A Note on Including Phenotypic Information from Monozygotic Twins in Variance Components QTL Linkage Analysis

D. M. Evans* and S. E. Medland
Queensland Institute of Medical Research and Joint Genetics Program, University of Queensland, Brisbane, Australia

Summary

Williams & Blangero (1999) derived closed form expressions for the power of a univariate variance components test of linkage for a variety of pedigree structures. We have extended their results by investigating the effect of including monozygotic twins in the design on the power to detect linkage. Specifically, we determined the power associated with a pedigree of size three, where individuals one and two were monozygotic twins and individual three was a full sibling to the twins. The power of this sampling unit was uniformly greater than the power obtained from a sib-pair under the same genetic model. The reason for this was that addition of a second monozygotic twin provided another estimate of the sibling correlation for the particular IBD class. In addition, when the total heritability of the trait was <50%, the number of individuals that needed to be phenotyped was less than that with sib-pairs alone. However, a pedigree consisting of a monozygotic pair and sibling was never as informative as a sib-trio, presumably because the sib-trio provided information about allele sharing between three individuals, whereas the monozygotic twins and sibling unit only provided one such relationship. We conclude that including a monozygotic twin in the analysis is an economical strategy, since only one twin needs to be genotyped.

In a previous edition of the Annals of Human Genetics, Williams & Blangero (1999) derived closed form expressions for the power of a univariate variance components (VC) test of linkage. The authors expressed the non-centrality parameter (NCP) of the test (and hence its theoretical power) in terms of variance components due to the quantitative trait locus (QTL) and the total additive genetic variance. They provided expressions for several relationships including sib-ships, arbitrary relative pairs, nuclear families and larger pedigrees. The authors found that larger pedigrees provided greater power per sampling unit than did smaller sampling structures, consistent with the results from previous data simulations (Schork, 1993; Dolan et al. 1999; Williams & Blangero, 1999).

In this manuscript we extended the results from Williams & Blangero (1999) to investigate the effect of including monozygotic (MZ) twins on the power of VC linkage analysis. The question of whether MZ twins provide increased power to detect QTLs is particularly relevant, since it is often the case that phenotypic and genotypic information are collected from MZ twins throughout the course of a twin study. In fact many of the world’s twin registries have recorded these data not only from twins, but also from the twins’ siblings and parents as well (e.g. Zhu et al. 1999). One might assume that since MZ twins always share both their alleles identical by descent (IBD), they are uninformative for linkage and therefore do not contribute to the power of the linkage test. While this is true in the case of a pair of individuals, it is not the case when the MZ twins form part of a larger sibship. In these cases it is vitally important to include phenotypic information from both MZ twins in the analysis.

* Address for correspondence: David Evans, The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, United Kingdom. Phone: +44 (0)1865 287598. Fax: +44 (0)1865 287698. E-mail: davide@well.ox.ac.uk
Following the methodology of Williams & Blangero (1999), consider a vector \( x = [x_1, \ldots, x_n] \) of quantitative trait values for a pedigree of \( n \) individuals. Assuming that the trait values are distributed multivariate normally, the log-likelihood \( l \) of the pedigree is given by:

\[
l = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma| - \frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu)
\]

where \( x \) is a vector of trait observations, and \( \mu \) and \( \Sigma \) are the vector of expected means and the expected covariance matrix of the pedigree, respectively (Lange et al. 1976). The covariance matrix \( \Sigma \) is a function of the genetic and environmental variance components to be modelled. In the simple case of a major additive gene acting against a background of additive polygenic and random environmental effects, the expected covariance matrix is:

\[
\Sigma = \hat{\Pi} \sigma_q^2 + 2 \Phi \sigma_a^2 + I_n \sigma_e^2
\]

where \( \sigma_q^2 \) is the additive genetic variance due to the QTL, \( \sigma_a^2 \) is the (residual) polygenic additive genetic variance, and \( \sigma_e^2 \) is the unique environmental variance (Amos, 1994; Hopper & Matthews, 1982). \( \hat{\Pi} \) is a matrix containing the elements \( \hat{\pi}_{ij} \) that denote the estimated proportion of alleles shared IBD at the trait locus by individuals \( i \) and \( j \), \( \Phi \) is the kinship matrix containing elements \( \phi_{ij} \) - the kinship coefficients between individuals \( i \) and \( j \) (Jacquard, 1974), and \( I_n \) is an identity matrix of order \( n \).

In order to illustrate the effect of including phenotypic information from MZ twins in a linkage analysis, consider a pedigree of size three, where individuals one and two are MZ twins and individual three is a full sibling to the twins. Assuming that the trait is standardized to unit variance, the expected covariance matrix of this pedigree is:

\[
\Sigma = \hat{\Pi} q^2 + 2 \Phi a^2 + I_3 e^2
\]

where \( q^2 \), \( a^2 \), and \( e^2 \) are the proportions of variance due to the QTL, residual additive polygenic sources, and non-shared environmental sources respectively. The kinship matrix \( \Phi \) is given by:

\[
\Phi = \begin{pmatrix}
.5 & .5 & .25 \\
.5 & .5 & .25 \\
.25 & .25 & .5 \\
\end{pmatrix}
\]

while the IBD sharing matrix takes one of three forms depending on the number of alleles shared IBD at the marker locus between the sibling and the MZ twins:

\[
\Pi_0 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \Pi_1 = \begin{pmatrix} 1 & 1 & .5 \\ 1 & 1 & .5 \\ 0 & 0 & 1 \end{pmatrix} \quad \Pi_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}
\]

with associated probabilities \( \pi_0 = \frac{1}{4}, \pi_1 = \frac{1}{2} \) and \( \pi_2 = \frac{1}{4} \) under the null hypothesis of no linkage. Notice how the four upper left elements of the matrix are always equal to one, since MZ twins share all their genes IBD. Under the null hypothesis, the QTL specific portion of the variance \( (q^2) \) is absorbed into the residual additive polygenic component so that the covariance matrix may be expressed as:

\[
\Sigma = 2 \Phi (q^2 + a^2) + I_3 e^2 = \begin{pmatrix} 1 & h^2 & .5h^2 \\ h^2 & 1 & .5h^2 \\ .5h^2 & .5h^2 & 1 \end{pmatrix}
\]

where \( h^2 = a^2 + q^2 \) is the total heritability of the trait.

Asymptotically, and under standard conditions, the likelihood ratio test under the alternative hypothesis of linkage is distributed as a non-central \( \chi^2 \) distribution with non-centrality parameter given by the formula:

\[
\zeta = (\hat{\theta}_1 - \hat{\theta}_0)' T (\hat{\theta}_1 - \hat{\theta}_0)
\]

where \( \theta_0 = [0, q^2 + a^2, e^2]' \) is the vector of parameter estimates under the null hypothesis (i.e. the hypothesis of no linkage), \( \theta_1 = [q^2, a^2, e^2]' \) is the vector of parameter estimates under the alternative hypothesis, and \( T \) is the Fisher information matrix (i.e. the negative of the expected value of the matrix of second order partial derivatives of the likelihood function with respect to the estimated parameters). Under a random effects model (i.e. where covariate effects are not considered in the exponential part of the likelihood), the information matrix has the following structure:

\[
T = E \left( - \frac{\partial^2 l}{\partial \sigma_q^2 \partial \sigma_j^2} \right) = \frac{1}{2} \text{Tr} \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_q^2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_j^2} \right)
\]

where \( l \) is the log likelihood function, \( \text{Tr} \) denotes the trace operator, \( \Sigma \) is the expected covariance matrix, and \( \sigma_q^2 \) and \( \sigma_j^2 \) are variance components. In the present
example, $T$ is a $3 \times 3$ matrix, whose elements are determined by the pedigree structure (i.e. the kinship matrix $\Phi$) as well as the IBD sharing (the $\hat{\Pi}$ matrix) among individuals at the QTL (Williams & Blangero, 1999):

$$t_{11} = \frac{1}{2} \text{Tr}(\Sigma^{-1} \hat{\Pi} \Sigma^{-1} \hat{\Pi}) = \frac{7h^4 - 10h^2 + 26}{2(h^4 - 2h^2 - 2)^2}$$

$$t_{12} = t_{21} = \frac{1}{2} \text{Tr}(\Sigma^{-1} \hat{\Pi} \Sigma^{-1} 2\Phi) = \frac{3h^4 - 6h^2 + 12}{(h^4 - 2h^2 - 2)^2}$$

$$t_{13} = t_{31} = \frac{1}{2} \text{Tr}(\Sigma^{-1} \hat{\Pi} \Sigma^{-1}) = \frac{3h^4 + 6}{(h^4 - 2h^2 - 2)^2}$$

$$t_{22} = \frac{1}{2} \text{Tr}(\Sigma^{-1} 2\Phi \Sigma^{-1} 2\Phi) = \frac{3h^4 + 6}{(h^4 - 2h^2 - 2)^2}$$

$$t_{23} = t_{32} = \frac{1}{2} \text{Tr}(\Sigma^{-1} 2\Phi \Sigma^{-1}) = \frac{3h^4 + 6}{(h^4 - 2h^2 - 2)^2}$$

$$t_{33} = \frac{1}{2} \text{Tr}(\Sigma^{-1} \Sigma^{-1}) = \frac{9h^8 - 12h^6 + 12}{2(h^2 - 1)^2(h^4 - 2h^2 - 2)^2}$$

Using equation (5) it is now a simple (albeit tedious) matter to calculate the NCP for the test of linkage in a single pedigree:

$$\Lambda = \frac{q^4(h^4 + 2h^2 + 2)}{2(h^4 - 2h^2 - 2)^2} \quad (7)$$

Note how this differs from the NCP for a single sib-pair:

$$\Lambda = \frac{q^4(h^4 + 4)}{2(h^4 - 4)^2} \quad (8)$$

and a sib-trio as calculated in Williams & Blangero (1999):

$$\Lambda = \frac{3q^4(h^4 + 2h^2 + 2)}{4(h^2 + 1)^2(h^2 - 2)^2} \quad (9)$$

Using the NCP, it is possible to calculate the number of sampling units (and hence the total number of individuals) required to achieve a desired level of power at a given level of significance (c.f. Williams & Blangero, 1999; Sham et al. 2000). Tables 1 and 2 display the number of sampling units and individuals respectively required

<table>
<thead>
<tr>
<th>QTL Variance ($q^2$)</th>
<th>$h^2$</th>
<th>.05</th>
<th>.10</th>
<th>.15</th>
<th>.20</th>
<th>.30</th>
<th>.40</th>
<th>.50</th>
<th>.60</th>
<th>.70</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>79801</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td>41825</td>
<td>26370</td>
<td>79353</td>
<td>19838</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.10</td>
<td>43377</td>
<td>10844</td>
<td>26338</td>
<td>6584</td>
<td>78612</td>
<td>19653</td>
<td>8734</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.15</td>
<td>44640</td>
<td>11160</td>
<td>4960</td>
<td></td>
<td>25969</td>
<td>6492</td>
<td>2885</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.20</td>
<td>45625</td>
<td>11406</td>
<td>5069</td>
<td>2852</td>
<td>25480</td>
<td>6370</td>
<td>1592</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.30</td>
<td>46812</td>
<td>11703</td>
<td>5201</td>
<td>2926</td>
<td>24194</td>
<td>6048</td>
<td>2688</td>
<td>1512</td>
<td>672</td>
</tr>
<tr>
<td></td>
<td>.40</td>
<td>47063</td>
<td>11765</td>
<td>5229</td>
<td>2941</td>
<td>22589</td>
<td>5647</td>
<td>2510</td>
<td>1412</td>
<td>627</td>
</tr>
<tr>
<td></td>
<td>.50</td>
<td>46510</td>
<td>11628</td>
<td>5168</td>
<td>2907</td>
<td>20757</td>
<td>5189</td>
<td>2306</td>
<td>1297</td>
<td>672</td>
</tr>
<tr>
<td></td>
<td>.60</td>
<td>45285</td>
<td>11321</td>
<td>5032</td>
<td>2830</td>
<td>20737</td>
<td>5189</td>
<td>2306</td>
<td>1297</td>
<td>672</td>
</tr>
<tr>
<td></td>
<td>.70</td>
<td>43511</td>
<td>10878</td>
<td>4835</td>
<td>2719</td>
<td>16731</td>
<td>4183</td>
<td>1859</td>
<td>1046</td>
<td>465</td>
</tr>
</tbody>
</table>

Table 1 Number of sampling units required for 90% power to detect linkage with a LOD score of 3. Results are shown for sib-pairs (normal font), MZ twins plus sib (bold font) and sib-trios (italics)
Table 2 Number of individuals needing to be phenotyped for 90% power to detect linkage with a LOD score of 3. Results are shown for sib-pairs (normal font), MZ twins plus sib (bold font) and sib-trios (italics)

<table>
<thead>
<tr>
<th>QTL Variance ($q^2$)</th>
<th>.05</th>
<th>.10</th>
<th>.15</th>
<th>.20</th>
<th>.30</th>
<th>.40</th>
<th>.50</th>
<th>.60</th>
<th>.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.05</td>
<td>125474</td>
<td>79709</td>
<td>138707</td>
<td>39677</td>
<td>130131</td>
<td>32533</td>
<td>79012</td>
<td>19753</td>
<td>157224</td>
</tr>
<tr>
<td>.10</td>
<td>136874</td>
<td>34218</td>
<td>15208</td>
<td>8555</td>
<td>76438</td>
<td>19110</td>
<td>8493</td>
<td>4777</td>
<td>149425</td>
</tr>
<tr>
<td>.15</td>
<td>140437</td>
<td>35109</td>
<td>15604</td>
<td>8777</td>
<td>72581</td>
<td>18145</td>
<td>8065</td>
<td>4536</td>
<td>2016</td>
</tr>
<tr>
<td>.20</td>
<td>139530</td>
<td>34883</td>
<td>15503</td>
<td>8158</td>
<td>62270</td>
<td>15567</td>
<td>6919</td>
<td>3892</td>
<td>1730</td>
</tr>
<tr>
<td>.30</td>
<td>135854</td>
<td>33964</td>
<td>15095</td>
<td>8491</td>
<td>56343</td>
<td>14086</td>
<td>6260</td>
<td>3521</td>
<td>1565</td>
</tr>
<tr>
<td>.40</td>
<td>130533</td>
<td>32633</td>
<td>14503</td>
<td>8158</td>
<td>50191</td>
<td>12548</td>
<td>6577</td>
<td>3137</td>
<td>1394</td>
</tr>
<tr>
<td>.50</td>
<td>12188</td>
<td>30370</td>
<td>13498</td>
<td>7593</td>
<td>4523</td>
<td>11512</td>
<td>5777</td>
<td>3067</td>
<td>1264</td>
</tr>
<tr>
<td>.60</td>
<td>113054</td>
<td>28123</td>
<td>12422</td>
<td>6919</td>
<td>4028</td>
<td>10406</td>
<td>5177</td>
<td>2867</td>
<td>1134</td>
</tr>
<tr>
<td>.70</td>
<td>105102</td>
<td>26170</td>
<td>11362</td>
<td>6365</td>
<td>3534</td>
<td>9310</td>
<td>4577</td>
<td>2477</td>
<td>1008</td>
</tr>
</tbody>
</table>

to achieve a LOD score of 3 with 90 percent power (i.e. corresponding to a critical value of 24.9847 for the likelihood ratio statistic) using sib-pairs, sib-trios, and pedigrees consisting of MZ twins and a sibling under a variety of genetic models (c.f. Williams & Blangero, 1999).

Table 1 shows that the addition of a MZ twin to the pedigree results in a sampling unit which is uniformly more powerful than performing linkage with sib-pairs alone. What is more, when the proportion of variance due to the QTL is small and the total heritability of the trait is low to moderate (i.e. < 50%), the total number of individuals needing to be phenotyped is actually less (Table 2). It is well known that in the case of sib-pairs, as the proportion of variance due to residual polygenic sources increases (and therefore the total heritability of the trait), the power to detect linkage also increases (Williams & Blangero, 1999; Sham et al. 2000). This is simply because as the error variance decreases, the sib-pair difference is more likely to accurately reflect IBD sharing at the trait locus.

However, this simple relationship is complicated when a MZ twin is added to the sib-pair. Addition of a second MZ twin essentially provides another estimate of the sibling correlation for the particular IBD class. This extra information is most beneficial when the unique environmental variance is high, when a single sibling correlation may not accurately reflect IBD sharing at the trait locus. In contrast, when the error variance is low (i.e. due to high heritability), a single sibling correlation may not accurately reflect IBD sharing at the trait locus, and the gain in information from having a second MZ twin phenotyped will not be as great. The net result of these two processes is a quadratic relationship between total heritability and power, so that most biologically plausible QTLs (i.e. < 50% of the total variance) are best detected when the total heritability is either very low or very high.
The increase in power achieved from including a MZ co-twin in the analysis is similar to the gain in power achieved from using repeated (phenotypic) measurements of the same individual (see e.g. Boomsma (2000) for an example of this strategy). However, we expect the gain in power from including a MZ co-twin to be greater. Similar to the case of the MZ co-twin, repeated measurements of the same individual provide additional estimates of the sibling correlation for the particular IBD class. However, repeated measurements differ only because of measurement error, whereas MZ twins differ from one another because of measurement error and (other) unique environmental influences. In other words, in the case of the repeated measures design, all of the correlations between siblings are contaminated with the same unique environmental influences. In contrast, the addition of an MZ co-twin provides another (potentially more accurate) estimate of the sibling correlation, independent of the unique environmental influences affecting the other twin.

Finally, we note that a pedigree consisting of a MZ pair and sibling is never as informative as a sib-trio (Tables 1 and 2). This is not surprising since the sib-trio provides information about IBD sharing between three individuals (i.e. sib 1 vs sib2, sib2 vs sib3, and sib1 vs sib3), whereas the MZ twins and sibling unit only provides one such relationship (i.e. twins one and two vs sib one). However, including an MZ twin in the analysis is economical since only one twin will need to be genotyped. In the end it will be up to the investigator to determine whether the cost of phenotyping an additional individual is outweighed by the gain in power to detect linkage.

Acknowledgements

We would like to thank Jeff Williams, John Blangero, Peter Visscher, David Duffy, Nick Martin and Pak Sham for helpful discussions.

References


Received: 27 March 2003
Accepted: 21 August 2003