Analysis of Case-Parental Control Studies: Method for the Study of Associations between Disease and Genetic Markers

W. Dana Flanders and Muin J. Khoury

Case-control studies using parents of case subjects as the control subjects provide an innovative way to study associations of genetic markers with disease risk. This approach, sometimes called the haplotype-relative risk method, has received recent attention because the use of parents as control subjects may reduce or eliminate the confounding associated with differences in race, ethnicity, or genetic background. We provide a new method for analysis of such case-parental control studies. The method of analysis is noniterative and yields simple estimates of risk ratios associated with genetic markers. It easily accommodates the situation in which data are available from only one parent. Although we illustrate the approach for a locus with two alleles, the analyses extend immediately to loci with multiple alleles. Am J Epidemiol 1996;144:696-703.

In genetic epidemiology, association studies seek to determine the association between a specific genetic marker and disease risk. This marker may be a genetic mutation or a resulting alteration in the structure or activity of enzymes or other proteins. As the ability to search for DNA markers at several sites improves, the potential for use of case-control studies to assess the association between such markers and disease risk is increasing (1-4).

In case-control studies, however, confounding and selection bias due to inappropriate choice of controls may distort results (1, 5, 6). For example, if race, ethnicity, or ancestry correlates with disease risk, perhaps because of environmental factors, and also with variations in the prevalence of the genetic marker, bias may occur and distort estimates of association. Therefore, selection of control subjects is a key design feature of such case-control studies. Selection of controls from a group with similar racial, ethnic, and ancestral backgrounds as that from which the case subjects arose can reduce this sort of confounding bias.

Rubinstein et al. (7) and Falk and Rubinstein (8) proposed an approach to selection of controls for case-control studies of association. In essence, they proposed using parents of cases as a sort of control group. They proposed comparing the distribution of the non-transmitted alleles in the parents of cases with the corresponding distribution in cases. This approach has the advantage that the genotype characterizing the controls comes from the same genetic population that gave rise to the cases. A major advantage of this study design using parents as controls is the elimination of the possibility that case-control differences are due to selection of controls whose genetic backgrounds differ systematically from those of cases. We will refer to this study design descriptively as the case-parental control study.

This case-parental control study, sometimes called the haplotype-relative risk method, has received attention recently (9-12). Ott (9) derived properties of the haplotype-relative risk method. Schaid and Sommer (10) derived likelihood methods for estimating relative risk parameters. Knapp et al. (11) showed that the odds ratio from such studies should not overestimate the true risk ratio comparing risk among those with a specific genotype with the risk among those with a comparison genotype.

Here we describe an alternative method for analysis of case-parental control studies. The approach some-
what resembles the case-base study design in which the exposure odds of cases are compared with those of controls sampled from the entire risk set. In our approach to analysis of case-parental control studies, the exposure odds of cases are compared with the exposure odds of all genotypes that would be expected from each mating type. The method is not iterative, does not require maximization of a likelihood, and leads to closed-form estimates of the risk ratio comparing the risk among those with a specific genotype with the risk among those with a comparison genotype. We also show how the method applies when information is available from more than one offspring per family. We discuss how one can stratify on exposure to an environmental factor to assess potential gene-environment interactions.

MATERIALS AND METHODS

We consider a case-parental control study in which we sample nuclear families with an affected offspring and determine the genotype of the offspring and his or her parents. We show how to estimate the risk ratio comparing the risk of disease among those with a particular genotype with the risk among those with other genotypes.

We derive results for a disease susceptibility locus that has two alleles, although the extension to multiple alleles is immediate. Denote the normal allele by \( N \) and the susceptibility allele by \( S \). Let \( D \) denote disease status (1 if disease present, 0 otherwise), \( g \) the offspring genotype \((g = 1, 2, \text{or 3 for genotype } NN, NS, \text{or SS, respectively})\), and \( m \) the parental mating type \((m = 1, 2, \text{or 3 for } NN \times NS, NS \times NS, \text{and } NS \times SS; \text{table 1})\). We do not consider other parental mating types since, for the other types, only one offspring genotype can occur so that such mating types are not informative. Also, let \( R_g \) denote the disease risk for subjects with genotype \( g \). The \( R_g \) represent average risks among those with each genotype. We assume that unmeasured (environmental) exposures do not correlate with mating type, i.e., no unmeasured confounders. The observed data consist of the number of cases by genotype and mating type, as summarized in table 1. Here, \( A_{ij} \) denotes the number of cases with mating type \( i \) and genotype \( j \), for \( i = 1, 2, 3 \) and \( j = 1, 2, 3 \). Throughout, we use the notation that a subscript replaced by a dot implies summation over the missing subscript.

First, consider the expected value of \( A_{ij} \). Conditional on having sampled \( A_i \) cases with parental mating type \( i \), the expected value is \( E(A_{ij}) = A_i \cdot P(g = j|D = 1, m = i) \cdot P(D = 1|g = j, m = i) = A_i \cdot P(g = j|m = i) \cdot R_j / P(D = 1|m = i) \). We can easily calculate \( P(g = j|m = i) \) for Mendelian transmission. For example, offspring of mating type \( m = 2 \) should have genotypes \( NN, NS, \text{and SS} \) with relative frequencies \( 1/4, 1/2, \text{and 1/4, respectively} \).

On the other hand, if disease were unrelated to genotype, we would expect the value of \( A_{ij} \) to be \( E_0[A_{ij}] = A_i \cdot P(g = j|m = i) \), where \( E_0[A_{ij}] \) denotes the expected value of \( A_{ij} \) under the null hypothesis: \( R_1 = R_2 = R_3 \). We indicate these expected frequencies in table 1. For example, \( E_0[A_{13}] = A_3/4 \).

Now, we define "odds ratios" as:

\[
\psi_{ij} = \frac{\sum_{i \in I} A_{ij}}{\sum_{i \in I} E_0[A_{ii}]},
\]

for \( j = 2, 3 \), where \( S_2 = \{1, 2\} \) and \( S_3 = \{2\} \). These sets represent values of \( i \) that are informative: the values for which \( E_0[A_{ij}] \) in the numerator and \( E_0[A_{ii}] \) in the denominator are both nonzero.

### TABLE 1. Observed and expected cell counts by mating type with both parents available

<table>
<thead>
<tr>
<th>Parental mating type</th>
<th>Offspring genotype*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( g = 1 \text{ (NM)} )</td>
</tr>
<tr>
<td></td>
<td>Observed</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------</td>
</tr>
<tr>
<td>( m = 1 \text{ (NN x NS)} )</td>
<td>( A_{11} )</td>
</tr>
<tr>
<td>( m = 2 \text{ (NS x NS)} )</td>
<td>( A_{21} )</td>
</tr>
<tr>
<td>( m = 3 \text{ (NS x SS)} )</td>
<td>( 0 )</td>
</tr>
</tbody>
</table>

* \( NN \), genotype with two normal alleles; \( NS \), genotype with one normal and one susceptibility allele; \( SS \), genotype with two susceptibility alleles.

†"Observed" denotes observed cell count; "expected" denotes expected cell count under the null hypothesis.

‡ Cells with a 0 must be 0.
On the basis of expected cell counts, $\Psi_{21}$ should estimate

$$
\Psi_{21} = \frac{\sum_{i=1}^{2} A_i \cdot P(g = 2|m = i) \cdot P(D = 1|g = 2)/(P(D = 1|m = i) \cdot A_i \cdot P(g = 2|m = i))}{\sum_{i=1}^{2} A_i \cdot P(g = 1|m = i) \cdot P(D = 1|g = 1)/(P(D = 1|m = i) \cdot A_i \cdot P(g = 1|m = i))}
$$

(2)

$$
= \frac{\sum_{i=1}^{2} A_i \cdot P(g = 1|m = i) \cdot P(D = 1|m = i)}{\sum_{i=1}^{2} A_i \cdot P(g = 2|m = i) \cdot P(D = 1|m = i)}
$$

Similarly results show that $\Psi_{31}$ estimates $R_3/R_1$, the risk ratio comparing genotype $SS$ with $NN$.

Standard Taylor series results (8) show that, for large samples, we can estimate the variance of the logarithm of $\Psi_{j1}$ as:

$$
\text{Var}(\ln(\Psi_{j1})) 
= \frac{1}{\Sigma_1} \sum_{i \in S_j} A_{ij}(A_i - A_{ij})/(A_i \cdot E_0[A_{ij}]^2) 
+ \frac{1}{\Sigma_2} \sum_{i \in S_j} A_{ii}(A_i - A_{ii})/(A_i \cdot E_0[A_{ii}]^2) 
+ \frac{2}{\Sigma_1 \cdot \Sigma_2} \sum_{i \in S_j} A_{ij} \cdot A_{ii} / (A_i \cdot E_0[A_{ij}] \cdot E_0[A_{ii}])
$$

(3)

where $\Sigma_1$ and $\Sigma_2$ denote the numerator and denominator of equation 1. We have treated $A_{ij}$ conditionally on $A_i$ as multinomial.

We then calculate approximate confidence limits by treating the logarithm of the odds ratio as though it were normally distributed (8). We give a similar, but slightly more complicated estimator of $\Psi_{31}$ in appendix 1. This estimator should have more statistical stability than that given by equation 1, but it loses some of the simplicity.

**Autosomal recessive and autosomal dominant models**

If an autosomal recessive model is postulated, we can estimate the risk ratio, $\Psi_{AR}$, comparing risk among those with genotype $SS$ with that among those with genotype $NS$ or $AW$ as:

$$
\Psi_{AR} = \frac{\sum_{i=2}^{3} A_{i3}/E_0[A_{i3}]}{\sum_{i=2}^{3} (A_{i2}/A_{i1})(E_0[A_{i2}]/E_0[A_{i1}])}
$$

(4)

We can estimate the variance of the logarithm of $\Psi_{AR}$ using equation 3, after replacing $A_{i1}$ with $(A_{i1} + A_{i2})$, $E_0[A_{i1}]$ with $(E_0[A_{i1}] + E_0[A_{i2}])$, and taking $S_j = \{2, 3\}$.

If an autosomal dominant model is postulated, we can estimate the risk ratio, $\Psi_{AD}$, comparing risk among those with genotype $SS$ or $NS$ with risk among those with genotype $NN$ as:

$$
\Psi_{AD} = \frac{\sum_{i=1}^{2} (A_{i0} + A_{i2})/(E_0[A_{i0}] + E_0[A_{i2}])}{\sum_{i=1}^{2} A_{i1}/E_0[A_{i1}]}
$$

(5)

We can estimate the variance of the logarithm of $\Psi_{AD}$ using equation 3, after replacing $A_{ij}$ with $(A_{i2} + A_{i3})$ and $E_0[A_{ij}]$ with $(E_0[A_{i2}] + E_0[A_{i3}])$, and taking $S_j = \{1, 2\}$.

**Only one parental genotype known**

Often, the investigator will know the genotype of only one parent. Our analytical approach extends easily to allow estimation of risk ratios in this situation, provided that we make an additional assumption: we assume non-assortive mating, i.e., that the probability that the unavail-
able parent contributes a particular allele is independent of the genotype of the available parent.

We summarize relevant data as in table 2, where $M$ denotes the partially known mating types ($M = 1$, 2, and 3 for types $NN$, $NS$, and $SS$, respectively). Let $p$ denote the probability that the unknown parent passes allele $S$ to the offspring and $q = 1 - p$ the probability that some other allele ($N$) is passed. By assumption, these probabilities remain constant for all genotypes of the available parent. Conditional on the genotype of the available parent, the expected value of $B_{ij}$ is $B_i * P(g = j|M = i) = B_i * P(g = j|M = i) P(D = 1|g = j, M = i) P(D = 1|M = i) = B_i * P(g = j|M = i) R_i/P(D = 1|M = i)$. Given our assumptions, calculation of the probabilities $P(g = j|M = i)$ is straightforward. For example, the expected value of $B_{13} = P(g = 1|M = NN) * B_1 = q * B_1$, under the null hypothesis (table 2).

Now, define the odds ratios:

\[ *\Psi_{32} = (B_{22}/B_{21}) = (B_{12}/B_{11}) \]  

(6)

and

\[ *\Psi_{21} = (B_{22}/B_{21}) = (B_{33}/B_{32}). \]  

(7)

One the basis of expected values, $*\Psi_{31}$ should estimate

\[ *\Psi_{32} = (E[B_{22}]/E[B_{21}]) = (E[B_{12}]/E[B_{11}]), \]

\[ = (B_i * p * R_3 + P(D = 1|M = NS)/B_2 * q * R_1 + P(D = 1|M = NS))/B_2 \]

\[ + (B_i * p * R_2 + P(D = 1|M = NN)/B_1) \]

\[ = R_3/R_2. \]

Similar results show that $*\Psi_{21}$ consistently estimates $R_2/R_1$ and that $*\Psi_{32}/*\Psi_{21}$ estimates $R_3/R_1$. Standard Taylor series results show that:

\[ \text{Var}(\log(*\Psi_{32})) = 1/B_{22} + 1/B_{21} + 1/B_{33} + 1/B_{32}. \]

(9)

and

\[ \text{Var}(\log(*\Psi_{21})) = 1/B_{22} + 1/B_{21} + 1/B_{33} + 1/B_{32}. \]

(10)

If data for some offspring are available from only one parent and for others from both parents, we may combine the corresponding estimates after first assessing the appropriateness of combining estimates. For example, combining estimates might be inappropriate if assortive mating occurred so that the assumptions used for equation 1, but not those for equations 7 and 8, were reasonable. If combination is appropriate, then we can obtain a summary estimate using a precision-based, weighted average (8):

\[ \text{ln}(RR_f) = (\text{ln}(\Psi_{j1})/\text{Var}(\text{ln}(\Psi_{j1}))) \]

\[ + \text{ln}(\log(*\Psi_{j1})) \]

\[ + 1/\text{Var}(\log(*\Psi_{j1})) \]

(11)

for $j = 1, 2$.

**Multiply affected families**

If multiple diseased offspring occur in each family, our previous estimates continue to apply. Specifically, we show in appendix 2 that we can use equation 1 to estimate risk ratios when we know the genotypes of both parents and equations 6 and 7 when we know only one parental genotype. The possible lack of independence created by use of multiple offspring per family, however, implies that we should use the alternative variance estimates given in appendix 2.

**RESULTS AND DISCUSSION**

Our approach allows direct estimation of risk ratios in case-parental control studies of the association between disease and genetic markers, and it accommodates the situation in which the genotype of only one parent is known for some offspring. A major advantage of the case-parental controls study using parents as controls is the elimination of the possibility that

<table>
<thead>
<tr>
<th>Mating type (genotype of available parent)</th>
<th>Offspring genotype*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>g = 1 (NN)</td>
</tr>
<tr>
<td></td>
<td>Observed†</td>
</tr>
<tr>
<td>M = 1 (NN)</td>
<td>B_{11}</td>
</tr>
<tr>
<td>M = 2 (NS)</td>
<td>B_{21}</td>
</tr>
<tr>
<td>M = 3 (SS)</td>
<td>0</td>
</tr>
</tbody>
</table>

* NN, genotype with two normal alleles; NS, genotype with one normal and one susceptibility alleles; SS, genotype with two susceptibility alleles.
† "Observed" denotes observed cell count; "expected" denotes expected cell count under the null hypothesis.
† Cells with a 0 must be 0.
case-control differences are due to (unrecognized) systematic selection of controls with different genetic backgrounds.

Compared with most other approaches to the analysis of case-parental studies, our approach and that of Schaid and Sommer (10) have the important advantage of not needing an assumption of Hardy-Weinberg equilibrium. Our approach differs from that of Schaid and Sommer, in part, because it is noniterative and yields direct estimates of risk ratios. The approach also differs from that of Schaid and Sommer when only one parental genotype is known, because we justify our approach without assuming Hardy-Weinberg equilibrium, though we do assume nonassortive mating.

We have assumed the absence of confounding. If a measured risk factor affects disease risks and associates with genotype, we can stratify on exposure and obtain separate estimates for each stratum. The stratum-specific estimates will, of course, not suffer from confounding by the risk factor. When appropriate, we can obtain summary estimates by standardization or by use of a weighted average (13).

We have illustrated how the method of analysis allows estimation of risk ratios for autosomal recessive and autosomal dominant patterns of inheritance. Although not shown here, the approach also generalizes easily to allow estimation of risk ratios for loci with multiple alleles. The analyst simply 1) defines the appropriate additional mating types and genotypes of offspring, 2) calculates expected values and combines them to form marginal totals, and 3) calculates odds ratios and variance estimates as before.

The use of parental controls provides an innovative approach to the selection of controls in case-control studies of genetic markers. They come from a population genetically similar to that from which the cases arose, thereby reducing or eliminating potential confounding associated with differences in ethnic or genetic backgrounds. Their use allows estimation of genetic marker disease associations and, as noted by Khoury et al. (1, 12), of gene-environment interactions (15). Inclusion of at least two affected siblings in each sibship would also allow linkage analyses, making use of sibs an attractive alternative to use of parental controls.

In summary, we have presented a new method for analysis of case-parental control studies. The method is noniterative, leads to closed-form estimates of risk ratios, and does not depend on Hardy-Weinberg equilibrium. If the genotype of only one parent is known, the method does depend on an assumption of nonassortive mating. The method extends immediately to study of multiple alleles at a given locus and allows for assessment of gene-environment interactions.

REFERENCES
APPENDIX 1

Alternative Estimator for $R_3/R_1$

We first estimate $\Psi_{32} = R_3/R_2$ as:

$$\Psi_{32} = \frac{\sum_{i=2}^{3} A_{i}/E_0[A_{i}] - \sum_{i=2}^{4} A_{i2}/E_0[A_{i2}]}{\sum_{i=2}^{3} A_{i}/E_0[A_{i}]}.$$  \hspace{1cm} \text{(A1)}

The ratio of $\Psi_{32}$ to $\Psi_{21}$ estimates $R_3/R_1$:

$$\Psi_{31} = \frac{\Psi_{32}}{\Psi_{21}}.$$  \hspace{1cm} \text{(A2)}

This last estimator involves more cells than the estimator in equation 1 and so should have greater statistical stability. The cost is loss of simplicity.

Standard Taylor series results (8) show that, for large samples, we can estimate the variance of the logarithm of $\Psi_{32}$ as:

$$\text{Vár}(\ln(\Psi_{32})) \approx \frac{1}{2} \sum_{i=2}^{3} A_{i}(A_{i} - A_{0})/(A_{i} \cdot E_0[A_{i}]^2)$$

$$+ \frac{1}{2} \sum_{i=2}^{4} A_{i2}(A_{i} - A_{0})/(A_{i} \cdot E_0[A_{i2}]^2)$$

$$+ \frac{2}{3^*} \sum_{i=2}^{3} A_{i3} \cdot A_{i2}/(A_{i} \cdot E_0[A_{i3}] \cdot E_0[A_{i2}])$$  \hspace{1cm} \text{(A3)}

where $3^*$ and $4^*$ denote the numerator and denominator of equation A1. We now estimate the variance of $\Psi_{31}$ using

$$\text{Var}(\ln(\Psi_{31})) = \text{Var}(\ln(\Psi_{32})) + \text{Var}(\ln(\Psi_{21})) - 2 \cdot \text{Cov}(\ln(\Psi_{32}), \ln(\Psi_{21}))$$  \hspace{1cm} \text{(A4)}

where the covariance term is given by

$$\text{Cov}(\ln(\Psi_{32}), \ln(\Psi_{21})) \approx \frac{1}{A_2} \left[ - \frac{A_{23}A_{22}}{3^*} \cdot E_0[A_{23}] \cdot E_0[A_{22}] + \frac{A_{23}A_{21}}{2^*} \cdot E_0[A_{23}] \cdot E_0[A_{21}] \
+ \frac{A_{22}A_{22}}{4^*} \cdot E_0[A_{22}] \cdot E_0[A_{22}] - \frac{A_{22}A_{21}}{2^*} \cdot E_0[A_{22}] \cdot E_0[A_{21}] \right]$$  \hspace{1cm} \text{(A5)}

(Appendix 2 follows)
APPENDIX 2

Estimation with Multiple Offspring per Family

We show that we can still estimate \( R_j/R_i \) by using equation 1 when both parental genotypes are known. Validity of other estimators presented here follows with similar logic. Let \( K_i \) denote the number of families of mating type \( i \), \( H_{ik} \) denote the number of offspring from the \( k \)th family of mating type \( i \), and \( X_{hijk} \) be 1 if the genotype of offspring \( h \) (\( h = 1, 2, \ldots, H^i \)) from the \( k \)th family (\( k = 1, 2, \ldots, K_i \)) with mating type \( i \) has genotype \( j \), and 0 otherwise.

Now, under the null hypothesis (\( R_1 = R_2 = R_3 \)), the expected value of \( X_{hijk} \) is \( E_0(X_{hijk}) = P(g = j|m = i) \). We can express the odds ratio in equation 1 as:

\[
RR_{21} = \frac{\sum_{h=1}^{H_i} \sum_{k=1}^{K_i} X_{h2k} E_0[X_{h2k}]}{\sum_{h=1}^{H_i} \sum_{k=1}^{K_i} X_{h1k} E_0[X_{h1k}]} \cdot \frac{\sum_{i=1}^{2} \sum_{j=1}^{K_i} Y_{j} E_0[Y_{j}]}{Y_{ik} E_0[Y_{ik}]} \tag{A6}
\]

where equality of \( E_0(X_{hijk}) \) and \( E_0[X_{i1j}] \) for all \( h \) and \( k \) justifies the second and third equalities, and where we have defined \( Y_{hijk} = X_{hijk}/E_0[X_{i1j}] \).

Now, conditional probabilities imply \( E(X_{hijk}) = P(X_{hijk} = 1) = P(g = j|m = i) \). Similarly, \( E_0[X_{i1j}] = P(g = j|m = i) \). Thus, \( E(Y_{hijk}) = R_j/P(D = 1|m = i) \). Now the expectation of a sum of random variables \( (Y_{hijk}) \) equals the sum of expectations, even if the variables are not independent. Thus,

\[
E[\sum_{k=1}^{K_i} X_{hijk}/E_0[X_{i1j}]] = \sum_{i=1}^{2} Y_{ij} = A_i * R_j/P(D = 1|m = i), \tag{A7}
\]

for \( i, j = 1, 2, 3 \), where \( \Sigma_{hk} \) and \( \Sigma_k \) denote summation over \( h \) and \( k \) and over \( k \), respectively.

Substituting this result with equation A6 shows that, on the basis of expected values, \( RR_{21} \) consistently estimates \( R_2/R_1 \).

However, \( X_{iijk} = A_{ijk} \), so that equations 1 and A6 are equivalent, after cancellation, and \( RR_{12} = \Psi_{12} \). Thus, we can use the same equations to estimate risk ratios when multiple-disease offspring are included from one family.

On the other hand, we must modify our variance estimates, since disease status of one offspring may not be independent of that of other offspring within the same sibship, even conditional on parental mating type. For example, dependency could reflect shared environmental exposures so that \( Y_{hijk} \) would correlate with \( Y_{h'j'k} \).

On the other hand, we can treat \( Y_{hijk} \) and \( Y_{h'i'j'k'} \) as independent for \( (i', k') \) and different from \( (i, k) \). Thus, Taylor series results show

\[
\text{Var}(\ln(\Psi_{12})) = \frac{1}{\Sigma_1} \sum_{i=1}^{2} \sum_{k=1}^{K_i} (Y_{ijk} - \bar{Y}_{ijk})^2/(K_i - 1) \]

\[
+ \frac{1}{\Sigma_2} \sum_{i=1}^{2} \sum_{k=1}^{K_i} (Y_{ijk} - \bar{Y}_{ijk})^2/(K_i - 1) \tag{A8}
\]

\[
+ \frac{2}{\Sigma_1 \cdot \Sigma_2} \sum_{i=1}^{2} \sum_{k=1}^{K_i} (Y_{ijk} - \bar{Y}_{ijk}) \cdot (Y_{ijk} - \bar{Y}_{ijk})/(K_i - 1)
\]

where \( \Sigma_1 \) and \( \Sigma_2 \) denote the numerator and denominator of equation 1, respectively, and where \( Y_{ijk} \) overscored with a “bar” denotes the mean value of the \( Y_{ijk} \), averaged over the \( K_i \) families with mating type \( i \). By using \( Y_{ijk} \), we have treated the sibship as the unit of analysis. A similar result holds for \( \text{Var}(\log(\Psi_{32})) \).

Similarly, for the situation in which we know the genotype of only one parent, we can estimate the log variances as:
\[
\text{Var}(\log(*\Psi_{32})) = \sum_{k=1}^{K_2} \left( \frac{(X_{23k} - \bar{X}_{23k})/B_{23} - (X_{21k} - \bar{X}_{21k})/B_{21}}{\bar{X}_{21k}/B_{21}} \right)^2/(K_2 - 1) \\
+ \sum_{k=1}^{K_1} \left( \frac{(X_{12k} - \bar{X}_{12k})/B_{12} - (X_{11k} - \bar{X}_{11k})/B_{11}}{(K_1 - 1)} \right)^2/(K_1 - 1)
\]

\[
\text{Var}(\log(*\Psi_{21})) = \sum_{k=1}^{K_3} \left( \frac{(X_{33k} - \bar{X}_{33k})/B_{33} - (X_{32k} - \bar{X}_{32k})/B_{32}}{(K_3 - 1)} \right)^2/(K_3 - 1) \\
+ \sum_{k=1}^{K_2} \left( \frac{(X_{23k} - \bar{X}_{23k})/B_{23} - (X_{21k} - \bar{X}_{21k})/B_{21}}{(K_2 - 1)} \right)^2/(K_2 - 1).
\]