpha = 0.05$ used by Hall in his simulations. These optimal sample sizes were $n^\circ = 24$, 43, 61, 76, 96, 125, 171, 246, and 384, with the corresponding values of $d = 0.4$,

Table 1. Asymptotic coverage and Type II error probabilities of triple and controlled triple sampling procedures: $\alpha = 0.05$, $\gamma = 0.50$.

	Uncontrolled optimal sample size n° in (2.1)								
	24	43	61	76	96	125	171	246	384
	Asymptotic coverage probability $\hat{P}_T$								
	0.9102	0.9278	0.9343	0.9374	0.94	0.9424	0.9444	0.9461	0.9475
Shifts k	Asymptotic Type II error probability $\hat{\beta}_T(k)$								
0.0	0.4997	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
0.1	0.4294	0.4262	0.4251	0.4245	0.4241	0.4236	0.4233	0.423	0.4227
0.2	0.3615	0.3553	0.353	0.3519	0.351	0.3502	0.3495	0.3489	0.3484
0.3	0.2981	0.2894	0.2861	0.2845	0.2832	0.2821	0.2811	0.2802	0.2795
0.4	0.241	0.2302	0.2262	0.2243	0.2226	0.2212	0.22	0.2189	0.218
0.5	0.191	0.1789	0.1744	0.1722	0.1704	0.1688	0.1674	0.1662	0.1653
0.6	0.1486	0.1359	0.1311	0.1289	0.127	0.1253	0.1238	0.1226	0.1216
0.7	0.1134	0.1009	0.0962	0.094	0.0921	0.0905	0.089	0.0878	0.0868
0.8	0.085	0.0732	0.0689	0.0668	0.0651	0.0635	0.0622	0.061	0.0601
0.9	0.0624	0.052	0.0481	0.0463	0.0448	0.0434	0.0422	0.0412	0.0403
1.0	0.0449	0.0361	0.0328	0.0313	0.03	0.0288	0.0278	0.0269	0.0262
1.5	0.0059	0.004	0.0033	0.003	0.0027	0.0025	0.0022	0.0021	0.0019
2.0	0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

	Controlled optimal sample size n^* in (2.3) for $\beta = 0.05$								
	81	145	206	257	324	422	578	832	1299
	Asymptotic Coverage Probability $\hat{P}_{TC}$								
	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997
Shifts k	Asymptotic Type II error probability $\hat{\beta}_{TC}(k)$								
0.0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
0.1	0.3593	0.3593	0.3592	0.3592	0.3592	0.3592	0.3592	0.3592	0.3592
0.2	0.2355	0.2355	0.2355	0.2355	0.2355	0.2355	0.2355	0.2355	0.2355
0.3	0.1398	0.1398	0.1398	0.1397	0.1397	0.1397	0.1397	0.1397	0.1397
0.4	0.0747	0.0747	0.0747	0.0747	0.0747	0.0747	0.0747	0.0747	0.0747
0.5	0.0358	0.0358	0.0357	0.0357	0.0357	0.0357	0.0357	0.0357	0.0357
0.6	0.0153	0.0153	0.0153	0.0153	0.0153	0.0153	0.0153	0.0153	0.0153
0.7	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058
0.8	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
0.9	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
1.0	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
1.5	0	0	0	0	0	0	0	0	0
2.0	0	0	0	0	0	0	0	0	0

0.3, 0.25, 0.225, 0.2, 0.175, 0.15, 0.125, and 0.1. We fixed the controlled probability of Type II error at $\beta = 0.05$ and employed the same value of α and values of d to obtain the corresponding values of $n^* = 81$, 145, 206, 257, 324, 422, 578, 832, and 1299. We set the design factor $\gamma = 0.3$, 0.5, and 0.8, and took the shift factor $k = 0(0.1)1.0$ and $1.5(0.5)2.5$ to study the impacts of these factors on both the coverage and the Type II error probabilities for both procedures.

The numerical computations based on Theorems 3.2 and 4.1 are presented in Table 1 for values of $k \leq 2$ and the case $\gamma = 0.50$ only for brevity (the effects of γ were negligible for the controlled procedure and slight for the uncontrolled procedure since γ only appears in the terms $Q_0(\cdot,\gamma)$ in Theorem 3.1 and $Q_1(\cdot,\gamma,k) - Q_2(\cdot,\gamma,k)$ in Theorem 4.1). The uncontrolled coverage probability was always less than the targeted value and approached it only asymptotically, while the controlled coverage probability was always substantially larger than the targeted value. As expected, as k increased the probabilities of Type II errors decreased for both procedures. However, $\hat{\beta}_{TC}(k) < \hat{\beta}_T(k)$ uniformly in k. It was also evident that attainment of the targeted value of $\beta = 0.05$ occurred for k between 0.4 and 0.5 for the controlled procedure *versus* for k between 0.8 and 0.9 for the uncontrolled procedure.

6. Discussion

It is worth reemphasizing that in considering *where* the control of Type II error of a fixed precision confidence interval for a parameter is to be achieved, *only* potential values of the parameter occurring *outside* the interval are legitimate candidates. Again, this is so because all potential values lying *within* the confidence interval comprise the *null* hypothesis. It is only when studying the *sensitivity* of the interval to *any* shifts in the true value of the parameter (i.e., without regard to control of Type II error) that the designation "Type II error" may be appropriately interpreted as applying to values of the parameter lying inside the interval, as was done in our previous work (Costanza *et al.* (1995)).

Regardless of the type of fixed width confidence interval to be employed to test the hypotheses considered in this paper, it is obvious that to satisfy *both* the previous requirements of controlling the coverage probability and the precision of estimation, as well as the additional requirement of controlling the probabilities of Type II error, the corresponding optimal sample size *must* be larger than it would be if Type II error probabilities were not required to be controlled. This is also the case for the usual classical method of testing hypotheses based on the use of confidence intervals, as discussed by Kupper and Hafner (1989). They also pointed out that power of tests of hypotheses based on samples of sizes determined *only* through prespecifying the length of the confidence interval (as in (2.1)) is extremely unsatisfactory. Instead, they strongly recommended the use of sample sizes whose determination formulas allow for control of power (as do (2.2) or (2.3)). What is new in our approach is the formulation and solution of the problem based on triple sampling, both in terms of the specific representation of the hypotheses to be tested as well as the derivation of the required optimal sample size and a detailed evaluation of the asymptotic and small to moderate sample size performance of the proposed procedure.

For most reasonable choices of α (e.g., $\alpha \leq 0.10$) and β (e.g., $\alpha \leq \beta \leq 0.20$), the ratio of the corresponding required optimal sample size n^* in (2.3) to n° in (2.1) is on the order of two or three, indicating the dominating effect of the additional requirement of controlling Type II error. Although the required relative increase may seem large on intuitive grounds, and/or the required absolute increase may be prohibitive in terms of available resources for some applications, *the fact remains that an increase of (at most) this magnitude is necessary to achieve all the prespecified objectives.* There are, however, other considerations which can reasonably be viewed as lessening the potential negative impact of having to take larger samples. Of course, for applications where the cost of sampling is a relatively minor issue in comparison to achieving the objectives of having a fixed with confidence interval with better coverage and controlled for Type II error, this would not be of concern.

First, as in any application of sample size determination in the planning phase of a proposed study, the researcher has the option of specifying different values for α, d, σ, β, and the k at which the Type II error probability is to be controlled at the required β. In our experience, it is frequently the case that the results of such calculations are thought of as "ballpark" estimates anyway. Depending on the particular problem and field of application, there may be some flexibility in the requirements, so that certain combinations of the values of these quantities may lead to "optimal" sample sizes which can be "lived with" as acceptable compromises.

And second, the proposed n^* in (2.3) represents a worst case scenario in that it is an upper bound for the more specific $n^\bullet$ in (2.2). Thus, the researcher could opt to focus more directly on controlling only the Type II error probability β at the particular shift size k of interest, which would imply that an optimal sample size smaller than n^* in (2.3) would be required. (The relative magnitudes of $n^\bullet$ in (2.2) and n° in (2.1) depend on the mathematical relation between b and k, so it less clear which is larger than the other.) of course, such an approach would maintain the targeted coverage, but would entail some losses relative to Type II error probabilities for other shift sizes.

In comparison to testing hypotheses using Hall's (1981) (unmodified) confidence interval, which was not designated to control Type II error and which barely maintains the targeted coverage, the controlled triple sampling procedure proposed in this paper has a coverage probability always greatly exceeding the targeted value, even for small optimal sample sizes. Hall's modified procedure is based solely on improving the coverage probability. Although this modification does lead to improvements in *some* probabilities of Type II error, nonetheless it still has the drawback that it is *not* specifically focused on achieving a particular prespecified control of Type II error. When such control *is* required, as often occurs, for example, in monitoring a process mean in statistical quality assurance, then the use of our proposed triple sampling procedure would be the method of choice since it has uniformly better probabilities of Type II error for all potential shifts of the mean occurring outside the fixed precision confidence interval (i.e., the control limits).

Acknowledgements

This work was supported in part by National Cancer Institute Grant P30-CA22435 to the Vermont Comprehensive Cancer Center (Biostatistics Core Resource). The authors thank the editor and the referees for helpful comments.

REFERENCES

Anscombe, F. J. (1953). Sequential estimation, *J. Roy. Statist. Soc. Ser. B*, **15**, 1–29.

Bristol, D. R. (1989). Sample sizes for constructing confidence intervals and testing hypotheses, *Statistics in Medicine*, **8**, 803–811.

Brownlee, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*, *2/e*, Wiley, New York.

Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean, *Ann. Math. Statist.*, **36**, 457–462.

Costanza, M. C., Hamdy, H. I., Haugh, L. D. and Son, M. S. (1995). Type II error performance of triple sampling fixed precision confidence intervals for the normal mean, *Metron*, **53**, 69–82.

Ghosh, M. and Mukhopadhyay, N. (1981). Consistency and asymptotic efficiency of two-stage and sequential procedures, *Sankhya Ser. A*, **43**, 220–227.

Hall, P. (1981). Asymptotic theory of triple sampling for sequential estimation of a mean, *Ann. Statist.*, **9**, 1229–1238.

Hamdy, H. I. (1988). Remarks on the asymptotic theory of triple stage estimation of the normal mean, *Scand. J. Statist.*, **15**, 303–310.

Hamdy, H. I., Mukhopadhyay, N., Costanza, M. C. and Son, M. S. (1988). Triple stage point estimation for the exponential location parameter, *Ann. Inst. Statist. Math.*, **40**, 785–797.

Kupper, L. L. and Hafner, K. B. (1989). How appropriate are popular sample size formulas?, *Amer. Statist.*, **43**, 101–105.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, *2/e*, Wiley, New York.

Montgomery, D. C. (1982). The economic design of control charts: a review and literature survey, *Journal of Quality Technology*, **14**, 40–43.

Rahim, M. A. (1993). Economic design of $\bar{x}$ control charts assuming Weibull in-control times, *J. Qual. Tech.*, **25**, 296–305.

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance, *Ann. Math. Statist.*, **16**, 243–258.

Tukey, J. W. (1991). The philosophy of multiple comparisons, *Statist. Sci.*, **6**, 100–116.