

Sequential tests for gene–environment interactions in matched case–control studies

Ingeborg van der Tweel^{*,†} and Maria Schipper

Centre for Biostatistics, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands

SUMMARY

The sample size necessary to detect a significant gene \times environment interaction in an observational study can be large. For reasons of cost-effectiveness and efficient use of available biological samples we investigated the properties of sequential designs in matched case–control studies to test for both non-hierarchical and hierarchical interactions. We derived the test statistics Z and V and their characteristics when applied in a two-sided triangular test. Results of simulations show good agreement with theoretical values for V and the type I error. Power values were larger than their theoretical values for very large sample sizes. Median gain in efficiency was about 27 per cent. For a ‘rare’ phenotype gain in efficiency was larger when the alternative hypothesis was true than under the null hypothesis. Sequential designs lead to substantial efficiency gains in tests for interaction in matched case–control studies. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: sequential tests; gene–environment interactions; matching; case–control studies; sample size

1. INTRODUCTION

To study the association between a disease, for example cancer, and a genetic risk factor or an environmental factor case–control studies can be designed and analysed. To adjust for possible confounding factors a case can be matched to one or more controls by, for example, ethnicity or age. Both genetic and environmental factors may contribute to the susceptibility of disease and these factors may interact in their influence on the risk of disease [1]. Then a gene–environment interaction can be interesting to explore. Interactions can be tested both hierarchically by first including the main effects (for example G and E for a genetic and an environmental factor) in the model and subsequently testing for the interaction between these main effects ($G \times E$), and non-hierarchically, thus modelling the interaction but not the associated main effects. A non-hierarchical model can be of interest when both the genetic type

*Correspondence to: Ingeborg van der Tweel, Centre for Biostatistics, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands.

†E-mail: i.vandertweel@bio.uu.nl

Received October 2003

Accepted June 2004

and the exposure are required to increase the risk of disease [1, 2]. This kind of association is discussed as model D [1].

Sample sizes necessary to detect significant $G \times E$ interactions can be large when the occurrence of the genetic factor is rare [3]. Furthermore, when an increased risk does exist for a genetic factor, the interaction is likely to be very moderate [4] and thus, again, large sample sizes will be needed. Besides, epidemiological studies often use biological samples that are limited in quantity or costly to obtain. These considerations make it essential to search for study designs that make very efficient use of available resources [5–8].

Matched case–control studies already require a smaller sample size than unmatched studies under the same model specifications to study an association with disease [7, 9]. While matching already leads to more efficient designs in terms of fixed sample size, sequential methods may reduce the average sample size even further [6, 7]. A sequential analysis tests the cumulative data every time new information becomes available.

In the following we describe the test statistics we derived for the sequential analysis of both hierarchical and non-hierarchical $G \times E$ interactions in matched case–control studies. We investigated the efficiency of these sequential analyses by simulations.

2. MODELS AND HYPOTHESES

In this paper, we assume a recessive bi-allelic disease with prevalence of the recessive allele a equal to q_a . Then a dichotomous disease phenotype can be defined from the three genotypes AA , Aa and aa . The probability of the disease phenotype, $P(G=1)$, equals q_a^2 for aa and the probability of no disease phenotype, $P(G=0)$, equals $1 - q_a^2$ for AA or Aa . The probability distribution for the dichotomous exposure variable is p_e for exposure, i.e. $P(E=1)$, and $1 - p_e$ for no exposure, i.e. $P(E=0)$. (Note that for a different disease model the same methodology can be applied by adjusting the probabilities of (no) disease phenotype.)

- (a) $G \times E$ interaction in a non-hierarchical model: The likelihood for a non-hierarchical interaction in a conditional logistic regression model, i.e. without main effects G and E , can be written as

$$L(\beta_{GE}) = \prod_{i=1}^n \frac{e^{\beta_{GE} G_{i1} E_{i1}}}{e^{\beta_{GE} G_{i1} E_{i1}} + e^{\beta_{GE} G_{i2} E_{i2}}} \quad (1)$$

where G_{i1} and E_{i1} denote the case information for genetic and environmental factors, respectively, and G_{i2} and E_{i2} the control information, in the i th matched set.

- (b) $G \times E$ interaction in a hierarchical model: The likelihood for a hierarchical interaction in a conditional logistic regression model, i.e. with main effects G and E , can be written as

$$L(\beta_G, \beta_E, \beta_{GE}) = \prod_{i=1}^n \frac{e^{\beta_G G_{i1} + \beta_E E_{i1} + \beta_{GE} G_{i1} E_{i1}}}{e^{\beta_G G_{i1} + \beta_E E_{i1} + \beta_{GE} G_{i1} E_{i1}} + e^{\beta_G G_{i2} + \beta_E E_{i2} + \beta_{GE} G_{i2} E_{i2}}} \quad (2)$$

where G_{i1} and E_{i1} denote the case information for genetic and environmental factors, respectively, and G_{i2} and E_{i2} the control information, in the i th matched set.

3. THE SEQUENTIAL TESTS

For a sequential test as developed by Whitehead [10] to test the null hypothesis $H_0: \theta = 0$ versus $H_1: \theta \neq 0$ two test statistics are needed, the efficient score statistic Z and Fisher's information V . Z is a cumulative measure for the exposure effect; V reflects the amount of information about θ contained in Z . The parameter of interest θ is standardized such that under the null hypothesis it always equals 0. An example is the logarithm of the odds ratio for the interaction. When the information of the genetic and environmental factors for a new case-control set is obtained, new values for Z and V can be calculated. This is called a continuous sequential analysis. Z and V can also be calculated when information on more than one matched case-control set becomes available, a so-called group sequential analysis. Based on Z and V a decision with respect to H_0 is made. Therefore cumulative (Z, V) -values are plotted in a graph forming a pathway. Critical boundaries that enable to make the decision are fixed beforehand. These critical boundaries depend on the (two-sided) type I error 2α , the type II error β and the parameter θ_R . θ_R is the expected value of the standardized parameter under the alternative hypothesis H_1 . Each new (Z, V) -point is compared to the predefined boundaries. We investigated the behaviour of the double triangular test (TT) [10] as a sequential test for $G \times E$ interaction. This leads to one of three decisions (see Figure 1):

- (i) enough evidence is obtained to reject the null hypothesis when the upper or lower boundary is crossed, no more observations are necessary;
- (ii) enough evidence is obtained to accept the null hypothesis when the inner wedge is reached, no more observations are necessary;
- (iii) more evidence is needed to come to a decision when the new point is still within the boundaries, so more observations are necessary.

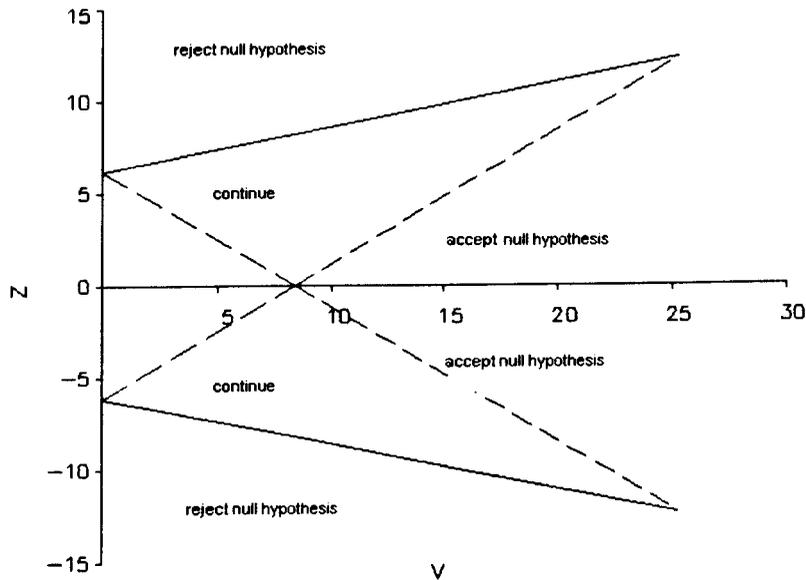


Figure 1. Decision areas for a double TT.

Table I. Gene (G) \times environment (E) interaction in a matched case-control set ($G = 0$: genetically not at risk, $G = 1$: genetically at risk, $E = 0$: not exposed, $E = 1$: exposed).

		Control			
		$G = 0$ $E = 0$	$G = 0$ $E = 1$	$G = 1$ $E = 0$	$G = 1$ $E = 1$
Case	$G = 0$				@
	$E = 0$				
	$G = 0$				@
	$E = 1$				
	$G = 1$				@
	$E = 0$				
	$G = 1$	@	@	@	
	$E = 1$				

@: case-control pairs informative for a test on interaction.

The test statistics Z and V for a sequential test on interaction in a non-hierarchical model on matched case-control data can be derived in a rather straightforward way [7, 10]. This derivation is given in Appendix A.

For a sequential test on interaction in a hierarchical model on matched-case-control data the test statistics Z and V are derived by first estimating the main effects in a conditional logistic regression model without an interaction term on the available cumulative data [10]. The estimated main effects are substituted as nuisance parameters so that Z and V can subsequently be derived for the sequential test on interaction (see Appendix B).

In a matched case-control design only discordant sets are informative for the relation to test. This means, for example, that when we study the effect of exposure, only matched pairs with the case exposed and the control not, or vice versa, are informative for a possible association between an exposure factor and a disease. This aspect applies also to tests on $G \times E$ interaction (see Table I). The fact that only discordant sets are informative influences the total number of matched sets necessary for a test on interaction. Based on the conditional logistic regression model, the probability of a discordant pair equals $\pi_{\text{disc}} = p_e q_a^2 (1 - p_e q_a^2) (\Psi + 1) / (1 + p_e q_a^2 (\Psi - 1))$ (see Appendix C), where Ψ is the odds ratio for the interaction term in the model, i.e. $\Psi = \exp(\beta_{GE})$. We assume that the odds ratios for the main effects of G and E are equal to 1.

4. FIXED SAMPLE SIZE DETERMINATION

Gauderman describes how the fixed sample size can be estimated for a test on $G \times E$ interaction in a hierarchical model [3]. His calculations are based on a likelihood ratio test statistic for a conditional logistic regression analysis of matched case-control data.

First the expected log-likelihood $\ell^1 = \mathcal{E}(\ln(L(\beta_G, \beta_E, \beta_{GE})))$, with L as defined in equation (2), is maximized with respect to the observable phenotype and exposure data. This results in expected MLEs $\hat{\beta}_G^1, \hat{\beta}_E^1, \hat{\beta}_{GE}^1$ and an expected log-likelihood $\hat{\ell}^1$. Then, the expected log-

likelihood $\ell^0 = \mathcal{E}(\ln(L(\beta_G, \beta_E)))$, under the null hypothesis (i.e. $H_0: \beta_{GE} = 0$) is maximized. This leads to expected MLEs $\hat{\beta}_G^0, \hat{\beta}_E^0$ and an expected log-likelihood $\hat{\ell}^0$. The likelihood ratio test statistic is defined as $\Lambda = 2(\hat{\ell}^1 - \hat{\ell}^0)$. For N matched sets $N\Lambda$ is the non-centrality parameter of the χ^2 -distribution under the alternative hypothesis H_1 . When both the genetic factor and the environmental factor are dichotomous, the test on interaction has one degree of freedom and N can be computed as $N = (z_x + z_\beta)^2 / \Lambda$ with 2α as the (two-sided) type I error, β as the type II error and z_x as the standardized normal deviate exceeded with probability x . For these calculations user-friendly software is provided [11].

We followed the same approach for a non-hierarchical model. Now the expected log-likelihoods $\ell^1 = \mathcal{E}(\ln(L(\beta_{GE})))$ and $\ell^0 = \mathcal{E}(\ln(L(\emptyset)))$ (the ‘null’ model), with L as defined in equation (1), are maximized leading to the MLE $\hat{\beta}_{GE}^1$ and expected log-likelihoods $\hat{\ell}^1$ and $\hat{\ell}^0$. Again the test on interaction has one degree of freedom and N can be calculated as before.

The necessary fixed sample size for a test on interaction in a non-hierarchical model can also be estimated using Whitehead’s test statistic V (Fisher’s information), $V = (z_x + z_\beta)^2 / (\beta_{GE})^2$ with $\beta_{GE} = \theta_R = \ln(\Psi)$ under H_1 [10]. The number of discordant matched case-control sets is equal to $N_{\text{disc}} = 4V$ (see Appendix A). The total number of matched case-control sets can be estimated by $N_{\text{tot}} = 4V / \pi_{\text{disc}}$ with π_{disc} as the probability of a discordant matched set.

5. SAMPLE SIZE DETERMINATION FOR A SEQUENTIAL TEST

When a sequential test is used, sample size is a stochastic variable and therefore cannot be determined beforehand. Only an average or median estimate or other characteristics of its distribution can be given. This estimate can be derived by multiplying the average or median value for V by 4 and dividing it by the probability of a discordant set. (The computer program PEST version 4 provides the average and median values for V [12].) A $G \times E$ interaction can be tested continuously or group sequentially. In genetic laboratory determinations a 96-wells plate is common for PCR-based genotyping. For a $G \times E$ interaction two wells are required for each matched case-control set, one to determine the phenotype of the case and one to determine the phenotype of the control, leading to a group size of, e.g. 44, when space is also reserved for blank and control samples. A group sequential analysis will in general be less efficient in terms of sample size than a continuous sequential analysis. However, it may be more efficient in terms of cost or logistic (laboratory) considerations.

For a ‘rare’ gene a group size of 44 often contains no information to be able to estimate the main effect of G in a hierarchical model. In these situations the group size was increased to $3 * 44$.

6. DATA GENERATION AND SIMULATIONS

To investigate the performance of the test statistics simulation studies were carried out both under $H_0: \Psi = 1.0$ and under H_1 with a Ψ of 1.5, 2.0 and 3.0, where Ψ is the odds ratio for the interaction term in the model. The odds ratios for the main effects G and E are assumed

to be equal to 1. The two-sided type I error 2α was set equal to 0.05, the power $1 - \beta$ to 0.80 or 0.90, $p_e = P(E = 1)$ to 0.25, and $q_a^2 = P(G = 1)$ to 0.01 for a 'rare' phenotype or 0.40 for a 'common' phenotype. For each combination of Ψ and $P(G = 1)$, a large population of matched cases and controls was simulated. The probability of occurrence of the several combinations of phenotype and exposure for both cases and controls was based on the law of Hardy–Weinberg and the conditional logistic regression model.

For each combination of the power, the Ψ and $P(G = 1)$ 2500 simulations were run on random samples from the corresponding large population both under H_0 and under H_1 . Simulations were performed using the so-called *Christmas tree correction* [10]. Every simulation run resulted in an average, median and 90th percentile for the information statistic V and for the total number of case–control sets N . The fraction of simulations that resulted in the rejection of H_0 is an estimate of the type I error and the power of the test.

7. RESULTS OF SIMULATIONS

For the non-hierarchical interaction model both continuous and group sequential analyses were simulated (Table II(a) and II(b)). For data with a 'common' phenotype ($P(G = 1) = 0.40$) and Ψ equal to 1.5 or 2.0 the simulation results agreed very well with the theoretical values for 2α and $1 - \beta$. For Ψ equal to 3.0 resulting significance levels were about 0.05, but power values were lower than the theoretical values. The median gain in number of matched case–control sets necessary for a sequential test compared to Gauderman's fixed sample size estimate was about 25 per cent. For data with a 'rare' phenotype ($P(G = 1) = 0.01$) power values were mostly larger than their theoretical values, especially for $\Psi = 1.5$. Median efficiency gain was about 22 per cent.

For the hierarchical interaction model only group sequential analyses were simulated (Table III(a) and III(b)). For the data with a 'common' phenotype simulation results for all Ψ s resembled their theoretical values. Only for $\Psi = 1.5$ power values were larger than the theoretical values. Median efficiency gain as compared to Gauderman's fixed sample size estimate was about 34 per cent. For data with a 'rare' phenotype power values were larger than theoretical values, but significance levels were about the theoretical value of 0.05. Median efficiency gain was about 30 per cent.

In general, for data generated under H_0 the median efficiency gain was smaller than that under H_1 .

8. AN EXAMPLE

Breast cancer is caused by genetic factors, environmental factors or a combination of these two in most cases. Van der Hel (submitted) investigated the combined effects of smoking and genetic polymorphisms in relevant metabolic genes. She also looked at the cumulative effect of putative at risk phenotypes on breast cancer risk. *N*-acetyltransferase 1 and 2 (NAT1, NAT2), glutathione S-Transferase M1 (GSTM1) and T1 (GSTT1) are enzymes, involved in carcinogen metabolisms. The genes coding for the NAT enzymes contain polymorphic sites, which cause variable enzymatic activity. GSTM1 or GSTT1 null phenotype results in a complete lack of enzymatic activity.

Table II. Results of 2500 simulations of sequential analyses on non-hierarchical interactions using a double triangular test.

	2α	$1 - \beta$	G_{fixed}	W_{fixed}		N_{av}	N_{med}	N_{90}	V_{av}	V_{med}	V_{90}	rej. fr.
(a) $P(G = 1) = 0.40$ and $P(E = 1) = 0.25$ (group size = 44 sets)												
OR = 1.5												
H_0	0.05	0.80			Cont.seq.	698.2	656	970	31.50	29.50	43.50	0.040
					Grp seq.	713.8	660	968	32.17	30.25	44.00	0.046
					<i>Theor.</i>	707.6	662	975	31.84	29.80	43.89	
		0.90			Cont.seq.	948.9	884	1315	42.67	39.75	58.75	0.046
					Grp seq.	960.3	924	1364	43.21	40.25	59.75	0.056
					<i>Theor.</i>	947.8	887	1306	42.65	39.91	58.78	
H_1	0.05	0.80	910	891	Cont.seq.	595.3	568	916	31.67	30.50	48.00	0.806
					Grp seq.	599.8	572	924	32.67	31.00	50.03	0.823
					<i>Theor.</i>	598.2	572	926	32.05	30.67	49.61	
		0.90	1218	1193	Cont.seq.	737.1	692	1180	39.23	36.75	62.00	0.900
					Grp seq.	722.2	660	1144	39.32	36.75	62.25	0.901
					<i>Theor.</i>	724.8	681	1154	38.83	36.46	61.82	
OR = 2.0												
H_0	0.05	0.80			Cont.seq.	242.0	229	338	10.91	10.25	14.75	0.050
					Grp seq.	255.4	220	352	11.55	11.00	15.75	0.044
					<i>Theor.</i>	242.2	227	334	10.90	10.20	15.02	
		0.90			Cont.seq.	325.2	307	457	14.62	13.75	20.25	0.046
					Grp seq.	339.8	308	484	15.24	14.25	20.75	0.048
					<i>Theor.</i>	324.2	304	447	14.59	13.66	20.11	
H_1	0.05	0.80	282	266	Cont.seq.	184.5	175	284	11.32	10.75	17.00	0.802
					Grp seq.	193.2	176	308	11.97	11.50	18.00	0.820
					<i>Theor.</i>	178.8	171	277	10.97	10.50	16.98	
		0.90	378	357	Cont.seq.	222.7	211	350	13.68	13.00	21.25	0.909
					Grp seq.	293.1	220	352	14.70	14.00	22.50	0.907
					<i>Theor.</i>	216.6	203	345	13.29	12.48	21.15	
OR = 3.0												
H_0	0.05	0.80			Cont.seq.	97.8	92	143	4.43	4.00	6.00	0.048
					Grp seq.	112.7	88	176	5.09	5.00	6.75	0.053
					<i>Theor.</i>	96.4	90	133	4.34	4.06	5.98	
		0.90			Cont.seq.	129.5	122	189	5.79	5.50	7.75	0.046
					Grp seq.	144.4	132	220	6.49	6.25	9.00	0.050
					<i>Theor.</i>	129.1	121	178	5.81	5.44	8.01	
H_1	0.05	0.80	100	87	Cont.seq.	63.9	61	99	4.49	4.75	7.25	0.756
					Grp seq.	73.6	88	88	5.60	5.75	8.00	0.774
					<i>Theor.</i>	58.3	56	90	4.37	4.18	6.76	
		0.90	134	116	Cont.seq.	79.7	76	124	5.96	5.75	8.75	0.858
					Grp seq.	89.9	88	132	6.81	6.75	10.25	0.882
					<i>Theor.</i>	70.5	66	112	5.29	4.97	8.42	

Table II. *Continued.*

2α	$1 - \beta$	G_{fixed}	W_{fixed}		N_{av}	N_{med}	N_{90}	V_{av}	V_{med}	V_{90}	rej. fr.	
(b) $P(G=1)=0.01$, $P(E=1)=0.25$ (group size = 3×44 sets)												
OR = 1.5												
H_0	0.05	0.80			Cont.seq.	27 034.2	25 564	37 799	31.97	30.25	44.75	0.060
					Grp seq.	26 392.1	24 684	37 536	32.20	30.00	45.00	0.062
					<i>Theor.</i>	<i>25 472.0</i>	<i>23 840</i>	<i>35 112</i>	<i>31.84</i>	<i>29.80</i>	<i>43.89</i>	
		0.90			Cont.seq.	36 413.8	34 522	50 558	42.99	40.50	59.28	0.048
					Grp seq.	35 158.2	32 846	49 148	42.79	49.75	60.00	0.064
					<i>Theor.</i>	<i>34 120.0</i>	<i>31 928</i>	<i>47 024</i>	<i>42.65</i>	<i>39.91</i>	<i>58.78</i>	
H_1	0.05	0.80	31312	30679	Cont.seq.	19 983.3	18 950	31 333	31.46	29.75	49.25	0.852
					Grp seq.	21 614.6	20 856	32 736	33.64	32.50	50.75	0.748
					<i>Theor.</i>	<i>20 677.4</i>	<i>19 787</i>	<i>32 006</i>	<i>32.05</i>	<i>30.67</i>	<i>49.61</i>	
		0.90	41918	41070	Cont.seq.	23 771.7	22 095	37 519	37.41	35.25	58.30	0.942
					Grp seq.	27 208.9	25 828	42 284	42.54	40.50	65.75	0.846
					<i>Theor.</i>	<i>25 051.6</i>	<i>23 523</i>	<i>39 884</i>	<i>38.83</i>	<i>36.46</i>	<i>61.82</i>	
OR = 2.0												
H_0	0.05	0.80			Cont.seq.	9303.1	8833	13 152	10.98	10.25	15.00	0.058
					Grp seq.	8776.5	8338	12 408	11.01	10.25	15.00	0.046
					<i>Theor.</i>	<i>8720.0</i>	<i>8160</i>	<i>12 016</i>	<i>10.90</i>	<i>10.20</i>	<i>15.02</i>	
		0.90			Cont.seq.	12 485.8	11 838	17 584	14.70	13.75	20.25	0.052
					Grp seq.	11 689.6	11 000	16 192	14.58	13.38	20.25	0.050
					<i>Theor.</i>	<i>11 672.0</i>	<i>10 928</i>	<i>16 088</i>	<i>14.59</i>	<i>13.66</i>	<i>20.11</i>	
H_1	0.05	0.80	9289	8759	Cont.seq.	5908.6	5655	9215	11.08	10.25	17.00	0.839
					Grp seq.	6377.6	6116	9724	11.55	11.00	17.00	0.782
					<i>Theor.</i>	<i>5850.7</i>	<i>5600</i>	<i>9056</i>	<i>10.97</i>	<i>10.50</i>	<i>16.98</i>	
		0.90	12436	11726	Cont.seq.	7065.8	6662	11 113	13.21	12.25	20.75	0.932
					Grp seq.	7702.5	7260	11 968	13.93	13.25	21.53	0.886
					<i>Theor.</i>	<i>7088.0</i>	<i>6656</i>	<i>11 280</i>	<i>13.29</i>	<i>12.48</i>	<i>21.15</i>	
OR = 3.0												
H_0	0.05	0.80			Cont.seq.	3741.1	3573	5426	4.40	4.00	6.00	0.047
					Grp seq.	3524.3	3256	5324	4.42	4.00	6.25	0.054
					<i>Theor.</i>	<i>3472.0</i>	<i>3248</i>	<i>4784</i>	<i>4.34</i>	<i>4.06</i>	<i>5.98</i>	
		0.90			Cont.seq.	4838.0	4586	7020	5.75	5.50	7.50	0.052
					Grp seq.	4616.7	4400	6644	5.84	5.50	8.00	0.054
					<i>Theor.</i>	<i>4648.0</i>	<i>4352</i>	<i>6408</i>	<i>5.81</i>	<i>5.44</i>	<i>8.01</i>	
H_1	0.05	0.80	3024	2622	Cont.seq.	1860.9	1758	2961	4.62	4.25	7.25	0.818
					Grp seq.	1913.8	1804	2992	4.75	4.75	7.00	0.802
					<i>Theor.</i>	<i>1765.7</i>	<i>1689</i>	<i>2731</i>	<i>4.37</i>	<i>4.18</i>	<i>6.76</i>	
		0.90	4048	3510	Cont.seq.	2272.7	2120	3589	5.58	5.00	8.75	0.919
					Grp seq.	2337.6	2200	3652	5.80	5.50	8.75	0.901
					<i>Theor.</i>	<i>2137.4</i>	<i>2008</i>	<i>3402</i>	<i>5.29</i>	<i>4.97</i>	<i>8.42</i>	

Notation: G_{fixed} : fixed sample size estimate according to Gauderman; W_{fixed} : fixed sample size estimate according to Whitehead; N_{av} : average number of case-control sets; N_{med} : median number of case-control sets; N_{90} : 90th percentile of the number of case-control sets; V_{av} : average value for V ; V_{med} : median value for V ; V_{90} : 90th percentile value for V ; rej. fr.: fraction of simulations that rejected H_0 .

Table III. Results of 2500 simulations of sequential analyses on hierarchical interactions using a double triangular test.

	2α	$1 - \beta$	G_{fixed}		N_{av}	N_{med}	N_{90}	V_{av}	V_{med}	V_{90}	rej. fr.
(a) $P(G = 1) = 0.40, P(E = 1) = 0.25$ (group size = 44 sets)											
OR = 1.5											
H_0	0.05	0.80		Grp seq.	1431.4	1364	1980	32.00	30.05	44.28	0.052
				Theor.				31.84	29.80	43.89	
		0.90		Grp seq.	1933.6	1804	2684	43.29	40.30	60.01	0.061
				Theor.				42.65	39.91	58.78	
H_1	0.05	0.80	2007	Grp seq.	1283.8	1232	2024	30.68	29.21	48.15	0.868
				Theor.				32.05	30.67	49.61	
		0.90	2687	Grp seq.	1505.9	1408	2376	36.02	33.77	57.58	0.945
				Theor.				38.83	36.46	61.82	
OR = 2.0											
H_0	0.05	0.80		Grp seq.	496.8	484	660	11.10	10.40	15.24	0.053
				Theor.				10.90	10.20	15.02	
		0.90		Grp seq.	666.8	616	924	14.92	14.04	20.23	0.054
				Theor.				14.59	13.66	20.11	
H_1	0.05	0.80	674	Grp seq.	461.0	440	704	11.29	10.81	17.33	0.802
				Theor.				10.97	10.50	16.98	
		0.90	902	Grp seq.	538.7	484	880	13.20	12.45	20.94	0.912
				Theor.				13.29	12.48	21.15	
OR = 3.0											
H_0	0.05	0.80		Grp seq.	213.0	220	308	4.66	4.44	6.31	0.051
				Theor.				4.34	4.06	5.98	
		0.90		Grp seq.	277.7	264	396	6.11	5.72	8.40	0.045
				Theor.				5.81	5.44	8.01	
H_1	0.05	0.80	270	Grp seq.	191.5	176	308	4.69	4.48	7.12	0.815
				Theor.				4.37	4.18	6.76	
		0.90	361	Grp seq.	230.3	220	352	5.66	5.35	8.89	0.908
				Theor.				5.29	4.97	8.42	
(b) $P(G = 1) = 0.01, P(E = 1) = 0.25$ (group size = 3×44 sets)											
OR = 1.5											
H_0	0.05	0.80		Grp seq.	36 336.5	34 320	51 084	32.80	30.89	46.08	0.066
				Theor.				31.84	29.80	43.89	
		0.90		Grp seq.	48 906.2	45 672	69 300	44.24	41.34	62.48	0.074
				Theor.				42.65	39.91	58.78	
H_1	0.05	0.80	44 614	Grp seq.	28 704.1	27 324	44 484	30.97	29.22	48.14	0.888
				Theor.				32.05	30.67	49.61	
		0.90	59 726	Grp seq.	33 674.7	32 076	52 549	36.35	34.65	56.48	0.968
				Theor.				38.83	36.46	61.82	
OR = 2.0											
H_0	0.05	0.80		Grp seq.	12 380.0	11 748	17 292	11.14	10.50	15.33	0.057
				Theor.				10.90	10.20	15.02	
		0.90		Grp seq.	16 478.5	15 444	22 968	14.86	13.83	20.64	0.064
				Theor.				14.59	13.66	20.11	

Table III. *Continued.*

	2α	$1 - \beta$	G_{fixed}		N_{av}	N_{med}	N_{90}	V_{av}	V_{med}	V_{90}	rej. fr.
H_1	0.05	0.80	13 942	Grp seq.	8513.5	8052	13 332	10.13	9.59	15.89	0.915
				Theor.				10.97	10.50	16.98	
		0.90	18 664	Grp seq.	11 545.6	11 088	17 952	14.07	13.48	21.87	0.887
				Theor.				13.29	12.48	21.15	
OR = 3.0											
H_0	0.05	0.80		Grp seq.	4869.7	4620	6864	4.38	4.09	5.99	0.039
				Theor.				4.34	4.06	5.98	
		0.90		Grp seq.	6487.0	6204	8976	5.88	5.53	8.10	0.043
				Theor.				5.81	5.44	8.01	
H_1	0.05	0.80	4965	Grp seq.	3355.4	3168	5148	4.46	4.28	6.86	0.850
				Theor.				4.37	4.18	6.76	
		0.90	6646	Grp seq.	3914.3	3564	6204	5.23	4.85	8.30	0.935
				Theor.				5.29	4.97	8.42	

Notation: G_{fixed} : fixed sample size estimate according to Gauderman; N_{av} : average number of case-control sets; N_{med} : median number of case-control sets; N_{90} : 90th percentile of the number of case-control sets; V_{av} : average value for V ; V_{med} : median value for V ; V_{90} : 90th percentile value for V ; rej. fr.: fraction of simulations that rejected H_0 .

Follow-up of a population-based screening program for early detection of breast cancer in the Netherlands (DOM) revealed 942 women with incident breast cancer. One thousand control women were randomly selected from the DOM-cohorts. As environmental factor the smoking status of the women was assessed at baseline by a self-administered questionnaire. Women were classified as 'never' smokers or 'ever' smokers. The probability of being an 'ever' smoker was estimated as 0.30. When a woman had three or four putative phenotypes at risk she was considered 'susceptible', otherwise she was not. The probability to be 'susceptible' was estimated as 0.30 (Van der Hel, submitted).

A total of 579 cases and controls could be matched on age. To detect an odds ratio equal to 2 with a hierarchical test on interaction, a two-sided significance level $2\alpha = 0.05$ and a power $1 - \beta = 0.80$ a fixed sample size of at least 674 matched sets (according to Gauderman) would be required. For a non-hierarchical test on interaction with the same specifications at least 290 (according to Whitehead) to 308 matched sets (according to Gauderman) would be required. Cumulative data on cases and controls were analysed in the chronological order the cases became apparent. Data were analysed sequentially with group sizes of 44 case-control sets.

The non-hierarchical test on interaction led to the acceptance of the null hypothesis (i.e. no interaction between G and E) after 5 groups of data (= 220 case-control sets of which 33 informative sets) were analysed (Figure 2(a)). Z was equal to -0.5 and V was equal to 8.25. The median unbiased estimate for Ψ was 0.92 (95 per cent C.I. 0.46; 1.84).

Thus an efficiency gain of about 29 per cent with respect to the fixed sample size was reached by using a group sequential test.

The hierarchical test on interaction also concluded that there was no evidence for an interaction between G and E after the analysis of nine groups of data (= 396 case-control sets

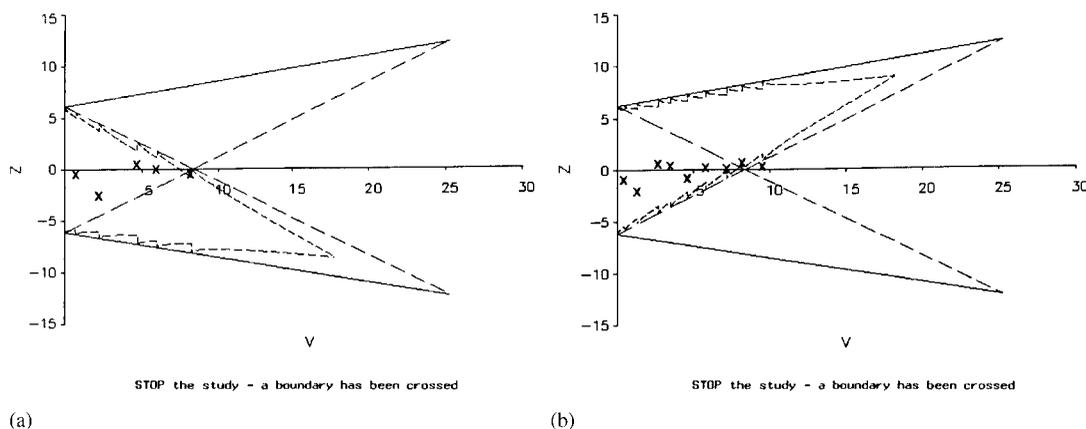


Figure 2. Results of a sequential double TT with $2\alpha = 0.05$ and $1 - \beta = 0.80$ to detect an odds ratio $\Psi = 2$. Z is the efficient score statistic; V is Fisher's information. Each \times stands for the cumulative results of a group of 44 case-control sets. The curved inner boundaries reflect the *Christmas tree correction* [10]. (a) a non-hierarchical interaction; and (b) a hierarchical interaction.

of which 73 were informative) (Figure 2(b)). Z was equal to 0.24 and V was equal to 9.56. The median unbiased estimate for Ψ was 1.11 (95 per cent C.I. 0.57; 2.19). The main effects were estimated as odds ratios for G and E , $OR_G = 1.173$ and $OR_E = 1.167$. Here an efficiency gain of 41 per cent was reached by using a group sequential test instead of a fixed sample test.

9. DISCUSSION

In this paper sequential tests for $G \times E$ interactions were developed. We emphasize that these tests can also be applied in situations where gene \times gene interactions are of interest. If the prevalence of the susceptible gene is rare, the sample size necessary to detect a significant $G \times E$ interaction in an observational study can be (very) large. Note that sample sizes will become even larger in case of gene \times gene interactions.

If an increased risk does exist for the combination of a genetic and an environmental factor, its size is more likely to be moderate (odds ratio of 1.2–1.6) than large, thus requiring a large number of observations [4]. As an alternative to a large study, many small studies can be pooled. But pooling of small studies may be hampered by publication bias (i.e. positive findings are more likely to be published than negative findings) [4]. Therefore a large, conclusive study to detect a moderate interaction would be preferred to a meta-analysis of small studies.

For reasons of cost-effectiveness and efficient use of available resources, like biological samples, we investigated the properties of sequential designs in matched case-control studies to test for interaction. Matched study designs already require smaller sample sizes than un-

matched designs. Sequential tests require, on average, a smaller sample size than their fixed sample size counterparts.

We derived the test statistics for sequential tests on hierarchical and non-hierarchical interactions in matched case-control studies. For non-hierarchical interactions we compared results of continuous and group sequential tests. The continuous sequential analyses reflect the theoretical properties of the tests, while the group sequential tests reflect more the way laboratory analyses will be performed in practice.

For the non-hierarchical models we estimated fixed sample sizes according to Gauderman and by using Whitehead's expression for V . Differences between the two estimates arise because Gauderman bases his calculations on the likelihood ratio test, while Whitehead uses the score test. For the hierarchical models we could only estimate sample size following Gauderman's formula. Efficiency gains for the sequential tests were calculated with respect to Gauderman's fixed sample size estimate.

Results of our simulations for sequential tests showed a good agreement with theoretical values for both types of interaction when a 'common' phenotype was assumed. Efficiency gains ranged from 19 to 48 per cent for Ψ s less than or equal to 2. Only for very small studies ($\Psi = 3$) the gain was obviously less. For the 'rare' phenotype the gain in efficiency was largest for data generated under the alternative hypothesis. The probability of a discordant, and thus informative, pair depends on the odds ratio and, especially when the phenotype is 'rare', this leads to smaller values for π_{disc} under H_0 (when $\Psi = 1$) than under H_1 (when $\Psi > 1$). This results sometimes in median estimates of the number of matched sets necessary that are only slightly smaller or even somewhat larger than the fixed sample size estimate. Because π_{disc} is smaller under H_0 than under H_1 , more case-control sets are needed under H_0 than under H_1 to get the same amount of information V .

In general, the larger studies in terms of sample size led to power values that were larger than their theoretical values. Resulting significance levels agreed well with their theoretical values.

The ability of a sequential test to accept the null hypothesis was illustrated by the example of the breast cancer data. A possible interaction with $\Psi = 2$ between genetic susceptibility and smoking was tested non-hierarchically and hierarchically. Both sequential tests accepted the null hypothesis without using all the available data. Gains of 29 and 41 per cent in the number of matched sets (as compared to the fixed sample size) necessary to come to a decision were reached for the non-hierarchical and hierarchical test, respectively.

Case-only designs are mentioned as alternative to matched case-control designs [2, 13, 14]. Gauderman shows that case-only designs can be more efficient than matched case-control designs to study (gene \times gene) interaction [14]. Case-only designs require no selection of controls, but they are only useful to test an interaction in the cases. They depend strongly on the assumption that the genetic and the environmental factor are independent in the large population. If that association has still to be examined, a (matched) case-control study yields more information. If genetic and environmental main effects are also of interest or have to be adjusted for, case-only designs are no option. Further work will show efficiency gains of the use of sequential designs in case-only studies and compare these efficiency gains to those from non-hierarchical models in matched case-control studies.

When biological samples are scarce or laboratory examinations are costly savings in samples, labour, and/or costs can be very valuable. Sequential tests can be very useful to handle the available data efficiently and can lead to considerable savings.

When biological samples for controls are abundant, but those for cases are scarce, still more efficiency can be obtained by matching more than one control to a case [7, 15].

APPENDIX A: A NON-HIERARCHICAL INTERACTION MODEL

The likelihood for a non-hierarchical interaction in a conditional logistic regression model has the form

$$L(\beta_{GE}) = \prod_{i=1}^n \frac{e^{\beta_{GE} G_{i1} E_{i1}}}{e^{\beta_{GE} G_{i1} E_{i1}} + e^{\beta_{GE} G_{i2} E_{i2}}}$$

where G_{i1} and E_{i1} denote the case information for genetic and environmental factors, respectively, and G_{i2} and E_{i2} the control information, in the i th matched set.

The logarithm of $L(\beta_{GE})$ is equal to

$$\ell(\beta_{GE}) = \sum_{i=1}^n \{ \beta_{GE} G_{i1} E_{i1} - \ln(e^{\beta_{GE} G_{i1} E_{i1}} + e^{\beta_{GE} G_{i2} E_{i2}}) \}$$

Taking the first derivative with respect to β_{GE} leads to

$$\partial \ell / \partial \beta_{GE} = \sum_{i=1}^n \left\{ G_{i1} E_{i1} - \frac{G_{i1} E_{i1} e^{\beta_{GE} G_{i1} E_{i1}} + G_{i2} E_{i2} e^{\beta_{GE} G_{i2} E_{i2}}}{e^{\beta_{GE} G_{i1} E_{i1}} + e^{\beta_{GE} G_{i2} E_{i2}}} \right\}$$

Substituting $\beta_{GE} = 0$ gives

$$\partial \ell / \partial \beta_{GE} (\beta_{GE} = 0) = \sum_{i=1}^n \left\{ G_{i1} E_{i1} - \frac{G_{i1} E_{i1} + G_{i2} E_{i2}}{2} \right\} = Z$$

Note that Z is the efficient score statistic evaluated under the null hypothesis ($\beta_{GE} = 0$). Note also that for the i th matched set

$$Z_i = G_{i1} E_{i1} - \frac{1}{2}(G_{i1} E_{i1} + G_{i2} E_{i2}) = \frac{1}{2}(G_{i1} E_{i1} - G_{i2} E_{i2})$$

For concordant case-control pairs both $G_{i1} E_{i1}$ and $G_{i2} E_{i2}$ equal 0 or both equal 1 (see Table I) and Z_i becomes 0; for discordant pairs Z_i becomes $+\frac{1}{2}$ or $-\frac{1}{2}$.

The negative of the second derivative with respect to β_{GE} is equal to

$$\partial^2 \ell / \partial \beta_{GE}^2 = \sum_{i=1}^n \left\{ \frac{\sum_{j=1,2} (G_{ij} E_{ij})^2 e^{\beta_{GE} G_{ij} E_{ij}} \cdot \sum_{j=1,2} e^{\beta_{GE} G_{ij} E_{ij}} - \left(\sum_{j=1,2} G_{ij} E_{ij} e^{\beta_{GE} G_{ij} E_{ij}} \right)^2}{\left(\sum_{j=1,2} e^{\beta_{GE} G_{ij} E_{ij}} \right)^2} \right\}$$

Substituting $\beta_{GE} = 0$ gives

$$V = \sum_{i=1}^n \left\{ \frac{\sum_j (G_{ij} E_{ij})^2 \cdot 2 - \left(\sum_j G_{ij} E_{ij} \right)^2}{4} \right\}$$

$$= \frac{1}{4} \sum_{i=1}^n (G_{i1} E_{i1} + G_{i2} E_{i2})(2 - (G_{i1} E_{i1} + G_{i2} E_{i2})).$$

Note that V is Fisher's information about θ contained in Z . For concordant case-control pairs both $G_{i1}E_{i1}$ and $G_{i2}E_{i2}$ equal 0 or both equal 1 (see Table I) and the contribution of the i th matched set to V is therefore 0; for discordant pairs the contribution of the i th matched set to V is $\frac{1}{4}$.

APPENDIX B: A HIERARCHICAL INTERACTION MODEL

The likelihood for a hierarchical interaction in a conditional logistic regression model has the form

$$L(\beta_G, \beta_E, \beta_{GE}) = \prod_{i=1}^n \frac{e^{\beta_G G_{i1} + \beta_E E_{i1} + \beta_{GE} G_{i1} E_{i1}}}{e^{\beta_G G_{i1} + \beta_E E_{i1} + \beta_{GE} G_{i1} E_{i1}} + e^{\beta_G G_{i2} + \beta_E E_{i2} + \beta_{GE} G_{i2} E_{i2}}}$$

where G_{i1} and E_{i1} denote the case information for genetic and environmental factors, respectively, and G_{i2} and E_{i2} the control information, in the i th matched set.

This likelihood is identical to the likelihood of fitting a logistic regression model to a set of data with constant response, no intercept and differences between the corresponding values for case and control for G , E and GE :

$$L(\beta_G, \beta_E, \beta_{GE}) = \prod_{i=1}^n \frac{e^{\beta_G \Delta G_i + \beta_E \Delta E_i + \beta_{GE} \Delta GE_i}}{1 + e^{\beta_G \Delta G_i + \beta_E \Delta E_i + \beta_{GE} \Delta GE_i}}$$

where $\Delta G_i = G_{i1} - G_{i2}$ (the difference between the case value G_{i1} and the control value G_{i2}), $\Delta E_i = E_{i1} - E_{i2}$ (the difference between the case value E_{i1} and the control value E_{i2}) and $\Delta GE_i = G_{i1}E_{i1} - G_{i2}E_{i2}$ (the difference between the case value $G_{i1}E_{i1}$ and the control value $G_{i2}E_{i2}$).

The logarithm of $L(\beta_G, \beta_E, \beta_{GE})$ is equal to

$$\ell(\beta_G, \beta_E, \beta_{GE}) = \sum_{i=1}^n \{ \beta_G \Delta G_i + \beta_E \Delta E_i + \beta_{GE} \Delta GE_i - \ln(1 + e^{\beta_G \Delta G_i + \beta_E \Delta E_i + \beta_{GE} \Delta GE_i}) \}$$

The first-order derivative with respect to β_{GE} is

$$\partial \ell / \partial \beta_{GE} = \sum_{i=1}^n \left\{ \Delta GE_i - \frac{\Delta GE_i e^{\beta_G \Delta G_i + \beta_E \Delta E_i + \beta_{GE} \Delta GE_i}}{1 + e^{\beta_G \Delta G_i + \beta_E \Delta E_i + \beta_{GE} \Delta GE_i}} \right\} \quad (\text{B1})$$

To test the null hypothesis $H_0 : \beta_{GE} = 0$ sequentially, we will have to derive Z and V statistics, that are adjusted for the main gene (G) and environmental (E) effects. Therefore, we derived estimates for β_G and β_E by fitting the reduced model (i.e. a model without the interaction term) in a logistic regression analysis [10, Chapter 7.5].

Substituting the obtained estimates b_G and b_E for β_G and β_E and $\beta_{GE} = 0$ in (B1) leads to

$$\begin{aligned} Z &= \sum_{i=1}^n \left\{ \Delta GE_i - \frac{\Delta GE_i e^{b_G \Delta G_i + b_E \Delta E_i}}{1 + e^{b_G \Delta G_i + b_E \Delta E_i}} \right\} \\ &= \sum_{i=1}^n \left\{ \Delta GE_i \left(1 - \frac{e^{b_G \Delta G_i + b_E \Delta E_i}}{1 + e^{b_G \Delta G_i + b_E \Delta E_i}} \right) \right\} \end{aligned}$$

Note that Z equals the sum over all non-zero values for the variable of interest (ΔGE_i) times the difference between the constant response ($=1$) and the fitted value for a logistic regression model with covariates ΔG_i and ΔE_i , a constant response and no intercept term.

To derive Fisher's information V we have to work out the second derivatives of the log-likelihood with respect to the parameter of interest β_{GE} .

We simplify the notation by denoting η_i for $\exp(\beta_G \Delta G_i + \beta_E \Delta E_i + \beta_{GE} \Delta GE_i)$.

$$\begin{aligned} \partial^2 \ell / \partial \beta_{GE}^2 &= \sum_{i=1}^n \left\{ -\frac{(\Delta GE_i)^2 \eta_i (1 + \eta_i) - (\Delta GE_i \eta_i)^2}{(1 + \eta_i)^2} \right\} \\ &= \sum_{i=1}^n \left\{ (\Delta GE_i)^2 \frac{\eta_i^2 - \eta_i (1 + \eta_i)}{(1 + \eta_i)^2} \right\} = \sum_{i=1}^n \left\{ (\Delta GE_i)^2 \frac{-\eta_i}{(1 + \eta_i)^2} \right\} \end{aligned} \tag{B2}$$

$$\begin{aligned} \partial^2 \ell / \partial \beta_{GE} \partial \beta_G &= \sum_{i=1}^n \left\{ -\frac{(\Delta GE_i) \Delta G_i \eta_i (1 + \eta_i) - (\Delta GE_i \eta_i) (\Delta G_i \eta_i)}{(1 + \eta_i)^2} \right\} \\ &= \sum_{i=1}^n \left\{ \Delta GE_i \Delta G_i \frac{\eta_i^2 - \eta_i (1 + \eta_i)}{(1 + \eta_i)^2} \right\} = \sum_{i=1}^n \left\{ \Delta GE_i \Delta G_i \frac{-\eta_i}{(1 + \eta_i)^2} \right\} \end{aligned} \tag{B3a}$$

$$\begin{aligned} \partial^2 \ell / \partial \beta_{GE} \partial \beta_E &= \sum_{i=1}^n \left\{ -\frac{(\Delta GE_i) \Delta E_i \eta_i (1 + \eta_i) - (\Delta GE_i \eta_i) (\Delta E_i \eta_i)}{(1 + \eta_i)^2} \right\} \\ &= \sum_{i=1}^n \left\{ \Delta GE_i \Delta E_i \frac{\eta_i^2 - \eta_i (1 + \eta_i)}{(1 + \eta_i)^2} \right\} = \sum_{i=1}^n \left\{ \Delta GE_i \Delta E_i \frac{-\eta_i}{(1 + \eta_i)^2} \right\} \end{aligned} \tag{B3b}$$

The information V is equal to $-[l_{\theta\theta}(\theta, \varphi) - \{l_{\theta\varphi}(\theta, \varphi)\}' \cdot \{l_{\varphi\varphi}(\theta, \varphi)\}^{-1} \cdot \{l_{\theta\varphi}(\theta, \varphi)\}]$ [10], where $\theta = \beta_{GE}$ is the parameter of interest and $\varphi = (\beta_G, \beta_E)'$ is the vector of nuisance parameters. Thus

$$\begin{aligned} l_{\theta\theta}(\theta, \varphi) &= \partial^2 \ell / \partial \beta_{GE}^2 \\ \{l_{\theta\varphi}(\theta, \varphi)\}' &= \{\partial^2 \ell / \partial \beta_{GE} \partial \beta_G; \partial^2 \ell / \partial \beta_{GE} \partial \beta_E\}' \\ l_{\varphi\varphi}(\theta, \varphi) &= \begin{pmatrix} \partial^2 \ell / \partial \beta_G^2 & \partial^2 \ell / \partial \beta_G \partial \beta_E \\ \partial^2 \ell / \partial \beta_E \partial \beta_G & \partial^2 \ell / \partial \beta_E^2 \end{pmatrix} \end{aligned} \tag{B4}$$

Substituting $\theta=0$ and $\varphi^* = (b_G, b_E)'$ into the expression for V leads for the separate parts of this equation to: (B2) this second derivative is equal to the negative of the sum over all non-zero values for ΔGE_i times $p_i(1 - p_i)$ with $p_i = \hat{\eta}_i / (1 + \hat{\eta}_i)$; (B3a) and (B3b) these second derivatives are the negative of the sum over all non-zero values for ΔGE_i times the

covariates ΔG_i and ΔE_i , respectively, times $p_i(1 - p_i)$; (B4) this matrix is the covariance matrix of the parameter estimates for the fit of a logistic regression model with covariates ΔG_i and ΔE_i , a constant response and no intercept term. (See also Reference [10, Section 7.6.2] for the derivation of Z and V).

(N.B. The same values for Z and V are derived for the conditional logistic regression model when a Cox PH regression model with strata is used to estimate the nuisance parameters.)

APPENDIX C

We define $p_e = P(E = 1)$, $q_a^2 = P(G = 1)$ and $\Psi = \exp(\beta_{GE})$.

We assume that the odds ratios for the main effects of G and E are equal to 1.

For the controls the following holds:

$$P_2(G = 0 \cap E = 0) = (1 - p_e)(1 - q_a^2)$$

$$P_2(G = 0 \cap E = 1) = p_e(1 - q_a^2)$$

$$P_2(G = 1 \cap E = 0) = (1 - p_e)q_a^2$$

$$P_2(G = 1 \cap E = 1) = p_e q_a^2$$

For the cases the following holds:

$$P_1(G = 0 \cap E = 0) = (1 - p_e)(1 - q_a^2)e^{b_0}$$

$$P_1(G = 0 \cap E = 1) = p_e(1 - q_a^2)e^{b_0}$$

$$P_1(G = 1 \cap E = 0) = (1 - p_e)q_a^2 e^{b_0}$$

$$P_1(G = 1 \cap E = 1) = p_e q_a^2 \Psi e^{b_0}$$

with

$$e^{b_0} = \frac{1}{1 + p_e q_a^2 (\Psi - 1)}$$

resulting in

$$\begin{aligned} \pi_{\text{disc}} &= P_1(G = 1 \cap E = 1)[P_2(G = 0 \cap E = 0) + P_2(G = 0 \cap E = 1) + P_2(G = 1 \cap E = 0)] \\ &\quad + P_2(G = 1 \cap E = 1)[P_1(G = 0 \cap E = 0) + P_1(G = 0 \cap E = 1) + P_1(G = 1 \cap E = 0)] \\ &= \frac{p_e q_a^2 (1 - p_e q_a^2)(\Psi + 1)}{1 + p_e q_a^2 (\Psi - 1)} \end{aligned}$$

ACKNOWLEDGEMENTS

We acknowledge the help from professor Theo Stijnen in the derivation of the test statistics for the hierarchical model and thank him for critically reading the manuscript. We thank Olga van der Hel for providing the data and the information on the research questions for the example.

REFERENCES

1. Ottman R. Theoretical epidemiology: gene-environment interaction: definitions and study designs. *Preventive Medicine* 1996; **25**:764-770.
2. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statistics in Medicine* 1994; **13**:153-162.
3. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. *Statistics in Medicine* 2002; **21**:35-50.
4. Brennan P. Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? *Carcinogenesis* 2002; **23**:381-387.
5. Van der Tweel I, Van Noord PAH, Kaaks R. Application of a sequential *t*-test in a cohort nested case-control study with multiple controls per case. *Journal of Clinical Epidemiology* 1993; **46**:253-259.
6. Van der Tweel I, Kaaks R, Van Noord PAH. Comparison of one-sample two-sided sequential *t*-tests for application in epidemiological studies. *Statistics in Medicine* 1996; **15**:2781-2795.
7. Van der Tweel I, Van Noord PAH. Sequential analysis of matched dichotomous data from prospective case-control studies. *Statistics in Medicine* 2000; **19**:3449-3464.
8. Aplenc R, Zhao H, Rebbeck TR, Prokert KJ. Group sequential methods and sample size savings in biomarker-disease association studies. *Genetics* 2003; **163**:1215-1219.
9. Schork NJ, Fallin D, Tiwari HK, Schork MA. In *Handbook of Statistical Genetics*, Balding DJ *et al.* (eds). Wiley: Chichester, 2001; 741-764.
10. Whitehead J. *The Design and Analysis of Sequential Clinical Trials* (revised 2nd edn). Wiley: Chichester, 1997.
11. Morrison J, Gauderman WJ. *Quanto, version 0.4*. University of Southern California, 2002.
12. The University of Reading. MPS Research Unit. *PEST4: Operating Manual*. Wiley: Chichester, 2000.
13. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls! *American Journal of Epidemiology* 1996; **144**:207-213.
14. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *American Journal of Epidemiology* 2002; **155**:478-484.
15. Ury HK. Efficiency of case-control studies with multiple controls per case: continuous or dichotomous data. *Biometrics* 1975; **31**:643-649.