CHAPTER 1

Application of a sequential $t$-test in a cohort nested case-control study with multiple controls per case

I. van der Tweel*, P.A.H. van Noord#, R. Kaaks§

*Centre for Biostatistics, Utrecht University, Utrecht, the Netherlands
#Department of Epidemiology, Utrecht University, Utrecht, the Netherlands
§International Agency for Research on Cancer, Lyon, France

J Clin Epidemiol 1993; 46: 253-259
Abstract

Application of sequential analysis may avoid unnecessary experimentation and achieve economical use of available biomaterial stored in biological banks. When, as often in cohort case-control studies, cases are scarce, it may be possible to use multiple control observations per case to increase the power of a test for detecting differences between cases and controls.

Samples from a biological data bank were analysed. We compared results of a non-sequential analysis with results of sequential $t$-tests for 1 to 5 controls matched per case in a cohort nested case-control study. Simulations are performed to get an idea of the unreliability and the power of the sequential test.

In general the sequential $t$-tests are too conservative with respect to the achieved power. Average sample numbers are lower for the sequential tests and decrease with multiple controls. More than 3 or 4 controls per case does not give a meaningful increase in efficiency.

Keywords: sequential $t$-test, multiple controls, simulations, efficiency, biobanking, cohort nested studies

1.1 Introduction

Sequential analysis of quantitative data has never found wide application in clinical trial practice, even though considering its use might be worthwhile. For ethical reasons alone one may wish to minimize the expected number of exposed patients. From an experimental point of view, one may wish to avoid unnecessary experimentation. In cohort nested case-control studies exposures may be assessed in biological samples stored in a biological bank. In this situation, economy with material from the biological bank may be a reason to choose a sequential type of analysis. In a prospective study, cases are often detected sequentially during follow-up. A sequential analysis could then limit the total duration of the study.

In a sequential case-control analysis, the response of a case is compared with the response of a single control. O’Neill describes in a detailed way a sequential analysis of a matched pair case-control study with a dichotomous response.\(^1\)

In a cohort study, usually there is only a limited amount of biological material per subject, and there are far more controls in the biobank for which such material can be analyzed than cases. Therefore it may be desirable to compensate for the loss of statistical power by comparing each case with more than one control.\(^2\)
Ury \(^3\) showed that, for non-sequential case-control studies with continuously distributed data, the efficiency of multiple \((k > 1)\) controls relative to matched pairs \((k = 1)\) is equal to \(2k/(k+1)\) for equal case and control variability.

Gail et al.\(^4\) show that in (non-sequential) situations with a limited number of cases, more than four controls per case (or vice versa) gives no more meaningful power increase.

We are unaware of literature about the efficiency of multiple controls per case in sequential analyses. Therefore, we compared the effect of more controls per case in a sequential design with the results of a non-sequential analysis.

### 1.2 Materials and patients

We performed retrospective analyses on data from a cohort nested case-referent (control) study on breast cancer and the selenium content in ppm of toenails (Van Noord\(^5\)). The aim of the study was to determine whether selenium, as available in the body, is already decreased before tumour occurrence.

Nail clippings had been collected since 1982 in a cohort of 8760 premenopausal (i.e. without menopausal signs) women (42-52 years of age), who attended to a breast cancer screening program. A total number of 64 premenopausal breast cancer cases were detected in this cohort. Controls were matched to cases for age. For 57 cases 5 controls per case were available; to 7 cases 3 or 4 controls could be matched per case.

Selenium content in the nails did not depend on age, probably due to the relatively small age-range in our data. No seasonal or other time trends were found in nail selenium contents during three years of investigation (unpublished results). The data were analysed in the order the cases became available over time.

### 1.3 Statistical analysis

#### 1.3.1 Non-sequential analysis

For matched case-control observations the minimal sample size \(n_1\) (i.e. the number of case-control pairs necessary) for detecting a true difference between case and control observations of at least \(\mu\) with a (two-sided) type I probability (or unreliability) \(\alpha\) and a type II probability \(\beta\) (i.e. power \(1-\beta\)) is\(^6\)

\[
n_1 = (t_\alpha + t_\beta)^2 \times \frac{\sigma_1^2}{\mu^2}
\]

where

\(\sigma_1^2\) is the variance of the difference between a case and a control observation,

\(t_\alpha\) and \(t_\beta\) are values from the \(t\)-table with \(n_1-1\) \(df\) corresponding to probabilities of \(\alpha/2\) and \(\beta\) respectively.
The type I probability $\alpha$ is the risk one wants to accept that the null hypothesis of no difference between case and control observations is falsely rejected; the type II probability $\beta$ is the risk of falsely not rejecting the null hypothesis when a true difference of at least $\mu$ exists between case and control observations.

In case of multiple (say $k$) control observations per case, assuming equal variances for cases and controls and, for the sake of argument, a negligible correlation between case and control observations, the variance of the difference between a case observation and the mean of the $k$ control observations becomes

$$
\sigma_k^2 = \left\{(k+1)/k\right\} \times \sigma^2 = \left\{(k+1)/2k\right\} \times \sigma_1^2,
$$

($$\sigma_1^2 = 2\sigma^2$$, where $\sigma^2$ is the variance of a single case or control observation).

The minimal number of case-control sets for detecting the same difference $\mu$ then becomes

$$
n_k = (t_{\alpha} + t_{\beta})^2 \times \sigma_k^2 / \mu^2 = n_1 \times \left\{(k+1)/2k\right\}.
$$

N.B. We assumed (near) independence of case and control observations. In case of a positive correlation between case and control observations, the result will be a smaller $\sigma_1^2$ and $\sigma_k^2$ and a smaller sample size needed to detect the same difference $\mu$.

### 1.3.2 Sequential analysis

Wald\(^7\) developed the theory for the ‘sequential probability ratio test’ (SPRT). Rushton\(^8\) further developed this theory to the one-sample, two-sided sequential $t$-test. This test is based on the probability ratio

$$
l_n = \frac{\text{probability of observed results given } H_1 \text{ true}}{\text{probability of observed results given } H_0 \text{ true}}
$$

for $n$ observations processed so far. For our situation with case-control sets, we pose as null hypothesis $H_0$ :

$$
\delta = \mu / \sigma_k = 0
$$

and as alternative hypothesis $H_1$ :

$$
|\delta| > 0
$$

where $\mu$ is the minimal mean difference to be detected and $\sigma_k$ is the theoretical standard deviation of the differences between the case and control observations. Because in most practical situations $\sigma_k$ will be unknown and needs to be estimated from the data, the parameter $\delta = \mu / \sigma_k$ is used in the test. The test operates as follows :

- continue sampling as long as $B < l_n < A$
- stop sampling and decide for $H_0$ as soon as $l_n < B$
- stop sampling and decide for $H_1$ as soon as $l_n > A$
To obtain approximately the a priori specified error probabilities $\alpha$ (two-sided type I error) and $\beta$ (type II error), Wald stated the theorem that $A \sim (1-\beta)/\alpha$ and $B \sim \beta/(1-\alpha)$. The logarithm of the probability or likelihood ratio $l_n$ can be calculated exactly using the series expansion of Kummer’s function.9

Rushton8 obtained a practical approximation to the logarithm of the likelihood ratio. See Appendix I for more details on Kummer’s function, Rushton’s approximation and our adaptation of the test statistic for $k$ control observations per case.

**1.3.3 Simulations**

To examine the effect of multiple controls per case in a sequential $t$-test on its overall type I and type II error, simulation studies were performed. A simulation program was written in Turbo Pascal Version 5.0 (Borland). Random case and control observations were generated following a normal distribution with expectation $\mu_0$ or $\mu_1$ and theoretical standard deviation $\sigma$. The values chosen for $\mu_0$, $\mu_1$, $\sigma$ and $\delta$ under $H_1$ are based on population values and a desirable shift in ppm of the selenium content (see Van Noord10). Both for case and control observations $\sigma$ was chosen equal to 0.15. Under $H_0$: $\delta = 0$, $\mu_0$ was chosen equal to 0.8. Under $H_1$: $|\delta| = \delta$, $\mu_1$ was equal to $0.8 + \delta * \sigma * \sqrt{2}$.

Both under $H_0$: $\delta = 0$ and under $H_1$: $|\delta| = \delta$ ($\delta = 0.3, 0.4$ and $0.5$ respectively), and with 1 to 5 controls per case, we ran a 1000 simulation runs ($\alpha = 0.05$, $1-\beta = 0.80$).

Per run the resulting decision (‘accept $H_0$’ or ‘reject $H_0$ in favour of $H_1$’) and the number of case-control sets necessary to come to that decision were recorded. Simulations were performed using both Rushton’s approximation to the logarithm of the likelihood ratio and the series expansion of Kummer’s function.

**1.4 Results**

**1.4.1 Non-sequential analysis**

The results of a randomized blocks analysis of variance on the ‘selenium and breast cancer’ data for $n = 57$ cases and 5 control observations per case are shown in Table 1. The mean difference between a case and the mean of the corresponding 5 control observations was 0.018 ppm with a SE = 0.029 ppm (NS).
A sequential t-test with multiple controls per case

Table 1 ‘Selenium and breast cancer’ study; descriptive statistics and ANOVA table for 57 cases with 5 controls per case

<table>
<thead>
<tr>
<th></th>
<th>mean (ppm)</th>
<th>SD (ppm)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>0.790</td>
<td>0.156</td>
<td>57</td>
</tr>
<tr>
<td>Controls</td>
<td>0.772</td>
<td>0.207</td>
<td>285</td>
</tr>
</tbody>
</table>

ANOVA table

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of squares</th>
<th>Degrees of freedom</th>
<th>Mean squares</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between matched sets</td>
<td>2.20</td>
<td>56</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within matched sets</td>
<td>0.16</td>
<td>5</td>
<td>0.03</td>
<td>&lt;1</td>
<td>NS</td>
</tr>
<tr>
<td>Case-controls*</td>
<td>0.02</td>
<td>1</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between controls</td>
<td>0.14</td>
<td>4</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>11.24</td>
<td>280</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Due to the difference between cases and the mean of the matched control observations.

Means, standard deviations and a randomized-blocks analysis of variance (ANOVA) table for \( n=57 \) cases with 5 controls per case. Data are the selenium content in ppm in toenails from the ‘selenium and breast cancer’ study.

Within matched sets the sum of squares, degrees of freedom and mean square are subdivided into two components: one that measures the variation because of a difference between cases and the mean of the matched control observations, and one that measures variation between controls. If we assume no differences between control observations, this last component can be combined with the residual sum of squares to give a (slightly) improved estimate of the residual mean square or error variance.

1.4.2 Sequential analysis

Sequential t-tests were performed on the ‘selenium and breast cancer’ data, using the available cases and a random sample of \( k (k = 1,\ldots,5) \) control observations available in the matched set. (For each sequential test performed, control observations were replaced.) Both Kummer’s function and Rushton’s approximation were applied.

The number of cases \( (n) \) at which the decision ‘\( H_0 \) cannot be rejected’ was reached, is tabulated in Table 2 for several alternative hypotheses \( (|\delta| = 0.3, 0.4, 0.5) \).

N.B. None of the tests led to rejection of \( H_0 \); in case of \( H_1; |\delta| = 0.3 \), for some tests no conclusion could be reached with the available number of case-control sets.
Table 2  ‘Selenium and breast cancer’ study; results of sequential $t$-tests for $k$ controls per case

| $k$ | R  | K  | $|\delta| = 0.3$ | R  | K  | $|\delta| = 0.4$ | R  | K  | $|\delta| = 0.5$ |
|-----|----|----|----------------|----|----|----------------|----|----|----------------|
| 1   | 21 | -- | 12             | 23 | 9  | 13             | 1   | 25 | 30             | 30 | 11 | 15             |
| 1   | 27 | 62 | 23             | 24 | 10 | 13             | 1   | 22 | 23             | 24 | 10 | 14             |
| 1   | 26 | 50 | 13             | 25 | 8  | 18             | 2   | 25 | 17             | 21 | 9  | 14             |
| 3   | 22 | -- | 18             | 21 | 12 | 16             | 4   | 22 | 12             | 21 | 8  | 13             |
| 5   | 22 | -- | 13             | 21 | 9  | 13             |

Results of the sequential $t$-tests, given 57-64 cases and random samples of $k$ controls per case, on the ‘selenium and breast cancer’ study ($\alpha = 0.05$ and $1-\beta = 0.80$); R, Rushton’s approximation; K, Kummer’s function.

1.4.3 Simulations

The relative efficiency of more ($k$) controls per case is depicted graphically in Figures 1 and 2 for $\delta = 0.4$. (For $\delta = 0.3$ and $\delta = 0.5$ the course of the relative efficiency is similar). There the relative sample size $n_k/n_1$ is plotted against $k$ for the median, mean and 95th-percentile number of cases required to reject $H_0$ in favour of $H_1$. The theoretical expected efficiency $(k+1)/2k$ is plotted as a comparison.

![Figure 1](image1.png)

Figure 1  Relative sample size ($n_k/n_1$) for mean (▲), median (●) and 95-th percentile (■) number of cases necessary to reject $H_0$ in favour of $H_1$: $|\delta| = 0.4$ compared to the theoretical expected value $(k+1)/2k$ (x), using Rushton’s approximation.
A sequential t-test with multiple controls per case

Figure 2  Relative sample size ($n_k/n_1$) for mean (▲), median (●) and 95-th percentile (■) number of cases necessary to reject $H_0$ in favour of $H_1$: $|\delta| = 0.4$ compared to the theoretical expected value $(k+1)/2k$ (x), using Kummer’s function.

Appendix II shows data and calculations of one of the simulations as an example.

1.5 Discussion

Biological data banks contain valuable material that can be analysed to explore new hypotheses with possible important public health consequences. But, with most chemical analyses, these unique biological samples are destroyed and thus economical tests are preferable.11

While, in case-control studies, cases are mostly scarce, but control samples abundant, statistical efficiency of non-sequential tests can be increased by including multiple controls per case. If the power using equal allocation ($k = 1$) is greater than 0.9, this is of no practical importance. If the equal allocation power is less than 0.9, meaningful power increases may be obtained, but more than 4 controls per case are seldom worthwhile.4

Retrospective analyses as well as prospective studies justify the use of sequential investigation to avoid unnecessary destruction of the biological material and to limit the total duration of the study. In prospective clinical trials ethical aspects may play a role. For example when chemotherapy is one of the trial arms in a trial comparing two cancer therapies, one wishes to expose as few patients as necessary in coming to a decision.

From an economical point of view we performed sequential t-tests with multiple control observations per case and compared the results with those of a non-sequential analysis and of simulation studies.
Chapter 1

Table 3  Comparison of expected and observed sample size for one control matched per case ($k = 1$)

|                      | $|\delta| = 0.3$ | $|\delta| = 0.4$ | $|\delta| = 0.5$ |
|----------------------|----------------|----------------|----------------|
| **Fixed**            |                |                |                |
| Paired $t$-test      | 88             | 50             | 32             |
| **Sequential**       |                |                |                |
| Expected:            |                |                |                |
| Cox’ approximation   | 57/34          | 34/20          | 22/14          |
| Observed:            |                |                |                |
| Simulation results   |                |                |                |
| **Rushton**          |                |                |                |
| Mean                 | 57/44          | 36/27          | 25/18          |
| Median               | 50/33          | 31/20          | 21/13          |
| **Kummer**           |                |                |                |
| Mean                 | 64/54          | 39/31          | 26/21          |
| Median               | 57/43          | 35/25          | 23/17          |

Sequential sample sizes are expressed as ‘number of case-control pairs necessary to reject $H_0$/number of case-control pairs necessary to accept $H_0$’.

Expected sample size for a non-sequential paired $t$-test and expected and observed sample sizes for sequential $t$-tests with matched pairs (i.e. 1 control per case).

The expected average sample numbers (ASN) for a sequential $t$-test with one control observation per case are already smaller than the minimal sample size required for a corresponding non-sequential (=fixed sample size) paired $t$-test (Table 3). (See Appendix III for the calculation of the ASN according to Cox’ approximation.) Notable in Table 3 is the fact that both the mean number of case-control pairs required to reject $H_0$ using Rushton’s approximation and the median number using Kummer’s function almost equal Cox’ approximated ASN. Only the median number of cases necessary to accept $H_0$ using Rushton’s approximation resembles the corresponding ASN according to Cox. Our simulations indicate that Cox’ approximation probably underestimates the average sample size, especially the expected ASN needed to accept $H_0$.

Most sequential $t$-tests of our ‘selenium and breast cancer’ data (Table 2) resulted in acceptance of $H_0$ at a considerably smaller number of case-control sets than necessary for a non-sequential analysis.

The simulations confirm these results even better. The largest gain in efficiency as compared to matched pairs is reached with 2 controls per case, when $H_0$ is rejected. When $H_0$ cannot be rejected, the gain in efficiency is smaller. The simulated power values are closer to each other for different values of $\delta$ using the exact Kummer function than they are using Rushton’s approximation.
Rushton’s approximation, on the other hand, is less conservative with respect to the simulated power and thus more economical in its use of case-control sets. Only with the matched-pairs simulations Rushton’s approximation yields a simulated power significantly less than the theoretical power of 0.80. In general, the simulated unreliability using Rushton’s approximation is larger than that using Kummer’s function and more often even larger than the theoretical unreliability of 0.05.

Skovlund and Walløe\textsuperscript{13} already drew attention to the conservatism of the sequential $t$-test when applied as a two-sample sequential test. Their smallest value for $\delta$ studied was 0.5, however. Neither did they simulate with more than 1 control matched per case.

In theory it is possible that a sequential test continues infinitely. To warrant that a decision is reached, albeit ‘no decision can be made’, it is recommended to set a restriction (e.g. once or twice the fixed sample size) to the total number of cases available for the test.

Our simulations illustrate that there is hardly any effect on the simulated power and unreliability when the sequential test procedure is truncated at twice the fixed sample size.

Truncating the procedure at the fixed sample size results in a simulated power that is still too large, except for the matched-pairs situation using Rushton’s approximation where it is too small. The unreliability resulting from the simulations using Rushton’s approximation with more than one control per case is often (significantly) too large.

When a sequential test is terminated at a small number of observations, point and interval estimates of the case-control difference are rather imprecise. We hold the view that these objections play a less important role when, as in our experimental set-up, a rather ‘qualitative’ answer (‘$H_0$ can/cannot be rejected’) suffices to distinguish promising new hypotheses from unfruitful ones (see for an example Van Noord\textsuperscript{10}).

Group sequential procedures (for matched case-control sets)\textsuperscript{15-18} also have the advantage of a reduction in the average sample size as compared to fixed-sample-size plans. There are some differences between group sequential procedures and a one-at-a-time SPRT, however. A one-at-a-time sequential approach can be stopped after every new case-control set, while a group sequential procedure can only be stopped after the next planned inspection. Furthermore, a group sequential procedure cannot come to the decision to accept the null hypothesis until after the last planned inspection. A SPRT can be stopped the very moment that evidence exists that the null hypothesis cannot be rejected anymore.

Therefore, the authors prefer a one-at-a-time SPRT over the group sequential procedure when ethical and/or economical motives play a role. Promising hypotheses as well as unfruitful ones are to be distinguished with as little as possible biological material destroyed or, for that matter, time and/or money spent.

Following Skovlund and Walløe\textsuperscript{14}, we hold the view that a sequential design might be
considered more often in prospective clinical trials as well as in (cohort-nested) case-control studies.

Furthermore, we are of opinion that a sequential $t$-test with 2-4 controls per case is appropriate in case-control studies and other experimental designs where the case material must be used economically and the response is available (almost) immediately. In general the investigation can then be stopped at a lower average sample size as compared to one control per case or a non-sequential test.

The use of exact calculations (the series expansion of Kummer’s function) is recommended, although less conservative procedures are to be developed.

Tables and figures summarizing the results from the computer simulations are available from the authors by written request.

1.6 Conclusions

1) A sequential $t$-test with 2-4 controls matched per case in general leads to lower average sample sizes than a matched-pairs sequential $t$-test or a non-sequential analysis. The largest gain in efficiency as compared to matched pairs is reached with 2 controls per case.

2) Rushton’s approximation to the logarithm of the likelihood ratio is rather inaccurate and leads to a power that is significantly too small in case of a matched-pairs analysis.

3) The use of Kummer’s function (the exact calculation) results in power values which are too conservative.

4) Cox’ approximation to the expected average sample number probably underestimates the expected sample size needed to accept $H_0$.

1.7 Acknowledgement

The study was supported by the Praeventiefonds (the Netherlands), grant No 28-1560.
1.8 References

APPENDIX I

The logarithm of the likelihood ratio \( l_n \) is a function of \( \delta, n \) and \( u^2 \) and equal to

\[
L = \ln(l_n) = \ln M(n/2; \frac{1}{2}; \frac{1}{2} \cdot \delta^2 \cdot u^2) - \frac{1}{2} \cdot n \cdot \delta^2
\]  

(1)

For the \( n \)th case-control pair \((n = 1, 2, 3, \ldots \) successively and one control observation per case\) \( u^2 \) is equal to

\[
u^2 = \left( \sum d_i \right)^2 / \sum d_i^2 = n \cdot \bar{t}^2 / (n-1 + \bar{t}^2), \quad i = 1, \ldots, n
\]

where

\[
\bar{t}^2 = n \cdot \text{mean}(d)^2 / \text{var}(d),
\]

\( d_i \) is the difference between the observation for the case and the control observation, and \( \text{mean}(d) \) and \( \text{var}(d) \) stand for the mean and variance of these differences. For every \( n \) \( L \) is compared to \( \ln((1-\beta)/(1-\alpha)) \) and \( \ln((1/\beta)/\alpha) \). \( M(a; b; x) \) is the confluent hypergeometric function, which can be calculated using Kummer’s function, a series expansion:

\[
M(a; b; x) = 1 + ax/b + a(a+1)x^2/[b(b+1)2!] + a(a+1)(a+2)x^3/[b(b+1)(b+2)3!] + \ldots
\]

We involved 30 terms of this expansion. Rushton’s approximation to \( L \) is equal to

\[
l_1 = \frac{1}{2} \cdot \delta \cdot \bar{u}^3 / \sqrt{n} + \sqrt{(n \cdot \delta^2 \cdot \bar{u}^2)} - (\frac{1}{2} \cdot n \cdot \delta^2 + \ln(2)).
\]

(2)

For \( k \) control observations per case the variance of the difference between the case observation and the mean of the \( k \) control observations is estimated using the cumulating case-control variance-covariance matrix. This estimate is then substituted as \( s^2 \) in the equations mentioned below. (The variance-covariance matrix takes the correlations among cases and controls into account. If we assume negligible correlations among control observations, equal variance for the control observations and equal correlations between the case and each of the controls, \( s^2 \) can be approximated by the variance of the differences between the case and the mean of the control observations.) Then Rushton’s approximation to \( L \) can be calculated by

\[
l_1 = \frac{1}{2} \cdot \delta \cdot \bar{u}^3_k / \sqrt{n} + \sqrt{(n \cdot \delta^2 \cdot \bar{u}^2_k)} - (\frac{1}{2} \cdot n \cdot \delta^2 + \ln(2))
\]

(3)

with

\[
\bar{u}^2_k = n \cdot \bar{t}^2_k / (n-1 + \bar{t}^2_k)
\]

and

\[
\bar{t}^2_k = n \cdot \text{mean}(d)^2 / s^2
\]

N.B. For matched case-control observations \((k = 1)\) equation (3) is equal to equation (2).
APPENDIX II

Data and calculations of one of the simulations with $\alpha = 0.05$, $1-\beta = 0.80$, $\delta = 0.5$, $\mu_0 = \mu_1 = 0.8$, $\sigma = 0.15$ and 2 controls per case (see Appendix I for the notation used):

<table>
<thead>
<tr>
<th>$n$</th>
<th>Case</th>
<th>Control</th>
<th>Control</th>
<th>$s^2$</th>
<th>$t_2^*</th>
<th>u_2^*</th>
<th>M</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.911</td>
<td>0.912</td>
<td>0.891</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.919</td>
<td>1.044</td>
<td>0.518</td>
<td>0.008</td>
<td>1.312</td>
<td>1.135</td>
<td>1.312</td>
<td>0.022</td>
</tr>
<tr>
<td>3</td>
<td>0.628</td>
<td>0.867</td>
<td>1.029</td>
<td>0.056</td>
<td>0.180</td>
<td>0.248</td>
<td>1.095</td>
<td>-0.284</td>
</tr>
<tr>
<td>4</td>
<td>0.781</td>
<td>0.759</td>
<td>0.861</td>
<td>0.037</td>
<td>0.273</td>
<td>0.334</td>
<td>1.174</td>
<td>-0.340</td>
</tr>
<tr>
<td>5</td>
<td>0.947</td>
<td>0.740</td>
<td>0.600</td>
<td>0.049</td>
<td>0.022</td>
<td>0.028</td>
<td>1.018</td>
<td>-0.608</td>
</tr>
<tr>
<td>6</td>
<td>0.527</td>
<td>0.728</td>
<td>0.791</td>
<td>0.050</td>
<td>0.084</td>
<td>0.099</td>
<td>1.075</td>
<td>-0.677</td>
</tr>
<tr>
<td>7</td>
<td>0.814</td>
<td>1.053</td>
<td>0.771</td>
<td>0.042</td>
<td>0.223</td>
<td>0.250</td>
<td>1.230</td>
<td>-0.668</td>
</tr>
<tr>
<td>8</td>
<td>0.784</td>
<td>0.730</td>
<td>0.877</td>
<td>0.036</td>
<td>0.263</td>
<td>0.290</td>
<td>1.308</td>
<td>-0.731</td>
</tr>
<tr>
<td>9</td>
<td>0.908</td>
<td>0.860</td>
<td>0.826</td>
<td>0.033</td>
<td>0.151</td>
<td>0.167</td>
<td>1.195</td>
<td>-0.947</td>
</tr>
<tr>
<td>10</td>
<td>0.745</td>
<td>0.637</td>
<td>0.580</td>
<td>0.032</td>
<td>0.018</td>
<td>0.020</td>
<td>1.025</td>
<td>-1.226</td>
</tr>
<tr>
<td>11</td>
<td>0.846</td>
<td>0.659</td>
<td>0.672</td>
<td>0.032</td>
<td>0.032</td>
<td>0.035</td>
<td>1.048</td>
<td>-1.328</td>
</tr>
<tr>
<td>12</td>
<td>0.650</td>
<td>0.762</td>
<td>0.919</td>
<td>0.032</td>
<td>0.019</td>
<td>0.020</td>
<td>1.031</td>
<td>-1.470</td>
</tr>
<tr>
<td>13</td>
<td>0.898</td>
<td>0.896</td>
<td>0.778</td>
<td>0.030</td>
<td>0.001</td>
<td>0.002</td>
<td>1.002</td>
<td>-1.623</td>
</tr>
</tbody>
</table>

After 13 case-control sets are evaluated, $M$ equals 1.002 and therefore $L = -1.623$ becomes smaller than the lower boundary, $\ln(\beta/(1-\alpha)) = -1.558$, and thus $H_0$ cannot be rejected.

When Rushton’s approximation to $L$ is applied, the sequential analysis can be stopped after the 10th case-control set, where $I_1 = -1.719$. 
APPENDIX III

For matched case-control observations, the average sample number (ASN) for a sequential \( t \)-test with unknown variance is approximately \((1+\delta^2/2)\) times the ASN for a test with known variance (Cox’ approximation, Wetherill and Glazebrook\(^{12}\)). Under \( H_0 \) this ASN (unknown variance) is about

\[
-2\delta^2 \cdot \alpha' \ln((1 - \beta)/\alpha') + (1 - \alpha') \cdot \ln((\beta)/(1 - \alpha'))
\]

and under \( H_1 \) this ASN is about

\[
(1 + 2\delta^2) \cdot \left\{ \beta \cdot \ln\left(\frac{\beta}{(1 - \alpha')}\right) + (1 - \beta) \cdot \ln\left(\frac{1 - \beta}{(\alpha')}\right) \right\}
\]

(with \( \alpha' = \alpha/2 \)).

We recognize that Cox’ approximation is an asymptotic result and that it is currently unknown how accurate it is.