CHAPTER 4

Sequential analysis of matched dichotomous data from prospective case-control studies

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

*Centre for Biostatistics, Utrecht University, Utrecht, the Netherlands
#Julius Center for Patient-Oriented Research, Department of Epidemiology, Utrecht University, Utrecht, the Netherlands

Statistics in Medicine 2000; 19: 3449-3464
Summary

Sequential analysis of randomized controlled clinical trials and epidemiological prospective (matched) case-control studies can have ethical or economical advantages above a fixed sample size approach. It offers the possibility to stop early when enough evidence for an apparent effect of the risk factor or lack of the expected effect is achieved. In clinical trials it is well accepted to stop the trial early in favour of the alternative hypothesis. In epidemiological studies, in general, the need is not felt to stop early in case of a clear exposure effect. Little attention has been paid, however, to early stopping and accepting the null hypothesis. In metabolic epidemiological studies where analysis destroys the biological material, the question of efficient use of samples, for example, those stored in a biobank, becomes crucial. Also a slow accrual of cases or the costs of follow-up of a cohort nested study can make it desirable to stop a study early when it becomes clear that no relevant exposure effect will be found. Matching can further reduce the amount of information necessary to reach a conclusion. We derived test statistics $Z$ (efficient score) and $V$ (Fisher's information) for the sequential analysis of studies with dichotomous data where each case can be matched to one or more controls. A variable matching ratio is allowed. These test statistics can be entered into the software PEST to monitor the course of the study. The double sequential probability ratio test and the double triangular test were evaluated with simulated data for odds ratios equal to 1.5, 2.0 and 2.5 and various type I and type II error probabilities both under $H_0$ and under $H_1$. Our simulations resulted in average and median values for the amount of information ($V$), that are far less than those for a fixed sample size study. Efficiency gain can range from 32 per cent to 60 per cent. The proposed sequential analysis was applied in an investigation on the possible relationship between the polymorphism of the MTHFR-gene and rectal cancer in a cohort of women with cases matched by age to one and to three controls. A sequential analysis of matched data can lead to early stopping in favour of $H_0$ or $H_1$, thus conserving valuable resources for future testing. A sequentially designed study can be more economical and less arbitrary than a study that makes use of conditional power or conditional coverage probability calculations to decide early stopping.

4.1 Introduction

In randomized controlled clinical trials and epidemiological (prospective) case-control studies it can be desirable to have at one’s disposal a statistical analysis that uses the least possible number of observations to come to a decision.

In a clinical trial it may be unethical to subject more patients than necessary to a treatment that turns out to be inferior. For example, Newman et al describe how a sequential analysis showed a poorer survival under radiotherapy plus razoxane than under
radiotherapy alone in patients with inoperable lung cancer. Thus the trial could be
stopped early, saving patients from an inferior treatment. Montaner et al\textsuperscript{2} showed that
oral corticosteroids can prevent early deterioration in patients with moderately severe
AIDS-related pneumonia. After 37 patients were analysed sequentially the null
hypothesis could be rejected. This result meant that 47 per cent (that is 33 out of 70) of
the foreseen number of patients did not have to be included in the trial. Moss et al\textsuperscript{3} used a
sequential design to demonstrate the superiority of an implanted defibrillator over
conventional medical treatment of patients at high risk for ventricular arrhythmia. All
clinical trials described\textsuperscript{1-3} were analysed using a so-called triangular test\textsuperscript{4}. In an
epidemiological prospective (cohort nested) case-control study, cases are often scarce or
accrue slowly and continuation of the follow-up of the cohort is frequently costly.
Especially in the emerging field of metabolic and genetic epidemiology with biosamples
from cohorts stored in a biobank, analysis of these biosamples is frequently destructive,
unlike analysis of questionnaire data. This introduces the need to be selective with regard
to the hypotheses that can be tested using these biosamples. Contrary to the samples of
the cases, control material is mostly abundantly available in cohort nested studies. Thus a
first step in the analysis of such a study is to increase the number of controls per case. A
second option is to analyse the data sequentially\textsuperscript{4-6}.

On average, a sequential analysis requires fewer observations than a fixed sample
analysis under the same design specifications\textsuperscript{4,5}. When, in an epidemiological study, each
case can be matched to a control, and in particular to multiple controls, a sequential
analysis of the matched data requires fewer case-control sets than a sequential analysis of
unmatched data with the same design specifications.

In an earlier paper we developed and compared one-sample two-sided sequential \textit{t}-
tests for epidemiological studies with more than one control matched per case and a
normally distributed exposure variable\textsuperscript{6}. In the present paper we derive the test statistics
for a sequential test with matched dichotomous data. A fixed and a variable matching
ratio are considered. Two sequential tests are compared: the sequential probability ratio
test and the triangular test. Comparisons are made by simulations.

The proposed sequential test was applied in a genetic epidemiological study
investigating the possible relationship between the polymorism of the MTHFR-gene and
rectal cancer in a cohort of women with cases matched by age to one and to three
controls.

When a sequential test is concluded, fixed sample estimation procedures can not be
applied, because the maximum likelihood estimate of the parameter is biased. Valid
estimation procedures lead to a median unbiased parameter estimate and corresponding
confidence interval\textsuperscript{4}. 

70
4.2 Example

We were confronted with the need for efficient use of biosamples in investigating data from the DOM cohort. Participants in this cohort were 50 to 69 years old. Between 1974 and 1984, women in this population-based breast cancer screening cohort volunteered to provide overnight urine samples. For each of more than 16,532 women 100 cm³ aliquots are stored at a temperature of -20°C in a biobank. It turned out that in these stored urine samples enough cells were available in the sediments to allow for PCR DNA probe analysis. Since the DNA was fragmented, it will not be possible to amplify full DNA, which otherwise could have solved the problem of limited biological material available for analysis.

Several hypotheses with respect to the roles of genetic polymorphism have been put forward in the literature that, when tested, will compete for the limited material. Since in particular the material of cases is limited (the number of potential controls is less of a problem given the cohort size), the question of efficient management of the samples emerged. We had earlier developed some strategies for handling this problem for a normally distributed exposure variable. For a dichotomous exposure new tests had to be developed.

Colorectal cancer was among the first tumours in this cohort where the problem of efficient management became pressing, given problems in the continuation of a complete follow-up of the entire cohort. A role for the methylene-tetra-hydrofolate-reductase (MTHFR) gene was claimed in the literature. Mutations in this gene occur in up to 13 percent of the population. The wild polymorphism was contrasted to the homozygote and heterozygote polymorphisms to explore the hypothesis that being a homozygote or heterozygote carrier of the MTHFR-gene mutation (677 C > T mutation) increases the risk of developing rectal cancer. Through follow-up by the regional cancer registration (IKMN) and a mortality register established with the general practitioners in the Utrecht region, incidence of or mortality due to rectal cancer was traced prospectively. Based on the date of incidence or mortality (whatever became first known), urine samples were thawed and tested. A total number of 72 cases were reported. For three of these cases no genetic information was available. Control women were matched to each case by age. We matched one and three controls per case in different analyses to study the efficiency aspect of multiple controls per case. So for 69 cases and 207 controls urine samples were retrieved from the biobank. Urines were treated according to a fixed protocol to obtain urine sediments and PCR gene product. All data were analysed sequentially in the chronological order in which the cases were notified.
4.3 The sequential tests

4.3.1 The test statistics $Z$ and $V$

Consider an epidemiological study where each case can be matched to one or more controls and the possible relation with an exposure variable is to be tested. Let $\pi_0$ be the probability of exposure for the controls and $\pi_1$ the probability of exposure for the cases. Only discordant matched sets provide information and thus are used in the analysis. Let $\pi$ be the probability that the case is exposed in a discordant pair of uncorrelated observations with

$$
\pi = \frac{\pi_1(1 - \pi_0)}{\pi_0(1 - \pi_1) + \pi_1(1 - \pi_0)}.
$$

Testing $H_0: \pi = 0.5$ against $H_1: \pi = \pi_R$ can be reparameterized as testing $H_0: \theta = 0$ against $H_1: \theta = \theta_R$. Then $\theta = \ln\{\pi/(1-\pi)\} = \ln\{(\pi_1(1-\pi_0))/(\pi_0(1-\pi_1))\} = \ln(\psi)$, where $\psi$ stands for the odds ratio (OR).

When in the $i$th matched set or stratum a case is matched to $M_i$ controls the conditional likelihood can be used to derive the test statistics $Z_i$ and $V_i$ per matched set (see Appendix I). The odds ratio $\psi$ is assumed to be the same in all matched sets or strata.

For the $i$th set

$$
Z_i = X_{1i} - \frac{1}{M_i + 1} \sum_{m=1}^{M_i+1} X_{mi} = \frac{M_i}{M_i + 1} [X_{i,case} - \bar{X}_{i,controls}]
$$

and

$$
V_i = \frac{1}{M_i + 1} \sum_{m=1}^{M_i+1} X_{mi}^2 - \left( \frac{1}{M_i + 1} \sum_{m=1}^{M_i+1} X_{mi} \right)^2 = \frac{1}{M_i + 1} \sum_{m=1}^{M_i+1} (X_{mi} - \bar{X}_i)^2
$$

(1)

where $X_{1i} = X_{i,case}$ is the value for the case in the $i$th set and $X_{2i}, ..., X_{M_i+1,i}$ are the values for the matched controls in the $i$th set (1 when exposed, 0 when not exposed) with $\bar{X}_{i,controls}$ as the average value for the controls in the $i$th set and $\bar{X}_i$ as the overall average value of the $i$th set. The test statistics $Z$ and $V$ are equal to $\Sigma Z_i$ and $\Sigma V_i$, respectively, over the successive sets $i = 1, 2, ..., n$, where $n$ is the number of sets observed so far.

For the $i$th matched set with all $M_i$ equal to 1

$$
Z_i = \frac{1}{2} (X_{1i} - X_{2i}) \quad \text{and} \quad V_i = \frac{1}{4} (X_{1i} - X_{2i})^2
$$

(that is, for each $i$, in fact $Z_i = \pm 0.5$ and $V_i = 0.25$) and $Z$ and $V$ become

$$
Z = \frac{1}{2} \sum_{i=1}^{n} (X_{1i} - X_{2i}) \quad \text{and} \quad V = \frac{1}{4} \sum_{i=1}^{n} (X_{1i} - X_{2i})^2
$$

(2)
The test statistics in (2) can also be written as \( Z = S_n - n/2 \) and \( V = n/4 \), with \( S_n \) as the number of exposed cases and \( n \) the number of discordant matched pairs so far observed (see also equations (3.47) and (3.48) in Whitehead\(^4\) with \( p_0=0.5 \)).

In general, for \( C_i \) cases matched to \( M_i \) controls the results in the \( i \)th matched set can also be tabulated as follows:

<table>
<thead>
<tr>
<th>( i )th matched set</th>
<th>cases</th>
<th>controls</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of persons</td>
<td>( C_i )</td>
<td>( M_i )</td>
<td>( N_i )</td>
</tr>
<tr>
<td>Number exposed</td>
<td>( S_i )</td>
<td>( T_i )</td>
<td>( E_i )</td>
</tr>
<tr>
<td>Number not exposed</td>
<td>( F_i )</td>
<td>( G_i )</td>
<td>( \bar{E}_i )</td>
</tr>
</tbody>
</table>

where \( S_i \) stands for the number of exposed cases, \( T_i \) for the number of exposed controls, \( E_i \) for the total number of exposed persons and \( \bar{E}_i \) for the total number of unexposed persons. The test statistic \( Z_i \) can be expressed as the difference between the observed number of exposed cases and the expected number under the null hypothesis, and the statistic \( V_i \) as the variance under the null hypothesis. Using this tabular notation,

\[
Z_i = S_i - \frac{E_i C_i}{N_i} = \frac{S_i M_i - T_i C_i}{M_i + C_i} \quad \text{and} \quad V_i = \frac{C_i M_i \bar{E}_i}{N_i^2 (N_i - 1)}
\]

For \( C_i = 1 \), these equations become

\[
Z_i = \frac{M_i}{M_i + 1} \left[ S_i - \frac{T_i}{M_i} \right] \quad \text{and} \quad V_i = \frac{E_i \bar{E}_i}{N_i^3}
\]

The above equations (3) for \( Z_i \) and \( V_i \) are equal to equations (1). (The variance formula as given here can be compared to formula (3.5) in Whitehead\(^4\). Our denominator is equal to \( N_i^2 (N_i - 1) \) due to the use of the conditional likelihood in our derivation, while that in formula (3.5) is \( N^3 \).)

The test statistic \( Z \) is the efficient score for \( \theta \) under the null hypothesis \( H_0: \theta = 0 \); \( V \) stands for Fisher's information about \( \theta \) contained in \( Z \). \( Z \) follows approximately a Normal distribution with expectation \( \theta V \) and variance \( V \) when samples are large and \( \theta \) is small\(^4,9\).

For 1:1 matching \( Z \) and \( V \) are equal to the numerator and the square root of the denominator, respectively, of McNemar's (fixed sample size) test for matched pairs without continuity correction. The test statistics \( Z \) and \( V \) for a sequential test with 1:M (\( M_i = M \) for all \( i \)) matching are equal to the numerator and the square root of the denominator, respectively, of the test statistic proposed by Miettinen\(^10\). They are related in the same way to the Mantel-Haenszel test statistic for matched data\(^11\) (without continuity correction) and to the logrank test statistic\(^12\).
When $M_i$ is equal to $M$ for all $i$, the ratio of the expectation of $V_i$ under the null hypothesis for $M = 2$ to the expectation of $V_i$ under the null hypothesis for $M = 1$ is equal to 1.33. Likewise, the ratio of the expectation of $V_i$ under the null hypothesis for $M = 3$ to the expectation of $V_i$ under the null hypothesis for $M = 1$ is equal to 1.50 (see Appendix II). This means that the amount of information for two controls matched per case is 1.33 times the amount of information obtained with matched pairs. For three controls matched per case, 1.50 times the amount of information contained in matched pairs is obtained.

In epidemiological studies a variable matching ratio can occur when not enough controls can be found to match to the same case, or when biological or genetic material stored in a biobank is either limited or not available (any more). Suppose a control is missing with probability $\pi_m$ (i.e. the number of controls matched to a case is now variable for the $i$th set). Under the null hypothesis and, for example, $\pi_m = 0.25$ the ratio of $E(V_i | M_i = 2; \pi_m)$ to $E(V_i | M = 1)$ is equal to 1.125, and the ratio of $E(V_i | M_i = 3; \pi_m)$ to $E(V_i | M = 1)$ is equal to 1.336 (see Appendix III). (When $M = 1$ and the control is missing, the case-control set is uninformative; likewise when the case is missing).

4.3.2 Two sequential tests

We compare the behaviour of the test statistics $Z$ and $V$ in two types of sequential tests: the sequential probability ratio test (SPRT) and the triangular test (TT). Both tests require critical boundaries to be specified in advance. Characteristics of both tests are described at length by Whitehead. For an open double SPRT the critical boundaries are as follows

$$Z = a + bV \quad (u_1)$$
$$Z = -a + bV \quad (u_2)$$
$$Z = a - bV \quad (l_2)$$
$$Z = -a - bV \quad (l_1)$$

The slopes $\pm b$ and intercepts $\pm a$ of these boundaries are functions of the error probabilities $2\alpha$ (the two-sided type I error) and $\beta$ (the type II error), as well as of the choice of a minimal relevant standardized exposure difference, $\theta_R$, that defines the (two-sided) alternative hypothesis $H_1: |\theta| = \theta_R$:

$$b = \frac{\theta_R}{\ln\left(\frac{1-\beta}{\beta}\right) + \ln\left(\frac{1-\alpha}{\alpha}\right)} \ln\left(\frac{1-\alpha}{\alpha}\right)$$
and
$$a = \frac{\ln\left(\frac{1-\beta}{\beta}\right) + \ln\left(\frac{1-\alpha}{\alpha}\right)}{2\theta_R}$$

For a double TT the critical boundaries are as follows
Sequential analysis of matched dichotomous data

\[ Z = a + cV \quad (u_1) \]
\[ Z = -a + 3cV \quad (u_2) \]
\[ Z = a - 3cV \quad (l_2) \]
\[ Z = -a - cV \quad (l_1). \]

The boundaries \( u_1 \) and \( u_2 \), and \( l_1 \) and \( l_2 \), cross at \( V_{\text{max}} = a/c \) and \( Z = \pm 2a \).

For \( \alpha \neq \beta \) no simple expressions can be given for \( a \) and \( c \).

Throughout this paper, it is assumed that the boundaries of the tests are placed symmetrically with respect to the horizontal axis. Both sequential tests are conveniently presented in the form of a graph, plotting \( Z \) against \( V \) (see Figures 1(a) and (b) for illustration). The testing process continues as long as the sample path formed by successive \( Z \)-values plotted versus corresponding \( V \)-values remains between the boundaries \( u_1 \) and \( u_2 \) or between \( l_1 \) and \( l_2 \). The sampling is stopped and \( H_0 \) rejected when the sample path crosses \( u_1 \) or \( l_1 \). The test is stopped with acceptance of \( H_0 \) when the sample path crosses \( u_2 \) or \( l_2 \). The test is also stopped with the acceptance of \( H_0 \) when the sample path has crossed the first parts of both \( u_2 \) and \( l_2 \). To adjust for the discrete monitoring of a strictly spoken continuous process, Whitehead recommends to replace the intercept \( a \) by \( a - 0.583\sqrt{V_i - V_{i-1}} \) for \( i = 1,2,... \) with \( V_0 = 0 \) (this adjustment is called the ‘Christmas tree’ correction).

Both tests are implemented in the computer program PEST\(^{13} \). Although PEST does not contain features for the design and analysis of matched data, it can be used by choosing the ‘DEFAULT-response’ option and entering precalculated values for \( Z \) and \( V \).

Figure 1  Double SPRT and double TT with \( \theta_R = \ln(2.0) \), \( 2\alpha = 0.05 \) and \( \beta = 0.20 \). The vertical dashed line denotes the fixed sample value for \( V \). (a) Double SPRT, values for the intercept \( a \) are \( \pm 3.643 \) and for the slope \( b \) are \( \pm 0.503 \). (b) Double TT, values for the intercept \( a \) are \( \pm 6.149 \), the slope of the boundaries \( u_1 \) and \( l_1 \) is \( \pm 0.244 \) and the slope of the boundaries \( u_2 \) and \( l_2 \) is \( \pm 0.731 \).
4.3.3 Simulations
The double SPRT and the double TT were evaluated with simulated data both under $H_0$: $\theta = 0$ and under $H_1$: $|\theta| = \theta_R$ with $\theta_R = \ln(\psi)$ for $\psi = 1.5$, $2.0$ and $2.5$ with $2\alpha = 0.10$, $\beta = 0.05$ and with $2\alpha = 0.05$, $\beta = 0.20$. Each evaluation consisted of 2500 simulation runs with $M = 1$, $2$ and $3$ controls matched per case. To evaluate the effect of a variable number $M_i$ of controls matched per case, data were simulated and analyzed by the double SPRT for $\psi = 1.5$ with $2\alpha = 0.10$, $\beta = 0.05$ and with $2\alpha = 0.05$, $\beta = 0.20$. For $M_i = 2$ and $M_i = 3$ the probability of a missing control ($\pi_m$) was chosen equal to 0.25. All evaluations of both the double SPRT and the double TT were performed with the so-called ‘Christmas tree’ correction of the boundaries.

4.3.4 Estimation of sample size
For a fixed sample design with matched case-control pairs the sample size can be estimated using the relation $V_{\text{fixed}} = (u_x + u_{\beta})^2 / \theta_R^2$ (Whitehead\textsuperscript{4}), where $u_x$ denotes the standardized normal deviate exceeded with probability $x$. The total sample size can be estimated by dividing the number of discordant pairs $n = V_{\text{fixed}} \times 4$ (see equation (2)) by the probability of a discordant pair of uncorrelated observations, $\pi_{\text{disc}} = \pi_0(1 - \pi_1) + \pi_1(1 - \pi_0)$. For our simulated data evaluated sequentially $\pi_{\text{disc}}$ can be approximated by

$$p_{\text{disc}} = \frac{4}{9} \times \frac{\psi^2 + \psi + 1}{(\psi + 1)^2}$$

(see Appendix IV). This probability is about 1/3 for an odds ratio $\psi$ in the range 1.0 to 2.5.

In practice, for a sequential study the expected average or median value for $V$ (as reported by the program PEST when designing the study) multiplied by 4 and divided by the probability of a discordant pair of observations estimates the average or median total number of case-control sets necessary for a study with 1 : 1 matching. For 1 : $M$ matching this number can be reduced by a factor $(M + 1) / 2M$ (Ury\textsuperscript{14}). (Note the reciprocal relation with the expectation of $V_i$ for $M = 2$ and $M = 3$.)

4.4 Results of the example
Homozygote or heterozygote carriers of the MTHFR-gene mutation were expected to have at least a twofold risk of developing rectal cancer compared to the carriers of the wild polymorphism. A double SPRT was designed with an OR equal to 2 as the alternative hypothesis ($\theta_R = \ln(2) = 0.69315$), a two-sided type I error $2\alpha = 0.05$ and a power $1-\beta = 0.80$. The expected average value for $V$ under $H_0$ is equal to 10.1; the expected median value for $V$ is equal to 8.8. The matched-pairs analysis ended without a decision after 69 matched pairs and 32 discordant pairs ($Z = -2.0$ and $V = 8.0$) (see Figure
2(a)). For three controls per case the sequential analysis could be terminated with accepting the null hypothesis after 41 matched sets and 35 discordant sets ($Z = -0.25$ and $V = 7.3$) (see Figure 2(b)). After stopping the sequential test, a median unbiased estimate\(^4\) can be given for the odds ratio: $\text{OR} = 0.87$ with its 95 per cent confidence interval $(0.39; 1.85)$.

The number of discordant pairs necessary for a fixed sample design with the design specifications as above is at least 66 ($V_{\text{fixed}} = 16.34$; $n \geq 16.34 \times 4$). Thus a saving of 47 per cent was reached in this study by a sequential analysis with three controls matched per case compared to a fixed sample design.

![Figure 2](image)

**Figure 2** Sample path of successive ($Z; V$)-values for (a) 1:1 matching, and for (b) 1:3 matching in a double SPRT with $\theta_R = \ln(2.0)$, $2\alpha = 0.05$ and $\beta = 0.20$. Data come from a cohort-nested case-control study relating exposure to the MTHFR-gene to the occurrence of rectal cancer in women (see text). The ‘curved’ lines indicate the ‘Christmas tree’ correction (see text and ref. 4).

### 4.5 Results of simulations

#### 4.5.1 Type I and type II errors

For all evaluated OR's the resulting type I errors were not significantly different from their theoretical values. Simulations with a theoretical $\beta$ equal to 0.05 resulted in acceptable type II errors. Only for $\psi \geq 2.0$ both the SPRT and the TT resulted in somewhat larger type II errors than the theoretical $\beta = 0.20$ for $M = 1$ or $M = 2$. (See Figures 3 and 4)
4.5.2 Sample size with a fixed number of controls per set

In Tables 1 and 2 observed and expected average and median values for $V$ and the observed total number of (concordant and discordant) case-control sets ($N$) are given. Values tabulated are for $1:1$ matching. Comparing the observed median values for $V$ to the expected fixed sample size value ($V_{\text{fixed}}$) shows that savings in the amount of information used by a sequential study can range from 32 per cent to 46 per cent when the null hypothesis is true and from 36 per cent to 60 per cent when the alternative hypothesis is true, irrespective of the value for the odds ratio.
For 1 : $M$ matching the expected reduction in total sample size $N$ by a factor $(M+1)/2M$ was found for most simulations. Small deviations were seen when the type II error was larger than $\beta = 0.20$.

4.5.3 SPRT with a variable number of controls per set

The effect of a variable number of controls per case on the results of the double SPRT is shown in Table 3. To achieve the same amount of 'information' with a variable number of controls per case as with a fixed number more case-control sets are needed. The type I and type II errors were not significantly different from their theoretical values.

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$\mathrm{H}_0$</th>
<th>$\mathrm{H}_1$</th>
<th>$V_{\text{fixed}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V$</td>
<td>$N$</td>
<td>$V$</td>
</tr>
<tr>
<td></td>
<td>average</td>
<td>median</td>
<td>average</td>
</tr>
<tr>
<td>1.5</td>
<td>49.55</td>
<td>42.75</td>
<td>594.4</td>
</tr>
<tr>
<td></td>
<td>47.82</td>
<td>41.59</td>
<td>32.57</td>
</tr>
<tr>
<td>2.0</td>
<td>16.71</td>
<td>14.50</td>
<td>200.6</td>
</tr>
<tr>
<td></td>
<td>16.36</td>
<td>14.23</td>
<td>11.14</td>
</tr>
<tr>
<td>2.5</td>
<td>9.57</td>
<td>8.25</td>
<td>9.36</td>
</tr>
<tr>
<td></td>
<td>6.38</td>
<td>5.18</td>
<td>12.89</td>
</tr>
</tbody>
</table>

(b) TT

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$\mathrm{H}_0$</th>
<th>$\mathrm{H}_1$</th>
<th>$V_{\text{fixed}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V$</td>
<td>$N$</td>
<td>$V$</td>
</tr>
<tr>
<td></td>
<td>average</td>
<td>median</td>
<td>average</td>
</tr>
<tr>
<td>1.5</td>
<td>47.29</td>
<td>44.50</td>
<td>568.3</td>
</tr>
<tr>
<td></td>
<td>47.50</td>
<td>44.69</td>
<td>35.33</td>
</tr>
<tr>
<td>2.0</td>
<td>16.30</td>
<td>15.25</td>
<td>195.7</td>
</tr>
<tr>
<td></td>
<td>16.25</td>
<td>15.29</td>
<td>12.09</td>
</tr>
<tr>
<td>2.5</td>
<td>9.11</td>
<td>8.50</td>
<td>109.6</td>
</tr>
<tr>
<td></td>
<td>6.92</td>
<td>6.37</td>
<td>12.89</td>
</tr>
</tbody>
</table>
Table 2  Observed and expected (italic) values for the amount of information ($V$) and the observed total number of case-control sets ($N$) for 1:1 matching, $\psi$ is the odds ratio, $2\alpha = 0.05$, $\beta = 0.20$ ($V_{\text{fixed}}$ is the expected fixed sample size value for $V$)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$H_0$</th>
<th>$H_1$</th>
<th>$V_{\text{fixed}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V$</td>
<td>$N$</td>
<td>$V$</td>
</tr>
<tr>
<td></td>
<td>average</td>
<td>median</td>
<td>average</td>
</tr>
<tr>
<td>(a) SPRT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>30.11</td>
<td>26.25</td>
<td>361.3</td>
</tr>
<tr>
<td></td>
<td>29.53</td>
<td>25.73</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>10.16</td>
<td>9.00</td>
<td>122.5</td>
</tr>
<tr>
<td></td>
<td>10.10</td>
<td>8.80</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>5.90</td>
<td>5.25</td>
<td>70.7</td>
</tr>
<tr>
<td></td>
<td>5.78</td>
<td>5.04</td>
<td></td>
</tr>
<tr>
<td>(b) TT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>31.76</td>
<td>30.00</td>
<td>381.5</td>
</tr>
<tr>
<td></td>
<td>31.81</td>
<td>29.79</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>10.95</td>
<td>10.25</td>
<td>132.0</td>
</tr>
<tr>
<td></td>
<td>10.89</td>
<td>10.19</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>6.27</td>
<td>6.00</td>
<td>74.8</td>
</tr>
<tr>
<td></td>
<td>6.23</td>
<td>5.83</td>
<td></td>
</tr>
</tbody>
</table>

Table 3  The ratio of the average and the median number of case-control sets for a variable number of controls ($M_i$) (with probability of a missing control $\pi_m = 0.25$) to those for a fixed number of controls for the double SPRT

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>$H_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_i = 2$</td>
<td>$M_i = 3$</td>
</tr>
<tr>
<td>average</td>
<td>median</td>
</tr>
<tr>
<td>(a) $\psi = 1.5$, $2\alpha = 0.10$, $\beta = 0.05$</td>
<td></td>
</tr>
<tr>
<td>1.168</td>
<td>1.171</td>
</tr>
<tr>
<td>1.113</td>
<td>1.125</td>
</tr>
<tr>
<td>(b) $\psi = 1.5$, $2\alpha = 0.05$, $\beta = 0.20$</td>
<td></td>
</tr>
<tr>
<td>1.173</td>
<td>1.174</td>
</tr>
<tr>
<td>1.126</td>
<td>1.128</td>
</tr>
</tbody>
</table>

4.6 Discussion

Sequential analysis of epidemiological studies can have advantages above a fixed sample size approach. It offers the possibility to stop early when enough evidence for an apparent effect of the risk factor or lack of the expected effect is achieved. Control for possible confounders or a more precise effect estimate could be an argument for not stopping early to reject the null hypothesis. On the other hand, studies on the relevance of certain
gene mutations for the development of tumours focus primarily on the exploration of hypotheses, using a limited number of biological samples, as proposed in this paper. Such studies are meant to build evidence and create prior information for further affirmative studies. Once a role for a gene mutation is detected in a sequential analysis a more elaborate (sequential) design including possible effect modifiers and/or confounders can be warranted. However, a possible lack of an exposure effect certainly is an argument in favour of early stopping and accepting the null hypothesis. Destructive or expensive laboratory tests can demand a minimization of the number of those tests performed. Strömberg\textsuperscript{15}, for example, calls for development of methods for early stopping of inconclusive epidemiological studies and further discussion.

In epidemiological studies, and especially in cohort nested studies, cases are often scarce while controls are abundant. Then, matching of controls and cases can increase the efficiency of an epidemiological study and multiple controls per case can improve the power of the study (Ury\textsuperscript{14}).

In clinical trials a matched data structure can arise when patients are subjected to two successive treatments in a random sequence (the so-called cross-over trials). A similar situation arises when two drugs or treatments are applied to different sides of the mouth (Fertig \textit{et al}\textsuperscript{16}) or two equivalent parts of the body. Most clinical cross-over trials have to do with the within-subject comparison of two treatments and thus matched pairs of observations.

In this paper, the test statistics $Z$ and $V$ were derived for the sequential analysis of studies with a dichotomous outcome where each case can be matched to one or more than one control. Using the computer program PEST, designed to perform sequential analysis, the test statistics can be entered into this computer program to monitor the course of the study. (It is thus possible to analyse stratified data, like the case-control study described by Strömberg\textsuperscript{15} (four strata) and matched case-control studies (many strata), sequentially using PEST\textsuperscript{4,13} by calculating the test statistics for each stratum and combining them.)

Our simulations resulted in acceptable type I errors for the evaluated OR's. The resulting type II errors were slightly larger than their theoretical values, especially for an OR = 2.5. Bellissant \textit{et al}\textsuperscript{17} performed some simulations to evaluate the small sample properties of an open SPRT and a TT applied to non-comparative phase II clinical trials. These non-comparative studies can be compared statistically to the results of our matched pair simulations. They performed simulations only with $\beta = 0.05$. Although they do not mention this explicitly, their simulations show higher type II errors for large $\theta_R$ ($= \ln(3.86) = 1.35$) especially with the TT.

The test statistics $Z$ and $V$ are derived assuming that the likelihood for the parameter $\theta$ resembles the normal likelihood. Sprott\textsuperscript{18} showed that among various reparametrizations of the binomial distribution the transformation
\eta(\pi) = \int_0^\pi \frac{dt}{[t(1-t)]^{2/3}}

completely removes the component of asymmetry. Therefore Whitehead\(^4\) suggests the use of \(\theta' = 4^{1/3}[\eta(\pi) - \eta(0.5)]\) instead of \(\theta\) for a sequential test with dichotomous data. The use of \(\theta'\) leads to the same test statistics under \(H_0\) as the use of \(\theta\) (equations (1), (2) and (3)). Only the critical boundaries of the sequential tests are adjusted. In theory, the use of \(\theta'\) instead of \(\theta\) should lead to type I and type II errors that are closer to their theoretical values \(2\alpha\) and \(\beta\). In fact, our simulations showed only negligible differences.

The statistic \(V\) is based on the ‘observed’ information about \(\theta\). For a dichotomous response the ‘observed’ and ‘expected’ information are equal. We considered the use of Cox’s test as a sequential analogue of Wald’s \(W_e\) test\(^19\). This test has the disadvantage that for every new case-control set not only \(Z\) and \(V\) but also the ML-estimate for \(\theta\) has to be calculated. For a sequential test, the ML-estimate is biased\(^4\). Although Cox’s test is asymptotically equivalent to the sequential tests described in this paper, some simulations for small samples (i.e. large \(\psi\), \(\psi=2.5\)) show type II errors larger than their theoretical values, but in addition far too small type I errors.

Two types of sequential tests were compared: an open sequential probability ratio test (SPRT) and a triangular test (TT). The TT is, by the shape of its boundaries, a ‘closed’ test. Theoretically, the SPRT can carry on infinitely without reaching a decision. The PEST-program also provides a possibility for a closed SPRT, the so-called ‘truncated SPRT’. The number of observations is then set to a limit to prevent this carrying on infinitely. Slopes and intercepts of the critical boundaries are adjusted to maintain the same significance level and power. In general, an SPRT is more efficient, in terms of the amount of information used, than a TT when the true parameter value \(\theta\) is equal to 0 (the null hypothesis) or when \(|\theta| \geq \theta_R\) (the alternative hypothesis). This is confirmed by our simulations. Median numbers of case-control sets were smaller for the double SPRT than for the double TT both under the null hypothesis and under the alternative hypothesis, when the same parameterization and type I and type II errors were applied.

Pasternack and Shore\(^20\) mentioned the possibility of applying a group sequential test to matched data from case-control studies. Laboratory or logistic conditions can require analysis of the data in groups or batches. Group sequential designs as well as designs that use \(\alpha\)-spending functions for repeated testing can be easily implemented in PEST\(^4,13\).

A number of (recent) publications have paid attention to the early stopping of a trial or study due to lack of treatment difference or exposure effect (Lan and Witte\(s\)^{21}; Hunsberger \textit{et al}^{22}; Betensky\(^{23}\); Strömberg\(^{15}\); Ware \textit{et al}^{24}). Lan and Witte\(s\)^{21} and Hunsberger \textit{et al}^{22} provide conditional power (CP) calculations for studies that were not sequential by design. Betensky\(^{23}\) developed lower boundaries for CP calculations in repeated significance and O’Brien-Fleming (group-sequential) designs. Strömberg\(^{15}\)
Sequential analysis of matched dichotomous data

considers the conditional coverage probability (CCP) instead of CP. (CP is the probability of rejecting the null hypothesis at the end of the study (when only part of the total amount of planned information is observed), conditional on the observed data and assuming an alternative hypothesis for the remainder of the study. CCP is the probability that a two-sided confidence interval around the final estimate given the observed data and an assumed alternative hypothesis for the remainder of the study, includes the parameter value under the null hypothesis). Betensky\textsuperscript{23} and Strömberg\textsuperscript{15} both observe that less attention is paid to designs for early acceptance of $H_0$ than to designs that allow early termination due to a clearly apparent result. Strömberg even calls for ‘further discussion concerning early stopping of epidemiologic studies’. We agree with both authors that early stopping in favour of $H_0$ can conserve valuable (time, financial, genetic, biological) resources for future testing (see for example\textsuperscript{6,8,25,26}). We emphasize, however, that CP and CCP calculations require extrapolation of the observed data and are based on rather arbitrary assumptions\textsuperscript{27}. We take the view that a (group-)sequential analysis can be more economical and more objective than an analysis based on CP or CCP calculations.

An alternative sequential approach that focuses on effect estimation more than on hypothesis testing is the repeated confidence interval (RCI) approach, as described by Jennison and Turnbull\textsuperscript{28}. RCIs can be constructed by inverting a group sequential test. Group sequential tests perform a number of interim analyses on accumulating data. Each interim analysis is performed using a nominal significance level that is smaller than the desired overall type I error $\alpha$ to guarantee that the overall significance level is not inflated by the multiple testing procedure. The critical values for such an interim analysis are used to construct the corresponding RCI for that interim inspection.

In general, we think that a clear distinction must be made between hypothesis testing and effect estimation. Repeated testing of the null hypothesis can lead to early stopping of a study either because there is evidence for a significant effect and the null hypothesis is rejected or because there is no such evidence given the data thus far and the null hypothesis can be accepted. Advantages of early stopping with acceptance of the null hypothesis lie in making the most efficient use of finite resources. Effect estimation can be carried out repeatedly during a sequential study or once at termination of the study when a final estimate with its confidence interval is desired. For example, in an epidemiological study monitoring the effects of environmental hazards, early stopping is mostly not relevant, but RCIs can be useful to contemplate the effect size in accumulated data. At termination of the study point and interval estimates should be adjusted according to the sequential stopping rule. Jennison and Turnbull do not indicate any adjustments, however. A limitation of the RCI approach is the lack of a power aspect when RCIs are used to stop a study early. The power guarantee is an essential part of the sequential design as suggested by Whitehead, which makes this design more appropriate for hypothesis testing.
In case of early stopping and accepting the null hypothesis emphasis lies more on hypothesis testing than on effect estimation. Nevertheless, a median unbiased estimate for the parameter $\theta$, and thus for the odds ratio, and a confidence interval can be obtained\textsuperscript{4,13}.

Average or median sample size for a study with 1:1 matching can be estimated from the average or median value for $V$ multiplied by 4 and divided by the probability of a discordant pair or set ($p_{\text{disc}}$). As $p_{\text{disc}}$ is often unknown, its expected value (see Appendix IV) can be used to get an approximate idea of the total number of matched sets necessary.

Our simulations confirm that a sequential analysis requires on average less case-control sets than a fixed sample analysis. For 1:1 matching savings in the amount of information used can range from 32 per cent to 60 per cent. Where efficiency is the aim, optimizing the case-control ratio in epidemiological studies is, in our opinion, an additional essential strategy whether a classical case-control or a cohort-nested study is concerned, since in epidemiological studies the collection or accrual of cases in a cohort or population may not be as simple or inexpensive as finding controls. When more than one control can be matched to each case, the amount of information per case-control set becomes larger. Also the probability of a discordant, informative set is larger. A variable matching ratio due to missing control information can easily be dealt with.

We further suggest that authors claiming efficiency gain give additional estimations of such gain (for example in per cent) to allow comparisons of different strategies for early stopping. We hope that this paper and further discussion may stimulate other investigators to consider these sequential designs as strategies for ethical or efficiency aspects in epidemiological studies, in order to pursue only the most promising hypotheses.

4.7 Acknowledgements

The study was supported by the Praeventiefonds (the Netherlands), grant nr 28-1560. We thank Charles Gimbrère, head of the Comprehensive Cancer Registration (IKMN), for providing the information about cases and controls, Olga van der Hel for laboratory examinations on the MTHFR-gene, Jan van den Broek for advice and stimulating discussions and Maria Schipper for carefully reading and commenting on the manuscript. Last, but certainly not least, we thank both referees for their valuable and helpful comments.
4.8 References


APPENDIX I

In case of 1 : $M_i$ matching the conditional likelihood for the $i$th matched set or stratum with $M_i$ controls matched to each case is equal to

$$L_i(\theta) = \frac{e^{\theta X_{1i}}}{e^{\theta X_{1i}} + \sum_{m=2}^{M_i+1} e^{\theta X_{mi}}}$$

The logarithm of $L_i(\theta)$ is equal to.

$$l_i(\theta) = \theta X_{1i} - \ln \left( \sum_{m=1}^{M_i+1} e^{\theta X_{mi}} \right)$$

Taking the first derivative with respect to $\theta$ and substituting $\theta = 0$ leads to

$$Z_i = X_{1i} - \frac{1}{M_i + 1} \sum_{m=1}^{M_i+1} X_{mi}$$

The negative of the second derivative with respect to $\theta$ and substitution of $\theta = 0$ leads to

$$V_i = \frac{1}{M_i + 1} \sum_{m=1}^{M_i+1} X_{mi}^2 - \left( \frac{1}{M_i + 1} \sum_{m=1}^{M_i+1} X_{mi} \right)^2$$
APPENDIX II

Let $\eta$ be the probability of exposure under the null hypothesis with $\eta = \pi_1 = \pi_0$. For $C_i$ cases matched to $M_i$ controls the results in the $i$th matched set can be tabulated as follows:

<table>
<thead>
<tr>
<th>$i$th matched set</th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of persons</td>
<td>$C_i$</td>
<td>$M_i$</td>
<td>$N_i$</td>
</tr>
<tr>
<td>Number exposed</td>
<td>$S_i$</td>
<td>$T_i$</td>
<td>$E_i$</td>
</tr>
<tr>
<td>Number not exposed</td>
<td>$F_i$</td>
<td>$G_i$</td>
<td>$E_i$</td>
</tr>
</tbody>
</table>

For the $i$th matched set, $V_i$ is equal to $\frac{C_i M_i E_i}{N_i^2 (N_i - 1)}$.

Because $E_i$ can be assumed to follow a binomial distribution with parameters $N_i$ and $\eta$, the expectation of $V_i$ is equal to

$$E(V_i) = \frac{C_i M_i}{N_i^2 (N_i - 1)} E(E_i) = \frac{C_i M_i}{N_i^2 (N_i - 1)} N_i (N_i - 1) \eta (1 - \eta) = \frac{C_i M_i}{N_i} \eta (1 - \eta)$$

For $C_i = 1$ and for $M_i = 1$, the expectation of $V_i$ is equal to $E(V_i) = \eta (1 - \eta)/2$, for $M_i = 2$ the expectation of $V_i$ is equal to $E(V_i) = 2\eta (1 - \eta)/3$, and for $M_i = 3$ the expectation of $V_i$ is equal to $E(V_i) = 3\eta (1 - \eta)/4$. 
APPENDIX III

Suppose, the probability of a missing control in a matched set is equal to $\pi_m$. For 1 case and $M_i$ controls in the $i$th matched set, the expectation of $V_i$ given $M_i$ becomes

$$E(V_i | M_i) = \sum_{k=1}^{M_i} \binom{M_i}{k} (1 - \pi_m)^k \pi_m^{M_i-j} E(V_i | k \text{ controls present in the } i\text{th set})$$

$E(V_i | k \text{ controls present in the } i\text{th set})$ is given in appendix II with $k = M_i$, $C_i = 1$ and $N_i = k+1$. 
APPENDIX IV

The probability of a discordant pair of uncorrelated observations \( \pi_{\text{disc}} \) is equal to \( \pi_0(1-\pi_1) + \pi_1(1-\pi_0) \). The probability \( \pi_1 \) can be substituted by a function of \( \pi_0 \) and \( \psi \). The probability \( \pi_0 \) is replaced in our simulations by \( U \), a random variable with a uniform distribution on \([0,1]\) and thus with expectation \( \mathcal{E}(U) = 1/2 \), variance \( \text{var}(U) = 1/12 \) and \( \mathcal{E}(U^2) = 1/3 \).

The probability of a discordant pair of uncorrelated observations \( \pi_{\text{disc}} \) can then be rewritten as

\[
\frac{\psi U(1-U) + U(1-U)}{\psi U + (1-U)} = (\psi + 1) \frac{U(1-U)}{\psi U + (1-U)} = (\psi + 1) \frac{\text{NUM}}{\text{DEN}}
\]

The expectation of \( \pi_{\text{disc}} \) is equal to \((\psi+1) \mathcal{E}(\text{NUM}/\text{DEN})\). The expected value of \( \text{NUM}/\text{DEN} \) can be approximated by\(^\text{30}\)

\[
\frac{\mathcal{E}(\text{NUM})}{\mathcal{E}(\text{DEN})} - \frac{\text{cov}(\text{NUM,DEN})}{(\mathcal{E}(\text{DEN}))^2} + \frac{\mathcal{E}(\text{NUM})\text{var}(\text{DEN})}{(\mathcal{E}(\text{DEN}))^3}
\]

with \( \mathcal{E}(\text{NUM}) = 1/6, \mathcal{E}(\text{DEN}) = (\psi+1)/2, \text{var}(\text{DEN}) = (\psi-1)^2/12 \) and

\[
\text{cov}(\text{NUM,DEN}) = \mathcal{E}\{U(1-U)[\psi U + (1-U)]\} - \mathcal{E}(\text{NUM}) \mathcal{E}(\text{DEN})
\]
\[
= \mathcal{E}\{\psi U^2(1-U) + U(1-U)^2\} - (\psi+1)/12
\]
\[
= \psi \mathcal{E}(U^2-U^3) + \mathcal{E}(U-2U^2+U^3) - (\psi+1)/12
\]
\[
= \psi(1/3-1/4) + (1/2-2/3+1/4) - (\psi+1)/12 = 0
\]

and thus becomes

\[
\frac{4}{9} \times \frac{\psi^2 + \psi + 1}{(\psi + 1)^3}.
\]

Then the expected value of \( \pi_{\text{disc}} \) becomes

\[
\frac{4}{9} \times \frac{\psi^2 + \psi + 1}{(\psi + 1)^3}.
\]