Web-based Supplementary Materials for
"Rapid Testing of SNPs and Gene-Environment Interactions in Case-Parent Trio Data Based on Exact Analytic Parameter Estimation"

by

Holger Schwender¹,²,*
Margaret A. Taub²
Terri H. Beaty³
Mary L. Marazita⁴
Ingo Ruczinski²

¹ Faculty of Statistics, TU Dortmund University, 44221 Dortmund, Germany
² Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205, USA
³ Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland 21205, USA
⁴ School of Dental Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15219, USA

*email: holger.schwender@udo.edu

Web Appendix A: Derivation of the Analytic Estimates for the Dominant Model

As discussed in the main text to this Supplementary Material, the first derivative of the log-likelihood of a conditional logistic regression model testing for a dominant effect is given by

\[ \frac{\partial \ell}{\partial \beta_{\text{dom}}} = d_2 - (d_2 + d_1) \times \frac{\exp(\beta_{\text{dom}})}{\exp(\beta_{\text{dom}}) + 1} + d_4 - (d_4 + d_3) \times \frac{\exp(\beta_{\text{dom}})}{\exp(\beta_{\text{dom}}) + 1/3}, \]

where the numbers \(d_1, j = 1, \ldots, 4\), of case-parent trios are defined in Table 2 of the main text.

In the following, we set \(z = \exp(\beta_{\text{dom}})\) for an easier notation in the derivation of the maximum-likelihood estimate for the parameter \(\beta_{\text{dom}}\), so that this first derivative becomes

\[ \frac{\partial \ell(\beta_{\text{dom}})}{\partial \beta_{\text{dom}}} = d_2 - (d_2 + d_1) \frac{z}{z + 1} + d_4 - (d_4 + d_3) \frac{z}{z + 1/3}. \]

Setting this derivative to zero leads to

\begin{align*}
(d_2 + d_1) \frac{z}{z + 1} + (d_4 + d_3) \frac{z}{z + 1/3} &= d_2 + d_4 \\
\iff (d_2 + d_1)z \left( \frac{z + 1}{3} \right) + (d_4 + d_3)z(z + 1) &= (d_2 + d_4)(z + 1) \left( \frac{z + 1}{3} \right) \\
\iff (d_1 + d_3)z^2 + \left( \frac{1}{3}d_1 - d_2 - d_3 + \frac{1}{3}d_4 \right)z &= \frac{1}{3}(d_2 + d_4). 
\end{align*}

Completing the square by adding

\[ h_{\text{dom}}^2 = \frac{1/3d_1 - d_2 + d_3 - 1/3d_4}{2(d_1 + d_3)} \]

to both sides of this equation results in

\[ (z + h_{\text{dom}})^2 = \frac{d_2 + d_4}{3(d_1 + d_3)} + h_{\text{dom}}^2, \]

and solving this equation for \(z\) leads to

\[ z = \pm \sqrt{\frac{d_2 + d_4}{3(d_1 + d_3)} + h_{\text{dom}}^2 - h_{\text{dom}}}. \]

Since the square-root term is always larger than or equal to \(h_{\text{dom}}\), there exists exactly one solution for \(\beta_{\text{dom}} = \log(z)\), unless \(d_2 + d_4 = 0\) or \(d_1 + d_3 = 0\), in which cases no maximum of the log-likelihood exists (see Li et al., 2009, for a related discussion). The maximum-likelihood estimator for \(\beta_{\text{dom}}\) is thus given by

\[ \hat{\beta}_{\text{dom}} = \log \left( \frac{\sqrt{d_2 + d_4}}{\sqrt{3(d_1 + d_3)} + h_{\text{dom}}^2} - h_{\text{dom}} \right) \]

with a variance estimated by the negative inverse of

\[ \frac{\partial ^2 \ell(\hat{\beta}_{\text{dom}})}{\partial \beta_{\text{dom}}^2} = -\left( d_2 + d_1 \right) \frac{\exp(\hat{\beta}_{\text{dom}})}{\exp(\hat{\beta}_{\text{dom}}) + 1} \]

\[ -\left( d_4 + d_3 \right) \frac{\exp(\hat{\beta}_{\text{dom}})}{3\left( \exp(\hat{\beta}_{\text{dom}}) + 1/3 \right)^2}. \]
Web Appendix B: Derivation of the Parameter Estimates for Testing Gene-Environment Interactions

When fitting a conditional logistic regression model

$$\beta_G X + \beta_{GE} (X \times E) \quad (A.1)$$

for testing an interaction between a SNP and a binary environmental variable $E$ (coded by zero and one), the contribution of a case-parent trio for which $e = 0$ to the conditional likelihood of this model is given by $w_j(\beta_G)$, and the weight of a trio for which $e = 1$ is $w_j(\beta_G + \beta_{GE})$, where — depending on whether $X$ codes for an additive, a dominant, or a recessive effect of the SNP — the values of the weight functions $w_j$ are specified either in Table 1, 2, or 3 of the main text. Denoting the numbers of trios showing the same case-pseudo-control combination and the same value $e \in \{0, 1\}$ of the environmental variable by $a_j^{(e)}$, the log-likelihood $\ell_{GXE}$ of the model (A.1) is — analogously to the log-likelihood (2) in the main text — given by

$$\ell_{GXE}(\beta_G, \beta_{GE}) = \sum_{j=1}^{10} a_j^{(0)} \log(w_j(\beta_G))
\quad + \sum_{j=1}^{10} a_j^{(1)} \log(w_j(\beta_G + \beta_{GE})).$$

Since

$$\ell(\beta^{(0)}) = \sum a_j^{(0)} \log(w_j(\beta_G))$$

and

$$\ell(\beta^{(1)}) = \sum a_j^{(1)} \log(w_j(\beta_G + \beta_{GE}))$$

with

$$\beta^{(1)} = \beta_G + \beta_{GE}, \quad (A.2)$$

and $\ell$ being the likelihood of the conditional logistic regression model considering only the main effect of $X$, it follows that

$$\frac{\partial \ell_{GXE}(\beta_G, \beta_{GE})}{\partial \beta_G} = \frac{\partial \ell(\beta^{(0)})}{\partial \beta_G} + \frac{\partial \ell(\beta^{(1)})}{\partial \beta_G}, \quad (A.3)$$

and

$$\frac{\partial \ell_{GXE}(\beta_G, \beta_{GE})}{\partial \beta_{GE}} = \frac{\partial \ell(\beta^{(1)})}{\partial \beta_{GE}}. \quad (A.4)$$

Since (A.4) is set to zero to determine the maximum-likelihood estimate for $\beta_{GE}$, (A.3) becomes

$$\frac{\partial \ell_{GXE}(\beta_G, \beta_{GE})}{\partial \beta_G} = \frac{\partial \ell(\beta^{(0)})}{\partial \beta_G}.$$

The maximum-likelihood estimate of $\beta_G$ is thus given by

$$\hat{\beta}_G = \beta^{(0)}.$$

With (A.2), it follows that

$$\hat{\beta}_{GE} = \beta^{(1)} - \beta^{(0)}.$$

The variances of $\hat{\beta}_G$ and $\hat{\beta}_{GE}$ can then be estimated by

$$\text{Var} \left( \hat{\beta}_G \right) = \text{Var} \left( \beta^{(0)} \right)$$

and

$$\text{Var} \left( \hat{\beta}_{GE} \right) = \text{Var} \left( \beta^{(1)} \right) + \text{Var} \left( \beta^{(0)} \right).$$

Web Appendix C: Testing Gene-Gene Interactions

C.1 Tests for Interactions

There are several (conditional) logistic regression models that can be used to detect two-way SNP interactions associated with disease. The simplest approach, which might be used for screening for interesting interactions, is to consider the model

$$\beta_{GC}(X_1 \times X_2), \quad (A.5)$$

where $X_1$ codes for either an additive, dominant, or recessive effect of one of the SNPs, and $X_2$ codes for one of these three effects for the other SNP. A more sophisticated model that allows to distinguish between a main effect of (at least) one of the SNPs and an interaction effect is

$$\beta_1 X_1 + \beta_2 X_2 + \beta_{GC}(X_1 \times X_2). \quad (A.6)$$

Alternatively, a likelihood ratio test might be employed to compare the likelihood of model (A.6) with the likelihood of a model only containing main effects. An extension of this procedure is the likelihood ratio test for epistatic interactions proposed by Cordell (2002), which compares the likelihood of

$$\eta_1 V_1 + \delta_1 Z_1 + \eta_2 V_2 + \delta_2 Z_2 + \gamma_1 (V_1 \times V_2) + \gamma_2 (V_1 \times Z_2) + \gamma_3 (Z_1 \times V_2) + \gamma_4 (Z_1 \times Z_2)$$

with the likelihood of

$$\eta_1 V_1 + \delta_1 Z_1 + \eta_2 V_2 + \delta_2 Z_2,$$

where values of $V_s$ are given by substracting 1 from the numbers of minor alleles shown at SNP $s$, $s = 1, 2$, and $Z_s = 0.5 - |V_s|$. One problem of testing two-way SNP interactions in case-parent trio data with a gTDT is that for each trio the genotypes of 15 pseudo-controls must be taken into account, as there are 16 combinations of the four genotype realizations possible at SNP 1 given the parents’ genotypes with the four possible genotype realizations at SNP 2, where the affected offspring shows one of these 16 genotype pairs.

C.2 Model $\beta_{GC}(X_1 \times X_2)$

If conditional logistic regression models of form (A.5) should be fitted in the conventional way, it becomes necessary to determine all these 15 pseudo-controls for each of the $n$ trios.
The pairs of case-pseudo-control-combinations that influence the maximization of the log-likelihood (A.8) of a recessive \times recessive model of type (A.5), their weights \( w_{jh} \) in this log-likelihood, and the numbers \( r_{jh} \) of trios showing the respective combination pairs, where the subscripts of the numbers \( r_{jh} \) refer to the subscripts of the numbers of trios in Table 3 of the main text to this supplementary material. For conciseness, the weights have been split into their denominator and numerator so that the table does not contain several rows for weights with the same denominator.

<table>
<thead>
<tr>
<th>Pair of Case-Control-Combination</th>
<th>Weight ( w_{jh} ) in Likelihood</th>
<th>Number of Trios: Numerator</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0, 1, 1)</td>
<td>(0, 0, 1, 1)</td>
<td>( 12 + 4 \exp(\beta) )</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>(0, 0, 0, 1)</td>
<td>( 14 + 2 \exp(\beta) )</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>(1, 1, 1, 1)</td>
<td>( 8 + \exp(\beta) )</td>
</tr>
<tr>
<td>(0, 0, 0, 1)</td>
<td>(0, 0, 0, 1)</td>
<td>( 15 + \exp(\beta) )</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>(1, 1, 1, 1)</td>
<td>( 12 + 4 \exp(\beta) )</td>
</tr>
</tbody>
</table>

so that the parameter \( \beta_{GG} \) of this model can be estimated by maximizing the likelihood

\[
L(\beta_{CG}) = \prod_{i=1}^{n} \left( \frac{\exp(\beta x_{1i0}x_{2i0})}{\sum_{k=0}^{15} \exp(\beta x_{1ik}x_{2ik})} \right), \tag{A.7}
\]

where \( x_{skh} \) is the value of \( X_s \) coding for a certain genetic effect of SNP \( s \) shown by the case \( (k = 0) \) or by one of the pseudo-controls \( (k = 1, \ldots, 15) \) in trio \( i = 1, \ldots, n \).

Analogously to the gTDTs for testing individual SNPs and gene-environment interactions, the likelihood (A.7) can be transformed into a log-likelihood of the form

\[
\ell(\beta_{CG}) = \sum a_{jh} \log(w_{jh}(\beta_{CG})), \tag{A.8}
\]

where \( a_{jh} \) are the numbers of trios showing genotype combination \( j \) of one case and three pseudo-controls at one of the SNPs and case-pseudo-control combination \( h \) at the other SNP (for these case-pseudo-control combinations in an additive, dominant, or recessive model, see Table 1, 2, or 3, respectively, in the main text to this Supplementary Material). Further, \( w_{jh}(\beta_{CG}) \) are the corresponding weights in the likelihood.

Since for the value of \( w_{jh}(\beta_{CG}) \) it does not matter which SNP shows combination \( j \) and which combination \( h \), the numbers \( a_{jh} \) are computed by counting how many trios belong to combination \( j \) at SNP 1 and combination \( h \) at SNP 2, and adding the number of trios showing combination \( h \) at SNP 1 and combination \( j \) at SNP 2. We therefore only consider numbers \( a_{jh} \) of trios and weights \( w_{jh}(\beta_{CG}) \) with \( j \leq h \). Moreover, employing this approach avoids the generation of the 15 pseudo-controls per trio.

In Web Tables 1, 2, and 3, the numbers of trios and weights are summarized for a recessive \times recessive model (i.e. a conditional logistic regression model in which for both SNPs a recessive effect is considered), a dominant \times dominant model, and an additive \times additive model, respectively, where the subscript \( j \) and \( h \) in these tables refer to the subscripts used in the corresponding Tables 3, 2, and 1, respectively, in the main text to this Supplementary Material. If we, for example, consider an additive \times additive model with \( j = 3 \) and \( h = 5 \), i.e. the case-pseudo-control combination \((1, 1, 2, 2)\) at one of the SNPs, and the combination \((0, 1, 1, 2)\) at the other SNP (see Table 1 in the main text to this Supplementary Material), then the weight \( w_{35}(\beta_{CG}) \) of each of the \( a_{35} \) trios showing this pair of case-pseudo-control combinations is determined by

\[
w_{35}(\beta_{CG}) = \frac{1}{4 + 4 \exp(2\beta_{CG}) + 2 \exp(4\beta_{CG})},
\]

as the 16 possible pairs of genotype realizations at the two SNPs are

\[(1, 0), (1, 0), (2, 0), (2, 0), (1, 1), (1, 1), (2, 1), (2, 1), (1, 1), (2, 1), (2, 1), (1, 2), (1, 2), (2, 2), (2, 2), \]

and the affected offspring exhibits the first of these pairs. For Web Tables 1, 2, and 3, note that the numerator of the weights in the likelihood (A.7) only depend on the genotypes of the affected offspring, while the denominator is a sum over all 16 possible pairs of genotypes.

Because the recessive \times recessive model has the smallest number of weights (which are summarized in Web Table 1), we start with this model. Using the above notation (i.e. denoting the number of trios showing the \( j \)th case-pseudo-control combination from Table 3 of the main text to this Supplementary Material at one of the SNPs, and the \( h \)th combination at the other SNP by \( r_{jh} \)), the log-likelihood (A.8) becomes

\[
\ell_r(\beta_{CG}) = \beta_{CG} \times (r_{22} + r_{24} + r_{25} + r_{45}) + \log(12 + 4 \exp(\beta_{CG})) \times (r_{11} + r_{12} + r_{22} + r_{35} + r_{45})
\]

\[
- \log(14 + 2 \exp(\beta_{CG})) \times (r_{13} + r_{14} + r_{23} + r_{24})
\]

\[
- \log(8 + \exp(\beta_{CG})) \times (r_{15} + r_{25})
\]

\[
- \log(15 + \exp(\beta_{CG})) \times (r_{33} + r_{34} + r_{44})
\]

\[
- \log(16) \times r_{55} + \sum_{p=1}^{6} r_{pq}
\]

for a recessive \times recessive model. Differentiating this log-
Web Table 2
The pairs of case-pseudo-control-combinations that influence the maximization of the log-likelihood (A.8) of a dominant × dominant model of type (A.5), their weights \( w_{jk} \) in this log-likelihood, and the numbers \( d_{jk} \) of trios showing the respective combination pairs, where the subscripts of the numbers \( d_{jk} \) refer to the subscripts of the numbers of trios in Table 2 of the main text to this supplementary material. For conciseness, the weights have been split into their denominator and numerator so that the table does not contain several rows for weights with the same denominator.

<table>
<thead>
<tr>
<th>Pair of Case-Control-Combination</th>
<th>Weight ( w_{jk} ) in Likelihood</th>
<th>Denominator</th>
<th>Number of Trios: Numerator</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0, 0, 1, 1)) (0, 0, 1, 1)</td>
<td>(12 + 4 \exp(\beta))</td>
<td>(d_{11}, d_{12} : 1; d_{22} : \exp(\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 0, 1, 1)) (0, 1, 1, 1)</td>
<td>(10 + 6 \exp(\beta))</td>
<td>(d_{13}, d_{14}, d_{23} : 1; d_{24} : \exp(\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 0, 1, 1)) (1, 1, 1, 1)</td>
<td>(8 + 8 \exp(\beta))</td>
<td>(d_{16} : 1; d_{26} : \exp(\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 1, 1, 1)) (0, 1, 1, 1)</td>
<td>(7 + 9 \exp(\beta))</td>
<td>(d_{33}, d_{34} : 1; d_{44} : \exp(\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 1, 1, 1)) (1, 1, 1, 1)</td>
<td>(4 + 12 \exp(\beta))</td>
<td>(d_{36} : 1; d_{46} : \exp(\beta))</td>
<td></td>
</tr>
</tbody>
</table>

Web Table 3
The pairs of case-pseudo-control-combinations that influence the maximization of the log-likelihood (A.8) of an additive × additive model of type (A.5), their weights \( w_{jk} \) in this log-likelihood, and the numbers \( a_{jk} \) of trios showing the respective combination pairs, where the subscripts of the numbers \( a_{jk} \) refer to the subscripts of the numbers of trios in Table 1 of the main text to this supplementary material. For conciseness, the weights have been split into their denominator and numerator so that the table does not contain several rows for weights with the same denominator.

<table>
<thead>
<tr>
<th>Pair of Case-Control-Combination</th>
<th>Weight ( w_{jk} ) in Likelihood</th>
<th>Denominator</th>
<th>Number of Trios: Numerator</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0, 0, 1, 1)) (0, 0, 1, 1)</td>
<td>(12 + 4 \exp(\beta))</td>
<td>(a_{11}, a_{12} : 1; a_{22} : \exp(\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 0, 1, 1)) (1, 1, 1, 1)</td>
<td>(8 + 4 \exp(\beta) + 4 \exp(2\beta))</td>
<td>(a_{13}, a_{14} : 1; a_{23} : \exp(\beta); a_{24} : \exp(2\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 0, 1, 1)) (0, 1, 1, 2)</td>
<td>(10 + 4 \exp(\beta) + 2 \exp(2\beta))</td>
<td>(a_{15}, a_{16}, a_{17}, a_{25} : 1; a_{26} : \exp(\beta); a_{27} : \exp(2\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 0, 1, 1)) (1, 1, 1, 1)</td>
<td>(8 + 8 \exp(\beta))</td>
<td>(a_{18} : 1; a_{28} : \exp(\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 0, 1, 1)) (1, 2, 2, 2)</td>
<td>(8 + 8 \exp(2\beta))</td>
<td>(a_{19} : 1; a_{29} : \exp(2\beta))</td>
<td></td>
</tr>
<tr>
<td>((1, 1, 2, 2)) (1, 1, 2, 2)</td>
<td>(4 \exp(\beta) + 8 \exp(2\beta) + 4 \exp(4\beta))</td>
<td>(a_{33} : \exp(\beta); a_{34} : \exp(2\beta); a_{44} : \exp(4\beta))</td>
<td></td>
</tr>
<tr>
<td>((1, 1, 2, 2)) (0, 1, 1, 2)</td>
<td>(4 + 4 \exp(\beta) + 6 \exp(2\beta) + 2 \exp(4\beta))</td>
<td>(a_{35}, a_{35} : 1; a_{36} : \exp(\beta); a_{37}, a_{46} : \exp(2\beta); a_{47} : \exp(4\beta))</td>
<td></td>
</tr>
<tr>
<td>((1, 1, 2, 2)) (1, 1, 1, 1)</td>
<td>(8 \exp(\beta) + 8 \exp(2\beta))</td>
<td>(a_{38} : \exp(\beta); a_{48} : \exp(2\beta))</td>
<td></td>
</tr>
<tr>
<td>((1, 1, 2, 2)) (2, 2, 2, 2)</td>
<td>(8 \exp(2\beta) + 8 \exp(4\beta))</td>
<td>(a_{39} : \exp(2\beta); a_{49} : \exp(4\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 1, 1, 2)) (0, 1, 1, 2)</td>
<td>(7 + 4 \exp(\beta) + 4 \exp(2\beta) + 4 \exp(4\beta))</td>
<td>(a_{55}, a_{56}, a_{57} : 1; a_{66} : \exp(\beta); a_{67} : \exp(2\beta); a_{77} : \exp(4\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 1, 1, 2)) (1, 1, 1, 1)</td>
<td>(4 + 8 \exp(\beta) + 4 \exp(2\beta))</td>
<td>(a_{58} : 1; a_{68} : \exp(\beta); a_{78} : \exp(2\beta))</td>
<td></td>
</tr>
<tr>
<td>((0, 1, 1, 2)) (2, 2, 2, 2)</td>
<td>(4 + 8 \exp(2\beta) + 4 \exp(4\beta))</td>
<td>(a_{59} : 1; a_{69} : \exp(2\beta); a_{79} : \exp(4\beta))</td>
<td></td>
</tr>
</tbody>
</table>
likelihood results in
\[
\frac{\partial \ell_n (\beta_{GC})}{\partial \beta_{GC}} = \frac{r_{22} + r_{24} + r_{25} + r_{44} + r_{45}}{n_0} - \frac{\exp (\beta_{GC})}{3 + \exp (\beta_{GC})} \left( \frac{r_{11} + r_{12} + r_{22} + r_{35} + r_{45}}{n_1} \right)
\]
\[
- \frac{\exp (\beta_{GC})}{7 + \exp (\beta_{GC})} \left( \frac{r_{13} + r_{14} + r_{23} + r_{24}}{n_2} \right)
\]
\[
- \frac{\exp (\beta_{GC})}{1 + \exp (\beta_{GC})} \left( \frac{r_{15} + r_{25}}{n_3} \right)
\]
\[
- \frac{\exp (\beta_{GC})}{15 + \exp (\beta_{GC})} \left( r_{33} + r_{34} + r_{44} \right).
\]  
(A.9)

Denoting, similar to Web Appendix A, \(z = \exp (\beta_{GC})\), and setting the first derivative (A.9) to zero leads to
\[
\frac{z}{3 + z} n_1 + \frac{z}{7 + z} n_2 + \frac{z}{15 + z} n_4 = n_0
\]
\[
\Leftrightarrow \frac{z}{3 + z} (7 + z)(15 + z)n_1 + z(3 + z)(15 + z)n_2 + z(3 + z)(7 + z)(15 + z)n_4 = (3 + z)(7 + z)(15 + z)n_0
\]
\[
\Leftrightarrow c_4 x^4 + c_3 x^3 + c_2 x^2 + c_1 x = 0
\]
with
\[
c_0 = 315 n_0
\]
\[
c_1 = 105 n_1 + 45 n_2 + 315 n_3 + 21 n_4 - 486 n_0
\]
\[
c_2 = 127 n_1 + 63 n_2 + 171 n_3 + 31 n_4 - 196 n_0
\]
\[
c_3 = 23 n_1 + 19 n_2 + 25 n_3 + 11 n_4 - 26 n_0
\]
\[
c_4 = \sum_{j=1}^{4} \sum_{k=1}^{5} r_{j,k} - n_0.
\]

The roots of this polynomial
\[
c_4 x^4 + c_3 x^3 + c_2 x^2 + c_1 x - c_0 = 0
\]
can then be determined analytically using a standard procedure for this purpose (see Web Appendix G). Since the first derivative (A.9) is continuous and monotonically decreasing in \(\beta_{GC}\) with
\[
\lim_{\beta_{GC} \to -\infty} \frac{\partial \ell_n (\beta_{GC})}{\partial \beta_{GC}} = n_0 \geq 0
\]
and
\[
\lim_{\beta_{GC} \to \infty} \frac{\partial \ell_n (\beta_{GC})}{\partial \beta_{GC}} = -c_4 \leq 0,
\]

there exists one real-valued root for the first derivative (A.9), and thus, one maximum for \(\ell_n (\beta_{GC})\), unless \(n_0 = 0\) or \(c_4 = 0\), i.e. unless none or all of the children are homozygous for the variant allele at both loci.

Such an analytic solution does not exist when considering a dominant \(\times\) dominant model of the form (A.5), as the first derivative of the corresponding log-likelihood
\[
\ell_{4d} (\beta_{GC}) = \beta_{GC} \times (d_{22} + d_{24} + d_{25} + d_{44} + d_{45})
\]
\[
- \log (12 + 4 \exp (\beta_{GC})) \times (d_{11} + d_{12} + d_{22})
\]
\[
- \log (10 + 6 \exp (\beta_{GC})) \times (d_{13} + d_{14} + d_{23} + d_{24})
\]
\[
- \log (8 + 8 \exp (\beta_{GC})) \times (d_{16} + d_{26})
\]
\[
- \log (7 + 9 \exp (\beta_{GC}) \times (d_{33} + d_{44})
\]
\[
- \log (4 + 12 \exp (\beta_{GC}) \times (d_{36} + d_{46})
\]
\[
- \log (16) \times \left( d_{66} + \sum_{p=1}^{6} d_{pG} \right)
\]

(cf. Web Table 2), which is given by
\[
\frac{\partial \ell_{4d} (\beta_{GC})}{\partial \beta_{GC}} = d_{22} + d_{24} + d_{25} + d_{44} + d_{45}
\]
\[
- \frac{\exp (\beta_{GC})}{3 + \exp (\beta_{GC})} (d_{11} + d_{12} + d_{22})
\]
\[
- \frac{\exp (\beta_{GC})}{5/3 + \exp (\beta_{GC})} (d_{13} + d_{14} + d_{23} + d_{24})
\]
\[
- \frac{\exp (\beta_{GC})}{1 + \exp (\beta_{GC})} (d_{16} + d_{26})
\]
\[
- \frac{\exp (\beta_{GC})}{7/9 + \exp (\beta_{GC})} (d_{33} + d_{34} + d_{44})
\]
\[
- \frac{\exp (\beta_{GC})}{1/3 + \exp (\beta_{GC})} (d_{36} + d_{46}),
\]

consists of five ratios of the form \(z/(b + z)\) leading to a polynomial of fifth degree. The roots of this polynomial and thus \(\beta_{GC}\) cannot be determined analytically, as no general procedure for finding the roots of a polynomial of a degree larger than 4 exists.

Similarly, the maximum-likelihood estimate \(\hat{\beta}_{GC}\) for an additive \(\times\) additive model needs to be determined analytically, where the log-likelihood (A.8) of this model has even a higher complexity than the log-likelihood for the dominant \(\times\) dominant model, Setting
\[
c_{pair} = a_{22} + a_{23} + 2a_{24} + 2a_{26} + 2a_{27} + a_{28} + 2a_{29}
\]
\[
+ a_{33} + 2a_{34} + a_{36} + 2a_{37} + a_{38} + 2a_{39} + 4a_{44}
\]
\[
+ 2a_{46} + 4a_{47} + 2a_{48} + 4a_{49} + a_{66} + 2a_{67} + a_{68}
\]
\[
+ 2a_{69} + 4a_{77} + 2a_{78} + 4a_{79}
\]

and ignoring trios exhibiting the tenth case-pseudo-control combination from Table 1 in the main text to this Supplementary Material, i.e. \((0,0,0,0)\), at least one of the two SNPs, which do not influence the maximization of the log-likelihood, the first derivative of this (reduced) log-likelihood
\[
\ell_{4a}^*(\beta_{GC}) = \beta_{GC} \times c_{pair}
\]
\[
- \log (12 + 4 \exp (\beta_{GC}) \times (a_{11} + a_{12} + a_{22})
\]
\[
- \log (8 + 8 \exp (\beta_{GC}) + 4 \exp (2\beta_{GC})) \times \sum_{j=1}^{4} \sum_{k=1}^{5} a_{j,k}
\]
\[ -\log(10 + 4 \exp(\beta_{GC} + \exp(2\beta_{GC}))) \times \sum_{j=1}^{2} \sum_{h=5}^{7} a_{jh} \]
\[ -\log(8 + 8 \exp(\beta_{GC})) \times (a_{18} + a_{28}) \]
\[ -\log(8 + 8 \exp(2\beta_{GC})) \times (a_{19} + a_{29}) \]
\[ -\log(4 \exp(\beta_{GC}) + 8 \exp(2\beta_{GC}) + 4 \exp(4\beta_{GC})) \times (a_{33} + a_{34} + a_{44}) \]
\[ -\log(4 + 4 \exp(\beta_{GC}) + 6 \exp(2\beta_{GC}) + 2 \exp(4\beta_{GC})) \times \sum_{j=1}^{4} \sum_{h=5}^{7} a_{jh} \]
\[ -\log(8 \exp(\beta_{GC}) + 8 \exp(2\beta_{GC})) \times (a_{38} + a_{48}) \]
\[ -\log(8 \exp(2\beta_{GC}) + 8 \exp(4\beta_{GC})) \times (a_{39} + a_{49}) \]
\[ -\log(7 + 4 \exp(\beta_{GC}) + 4 \exp(2\beta_{GC}) + \exp(4\beta_{GC})) \times \sum_{j=5}^{6} \sum_{h=5}^{7} a_{jh} \]
\[ -\log(4 + 8 \exp(\beta_{GC}) + 4 \exp(2\beta_{GC})) \times \sum_{j=5}^{7} a_{18} \]
\[ -\log(4 + 8 \exp(2\beta_{GC}) + 4 \exp(4\beta_{GC})) \times \sum_{j=5}^{7} a_{49} \]
\[ -\log(16) \times (a_{38} + a_{49} + a_{49}) \]
is given by
\[ \frac{\partial^2 \log L}{\partial \beta_{GC}^2} = \sum_{j=5}^{7} a_{jh} \]
\[ \frac{\exp(\beta_{GC}) + 2 \exp(2\beta_{GC}) + 4 \exp(4\beta_{GC})}{1.75 + \exp(\beta_{GC}) + \exp(2\beta_{GC}) + 0.25 \exp(4\beta_{GC})} \times \sum_{j=5}^{7} a_{jh} \]
\[ \frac{\exp(\beta_{GC}) + \exp(2\beta_{GC}) + 0.5 \exp(3\beta_{GC}) + 0.5 \exp(4\beta_{GC})}{0.5 + \exp(\beta_{GC}) + 0.5 \exp(2\beta_{GC}) + \exp(4\beta_{GC})} \times \sum_{j=5}^{7} a_{jh} \]
\[ \frac{\exp(2\beta_{GC}) + \exp(4\beta_{GC})}{0.25 + 0.5 \exp(2\beta_{GC}) + 0.25 \exp(4\beta_{GC})} \times \sum_{j=5}^{7} a_{jh} \]

(cf. Web Table 3). When set to zero, this first derivative cannot be solved for \(\beta_{GC}\) analytically so that no closed-form solution for \(\beta_{GC}\) exists in the additive × additive model.

Besides testing the same genetic effect for both SNPs in a two-way interaction, it might also be of interest to consider different genetic modes of inheritance for these two SNPs. In these situations, it can no longer be ignored which SNP shows which case-pseudo-control combination, in particular, when computing the numbers of trios showing a certain case-pseudo-control combination at the first SNP and another, differently coded case-pseudo-control combination at the second SNP. Using the weights from Web Table 4 and the numbers of trios exhibiting the corresponding case-pseudo-control combinations at the two SNPs, log-likelihoods of the form (A.8) and their first derivatives can be derived for an additive × dominant, an additive × recessive, and a dominant × recessive model. Analogously, the log-likelihoods and their first derivative for a dominant × additive, a recessive × additive, and a recessive × dominant model can be obtained. In any of these models, however, the maximum-likelihood estimator for \(\beta_{GC}\) needs to be determined numerically, as the log-likelihood and the first derivative is composed of weights/ratios with 12, 11, or 6 different denominators, when considering an additive × dominant, an additive × recessive, or a dominant × recessive model, respectively.

C.3 MAX gTDT for Interactions

As in the analysis of individual SNPs or gene-environment interactions, it is also possible to use the maximum over the values of the gTDT statistics for the nine possible models for testing a two-way interaction as test statistic. For the computation of permutation-based p-values for this MAX test, the procedure presented in Section 4 of the main text to this Supplementary Material can be adapted to test two-way interactions by drawing random numbers from several appropriate binomial and multinomial distributions with different parameter values that depend on the distributions of the genotypes in the different pairs of case-pseudo-control combinations as shown in Table 3.

If, for example, we consider the second pair of case-pseudo-control combinations in Table 3, i.e. \((0, 0, 1, 1)\) and \((1, 1, 2, 2)\), then under the null hypothesis the probability that for the affected offspring \(x_{10}x_{20} = 0\) is 8/16 = 0.5, while the probabilities for \(x_{10}x_{20} = 1\) and \(x_{10}x_{20} = 2\) are each 0.25.

Depending on whether the maximum gTDT value should be computed only over the three models considering the same genetic mode of inheritance for both SNPs or over all nine interaction models, it again can or cannot be ignored which...
SNP shows which case-pseudo-control combination. If thus a SNP 2, then we can draw from a binomial/multinomial distribution with the conditional logistic regression model (A.6) consisting of complex when considering more sophisticated models such as the conditional logistic regression model of type (A.6), i.e. add×dom, add×rec, dom×rec.

The pairs of case-pseudo-control combinations influencing the maximization of the log-likelihood (A.8) of an additive × dominant, an additive × recessive, or dominant × recessive model of type (A.5), the denominator of their weights in this log-likelihood, and the types of model in which the weights are used.

<table>
<thead>
<tr>
<th>Combination</th>
<th>Denominator of the weight in the Likelihood</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0, 1, 1)</td>
<td>12 + 4 exp(β)</td>
<td>add×dom, add×rec, dom×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>10 + 6 exp(β)</td>
<td>add×dom</td>
</tr>
<tr>
<td>(0, 0, 0, 1)</td>
<td>14 + 2 exp(β)</td>
<td>add×rec, dom×rec</td>
</tr>
<tr>
<td>(1, 1, 2, 2)</td>
<td>8 + 8 exp(β)</td>
<td>add×dom, add×rec, dom×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>8 + 4 exp(β) + 4 exp(2β)</td>
<td>add×dom, add×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>4 + 6 exp(β) + 6 exp(2β)</td>
<td>add×dom</td>
</tr>
<tr>
<td>(1, 1, 1, 1)</td>
<td>12 + 2 exp(β) + 2 exp(2β)</td>
<td>add×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>8 exp(β) + 8 exp(2β)</td>
<td>add×dom, add×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>10 + 4 exp(β) + 2 exp(2β)</td>
<td>add×dom, add×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>7 + 6 exp(β) + 3 exp(2β)</td>
<td>add×dom</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>13 + 2 exp(β) + exp(2β)</td>
<td>add×rec</td>
</tr>
<tr>
<td>(1, 1, 1, 1)</td>
<td>4 + 8 exp(β) + 4 exp(2β)</td>
<td>add×dom, add×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>10 + 6 exp(β)</td>
<td>dom×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>13 + 3 exp(β)</td>
<td>dom×rec</td>
</tr>
<tr>
<td>(1, 1, 1, 1)</td>
<td>4 + 12 exp(β)</td>
<td>dom×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>10 + 6 exp(β)</td>
<td>add×dom, add×rec, dom×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>4 + 12 exp(β)</td>
<td>add×dom</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>12 + 4 exp(β)</td>
<td>add×rec, dom×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>8 + 8 exp(2β)</td>
<td>add×dom, add×rec</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>4 + 12 exp(2β)</td>
<td>add×dom</td>
</tr>
<tr>
<td>(0, 0, 1, 1)</td>
<td>12 + 4 exp(2β)</td>
<td>add×rec</td>
</tr>
</tbody>
</table>

SNP shows which case-pseudo-control combination. If thus a \(a_{j}^{h}\) denotes the number of trios showing the \(j\)th case-pseudo-control combination at SNP 1 and the \(h\)th combination at SNP 2, then we can draw from a binomial/multinomial distribution with a \(a_{j}^{h}\) and \(a_{h}^{j}\) observations, if only the three models are considered in the MAX gTDT. Otherwise, we have to draw once from a distribution with a \(a_{j}^{h}\) observations, and once from a distribution with a \(a_{h}^{j}\) observations.

C.4 Model \(β_{1}X_{1} + β_{2}X_{2} + β_{G}G(X_{1} × X_{2})\)

The log-likelihood and its maximization become even more complex when considering more sophisticated models such as the conditional logistic regression model (A.6) consisting of three parameters (two for the main effects and one for the interaction) or the likelihood ratio test of Cordell (2002). In these models, not only 15 pseudo-controls per trio have to be taken into account, but the (log-)likelihood must also be maximized over three or more parameters. As exemplified in Web Appendix D considering a discordant sib-design, the latter can even make it impossible to derive analytic estimates when a conditional logistic regression model with three parameters and a 1:3 matching should be fitted.

We therefore only briefly discuss a conditional logistic regression model of type (A.6), i.e. \(β_{1}X_{1} + β_{2}X_{2} + β_{G}G(X_{1} × X_{2})\), in which both \(X_{1}\) and \(X_{2}\) code for a recessive effect of one of the two SNPs. Denoting the number of trios showing the
The first derivatives with respect to $\beta_1$, $\beta_2$, and $\beta_{GG}$ can hence be determined by
respectively.

For the existence of closed-form solutions for the maximum-likelihood estimates of $\beta_1$, $\beta_2$, and $\beta_{GE}$, it is necessary that – after setting the three derivatives to zero – two of these parameters can be eliminated from one of the equations (by, for example, inserting the other equations set to zero into this equation) so that this equation can be solved for the remaining parameter. This, however, does not seem to be possible in the first derivatives of the log-likelihood $\ell_i$, in particular, because of several denominators composed of the form

$$a + b \exp(\beta_1) + c \exp(\beta_2) + d \exp(\beta_1 + \beta_2 + \beta_{GE})$$

with $a, b, c, d \in \mathbb{N}$. Hence, the maximum-likelihood estimators for $\beta_1$, $\beta_2$, and $\beta_{GE}$ must be determined numerically.

Web Appendix D: Discordant Sib-Design under Gene-Environment Independence

Besides case-parent trio designs, there exist other family-based designs to which a conditional logistic regression can be applied, as, for example, in these designs each case is matched with a sibling or a cousin that does not show the disease affecting the case. Chatterjee et al. (2005), for example, consider a discordant sib-design in which the sibling (or cousin) shows another genotype at the SNP as the corresponding case, and propose a conditional logistic regression model for this design under the assumption that for pairs of relatives within each family the joint distribution of a SNP and the environmental variable is independent. Since – as we will see the following – their model is also based on a 1:3 matching, and in particular, since Chatterjee et al. (2005) show that their family-based case-control design can be more efficient than the case-parent trio design, we evaluate in the following whether it is also possible to derive closed-form maximum-likelihood estimates for the parameters in their model.

In contrast to the gTDT for gene-environment interactions (see Section 3 of the main text to this Supplementary Material), the value of the environmental variable can differ between the case and the matched control in the model of Chatterjee et al. (2005) so that this model is given by

$$\beta_G X + \beta_E E + \beta_{GE} (X \times E),$$

where $X$ codes either for an additive, a dominant, or a recessive effect of a SNP. Denoting the values of gene-environment pair $(X, E)$ for the case in family $i$ by $(x_{i1}, e_{i1})$, $i = 1, \ldots, n$, and for the matched control, i.e. the sibling or cousin, by $(x_{i2}, e_{i2})$ with $x_{i2} \neq x_{i1}$ (i.e. with discordant genotypes), Chatterjee et al. (2005) do not only consider these two individuals, but also two artificial pseudo-controls with values $(x_{i1}, e_{i2})$ and $(x_{i2}, e_{i1})$. The likelihood of their conditional logistic regression model with a 1:3 matching is thus given by

$$L(\beta_G, \beta_E, \beta_{GE}) = \prod_{i=1}^{n} \exp\left(\beta_G x_{i1} + \beta_E e_{i1} + \beta_{GE} (x_{i1} e_{i1})\right).$$

We first consider an additive mode of inheritance for the SNP in the model. Since the genotypes of the case and the matched control are discordant, there exist six genotype pairs $(x_{i1}, x_{i2})$:

$$(0, 1), (1, 0), (0, 2), (2, 0), (1, 2), (2, 1).$$

As in the main text to this Supplementary Material, we focus our interest on binary environmental variables so that there are four possible pairs of values $(e_{i1}, e_{i2})$ for the environmental variable:

$$(0, 0), (0, 1), (1, 0), (1, 1).$$

Thus, there are $6 \times 4 = 24$ possible combinations between the genotypes and the values of the environmental variable that have to be taken into account when transforming the likelihood (A.10) – analogously to the likelihoods in the main text to this Supplementary Material – into the log-likelihood

$$\ell_{add}(\beta_G, \beta_E, \beta_{GE}) = \sum_{p=1}^{4} \sum_{q=1}^{4} c_{pq} \log(w_{pq}),$$

where the weights $w_{pq}$ corresponding to the numbers $c_{pq}$ of case-control pairs showing a specific gene-environment combination are summarized in Table 5.

Setting

$$t_G = \sum_{q=1}^{4} (c_{2q} + 2c_{4q} + c_{6q}),$$

$$t_E = \sum_{p=1}^{6} c_{p3},$$

$$t_{GE} = c_{23} + 2c_{43} + c_{53} + 2c_{63} + c_{24} + 2c_{44} + c_{64},$$

as well as

$$f_1(\beta_G, \beta_E, \beta_{GE}) = 1 + \exp(\beta_G) + \exp(\beta_E) + \exp(\beta_G + \beta_E + \beta_{GE}),$$

$$f_2(\beta_G, \beta_E, \beta_{GE}) = 1 + \exp(2\beta_G) + \exp(\beta_E) + \exp(2\beta_G + \beta_E + 2\beta_{GE}),$$

and

$$f_3(\beta_G, \beta_E, \beta_{GE}) = 1 + \exp(\beta_G) + \exp(\beta_E + \beta_{GE}) + \exp(\beta_G + \beta_E + 2\beta_{GE}),$$
Weights and symbols for the numbers of case-control pairs showing the different genotype-exposure combinations \((X, E)\) in the conditional logistic regression model of Chatterjee et al. (2005), when an additive mode of inheritance is considered.

<table>
<thead>
<tr>
<th>(X)</th>
<th>(E)</th>
<th>Weights (w_{pq}) in the likelihood</th>
<th>Numbers (c_{pq})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 1)</td>
<td>(0, 0)</td>
<td>(\frac{1}{2 + 2 \exp(\beta_C)})</td>
<td>(c_{11})</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>(0, 0)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{21})</td>
</tr>
<tr>
<td>(0, 2)</td>
<td>(0, 0)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{31})</td>
</tr>
<tr>
<td>(2, 0)</td>
<td>(0, 0)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(2\beta_C) + \exp(2\beta_G + \beta_{GE})}{\exp(2\beta_C)})</td>
<td>(c_{41})</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>(0, 0)</td>
<td>(\frac{1 + \exp(\beta_E + \beta_{GE}) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{51})</td>
</tr>
<tr>
<td>(2, 1)</td>
<td>(0, 0)</td>
<td>(\frac{1 + \exp(\beta_E + \beta_{GE}) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{61})</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>(1, 0)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{12})</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>(1, 0)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{22})</td>
</tr>
<tr>
<td>(0, 2)</td>
<td>(1, 0)</td>
<td>(\frac{1 + \exp(\beta_E + \exp(2\beta_C) + \exp(2\beta_G + \beta_{GE})}{\exp(2\beta_C)})</td>
<td>(c_{32})</td>
</tr>
<tr>
<td>(2, 0)</td>
<td>(1, 0)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(2\beta_C) + \exp(2\beta_G + \beta_{GE})}{\exp(2\beta_C)})</td>
<td>(c_{42})</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>(1, 0)</td>
<td>(\frac{1 + \exp(\beta_E + \beta_{GE}) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{52})</td>
</tr>
<tr>
<td>(2, 1)</td>
<td>(1, 0)</td>
<td>(\frac{1 + \exp(\beta_E + \beta_{GE}) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{62})</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{13})</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{23})</td>
</tr>
<tr>
<td>(0, 2)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E + \exp(2\beta_C) + \exp(2\beta_G + \beta_{GE})}{\exp(2\beta_C)})</td>
<td>(c_{33})</td>
</tr>
<tr>
<td>(2, 0)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E + \exp(2\beta_C) + \exp(2\beta_G + \beta_{GE})}{\exp(2\beta_C)})</td>
<td>(c_{43})</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E + \beta_{GE}) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{53})</td>
</tr>
<tr>
<td>(2, 1)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E + \beta_{GE}) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{63})</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{14})</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{24})</td>
</tr>
<tr>
<td>(0, 2)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E + \exp(2\beta_C) + \exp(2\beta_G + \beta_{GE})}{\exp(2\beta_C)})</td>
<td>(c_{34})</td>
</tr>
<tr>
<td>(2, 0)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E + \exp(2\beta_C) + \exp(2\beta_G + \beta_{GE})}{\exp(2\beta_C)})</td>
<td>(c_{44})</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E + \beta_{GE}) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{54})</td>
</tr>
<tr>
<td>(2, 1)</td>
<td>(1, 1)</td>
<td>(\frac{1 + \exp(\beta_E + \beta_{GE}) + \exp(\beta_C) + \exp(\beta_G + \beta_{GE})}{\exp(\beta_C)})</td>
<td>(c_{64})</td>
</tr>
</tbody>
</table>
the log-likelihood (A.11) is given by

\[
\ell_{\text{add}}(\beta_G, \beta_E, \beta_{GE}) = \beta_G t_G + \beta_E t_E + \beta_{GE} t_{GE}
\]

\[
- \log(2 + 2 \exp(\beta_G)) \times (c_{11} + c_{21} + c_{51} + c_{61})
\]

\[
- \log(2 + 2 \exp(2\beta_E)) \times (c_{11} + c_{41})
\]

\[
- \log\left(\frac{f_1(\beta_G, \beta_E, \beta_{GE})}{1 + \exp(\beta_G)}\right) \times (c_{12} + c_{22} + c_{13} + c_{23})
\]

\[
- \log\left(\frac{f_2(\beta_G, \beta_E, \beta_{GE})}{1 + \exp(2\beta_E)}\right) \times (c_{32} + c_{42} + c_{33} + c_{43})
\]

\[
- \log\left(\frac{f_3(\beta_G, \beta_E, \beta_{GE})}{1 + \exp(\beta_{GE})}\right) \times (c_{52} + c_{62} + c_{53} + c_{63})
\]

\[
- \log(2 + 2 \exp(\beta_G + \beta_{GE})) \times (c_{14} + c_{24} + c_{54} + c_{64})
\]

\[
- \log(2 + 2 \exp(2\beta_E + \beta_{GE})) \times (c_{34} + c_{44})
\]

Differentially this log-likelihood with respect to \(\beta_G, \beta_E, \) and \(\beta_{GE}\) leads to

\[
\frac{\partial \ell_{\text{add}}(\beta_G, \beta_E, \beta_{GE})}{\partial \beta_G} = t_G
\]

\[
- \frac{\exp(\beta_G)}{1 + \exp(\beta_G)} \left( c_{11} + c_{21} + c_{51} + c_{61} \right)
\]

\[
- \frac{\exp(2\beta_G)}{1 + \exp(2\beta_G)} \left( c_{11} + c_{41} \right)
\]

\[
- \exp(\beta_G) + \exp(\beta_G + \beta_E + \beta_{GE}) \sum_{p=1}^{3} \sum_{q=2}^{3} c_{pq}
\]

\[
- \frac{2 \exp(2\beta_G) + 2 \exp(2\beta_G + \beta_E + 2\beta_{GE})}{f_3(\beta_G, \beta_E, \beta_{GE})} \sum_{p=3}^{4} \sum_{q=2}^{3} c_{pq}
\]

\[
- \frac{\exp(3\beta_G + \beta_E + \beta_{GE})}{1 + \exp(3\beta_G + \beta_{GE})} \left( c_{14} + c_{24} + c_{54} + c_{64} \right)
\]

\[
- \frac{2 \exp(2\beta_G + 2\beta_{GE})}{1 + \exp(2\beta_G + 2\beta_{GE})} \left( c_{31} + c_{41} \right)
\]

\[
\frac{\partial \ell_{\text{add}}(\beta_G, \beta_E, \beta_{GE})}{\partial \beta_E} = t_E
\]

\[
- \frac{\exp(\beta_E) + \exp(\beta_G + \beta_E + \beta_{GE})}{f_1(\beta_G, \beta_E, \beta_{GE})} \sum_{p=2}^{3} \sum_{q=3}^{3} c_{pq}
\]

\[
- \frac{\exp(2\beta_E)}{f_2(\beta_G, \beta_E, \beta_{GE})} \sum_{p=3}^{4} \sum_{q=2}^{3} c_{pq}
\]

\[
- \exp(\beta_E + \beta_{GE}) \left( 1 + \exp(\beta_G + \beta_{GE}) \right) \sum_{p=5}^{6} \sum_{q=2}^{3} c_{pq}
\]

\[
\frac{\partial \ell_{\text{add}}(\beta_G, \beta_E, \beta_{GE})}{\partial \beta_{GE}} = t_{GE}
\]

\[
- \frac{\exp(\beta_G + \beta_E + \beta_{GE})}{f_1(\beta_G, \beta_E, \beta_{GE})} \sum_{p=1}^{3} \sum_{q=2}^{3} c_{pq}
\]

\[
- \frac{2 \exp(2\beta_G + \beta_E + 2\beta_{GE})}{f_2(\beta_G, \beta_E, \beta_{GE})} \sum_{p=3}^{4} \sum_{q=2}^{3} c_{pq}
\]

\[
- \frac{\exp(\beta_E + \beta_{GE})}{f_3(\beta_G, \beta_E, \beta_{GE})} \sum_{p=5}^{6} \sum_{q=2}^{3} c_{pq}
\]

respectively.

As for the 3df model (A.6) in Web Appendix C.3, it does not seem to be possible to eliminate two of the parameters \(\beta_G, \beta_E, \) and \(\beta_{GE}\) from one of the first derivatives when they are set to zero. This is in particular due to the derivatives of log(\(f_k(\beta_G, \beta_E, \beta_{GE})\)) \(k = 1, 2, 3\), so that closed-form solutions for the maximum-likelihood estimates of \(\beta_G, \beta_E, \) and \(\beta_{GE}\) do not seem to exist, and the corresponding log-likelihood must be maximized numerically.

Although there are only three genotype pairs \((x_1, x_2)\) and thus 12 genotype-exposure combinations when considering either the dominant or the recessive mode of inheritance (see Web Table 6), there also does not seem to exist a closed-form solution in these two situations, as their log-likelihoods

\[
\ell_{\text{dom}}(\beta_G, \beta_E, \beta_{GE}) = \frac{1}{4} (d_{31} + d_{34})
\]

\[
+ \beta_G \sum_{q=1}^{4} d_{2q} + \beta_E \sum_{p=1}^{3} d_{p3} + \beta_{GE} (d_{23} + d_{24} + d_{33})
\]

\[
- \log(2 + 2 \exp(\beta_G)) \times (d_{11} + d_{21})
\]

\[
- \log\left(\frac{f_1(\beta_G, \beta_E, \beta_{GE})}{1 + \exp(\beta_G)}\right) \times \sum_{p=1}^{2} \sum_{q=2}^{3} d_{pq}
\]

\[
- \log(2 + 2 \exp(\beta_E + \beta_{GE})) \times (d_{14} + d_{24})
\]

\[
- \log(2 + 2 \exp(\beta_E + \beta_{GE})) \times (d_{32} + d_{33})
\]

and

\[
\ell_{\text{rec}}(\beta_G, \beta_E, \beta_{GE}) = \frac{1}{4} (r_{11} + r_{14})
\]

\[
+ \beta_G \sum_{q=1}^{4} r_{1q} + \beta_E \sum_{p=1}^{3} r_{p3} + \beta_{GE} (r_{33} + r_{34})
\]

\[
- \log(2 + 2 \exp(\beta_G)) \times (r_{12} + r_{13})
\]

\[
- \log\left(\frac{f_1(\beta_G, \beta_E, \beta_{GE})}{1 + \exp(\beta_G)}\right) \times (r_{21} + r_{31})
\]

\[
- \log(2 + 2 \exp(\beta_E + \beta_{GE})) \times (r_{24} + r_{34})
\]

and

\[
\ell_{\text{add}}(\beta_G, \beta_E, \beta_{GE}) = \frac{1}{4} (c_{11} + c_{31})
\]

\[
+ \beta_G \sum_{q=1}^{4} c_{2q} + \beta_E \sum_{p=1}^{3} c_{p3} + \beta_{GE} (c_{23} + c_{24} + c_{33})
\]

\[
- \log(2 + 2 \exp(\beta_G)) \times (c_{11} + c_{21})
\]

\[
- \log\left(\frac{f_1(\beta_G, \beta_E, \beta_{GE})}{1 + \exp(\beta_G)}\right) \times \sum_{p=1}^{2} \sum_{q=2}^{3} c_{pq}
\]

\[
- \log(2 + 2 \exp(\beta_E + \beta_{GE})) \times (c_{14} + c_{24})
\]

\[
- \log(2 + 2 \exp(\beta_E + \beta_{GE})) \times (c_{32} + c_{33})
\]
Web Table 6
Weights and symbols for the numbers $d_{pq}$ or $r_{pq}$ of case-control pairs showing the different genotype-exposure combinations $(X, E)$ in the conditional logistic regression model of Chatterjee et al. (2005), when a dominant or recessive mode of inheritance, respectively, is considered, where the column “Model” indicates under which mode of inheritance the genotype-exposure combinations occur.

<table>
<thead>
<tr>
<th>$X$</th>
<th>$E$</th>
<th>Model</th>
<th>Weight in likelihood</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 0)</td>
<td>(0, 0)</td>
<td>Recessive</td>
<td>$\frac{1}{4}$</td>
<td>$r_{11}$</td>
</tr>
<tr>
<td>(0, 0)</td>
<td>(0, 1)</td>
<td>Recessive</td>
<td>$\frac{1}{2 + 2 \exp(\beta_E)}$</td>
<td>$r_{12}$</td>
</tr>
<tr>
<td>(0, 0)</td>
<td>(1, 0)</td>
<td>Recessive</td>
<td>$\frac{\exp(\beta_E)}{2 + 2 \exp(\beta_E)}$</td>
<td>$r_{13}$</td>
</tr>
<tr>
<td>(0, 0)</td>
<td>(1, 1)</td>
<td>Recessive</td>
<td>$\frac{1}{4}$</td>
<td>$r_{14}$</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>(0, 0)</td>
<td>Both</td>
<td>$\frac{1}{2 + 2 \exp(\beta_G)}$</td>
<td>$d_{11}$ or $r_{21}$</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>(0, 1)</td>
<td>Both</td>
<td>$1 + \exp(\beta_G) + \exp(\beta_E) + \exp(\beta_G + \beta_E + \beta_{GE})$</td>
<td>$d_{12}$ or $r_{22}$</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>(1, 0)</td>
<td>Both</td>
<td>$\frac{\exp(\beta_E)}{1 + \exp(\beta_G) + \exp(\beta_G + \beta_E + \beta_{GE})}$</td>
<td>$d_{13}$ or $r_{23}$</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>(1, 1)</td>
<td>Both</td>
<td>$\frac{1}{2 + 2 \exp(\beta_G) + \beta_{GE}}$</td>
<td>$d_{14}$ or $r_{24}$</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>(0, 0)</td>
<td>Both</td>
<td>$\frac{\exp(\beta_G)}{2 + 2 \exp(\beta_G)}$</td>
<td>$d_{21}$ or $r_{31}$</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>(0, 1)</td>
<td>Both</td>
<td>$1 + \exp(\beta_G) + \exp(\beta_E) + \exp(\beta_G + \beta_E + \beta_{GE})$</td>
<td>$d_{22}$ or $r_{32}$</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>(1, 0)</td>
<td>Both</td>
<td>$\frac{\exp(\beta_G + \beta_{GE})}{1 + \exp(\beta_G) + \exp(\beta_E) + \exp(\beta_G + \beta_E + \beta_{GE})}$</td>
<td>$d_{23}$ or $r_{33}$</td>
</tr>
<tr>
<td>(1, 0)</td>
<td>(1, 1)</td>
<td>Both</td>
<td>$\frac{\exp(\beta_G + \beta_{GE})}{2 + 2 \exp(\beta_G + \beta_{GE})}$</td>
<td>$d_{24}$ or $r_{34}$</td>
</tr>
<tr>
<td>(1, 1)</td>
<td>(0, 0)</td>
<td>Dominant</td>
<td>$\frac{1}{4}$</td>
<td>$d_{31}$</td>
</tr>
<tr>
<td>(1, 1)</td>
<td>(0, 1)</td>
<td>Dominant</td>
<td>$\frac{1}{2 + 2 \exp(\beta_E + \beta_{GE})}$</td>
<td>$d_{32}$</td>
</tr>
<tr>
<td>(1, 1)</td>
<td>(1, 0)</td>
<td>Dominant</td>
<td>$\frac{\exp(\beta_E + \beta_{GE})}{2 + 2 \exp(\beta_E + \beta_{GE})}$</td>
<td>$d_{33}$</td>
</tr>
<tr>
<td>(1, 1)</td>
<td>(1, 1)</td>
<td>Dominant</td>
<td>$\frac{1}{4}$</td>
<td>$d_{34}$</td>
</tr>
</tbody>
</table>
Web Table 7
Top 5 SNPs found in the analysis of the case-parent trio data from the International Cleft Consortium with a gTDT testing each of the 569,187 autosomal SNPs for an additive effect.

<table>
<thead>
<tr>
<th>SNP</th>
<th>( \hat{\beta}_{add} )</th>
<th>Statistic</th>
<th>p-Value</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs987525</td>
<td>0.577</td>
<td>66.400</td>
<td>( 3.68 \times 10^{-16} )</td>
<td>1.781</td>
<td>[1.550, 2.047]</td>
</tr>
<tr>
<td>rs1784394</td>
<td>-0.999</td>
<td>61.204</td>
<td>( 5.15 \times 10^{-15} )</td>
<td>0.368</td>
<td>[0.287, 0.473]</td>
</tr>
<tr>
<td>rs10863790</td>
<td>-0.545</td>
<td>58.625</td>
<td>( 1.91 \times 10^{-14} )</td>
<td>0.580</td>
<td>[0.504, 0.667]</td>
</tr>
<tr>
<td>rs28470586</td>
<td>-0.702</td>
<td>53.762</td>
<td>( 2.26 \times 10^{-13} )</td>
<td>0.495</td>
<td>[0.411, 0.598]</td>
</tr>
<tr>
<td>rs1519847</td>
<td>0.443</td>
<td>51.514</td>
<td>( 7.11 \times 10^{-13} )</td>
<td>1.558</td>
<td>[1.380, 1.759]</td>
</tr>
</tbody>
</table>

indicate. Again, in particular \( f_1(\beta_G, \beta_E, \beta_{GE}) \) prevents a closed-form solution for the maximum-likelihood estimates of the parameters. For example, when considering the recessive conditional regression model, we have reanalyzed the genotype data on the 569,187 autosomal polymorphic markers measured in 1925 case-parent trios from the study of the International Cleft Consortium (Beaty et al., 2010). We have in particular tested these 569,187 SNPs individually and interactions between each of these SNPs and gender with the different versions of the gTDT.

While Section 6 of the main text to this supplementary material focuses on the comparison of the analytic and the iterative approach to the gTDT, we present in the following some results from the reanalysis of the individual SNPs and the gene-environment interactions. For more general results of the analysis of the case-parent trio data from the International Cleft Consortium, see Beaty et al. (2010). For analyses of other interactions between SNPs and environmental variables (maternal smoking, alcohol consumption, and multivitamin supplementation) in a subset of this case-parent trio data, see Beaty et al. (2011).

Web Appendix E: Results from the Analysis of the Trio Data from the International Cleft Study
To investigate how large the gain in computing time is when using the analytic solution for the gTDT instead of the conventional iterative fitting procedure for the corresponding conditional regression model, we have reanalyzed the genotype data on the 569,187 autosomal polymorphic markers measured in 1925 case-parent trios from the study of the International Cleft Consortium (Beaty et al., 2010). We have in particular tested these 569,187 SNPs individually and interactions between each of these SNPs and gender with the different versions of the gTDT.

Web Figure 1. QQ-plot of the values of the gTDT statistic for the parameter \( \beta_{GE} \) of the conditional logistic regression model for testing the interaction of each of the 569,187 autosomal SNPs from the case-parent trio study conducted by the International Cleft Consortium with the environmental variable “gender” under an additive mode of inheritance.
Web Figure 2. Manhattan plots for the gTDT analyses of 569,187 autosomal SNPs from the case-parent trio study of the International Cleft Consortium (upper panel) and the interactions between these SNPs and gender (lower panel), where in the latter application the p-values for testing the parameter $\beta_{GE}$ are displayed and in both applications an additive mode of inheritance is assumed for the SNPs.

In Web Table 7, the five SNPs with the largest gTDT statistics for testing an additive effect are presented along with their values for $\hat{\beta}_{\text{add}}$, their p-values, their odds ratios, and the 95% confidence intervals for these odds ratios. In Web Figure 1, a qq-plot of the values of the gTDT statistic for the parameter $\beta_{GE}$ is shown when testing the interaction between each SNP and the environmental variable “gender” under an additive mode of inheritance. This figure reveals that none of the interactions between gender and any SNP shows genome-wide significance (defined by an unadjusted p-value smaller than $0.05/569,187 = 8.8 \times 10^{-8}$). In fact, the smallest p-value is $2.74 \times 10^{-6}$. Finally, Web Figure 2 shows the Manhattan plots for the applications of the gTDT to the individual SNPs and the gene-environment interactions under an additive mode of inheritance.

Web Appendix F: Simulation Study
To assess the reduction in computing time for different numbers of SNPs and case-parent trios, we randomly drew ten data sets for each combination of 500, 1000, 1500, and 1925 case-parent trios with 100, 500, 1000, 5000, 10000, 15000, and 20000 SNPs from the trio data of the International Cleft Consortium, and applied both the analytic and the conventional iterative approach to the gTDT to these data sets, testing for both additive and dominant effects. We did not consider the recessive model, as the parameter estimate and thus the CPU time for this model are similar to the ones for the dominant model.

The computations were performed on 2.7 GHz machines with 40 GB of RAM in the statistical software environment R version 2.11.1, using the R package trio (version 1.1.17), which is available at the Comprehensive R Archive Network (http://cran.r-project.org).

The results of this simulation study are summarized in Web Table 8. This table shows that in general the improvements in computing time were about two orders of magnitude. It, for example, took less than a second to test up to 1000 SNPs using the analytic approach, compared to up to 76 seconds using the numeric parameter estimation. When testing 20,000 SNPs, all closed-form estimates and the corresponding gTDT statistics were derived in only a few seconds, while the conventional approach took around 25 minutes. In general, the computing time of the gTDT based on the closed-form maximum-likelihood estimates was always 75 times, usually 90 to 100 times, and frequently more than 100 times lower than the computing time of the conventional numerical approach, with a slight decrease in relative computational gain as the number of trios increased. The computing times of both approaches were rather stable across the replicates. The minimum and maximum computing times were always within five percent, and typically within one percent of the median computing time. Thus, details on the variations in computing time are not shown.

Web Appendix G: Computing the Roots of a Polynomial of Fourth Degree Analytically
The roots of the polynomial

$$c_4 x^4 + c_3 x^3 + c_2 x^2 + c_1 x + c_0$$

(A.12)

can be analytically determined using the following procedure:
Web Table 8

For each combination of several numbers of SNPs and case-parent trios, the median computing time (in seconds) of applications of the gTDT for testing either an additive or a dominant effect to ten data sets of the same size. The gTDT was performed by employing either the closed-form solutions derived in Section 2 of the main text or the conventionally used iterative fitting procedure, where the median computing time of the latter is shown in brackets.

<table>
<thead>
<tr>
<th>Trios</th>
<th>Effect</th>
<th>100 SNPs</th>
<th>500 SNPs</th>
<th>1000 SNPs</th>
<th>5000 SNPs</th>
<th>10,000 SNPs</th>
<th>15,000 SNPs</th>
<th>20,000 SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>Add</td>
<td>0.023 (2.56)</td>
<td>0.12 (12.6)</td>
<td>0.24 (25.1)</td>
<td>1.22 (130.5)</td>
<td>2.44 (306.7)</td>
<td>3.66 (476.6)</td>
<td>4.85 (543.3)</td>
</tr>
<tr>
<td>500</td>
<td>Dom</td>
<td>0.019 (2.56)</td>
<td>0.11 (12.7)</td>
<td>0.22 (25.5)</td>
<td>1.07 (118.9)</td>
<td>2.15 (311.9)</td>
<td>3.24 (498.2)</td>
<td>4.30 (654.2)</td>
</tr>
<tr>
<td>1000</td>
<td>Add</td>
<td>0.047 (4.56)</td>
<td>0.24 (21.7)</td>
<td>0.47 (42.0)</td>
<td>2.36 (219.5)</td>
<td>4.69 (492.4)</td>
<td>7.05 (818.5)</td>
<td>9.36 (901.6)</td>
</tr>
<tr>
<td>1000</td>
<td>Dom</td>
<td>0.036 (4.56)</td>
<td>0.21 (21.5)</td>
<td>0.41 (42.5)</td>
<td>2.01 (202.9)</td>
<td>4.07 (524.4)</td>
<td>6.19 (818.5)</td>
<td>8.24 (941.9)</td>
</tr>
<tr>
<td>1500</td>
<td>Add</td>
<td>0.084 (6.83)</td>
<td>0.38 (31.1)</td>
<td>0.74 (60.2)</td>
<td>3.67 (310.2)</td>
<td>7.31 (621.0)</td>
<td>10.95 (1112.6)</td>
<td>14.59 (1330.6)</td>
</tr>
<tr>
<td>1500</td>
<td>Dom</td>
<td>0.066 (6.80)</td>
<td>0.37 (30.4)</td>
<td>0.71 (60.8)</td>
<td>3.51 (286.3)</td>
<td>6.99 (736.4)</td>
<td>10.54 (1088.7)</td>
<td>14.05 (1267.2)</td>
</tr>
<tr>
<td>1925</td>
<td>Add</td>
<td>0.103 (9.02)</td>
<td>0.51 (38.9)</td>
<td>0.98 (75.3)</td>
<td>4.72 (381.8)</td>
<td>9.37 (757.0)</td>
<td>14.07 (1358.9)</td>
<td>18.68 (1509.3)</td>
</tr>
<tr>
<td>1925</td>
<td>Dom</td>
<td>0.085 (9.00)</td>
<td>0.45 (37.8)</td>
<td>0.93 (76.0)</td>
<td>4.60 (355.1)</td>
<td>9.08 (895.6)</td>
<td>13.62 (1343.5)</td>
<td>18.14 (1564.5)</td>
</tr>
</tbody>
</table>

(1) Set

\[ t_1 = \frac{3}{8} \left( \frac{c_3}{c_4} \right)^2 + \frac{c_2}{c_4} \]

\[ t_2 = \frac{1}{8} \left( \frac{c_3}{c_4} \right)^3 - \frac{c_3 c_2}{2 c_4^2} + \frac{c_1}{c_4} \]

\[ t_3 = -\frac{3}{256} \left( \frac{c_3}{c_4} \right)^4 + \frac{c_2 c_3^2}{16 c_4} - \frac{c_3 c_1}{4 c_4^2} + \frac{c_0}{c_4} \]

(2) If \( t_2 = 0 \), then the four roots are given by

\[ -\frac{c_3}{4 c_4} \pm \sqrt{0.5 \left( t_1 \mp \sqrt{t_1^2 - 4 t_3} \right)} \]

where all four combinations of the signs \( \pm \) and \( \mp \) have to be taken into account. Otherwise, continue with Step 3.

(3) Set

\[ p = \frac{t_1^2}{12} - t_3 \]

\[ q = -\frac{t_1^3}{108} + \frac{t_1 t_3}{3} - \frac{t_2}{8} \]

\[ d = \left( \frac{q}{2} \right)^2 + \left( \frac{t_3}{3} \right)^3 \]

(4) If \( d < 0 \), then a situation called casus irreducibilis is given and the roots can be computed as described in Steps 5 and 6. Otherwise, determine the four roots \( x_h, \ h = 1, \ldots, 4 \), of the polynomial (A.12) by

\[ x_1 = -\frac{c_3}{4 c_4} + \frac{1}{2} \left( \sqrt{t_1 + 2y} + \sqrt{-3t_1 - 2y - \frac{2t_2}{\sqrt{t_1 + 2y}}} \right) \]

\[ x_2 = -\frac{c_3}{4 c_4} + \frac{1}{2} \left( \sqrt{t_1 + 2y} - \sqrt{-3t_1 - 2y - \frac{2t_2}{\sqrt{t_1 + 2y}}} \right) \]

\[ x_3 = -\frac{c_3}{4 c_4} + \frac{1}{2} \left( -\sqrt{t_1 + 2y} + \sqrt{-3t_1 - 2y + \frac{2t_2}{\sqrt{t_1 + 2y}}} \right) \]

\[ x_4 = -\frac{c_3}{4 c_4} + \frac{1}{2} \left( -\sqrt{t_1 + 2y} - \sqrt{-3t_1 - 2y + \frac{2t_2}{\sqrt{t_1 + 2y}}} \right) \]

where

\[ y = \begin{cases} \frac{5}{6} t_1 - \sqrt{q}, & \text{if } P = 0 \\ \frac{5}{6} t_1 + u - \frac{p}{3u}, & \text{otherwise} \end{cases} \]

with

\[ u = -\sqrt{\frac{q}{2} + \frac{d}{2}}. \]

(5) If \( d < 0 \), set

\[ g_1 = 2 \sqrt{-\frac{4p}{3}} \]

\[ g_2 = \frac{1}{3} \arccos \left( \frac{4q}{(-4p/3)^{1/2}} \right) \]

and determine

\[ v_1 = g_1 \cos \left( g_2 \right) \]

\[ v_2 = -g_1 \cos \left( g_2 + \frac{\pi}{3} \right) \]

\[ v_3 = g_1 \cos \left( g_2 + \frac{2\pi}{3} \right) \]

as well as

\[ y_k = \sqrt{-v_k - \frac{2}{3} t_1}, \ k = 1, 2, 3. \]

(6) Further, set

\[ s = -\frac{t_2}{2 \Pi_{k=1} y_k} \in \{-1, 1\} \]

and determine the roots \( x_h, \ h = 1, \ldots, 4 \), of the polyno-
mial (A.12) by

\[ x_1 = s (y_1 + y_2 + y_3) - \frac{t_2}{4t_1} \]

\[ x_2 = s (y_1 - y_2 - y_3) - \frac{t_2}{4t_1} \]

\[ x_3 = s (-y_1 + y_2 - y_3) - \frac{t_2}{4t_1} \]

\[ x_4 = s (-y_1 - y_2 + y_3) - \frac{t_2}{4t_1} \]

References


