Meta-Analysis and Combining Information in Genetics

Rudy Guerra
&
David Allison

Contents

CHAPTER 1

Heterogeneity in Meta-Analysis of Quantitative Trait Linkage Studies

Hans C. van Houwelingen and Jérémie J. P. Lebrec
Leiden University Medical Center, The Netherlands

1.1 Introduction

In complex diseases where many genes might be involved in the genetic causation of the disease, individual loci influencing a quantitative trait are most likely to explain only a small proportion of its total variance. Consequently, there is a huge problem of lack of statistical power. Most linkage studies published to date only consist of a few hundred pedigrees with a limited number of individuals and, therefore, have little power to detect linkage of any but the ”largest” quantitative trait loci (QTL). In order to enhance power, it is now common practice to retrospectively pool evidence for linkage from several different studies. However, in pooling data from different studies, one should be aware of the possible heterogeneity between studies. The aim of this chapter is to present statistical models for describing this heterogeneity and approaches to analyze heterogeneous data

We distinguish two types of heterogeneity: locus and size heterogeneity. The populations used in each of the studies often have different genetic backgrounds and a locus affecting the trait of interest in one population might have no effect in another one; we will refer to this type of heterogeneity as *locus heterogeneity*. In other instances, the same locus may influence the trait in all populations, but there are many reasons to believe that the size of the effect will vary. For instance, the frequency of the causal allele may be much smaller in some populations or it may interact with other loci, or with environments and risk factors. We will refer to this type of heterogeneity as *size heterogeneity*.

Besides those biological sources of heterogeneity, some common logistic sources of variation often arise: typically, genotyping will have been carried out on different marker maps (and even when identical markers are used, their allele frequencies may vary across populations) and families may have been sampled according to different schemes. More simply, the phenotypes measured may vary in their method of collection from study to study.

When the raw data are available, one obvious way to gather evidence from several studies is to pool the data into a meta-file and proceed with an overall analysis. In the case of linkage studies with different marker maps, the data manipulations involved are very tedious. Moreover, the data sets become unnecessarily large because of the artificially created missing data on markers that are used in other studies. Furthermore, running standard methods of analysis on such large data files usually requires uncommon computing capacities. Therefore, we advocate the meta-analytic approach that collect all relevant summary information for each study and uses that as starting point for further analysis. Of course another simple reason for favoring meta-analysis is that researchers usually simply cannot access the raw data for each study and have to be content with individual test statistics along with (at best) parameter estimates.

We refer the reader to Dempfle and Loesguen (2003) and Rao and Province (2001) for recent overviews of meta-analytic methods for linkage studies. Most methods are in the spirit of the classical meta-analysis. An interesting, widely applicable, alternative are the rank-based methods such as the GSMA (Wise *et al.*, 1999). They might be sub-optimal compared to approaches based on the pooling of estimates of a common linkage parameter, but much more robust because of the built-in genomic control. Note that associated methods that assess heterogeneity have recently been developed (Zintzaras and Ioannidis, 2005). The idea of pooling different estimates of a common linkage effect across studies is not new although it has only been described for sib pair designs to date. Gu *et al.*(1998) use the excess identical-by-descent (IBD) sharing as a common effect, but their approach appears to be limited to studies with the same marker maps. Li and Rao (1996) and Etzel and Guerra (2002) both use the slope in a classical Haseman-Elston regression as a common effect, the former suffering the same restriction as Gu *et al.*(1998) regarding location of markers. Interestingly in the latter, the authors explicitly adjust for the (study-specific) marker to locus distance and allow for heterogeneity across studies by means of a random effect. Unfortunately, they do not seem to efficiently take into account the within-study dependence structure between markers.

Classical methods of meta-analysis originally introduced in the field of clinical trials (DerSimonian and Laird, 1986) can be adapted to linkage studies. The sufficient statistics used to perform such approaches are some measure of effect on a common grid of putative locations and its associated standard error. In the case of quantitative traits, a natural estimate of common linkage effect is the proportion of total variance explained by a putative location. We first describe the meta-analytic tools, assuming that QTL effect estimates and standard errors are available for all studies on a *common grid* of locations. In Section 1.2 the traditional meta-analytic approach in the context of linkage is reviewed, including how to test and allow for *size heterogeneity*, while in Section 1.2.4 we introduce a simple finite mixture model to account for potential *locus heterogeneity*. A complication that arises in both approaches for heterogeneous data is that variance components are nonnegative by definition. We will discuss the consequences of that for estimation and testing. In Section 1.3, we quickly review the methods which should be used for the analysis of individual stud-

ies in order to yield the relevant statistics required for meta-analysis as advocated in Sections 1.2. All methods are illustrated by means of four data sets used for a genome-wide scans for lipid levels in Section 1.4.

1.2 The classical meta-analytic method

Introductions to classical meta-analysis can be found in two Tutorials in Biostatistics in Statistics in Medicine, namely Normand (1999) and van Houwelingen *et al.*(2002). In this section, we recall briefly how meta-analysis is classically carried out and introduce some refinement that is specific to the variance component model used in linkage studies. We assume that at a given *common putative position*, each study (indexed by $i = 1, \ldots, K$) provides a consistent estimate $\hat{\gamma}_i$ of the true QTL effect γ_i of that locus and an associated standard error s_i. The link with the traditional lodscore is given by $\mathrm{LOD_i} = (\hat{\gamma_i}^2/\mathrm{s_i^2})/(2 \times \ln(10))$. Details of the definition of the variance component and its estimation are given in Section 1.3.

1.2.1 Analysis under homogeneity

The simplest approach to meta-analysis assumes that the effects γ_i's are all equal to a common value γ so that $\hat{\gamma}_i \sim N(\gamma, s_i^2)$. This is known as the ***homogeneity assumption*** In this situation the corresponding maximum likelihood estimator of γ is given by the weighted average

$$\hat{\gamma}_{\mathrm{hom}} = \frac{\sum_i \hat{\gamma}_i/s_i^2}{\sum_i 1/s_i^2} \text{ with standard error } SE_{\mathrm{hom}} = 1/\sqrt{\sum_i 1/s_i^2}\,. \tag{1.1}$$

The null hypothesis of no effect, that is $\gamma = 0$ versus the alternative $\gamma > 0$, can be tested by means of the one-sided statistic

$$\left(z^+_{\mathrm{hom}}\right)^2 = \begin{cases} \left(\hat{\gamma}_{\mathrm{hom}}/SE_{\mathrm{hom}}\right)^2, & \text{if} \hat{\gamma}_{\mathrm{hom}} > 0 \\ 0 & \text{if} \hat{\gamma}_{\mathrm{hom}} \leq 0 \end{cases}$$

which follows the mixture distribution $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ under the null hypothesis, where χ_0^2 denotes the degenerate density with all mass in 0. The corresponding $\mathrm{LOD_{hom}}$ score can be calculated as $\left(z^+_{\mathrm{hom}}\right)^2 / \left(2 \times ln(10)\right)$. Observe that we do not truncate the estimated $\hat{\gamma}_i$ at zero, if negative, because that would complicate the pooling considerably. However, truncation is no problem in the final stage.

1.2.2 Test for heterogeneity

Even when the same locus is affecting a trait in different populations, it seems difficult to believe, for reasons given in Section 1.1, that the QTL effects are all equal. In

the setting introduced earlier, this situation of *size heterogeneity* can be tested:

$$\begin{aligned} \mathrm{H}_0 &: \gamma_1 = \gamma_2 = \cdots = \gamma_K \equiv \gamma_{\mathrm{hom}} \\ \mathrm{H}_1 &: \text{at least one } \gamma_i \text{ is different ,} \end{aligned}$$

the hypothesis of homogeneity H_0 can be tested using the following statistic

$$X^2 = \sum_{i=1}^{K} \frac{(\hat{\gamma}_i - \hat{\gamma}_{\mathrm{hom}})^2}{s_i^2}$$

whose approximate null distribution is χ^2_{K-1}. In practice, any test for heterogeneity is likely to have little power because individual studies tend to have low precision. Nonetheless, the test can formally suggest heterogeneity in some instances, as will be seen in Section 1.4. Note that the X^2 statistic has an appealing interpretation (at least for researchers with experience in parametric linkage). Indeed, it can be re-written as

$$\begin{aligned} X^2 &= \sum_{i=1}^{K} \frac{\hat{\gamma}_i^2}{s_i^2} - \frac{\hat{\gamma}^2_{\mathrm{hom}}}{(\sum_i 1/s_i^2)^{-1}} \\ &= 2 \times \ln 10 \times \Big(\sum_{i=1,\ldots,K} \mathrm{LOD}_i - \mathrm{LOD}_{\mathrm{hom}} \Big) . \end{aligned}$$

In other words, the individual LODs add up only when the effect is perfectly homogeneous.

1.2.3 Modeling size heterogeneity

The classical way to allow for heterogeneity between studies is to introduce an additional layer in the earlier homogeneous model by assuming that the true study specific effects γ_i's themselves arise from some distribution. The usual model is a normal distribution with common mean γ and a between study variance σ^2. This is referred to as a normal mixture model (or random effect model) and results in marginal distributions for the observations given by $\hat{\gamma}_i \sim N(\gamma, s_i^2 + \sigma^2)$. If the between study variance σ^2 were known, the estimate of γ would be

$$\hat{\gamma}_{\mathrm{het}}(\sigma^2) = \frac{\sum_i w_i \hat{\gamma}_i}{\sum_i w_i} \text{ with } w_i = \frac{1}{\sigma^2 + s_i^2} \text{ and with standard error } SE_{\mathrm{het}} = 1/\sqrt{\sum_i w_i} ,$$

So, one way to carry out estimation is by maximization of the profile log-likelihood $pl(\sigma^2) = l(\hat{\gamma}_{\mathrm{het}}(\sigma^2), \sigma^2)$.

In the context of linkage where the actual effects γ_i's are standardized variance components themselves, this model only makes sense if the probability $\Phi(-\gamma/\sigma)$ of negative γ's is negligibly small. In practice that is achieved if the coefficient of variation $\sigma/\gamma < 1/2$. For the same reasons, the null hypothesis of no locus effect requires that all γ_i's should be equal to 0 with probability 1. Hence, the null hypothesis specifies both $\gamma = 0$ and $\sigma^2 = 0$, which is different from the usual situation in meta-analyzes of clinical trials. The test for linkage is then given by the corresponding

log-likelihood difference

$$2 \times \left[pl(\hat{\sigma}^2) - l\left(\gamma = 0, \sigma^2 = 0\right)\right]$$

so that evidence for heterogeneity potentially contributes to the rejection of the null hypothesis of no linkage. The use of the usual mixture $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ for the null distribution of this non-standard likelihood is anti-conservative, the correct asymptotic distribution is given by a mixture $(\frac{1}{2}-p)\chi_0^2+\frac{1}{2}\chi_1^2+p\chi_2^2$ (Self and Liang, 1987). However, asymptotic results are unlikely to be useful since we typically have very few observations (i.e. studies) to pool together. In practice, we use the anti-conservative limits dictated by the $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ mixture as a screening tool and resort to parametric bootstrapping for refinement of the level of significance once interesting positions have been identified.

1.2.4 A two-point mixture model for locus heterogeneity

In some cases, the previous model will not be adequate to model differences between studies because heterogeneity is qualitative rather than quantitative, in other words the locus influences the trait in some studies/populations and not at all in others. There is an indication of such qualitative heterogeneity when the normal mixture model yields a large coefficient of variation σ/γ allowing negative γ 's under the normal mixture . In analogy to what is done routinely at the family level in parametric linkage (e.g. Ott (1999), see also Holliday *et al.*(2005) for a recent application) and can be done in the variance components setting (Ekstrom and Dalgaard, 2003), one can fit a two-point mixture model at the study level as follows: $\hat{\gamma_i}|\gamma_i \sim N(\gamma_i, s_i^2)$ with

$$\gamma_i = \begin{cases} \gamma, & \text{with probability } \alpha; \\ 0, & \text{with probability } 1-\alpha \end{cases}$$

so that, marginally,

$$\hat{\gamma_i} \sim \alpha N(\gamma, s_i^2) + (1-\alpha)N(0, s_i^2) \; .$$

The basic idea is that only a proportion α of the studies show linkage to the putative locus and γ is the QTL effect among those studies only. (Hence, γ is not longer the mean value of the γ_i's as in the normal mixture model. Care is needed when comparing the models) . For estimation purposes, this mixture of normal distributions naturally lends itself to the EM algorithm (Dempster *et al.*, 1977). Denoting by $\phi(x; \mu, \sigma^2)$ the normal density function with mean μ and variance σ^2, the E (estimation) step at stage $k+1$ of the iterative procedure consists in calculating the posterior probabilities $\tau_i^{(k+1)}$'s that the $\hat{\gamma}_i$'s have arisen from a normal distribution with mean $\gamma^{(k)}$ given the prior mixing proportion $\alpha_{(k)}$ i.e.

$$\tau_i^{(k+1)} = \frac{\alpha^{(k)}\phi(\hat{\gamma}_i, \gamma^{(k)}, s_i^2)}{\alpha^{(k)}\phi(\hat{\gamma}_i, \gamma^{(k)}, s_i^2) + (1-\alpha^{(k)})\phi(\hat{\gamma}_i, 0, s_i^2)} \; ,$$

whereas the M (maximization) step gives the updated parameters $\alpha^{(k+1)}$ and $\gamma^{(k+1)}$ as

$$\begin{aligned} \alpha^{(k+1)} &= \sum_{i=1}^{K} \tau_i^{(k+1)}/K \\ \gamma^{(k+1)} &= \frac{\sum_{i=1}^{K} \hat{\gamma}_i \tau_i^{(k+1)}/s_i^2}{\sum_{i=1}^{K} \tau_i^{(k+1)}/s_i^2} \, . \end{aligned}$$

The model parameters α and γ are constrained in $[0,1]$ and $[0,+\infty[$ respectively and although the EM estimation procedure described above ensures that $\alpha \in [0,1]$, the estimate of γ will sometimes be negative in which case we set $\hat{\gamma} = 0$ and $\hat{\alpha} = 0$ too. Under usual regularity conditions, the corresponding likelihood ratio test would be asymptotically distributed as a $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ under the null hypothesis. However, here the situation is further complicated by the fact that the model parameters are not identifiable under the null hypothesis (indeed if $\gamma = 0$, any choice of α will give the same likelihood). One way to tackle this problem is to slightly modify the likelihood as done by Chen *et al.*(2001) and derive corresponding simple asymptotics, but for the same reason alluded to in Section 1.2, we prefer to resort to parametric bootstrapping techniques in order to assess significance of the likelihood ratio test.

The model for size heterogeneity and locus heterogeneity could be combined into a model where either $\gamma = 0$ with probability $1 - \alpha$ or γ follows a normal distribution with probability α .

1.3 Extracting the relevant information from the individual studies

As we described in Section 1.2 the basic ingredients of a classical meta-analysis are study specific QTL effects' estimates $\hat{\gamma}_i$'s in the $i = 1, \ldots, K$ studies available and their associated standard errors s_i's on a *common* fine *grid* of genome locations. In this section, we explain how to obtain these estimates in practice and how to adjust for varying information across studies.

1.3.1 General approach

For random samples of the trait values, the variance components method (Almasy and Blangero, 1998; Amos, 1994) is the standard way of testing for linkage to a quantitative trait. Unfortunately, the emphasis of most computer programs implementing the variance components method has been placed on testing rather than estimating and they rarely provide both QTL effect estimates and associated standard errors. In the context of linkage, two exceptions that we know of are the `MENDEL` (Lange, 2001) and `Mx` softwares (Neale *et al.*, 1999). However, in principle, this is not so much of a problem because asymptotic standard errors s can be obtained provided the QTL effect estimate $\hat{\gamma}$ is present (and differs from 0) in addition to its statistical significance, using the approximate relation $(\hat{\gamma}/s)^2 \simeq \chi^2$ with $\chi^2 = \text{LOD} \times 2 \times \ln(10)$.

At positions where the QTL estimate is 0, one could interpolate values of s at neighboring positions where $\hat{\gamma} \neq 0$. One problem with the variance components method, as far as pooling of estimates is concerned, is that $\hat{\gamma}$ is constrained to remain nonnegative and pooling of several imprecise estimates $\hat{\gamma}_i$'s could result in a positively biased estimate of the true QTL effect γ. Whenever possible, we would personally favor adequate regression or score test approaches (Lebrec *et al.*, 2004) to linkage whose slope is equal to $\hat{\gamma}$ and is allowed to be negative. As shown by Putter *et al.*(2002), such approaches are equivalent to the variance components method.

Often, data are selected based on phenotype values (selected sample such as affected sibpairs, extremely discordant pairs, etc ...), the variance components method is no longer valid and appropriate methods that take into account the sampling scheme need to be employed. These so-called inverse regression methods first introduced by Sham and Purcell (2001) have been implemented in `MERLIN-regress` (Sham *et al.*, 2002). A typical output from the software will provide a signed estimate of the QTL effect $\hat{\gamma}$ and associated standard error s at an arbitrary grid of positions. The software can also be used in case of random samples as an alternative for the variance components modules. Because of its very convenient output we advocate the use of `MERLIN-regress` when analyzing linkage data whenever suitable. One outstanding problem with `MERLIN-regress` is the use of an imputed covariance for IBD sharing which can lead to bias in estimation especially in genome areas where markers information is very low. In practice, one clear indication that the imputed covariance is not a good approximation is when the software either gives out QTL estimates larger than 1 with huge associated LOD scores or no estimates at all (NA). In practice, marker maps and densities vary widely and one often ends up with areas of the genome with scarce information. In this case, we advocate the use of a more reliable IBD covariance matrix which we calculate by Monte Carlo simulations. In Section 1.3.2, we provide more details on how we do this in practice.

1.3.2 Special case: sib pair designs

In order to show how we adjust for differing marker maps (or different allele frequencies on the same map), we now outline the inverse regression approach in the simplest and most widespread case of sib pair studies. The trait values $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)'$ are assumed to have been standardized and to follow the usual additive variance components model i.e. the vector $\boldsymbol{x}$ is assumed to follow a bivariate normal distribution with mean 0 and covariance matrix Σ

$$\Sigma = \begin{bmatrix} 1 & \rho + \gamma(\pi - \frac{1}{2}) \\ \rho + \gamma(\pi - \frac{1}{2}) & 1 \end{bmatrix} .$$

Here π is the proportion of alleles shared IBD measured exactly at the QTL position and γ therefore represents the proportion of total variance explained by the QTL, ρ is the marginal sib-sib correlation for the trait of interest. An extension of a relation shown in Putter *et al.*(2003) under complete information gives an approximate regression (valid for small values of γ) between excess IBD sharing and a function of

the phenotype trait values which is the basis of the inverse regression approach:

$$\mathrm{E}(\hat{\pi} - \frac{1}{2}|\boldsymbol{x}, \gamma) \simeq \gamma \ \mathrm{var}_0(\hat{\pi})\ C(\boldsymbol{x}, \rho)$$

where

$$\hat{\pi} = \frac{1}{2} \times \mathrm{P}_0(\pi = \frac{1}{2}|M) + 1 \times \mathrm{P}_0(\pi = 1|M)$$

is the usual estimate of IBD sharing given marker data M available while

$$C(\boldsymbol{x}, \rho) = \left[(1+\rho^2)x_1x_2 - \rho(x_1^2 + x_2^2) + \rho(1-\rho^2)\right] / (1-\rho^2)^2$$

and is sometimes referred to as the optimal Haseman-Elston function. For a sample of $j = 1, \ldots, N$ sib pairs, the method of least squares provides an approximately consistent estimate of γ given by

$$\hat{\gamma} = \frac{\sum_{j=1}^{N}(\hat{\pi}_j - \frac{1}{2})C(\boldsymbol{x}_j, \rho)}{\mathrm{var}_0(\hat{\pi}) \times \sum_{j=1}^{N} C^2(\boldsymbol{x}_j, \rho)}\ , \tag{1.2}$$

$$\text{with standard error } s = \left(\mathrm{var}_0(\hat{\pi}) \times \sum_{j=1}^{N} C^2(\boldsymbol{x}_j, \rho)\right)^{-1/2}\ . \tag{1.3}$$

Here $\mathrm{var}_0(\hat{\pi})$ represents the variance of $\hat{\pi}$ under the null hypothesis and would equal $\frac{1}{8}$ under complete information and although an exact calculation is extremely tedious it can be closely approximated by simple Monte Carlo simulations. For example, one can use the options `--simulate` and `--save` in `Merlin` (Abecasis *et al.*, 2002) to generate a large number of pedigrees with a given structure (sib pairs here), markers' characteristics (i.e. allele frequencies and inter-marker distances) and possibly missing pattern for genotypes, the true $\mathrm{var}_0(\hat{\pi})$ can then be accurately approximated by the sample variance of $\hat{\pi}$. We show in Figure 1.1 how widely this measure of marker information may vary within and between studies. It is therefore crucial to appropriately account for this variation when estimating γ, failure to do so may introduce bias in the QTL estimates. If no such information is available, it is possible in principle to calibrate scan by comparing mean or median QTL variance components over the whole genome between studies, but in small studies such methods might be prone to error.

1.3.3 Retrieving information on the common grid from an individual study

For the meta-analysis we need to define a common grid of locations and obtain QTL estimates on that grid for each study. However, it can happen that in the individual studies, the only data at hand are QTL estimates ($\hat{\gamma}$'s) and their standard errors (s's) on an original grid of locations which is not the common one we wish to use. Typically this original grid would be a set of say $t = 1, \ldots, M$ markers' positions. We show how to obtain QTL estimates and associated standard errors on this new common grid of locations, if the characteristics of the original map are available and from the IBD distribution for that map under the null hypothesis.

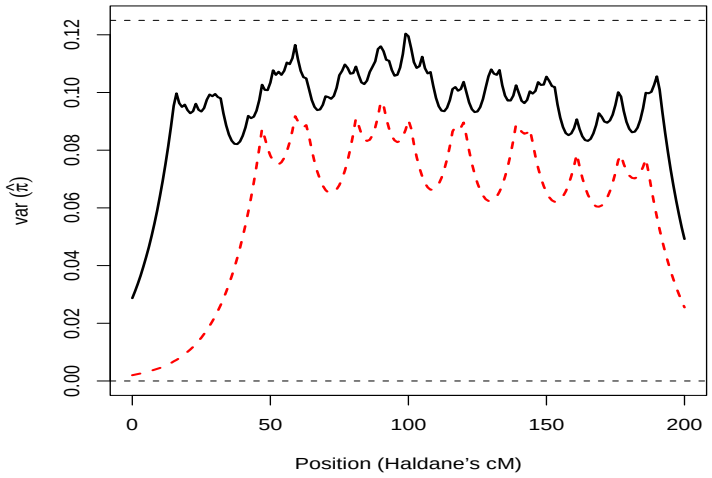

Figure 1.1 *Marker information (*$\text{var}_0(\hat{\pi})$*) in the Australian (continuous line) and Dutch (broken line) data sets Vs. position (Haldane's cM) - Chromosome 6*

For the sake of simplicity, we stick to sib-pair designs as in the previous section. Given the $M \times 1$ vector of original QTL effect estimates $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_t)_{t=1,\ldots,M}$ and associated standard errors $(s_t)_{t=1,\ldots,M}$, the best linear approximation of the QTL effect $\hat{\gamma}_q$ at an arbitrary position denoted q is given by a weighted least squares estimate

$$\begin{aligned} \hat{\gamma}_q &= \frac{\omega_q' V^{-1} \hat{\boldsymbol{\gamma}}}{\omega_q' V^{-1} \omega_q} , \\ \text{with standard error } s_q &= \left(\omega_q' V^{-1} \omega_q\right)^{-1/2} . \end{aligned}$$

Here $'$ denotes the transpose of a vector. The matrix V is proportional to the variance-covariance matrix of the vector $\hat{\boldsymbol{\gamma}}$ under the null hypothesis of no linkage and is given by

$$V_{kl} = \begin{cases} \text{var}_0(\hat{\pi}_k)^{-1} & \text{if } k = l \\ \text{Cov}_0(\hat{\pi}_k, \hat{\pi}_l)\ (\text{var}_0(\hat{\pi}_k)\ \text{var}_0(\hat{\pi}_l))^{-1} & \text{if } k \neq l \end{cases} ,$$

Furthermore, ω_q is the $M \times 1$ vector whose k^{th} element is given by

$$\omega_{q,k} = \frac{\text{Cov}_0(\hat{\pi}_k, \hat{\pi}_q)}{\text{var}_0(\hat{\pi}_k)} .$$

All the var_0 and Cov_0 terms can in principle be calculated by Monte Carlo simulations provided the map characteristics and pedigree structure are known.

In the idealized case of a saturated map which would supply perfect IBD knowledge at any location on a chromosome, all var_0 terms are equal to $\frac{1}{8}$ and $\text{Cov}_0(\hat{\pi}_{t_1}, \hat{\pi}_{t_2}) =$

$\frac{1}{8}(1 - 2\theta_{t_1,t_2})^2$, where θ_{t_1,t_2} is the recombination fraction between loci at t_1 and t_2 (Risch, 1990). Taking the off-diagonal terms in V to be equal to 0 (i.e. assuming that markers are not linked), one obtains the estimate of QTL effect advocated by Etzel and Guerra (2002) (in the special case that between-study variance $\sigma^2 = 0$). In the context of meta-analysis, it is important to properly account for differences in marker information between studies, unless the marker maps are close to saturated in all studies. Remarkably, the elements needed to calculate $\hat{\gamma}_q$ and s_q at any arbitrary location are just the corresponding estimates at M marker locations and map characteristics, none of the subject-specific data (traits values, individual IBD estimates $\hat{\pi}_i$) are needed.

1.4 Example

We applied the methods previously described to four data sets on lipid levels originating from Australia (aus), The Netherlands (nlj and nlo) and Sweden (swe). The full results are reported in Heijmans *et al.*(2005) and we have selected only one endpoint (LDL cholesterol levels) for illustration purposes. The data available for linkage analysis consisted almost entirely of sib pairs (371, 83, 110 and 36 pairs in the aus, nlj, nlo and swe data sets, respectively) with the exception of the Australian data set which also had 1 family with three siblings and 3 families with four siblings. Genotyping has been carried out using a common marker map for the nlj, nlo and swe data sets but with a different denser map for the aus data set. We actually had access to the raw data sets and could therefore easily obtain QTL estimates and standard errors on a common grid of positions.

Prior to linkage analysis (using `MERLIN-regress`), raw phenotypic data were adjusted for sex and age, within country. The analysis of the actual data revealed little differences between the three methods described in Section 1.2, this is partly due to the small sample sizes in the individual data sets which does not allow to clearly establish heterogeneity between studies. We present graphically the original results for two interesting chromosomes: chromosome 2 (Figure 1.2) and chromosome 13 (Figure 1.3). Note that the QTL variance estimates and LOD scores of the Pooled analyses have been multiplied by 0.95 and 1.05 for the random effect model (labelled 'het') and the two-point mixture model (labelled '2-p mixt') respectively, this was necessary to make all curves visible.

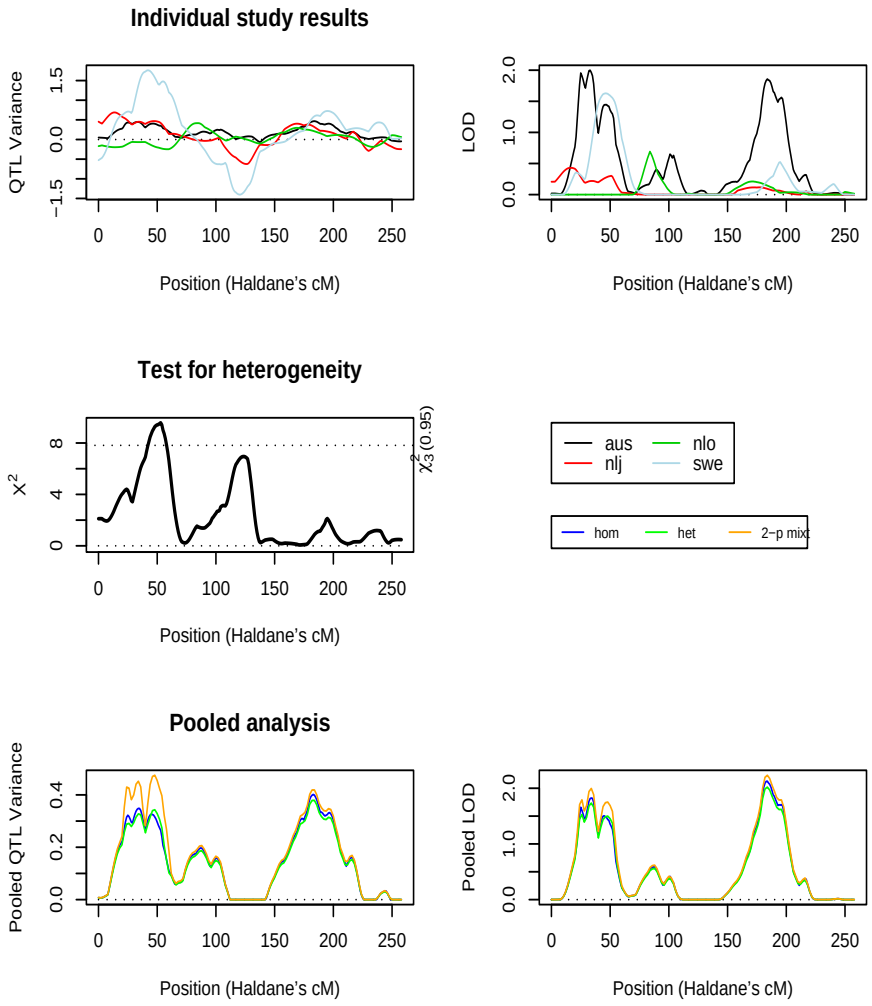

Figure 1.2 *Original data - Chromosome 2 - LDL cholesterol level*

In chromosome 2, the three pooled estimates of QTL variance coincide everywhere apart from the 20-60cM region where the two-point mixture model gives a higher estimate with corresponding estimate of proportion of study linked $\hat{\alpha} = 0.75$ (i.e. the 'nlo' data set is not linked) at 32cM where the maximum LOD score is attained. The corresponding pooled LOD score is roughly the same as the maximum LOD score obtained in the 'aus' data set and therefore there seems to be no gain in pooling the three linked data sets in this case. On chromosome 13, the pooling results in a very slight increase in LOD score in the region around 20cM compared to the maximum

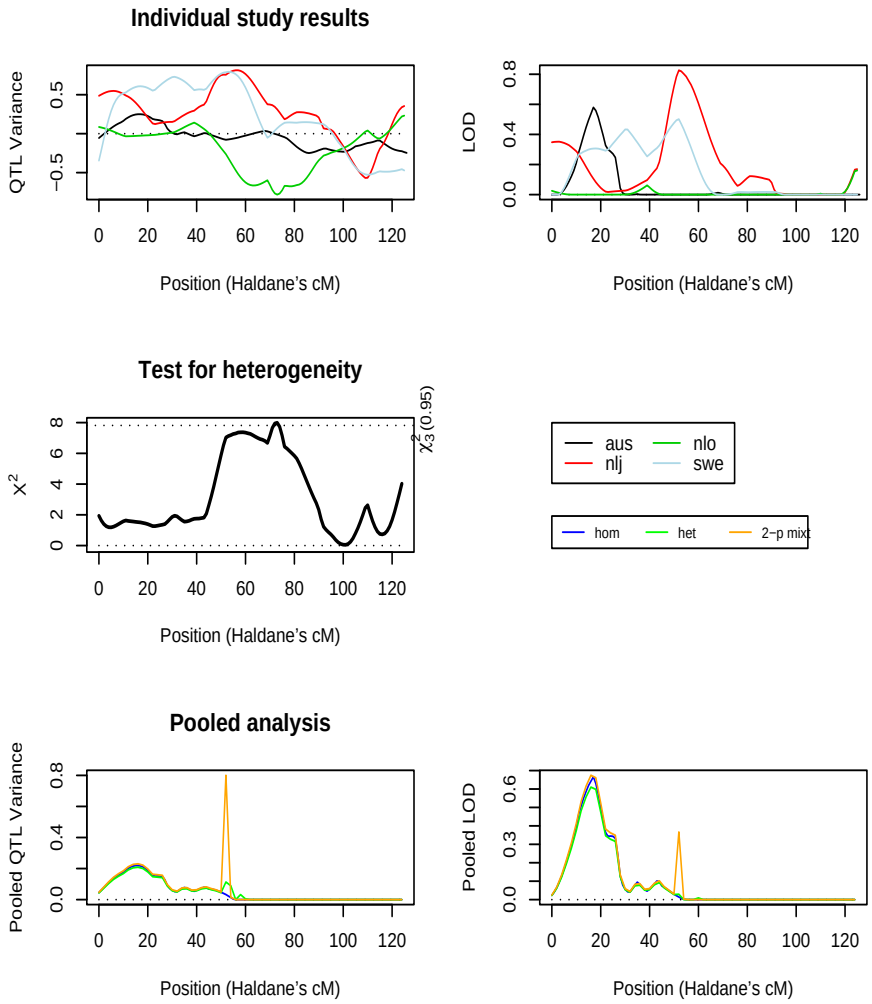

Figure 1.3 *Original data - Chromosome 13 - LDL cholesterol level*

of the individual LOD scores and the three methods give the same score. Note the sudden rise and fall in the estimate of QTL variance $\hat{\gamma}$ for the two-point mixture at 52cM which corresponds to a decrease in $\hat{\alpha}$ from 1.0 to 0.36. The fitting algorithm of the two-point mixture actually gave negative values for $\hat{\gamma}$ right of 54cM so the estimates were truncated to 0. Given those unconvincing real-life examples, one can legitimately asks the next two questions:

1. In practice, is there any gain in pooling data sets at all? I.e. can we obtain higher LOD scores than the maximum of the individual LOD scores?
2. Does allowance for heterogeneity help in enhancing statistical significance? I.e. are the LOD scores for the random effect model and/or the two-point mixture model ever higher than the LOD score of the homogeneity model?

The answer to question 1. is 'Yes' even when individual studies are small provided the QTL effects are more or less the same in all studies i.e. the assumption of homogeneity is verified. The answer to question 2. is also 'Yes' but only when the sample size in the individual studies are large enough as we show by means of a simulated example inspired from the original lipid levels data. We artificially increased the sample size of each of the four data sets by a factor 4 (i.e. the standard errors were divided by 2). The corresponding results are displayed graphically in Figure 1.4 for chromosome 2 and in Figure 1.5 for chromosome 13. In the 20-70cM region of chromosome 2, studies 'aus' and 'swe' both show clear linkage signals, QTL estimates vary quite widely across studies which is now unambiguously shown by the heterogeneity test. We are probably in presence of both quantitative and qualitative heterogeneity here since study 'nlo' shows no QTL effect at all. As a result, the significant signals observed in the 'aus' and 'swe' studies (maximum LOD score $\simeq 8$) weaken in the homogeneous model (maximum LOD score $\simeq 7$) while both the heterogeneity model and the two-point mixture enhance it further (maximum LOD score $\simeq 10$). Heterogeneity therefore contributes to the proof that a linkage effect is present. Similar outputs are displayed for chromosome 13 in Figure 1.5. In the 40-70cM region, heterogeneity of QTL effects is now clearly qualitative (both 'nlj' and 'swe' have similar QTL effects with corresponding suggestion for linkage) and the pooled homogeneous analysis is dominated by the large 'aus' study with QTL variance estimates close to 0 which entirely obliterates the individual linkage signals of 'nlj' and 'swe'. The two-point mixture works best here in pooling evidence from the two positive studies and enhancing the LOD score beyond 4 in a much narrower region (maximum LOD score $\simeq 3.5$ in individual studies).

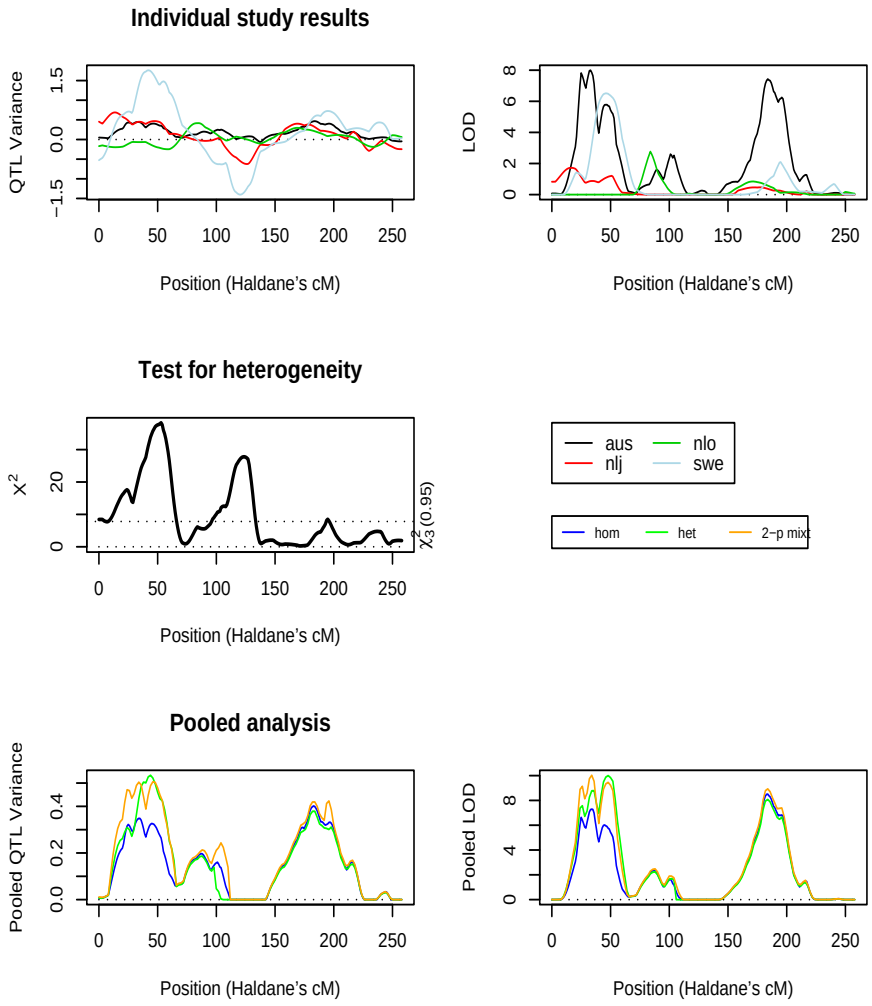

Figure 1.4 *Artificial data - Chromosome 2 - LDL cholesterol level*

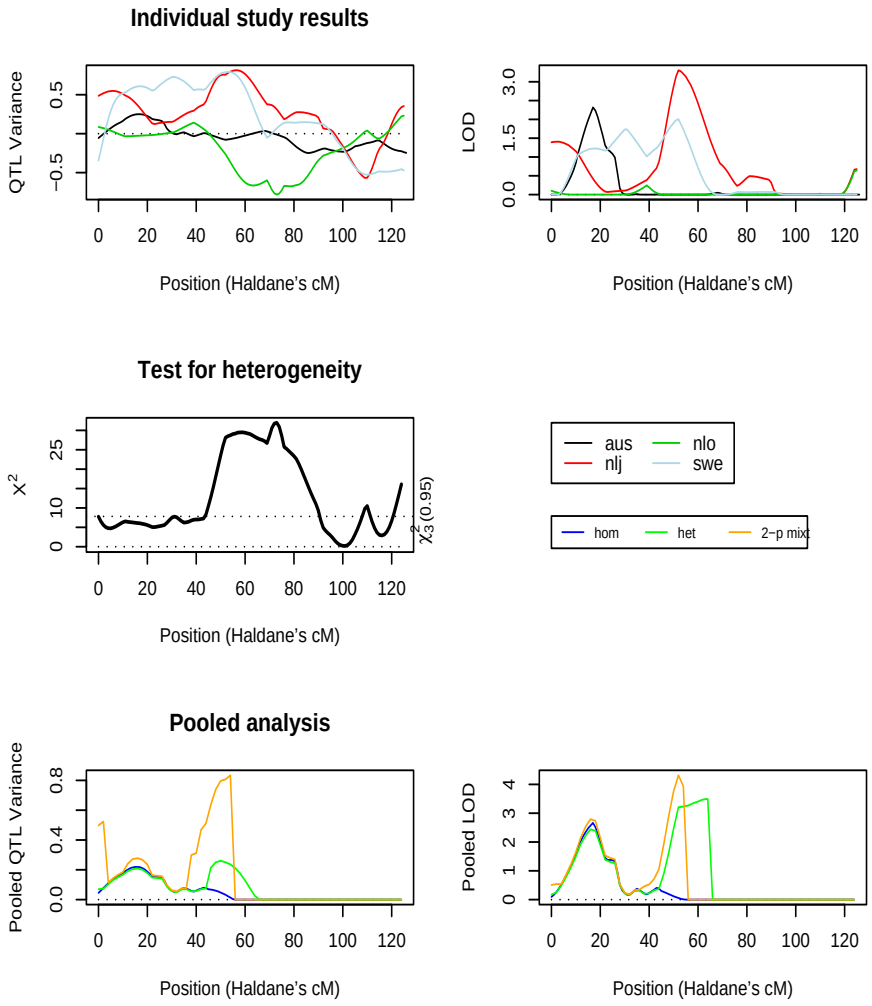

Figure 1.5 *Artificial data - Chromosome 13 - LDL cholesterol level*

1.5 Discussion

We have detailed how classical meta-analytic methods can be adapted to linkage provided consistent estimates of QTL effects along with standard errors are available for each study on a common grid of positions. The methods required to obtain such summary statistics are now well developed and their software implementation has been publicly available for a number of years. We realize, however, that most published

studies to date will not have sufficient information in order to carry out the method advocated here. Indeed, it is still common practice nowadays in the literature, even for QTL mapping where the effect to be estimated is fairly uncontroversial, to publish statistics conveying statistical significance only (i.e. LOD scores) without any idea of the actual effect estimate. This heavily hinders powerful pooling of the many small linkage studies available in the community. Gu *et al.*(1998) presented guidelines on how to report linkage studies that would enable future meta-analysis using IBD sharing as a common linkage parameter. Since the analysis tools are available (e.g. `MERLIN-regress`), it should be expected by journals that researchers publish QTL effects and associated standard errors (at least as add-on information) on a grid of locations.

Given the small individual study sizes one typically encounters, any test for heterogeneity of QTL effects across studies is bound to suffer from a lack of power. This was reflected in the test for heterogeneity of the real lipid levels data as well as in the estimate of the between study variance component σ^2 which very rarely differs from 0 (Heijmans *et al.*, 2005). Another way to test for heterogeneity in the random effect model setting is to test whether $\sigma^2 = 0$ and this is known to be asymptotically equivalent to the X^2 test that we have presented (Andersen *et al.*, 1999). Note that the classical random effects model is probably not the most appropriate in the case of linkage, indeed the fact that the QTL effect is a variance component precludes it from being negative (which is not impossible under the normal mixture model) and suggests that the random effects γ_i's could be more appropriately modelled as arising from a Γ distribution but estimation then becomes less straightforward.

The idea of applying the concept of finite mixture models to meta-analysis is also not new (Böhning *et al.*, 1998) although it is new for meta-analysis of linkage studies as far as we are aware. It is based on the simple idea that only studies with a positive effect should be pooled together to provide evidence for linkage. Instead of doing this by hand, we let the data decide which study exhibits positive linkage. Note that one can also formally test for locus heterogeneity by assessing whether α differs from 1. Ultimately, given a sufficiently large number of studies with decent precision, it would be possible to fit a model that adapts to both locus and size heterogeneity by combining the random effect and the two-point mixture models.

1.6 Acknowledgements

We wish to thank Bas Heijmans, Marian Beekman, Hein Putter, Nico Lakenberg, Henk Jan van der Wijk, John Whitfield, Danielle Posthuma, Nancy Pedersen, Nick Martin, Dorret Boomsma and Eline Slagboom for allowing us to use their lipid levels data as an example. Gathering these data was a tremendous effort of many individuals and in particular we would like to acknowledge the work of Kate van Duijn, Eka Suchiman and Dennis Kremer for expert technical assistance; Harry Beeby, Scott Gordon and Megan Campbell for assistance with cleaning Australian genotype data; Dr. Daniel O'Connor for his effort to complete the Australian genotype data set;

and the Mammalian Genotyping Service of the Center for Medical Genetics, Marshfield, WI, USA (director Dr. James Weber) for genotyping part of the Australian twin sample. This work was partially funded by the GENOMEUTWIN project which is supported by the European Union Contract No. QLG2-CT-2002-01254.

1.7 References

A. Dempfle and S. Loesguen, "Meta-analysis of linkage studies for complex diseases: an overview of methods and a simulation study", Ann. Hum. Genet., No. 68, pp 69–93, 2003.

D. C. Rao and M. A. Province, "Genetic dissection of complex traits", Academic Press, 2001.

L. H. Wise, J. S. Lanchbury and C. M. Lewis, "Meta-analyses of genome searches", Ann. Hum. Genet. , No. 63, pp. 263–272, 1999.

E. Zintzaras and J. P. A. Ioannidis, "Heterogeneity testing in meta-analyses of genome searches", Genet. Epidemiol. , No. 28, pp. 123–137, 2005.

C. Gu, M. Province, A. Todorov D. C. and Rao, "Meta-analysis methodology for combining non-parametric sibpair linkage results: genetic homogeneity and identical markers", Genet. Epi., No. 15, pp. 609–626, 1998.

Z. Li and D. C. Rao, "Random effects model for meta-analysis of multiple quantitative sib pair linkage studies", Genet. Epi., No. 13, pp. 377–383, 1996.

C. J. Etzel and R. Guerra, "Meta-analysis of genetic-linkage analysis of quantitiative-trait loci", Am. J. Hum. Genet., No. 71, pp. 56–65, 2002.

R. DerSimonian and N. Laird, "Meta-analysis in clinical trials", Controlled Clinical Trials, No. 7, pp. 177–188, 1986.

S. T. Normand, "Meta-analysis: formulating, evaluating, combining and reporting", Statistics in Medicine, No. 18, pp. 321–359, 1999.

J. C. van Houwelingen, L. R. Arends T. and Stijnen, "Advanced methods in meta-analysis: multivariate approach and meta-regression", Statistics in Medicine, No. 21, pp. 589–624, 2002.

J. Ott, "Analysis of human genetic linkage", 3^{rd} Ed., Baltimore: Johns Hopkins University Press.

E. Holliday, B. Mowry, D. Chant and D. Nyholt, "The importance of modelling heterogenity in complex disease: application to NIMH schizophrenia genetics initiative data", Human Heredity, To appear.

C. T. Ekstrom and P. Dalgaard, "Linkage analysis of quantitative trait loci in the presence of heterogeneity", Human Heredity, No. 55, pp. 16–26, 2003.

A. P. Dempster, N M. Laird D. B. and Rubin, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)", J. R. Stat. Soc. B, No. 39, pp. 1–38, 1977.

H. Chen, J. Chen and J. Kalbfleisch, "A modified likelihood ratio test for homogeneity in finite mixture models", J. R. Stat. Soc. B, No. 63, pp. 19–29, 2001.

L. Almasy and J. Blangero, "Multipoint quantitative-trait linkage analysis in general pedigrees", Am. J. Hum. Genet., No. 62, pp. 1198–1211, 1998.

C. I. Amos, "Robust variance-components approach for assessing genetic linkage in pedigrees", Am. J. Hum. Genet., No. 54, pp. 535–543, 1994.

K. Lange, R. Cantor, S. Horvath, M. Perola, C. Sabatti, J. Sinsheimer E. and Sobel, "Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets", Am. J. Hum Genetics, No. 69(supplement), pp. 504–504, 2001.

M. C. Neale, S. Boker, G. Xie and H. Maes, "Mx: Statistical Modeling", Department of Psychiatry, Box 126 MCV, Richmond, VA 23298, 1999.

J. Lebrec, H. Putter and J. C. van Houwelingen, "Score test for detecting linkage to complex traits in selected samples", Genet. Epidemiol., No. 27, pp. 97–108, 2004.

H. Putter, L. A. Sandkuijl and J. C. van Houwelingen, "Score test for detecting linkage to quantitative traits", Genet. Epidemiol., No. 22, pp. 345–355, 2002.

P. C. Sham and S. Purcell, "Equivalence between Haseman-Elston and Variance-Components linkage analyses for sib-pairs", Am. J. Hum. Genet., No. 68, pp. 1527–1532, 2001.

P. C. Sham, S. Purcell, S. S. Cherny and G. R. Abecassis, "Powerful regression-based quantitative-trait linkage analysis of general pedigrees", Am. J. Hum. Genet., No. 71, pp. 238–253, 2002.

H. Putter, J. Lebrec and J. C. van Houwelingen, "Selection strategies for linkage studies using twins", Twin Research, Vol. 6, No. 5, pp. 377–382, 2003.

G. R. Abecasis, S. S. Cherny, W. O. Cooksona and L. R. Cardon, "Merlin - rapid analysis of dense genetic maps using sparse gene flow trees", Nature Genet., Vol. 30, pp. 97–101, January 2002.

N. Risch, "Linkage strategies for genetically complex traits. ii. the power of affected relative pairs", Am. J. Hum. Genet., No. 46, pp. 229–241, 1990.

B. T. Heijmans, M. Beekman, H. Putter, N. Lakenberg, H. J. van der Wijk, J. B. Whitfield, D. Posthuma, N. L. Pedersen, N. G. Martin, D. I. Boomsma, P. E. Slagboom, "Meta-analysis of four new genome scans for lipid parameters and analysis of positional candidates in positive linkage regions", Eur. J. Hum. Genet., To appear.

P. K. Andersen, J. P. Klein and M. J. Zhang, "Testing for centre effects in multi-centre survival studies: A Monte Carlo comparison of fixed and random effects tests", Statistics in Medicine, Vol. 18, pp. 1489–1500, 1999.

D. Böhning, E. Dietz P. and P. Schlattmann, "Recent developments in computer-assisted analysis of mixtures", Biometrics, No.54, pp. 525–536, 1998.