Utilizing Information from Multiple Genome Scan Studies in an Empirical Bayes Framework for Strengthening Inferences and Narrowing QTL Location Estimates

---

Kui Zhang
Howard Wiener
T. Mark Beasley
Christopher I. Amos
&
David B. Allison

CHAPTER 1

Utilizing Information from Multiple Genome Scan Studies in an Empirical Bayes Framework for Strengthening Inferences and Narrowing QTL Location Estimates

Kui Zhang[1], Howard Wiener[2], T. Mark Beasley[1], Christopher I. Amos[3], David B. Allison[1,4]

[1]Department of Biostatistics, Section of Statistical Genetics, The University of Alabama at Birmingham, Birmingham, AL, 35294

[2]Department of Epidemiology, The University of Alabama at Birmingham, Birmingham, AL, 35294

[3]Department of Epidemiology, The U.T. MD Anderson Cancer Center, Houston, TX, 77030

[4]Department of Nutrition Sciences and Clinical Nutrition Research Center, The University of Alabama at Birmingham, Birmingham, AL, 35294

1.1 Abstract

Compared with an individual genome scan study for quantitative trait locus (QTL) mapping, meta-analysis, which can formally utilize the data from multiple genome scan studies, generally can provide higher statistical power, more precise estimates of QTL location and effect, and tighter confidence intervals for QTL location and effect. Therefore, meta-analysis is very useful for prioritizing regions for subsequent follow-up studies. In some situations, investigators who already have initialized a genome scan study would like to evaluate the ”believability” of apparent linkage signals by examining the results of other genome scan studies of the same trait and informally update their beliefs about which linkage signals in their own scan most merit follow-up via a subjective-intuitive integration approach. In this situation, meta-analysis methods may not be suitable because they treat all genome scan studies ”equally”, which is not subjective to the initial genome scan study conducted by the investigator. In the contrast, empirical Bayes (EB) based methods that can formally

borrow information from other genome scan studies to update the estimates and adjust the confidence in finding in an objective fashion could be useful. In empirical Bayes based methods, the linkage statistics from other genome scan studies are used as prior information to update the linkage statistics obtained from the genome scan study conducted by the investigator. The updated linkage statistics can then be used to estimate the QTL location and effect. In this chapter, we summarize the empirical Bayes based methods for multiple genome scan studies using sib pairs. We also evaluate their performance in terms of their power to and their accuracy to estimate the QTL location and effect, using extensive simulations based on actual marker spacing and allele frequencies from available data. Results indicate that the empirical Bayes based methods are insensitive to between-study heterogeneity. The empirical Bayes based methods can yield higher statistical power, generate more precise estimates for the QTL location and effect, and provide narrower confidence intervals than results from an individual study.

1.2 Introduction

Genome scan studies for linkage analysis have been widely used to search candidate regions containing quantitative trait loci (QTLs). However, most genome scan studies for QTL mapping are analyzed without formal consideration of information provided by other genome scan studies of the same trait. The resulting candidate regions often contain a large number of functional genes due to the lower power of individual genome scan studies. Consequently, the subsequent fine mapping and positional cloning for these candidate regions may be problematic. When multiple genome scan studies of the same trait are available, we may increase the power to detect the linkage between markers and QTLs by using information provided from all these studies. Methods that can formally integrate data from multiple genome scan studies have been emerging as useful and powerful tools in the field of linkage analysis for QTL mapping.

Marked heterogeneity can exist in multiple genome scan studies and pose daunting challenges in such analysis. Different genome scan studies can use different genetic marker loci and marker maps, different statistical methods to test for linkage, and different sampling schemes. Furthermore, the QTL effect can vary across studies because of disparate environmental effects and population substructures. The combination of raw data from all studies with a well-designed pre-analysis procedure would be a preferred approach to overcome such difficulties. However, in many situations this is not feasible because only some statistics, rather than the raw data, are available. For these reasons, two closely related but distinct groups of methods have been developed to test and map QTLs by integrating same type of statistics obtained from multiple genome scan studies: Meta-analysis and Empirical Bayes (EB).

The first group of methods is meta-analyses, which can be viewed as a set of statistical procedures designed to summarize statistics across independent studies that address similar scientific questions. Several meta-analysis methods have been developed to detect linkage between genetic markers and QTLs (Allison and Heo, 1998;

Etzel and Guerra, 2002; Gu et al., 1998; Guerra, 2002; Guerra et al., 1999; Hedges and Olkin, 1985; Li and Rao, 1996; Rice, 1997; Wise et al., 1999). For example, Allison and Heo (1998) used Fisher's method (Fisher, 1925) to show strong evidence of linkage in OB regions by combining p-values from five published linkage studies on these regions. They also illustrated this method's applicability in the presence of marked heterogeneity across studies. This technique has also been used by other researchers (e.g., Guerra, 2002; Wise et al., 1999). However, it is difficult to use this technique to estimate the parameters of interested in, such as the location and effect of a QTL, because of the method's nonparametric nature. At the same time, several meta-analysis methods that can estimate the parameters of interest across studies, such as the location and effect of a QTL, by combining estimates of Haseman-Elston regression slopes and associated variances at marker loci (Haseman and Elston, 1972) have been developed too (Etzel and Guerra, 2002; Gu et al., 1998; Li and Rao, 1996). The weighted least-square estimator (WLSE) developed by Etzel and Guerra (2002) does not require the same marker map or the same QTL effect across all studies. For a more detailed review of these meta-analysis methods, please refer to (Some References; BOOK CHAPTER in the Same Book).

The second group of methods is based on the EB framework (Beasley et al., 2005; Bonney et al. 1992; Zhang et al., 2005). In EB based methods, the linkage statistics (e.g., Haseman-Elston regression slopes and their associated variances at marker loci) are obtained from each individual genome scan study and then the linkage statistics from an individual study of interest are updated by incorporating the linkage statistics from other studies. The updated linkage statistics can be used for detecting the linkage between a marker and the QTL and mapping the location of the QTL. It is worth emphasizing the key difference between the EB based methods and the meta-analysis methods. In the empirical Bayes analysis, an individual genome scan study of interest is identified as the primary study and the rest of studies are considered as the background studies. Theoretically, each individual study can be claimed as the primary study. However, the study of primary interest to an investigator would be the study conducted by the investigator; presumably the investigator would be able to obtain further genotypes from the individuals in the primary study for fine mapping, while this type of information would not necessarily be available from the background studies. In the meta-analysis methods, each individual study is equivalent to the other studies used and investigators are interested in the overall results.

The rest of this paper is organized as follows: the EB based method for the genome scan studies using sib pairs and the simulation design are described in Methods section; the assessment of the performance of the empirical Bayes based methods for detecting linkage between markers and the QTL and mapping the location of the QTL are presented in Results section; the conclusions, the implications, and the possible extensions of the empirical Bayes based method are given in Discussion section.

1.3 Methods

1.3.1 Haseman-Elston Regression Analysis for a Single Genome Scan Study with m Markers using Sib Pairs

The Haseman-Elston Regression has been widely used to detect the linkage between genetic markers and QTLs using sib pairs. Suppose that the trait values, the squared trait difference, and the estimated proportion of alleles shared identical-by-descent (IBD) at a marker locus for the ith sib pair are denoted as $y_i = (y_{i1}, y_{i2})$, $Y_i^D = (y_{i1} - y_{i2})^2$, and π_i, respectively. Then the Haseman-Elston method can be represented by a simple regression of Y_i^D on π_i :

$$Y^D = \beta_0 + \beta\pi + \varepsilon$$

The regression slope β has the expectation $E(\beta) = -2(1-2\theta)^2\sigma_g^2$, where θ is the recombination fraction between the marker locus and the QTL, and σ_g^2 is the phenotypic variance explained by the additive effect of this QTL. Thus, the regression slope β is 0 under the null hypothesis of no linkage, and is negative under the alternative hypothesis. Specifically, if there are m markers and the estimates of the slope and its associated variance at each marker are denoted by $\hat{\beta}_j$ and $\hat{S}_j^2$, ($j = 1, \ldots, m$), then the t statistic $t_j = -\hat{\beta}_j/\sqrt{\hat{S}_j^2}$ asymptotically follows a standard normal distribution under the null hypothesis of no linkage. The null hypothesis is rejected with the 5% nominal level at the j th marker if t_j exceeds 1.645. It is also worth noting that the regression slope and its associated variance are only estimated at the marker loci with determined genotype in the original Haseman-Elston method. Due to the coarse marker map in linkage analysis, this method is more suitable for detecting linkage between markers and the QTLs rather than estimating the QTL location and effect.

The Haseman-Elston methods assumes the normality of trait values and is robust even this assumption is violated for a reasonably large sample size ($n > 100$ sib pairs) (Allison et al., 2000). However, the original Haseman-Elston regression tends to have lower power than the variance component method. Other modified Haseman-Elston regression methods were subsequently developed (Amos 1994; Drigalenko 1998; Elston et al., 2000; Feingold, 2002; Sham et al., 2001; Xu et al., 2000). For example, additional power can be acquired by regressing the mean corrected squared sums of trait values $Y_i^S = (y_{i1} - \bar{y} + y_{i2} - \bar{y})$ (Drigalenko 1998), the mean corrected cross product of trait values $Y_i^P = (y_{i1} - \bar{y})(y_{i2} - \bar{y}) = (Y_i^S - Y_i^D)^2$ (Drigalenko 1998; Elston et al., 2000), or a weighted combination of Y_i^D and Y_i^S (Xu et al., 2001) on π_i, where $\bar{y}$ is the mean trait value over all sib pairs.

1.3.2 The Interval Mapping (IM) Method to Detect Linkage between Markers and QTLs and Estimate the QTL Location and Effect Based on m Markers from a Single Genome Scan Study Using Sib Pair

Fulker at al. (1995) developed an interval mapping (IM) method to detect linkage between markers and QTLs and to estimate the QTL location and effect. They first used the estimated proportion of alleles shared IBD at all marker loci on a single chromosome and the genetic distance between these markers to estimate the proportion of IBD sharing at virtually any location on the chromosome and then perform the Haseman-Elston regression at this location (Fulker et al., 1995). Suppose that the estimates of regression slope and its associated variance at each analysis point q along the chromosome are denoted by $\hat{\beta}_q$ and $\hat{S}_q^2$, respectively. At any analysis point q, the null hypothesis of no linkage is rejected at the nominal 5% level if the value of test statistic $\hat{t}_q = -\hat{\beta}_q/\sqrt{\hat{S}_q^2}$ is greater than 1.645. The analysis point, $\hat{q}$, that gives the maximum value of the test statistic $\hat{t}_q = -\hat{\beta}_q/\sqrt{\hat{S}_q^2}$, is taken as the estimate for the QTL location. The point estimate of QTL effect, σ_g^2, is given by $\hat{\sigma}_g^2 = -\hat{\beta}_{\hat{q}}/2$.

1.3.3 Empirical Bayes Model (Bayesian Hierarchical Normal Model)

In Bayesian analysis, the choice of reasonable prior distribution for parameters is sometimes not obvious. However, if data from several independent studies are available, the prior information can be extracted from the data. Such approaches are called empirical Bayes methods (Carlin and Louis, 2000a). These methods can be viewed as approximations to a complete hierarchical Bayesian analysis; hybrid approaches between classical frequentist methods and fully Bayesian methods. Both parametric and non-parametric approaches exist (Carlin and Louis 2000b), but even the parametric varieties do not depend on strong distributional assumptions (Efron and Morris, 1973).

The empirical Bayes approach as proposed by Efron and Morris (1973; 1975) can be described by a widely used two-level hierarchical normal model. Suppose β is the parameter of interest and there are k populations available to estimate β_i in each population, where β_i can be different among k populations. At the first level, the maximum likelihood estimators $\hat{\beta}_i (i = 1, \ldots, k)$ for β_i can be obtained and we assume that $\hat{\beta}_i|\beta_i$ asymptotically follows a normal distribution, $N(\beta_i, S_i^2)$. At the second level, β_i is specified by a normal model with an r -dimensional predictor x_i, a common regression coefficient μ , and an unknown variance $A \geq 0$; i.e., $\beta_i|\mu \sim N(x_i'\mu, A)$. Using the Bayesian rule, it is easy to compute the marginal distribution of $\hat{\beta}_i$ (given μ and A) and conditional distribution of β_i (given $\hat{\beta}_i$, μ, and A):

$$\hat{\beta}_i|\mu, A \sim N(x_i'\mu, S_i^2 + A), i = 1, \ldots, k \tag{1.1}$$

and

$$\beta_i|\hat{\beta}_i, \mu, A \sim N((1-B_i)\hat{\beta}_i + B_i x_i'\mu, S_i^2(1-B_i)), i = 1, \ldots, k \qquad (1.2)$$

where $B_i = S_i^2/(S_i^2 + A)$ is an unknown shrinkage factor. Generally, S_i^2 is unknown and is replaced by $\hat{S}_i^2$, the estimates of associated variance of β_i. A and μ can be estimated by the maximum likelihood methods or by more advanced techniques developed by Tang and Morris (2003). Then we can use $\tilde{\beta}_i = (1-B_i)\hat{\beta}_i + B_i x_i'\mu$ and $\tilde{S}_i^2(1-B_i)$ as the final estimator for β_i and its associated variance, respectively.

1.3.4 Application of the Empirical Bayes Method to Each Marker Based on k Studies with m markers and Identical Marker Map

Empirical Bayes methods have been used in many contexts, including genetic research (Bonney et al., 1992; Li and Rao, 1996; Lockwood et al., 2001; Witte, 1997). We tailored the general empirical Bayes procedure for linkage analysis. Assume that data for the detection of linkage to the same QTL are available from k genome scans using sib pairs. Within each of the k studies, a set of m markers with the identical map are used. For each marker locus from each study the regression coefficient, β_{ij}, is the parameter of interest and describes the effect of the putative QTL on the phenotype. The expectation of β_{ij} equals to $-2(1-2\theta_{ij})^2\sigma_{gi}^2$ at marker locus $j(j = 1, \ldots, m)$ in study $i(i = 1, \ldots, k)$, where θ_{ij} is the recombination fraction between the QTL and the marker j in study i and σ_{gi}^2 is the total genetic variance of the QTL in study i. From the Haseman-Elston regression analysis, we can obtain the estimator $\hat{\beta}_{ij}$ for β_{ij} and its estimated sampling variance$\hat{S}_{ij}^2(i = 1, \ldots, k; j = 1, \ldots, m)$. For k available studies, all k studies are first used to estimate parameters μ_j and A_j $(j = 1, \ldots, m)$, then the empirical Bayes estimators $\tilde{\beta}_{ij}$ and $\tilde{S}_{ij}^2$ for β_{ij} and associated variance can be easily obtained using formulas (1.1) and (1.2) for each of k studies.

1.3.5 The IM-EB Method to Detect Linkage between Markers and QTLs and Estimate the QTL Location and Effect Based on m Markers and k Genome Scan Studies Using Sib Pairs

In this section, we give the detailed desription for the IM-EB method to detect linkage between markers and QTLs and estimate the QTL location and effect from multiple genome scan studies using sib pairs. We assume that data of genome scans using sib pairs with the same trait are available and consider the first study as the primary study. Within each of the studies, a set of markers are used within the same chromosomal region and are denoted as $M_{ij}(i = 1, \ldots, k; j = 1, \ldots, m)$.

For the IM-EB method, the estimates of the regression slope and its associated variance, $\hat{\beta}_{iq}$ and $\hat{S}_{iq}^2$ $(i = 1, \ldots, k$) at each analysis point q on the chromosome, are obtained using the IM method (Fulker et al., 1995). Then, the empirical Bayes estimates, $\tilde{\beta}_{iq}$ and $\tilde{S}_{iq}^2$ $(i = 1, \ldots, k$), are obtained from each of the k studies by using GRIMM (Tang and Morris, 2003). GRIMM is independently applied to each analysis

point along the chromosome. The test statistic for the primary study is then calculated on the basis of $\hat{t}_{1q} = -\hat{\beta}_{1q}/\sqrt{\hat{S}^2_{1q}}$and $\tilde{t}_{1q} = -\tilde{\beta}_{1q}/\sqrt{\tilde{S}^2_{1q}}$ at the analysis point q. The analysis point $\hat{q}$ having a maximum value $\hat{t}_{1\hat{q}}$ over the entire chromosome is considered as the IM estimate of QTL location and consequently, the IM estimate of σ^2_{1g} is given by $\hat{\sigma}^2_{1g} = -\hat{\beta}_{i\hat{q}}/2$. The same procedure can be applied to $\tilde{t}_{1q}$ to obtain $\tilde{q}$ and $\tilde{\sigma}^2_{1g}$, the IM-EB estimates of QTL location and effect, respectively. At each analysis point q , the null hypothesis of no linkage is rejected with the 5% nominal level by the IM estimator if the value of test statistic $\hat{t}_{1q} = -\hat{\beta}_{1q}/\sqrt{\hat{S}^2_{1q}}$ is greater than 1.645. Similarly, the null hypothesis is rejected at the 5% nominal level if the value of test statistic $\tilde{t}_{1q} = -\tilde{\beta}_{1q}/\sqrt{\tilde{S}^2_{1q}}$ is greater than 1.645.

1.3.6 Simulation Designs

To investigate the performance of the empirical Bayes method to incorporate data from multiple genome scan studies using sib pairs, we conducted the following simulations. We assumed that there is only one QTL with no background polygenic variation and no shared sib environment effect, or equivalently that such effects are subsumed into the residual variance. There were two alleles at the QTL with the high-risk allele having a frequency of 0.05. We chose 15 microsatellite markers on chromosome 11 that were used for a recent genome scan of Alzheimer's disease (Blacker et al., 2003) because it provides known parameters, including the location and the allele frequencies at each marker locus, for simulations. The trait value of each individual was generated according to the genetic model, $y = \mu + g + \varepsilon$, where μ is the overall trait mean across the population, g is additive effect of the high-risk allele, and ε is the normally distributed random error. We set $E(\varepsilon) = 0$, $cov(g, \varepsilon) = 0$, and $\mu = 70$ and set the total variance of g and ε, $\sigma^2_g + \sigma^2_\varepsilon$ as 1 for all studies.

For each simulation, 5, 10, or 15 studies were generated corresponding to a single study of interest with 4, 9, or 14 background studies, respectively. We generated 500 unrelated sib pairs in each study. We used the same marker map for all studies. For the primary study, the QTL was positioned 65cM from the p-terminus of the chromosome. The heritability of QTL was set either to 0 (without QTL effect) or 15% (with non-zero QTL effect). For background studies, the location and the heritability of QTL could be same as, or different from that of the primary study. The location of QTL in background studies was set either at 35cM or at 65 cM from the p-terminus of the chromosome. The marker locations along with the QTL location are shown in Figure 1.1. The heritability of QTL in each background study varied between 0 and 25% in increments of 5%, which represented variation from the primary study. The number of background studies with non-zero QTL effect varied but all background studies having a non-zero QTL effect were given the same value of heritability. This simulation strategy can accommodate different degrees of heterogeneity among the primary study and the background studies. It can also include a variety of combinations of weak to strong linkage signals among the primary study and background

studies. For example, we can set the QTL heritability in the primary study to 15%, the heritability of half of the background studies to 0, and the heritability of the other half of background studies to 25% to represent the situation that the primary study has the moderate linkage signal while some background studies show small to no QTL effect and some of background studies have stronger QTL effect. Other situations can be easily accommodated by varying the number of background studies with non-zero QTL effect and their heritability.

Once the genotypic and phenotypic data were generated, the estimates of the Haseman-Elston regression slopes and their associated variances at each marker or analysis point in each study were determined by regressing the weighted combination of Y^D and Y^S on π (Xu et al., 2001).

1.4 Results

To assess the performance of the empirical Bayes based method, IM-EB, in terms of its power to detect linkage between markers and QTLs and its accuracy to estimate the QTL location and effect, we adapted different simulation strategies and recorded and used different summary statistics.

1.4.1 The Type I Error Rate and Power of the IM-EB method to Detect Linkage between Markers and QTLs

We first investigate the type I error rate of the IM-EB method. It is important to understand that a null model in this context refers only to the study of interest, whether or not the background studies contain a linked QTL. We generated 1,000 data sets with 5, 10, and 15 studies. In all studies, the QTL was positioned 65cM from the p-terminus of the chromosome. In the primary study, the heritability of QTL was set to 0. In the background studies, the heritability was set either to 0 or some value between 5% and 25% in increments of 5%. The number of background studies with non-zero QTL effect varied. Under any particular condition, all background studies with non-zero QTL effect had the same heritability. In the primary study, the null hypothesis was rejected when the IM-EB statistics at 65cM from the p-terminus of the chromosome exceeded 1.645. Figure 1.2 shows the type I error rate of the IM-EB method, which is the proportion of simulations in which the null hypothesis was rejected.

For 5 studies, the number of background studies with non-zero QTL effect was set to 1, 2, 3, or 4. It can be seen when three or fewer background studies have non-zero QTL effect, the type I error rate stays below to the nominal 5% error rate. When all 4 background studies have a heritability of 10%, the type I error rate can be greater than the nominal 5% error rate. The highest type I error rate is 8.5% for all 4 background studies having a heritability of 25%. For 10 studies, the number of background studies with non-zero QTL effect was set to 2, 4, 6, 8, or 9. When there

are fewer than 4 background studies having non-zero heritability as high as 20%, the type I error rate is below to the nominal 5% error rate. The type I error rate is in¤ated when 8 or more background studies have a heritability greater than 10%. The highest type I error is 16%. For 15 studies, the number of background studies with non-zero QTL effect was set to 3, 6, 9, 12, or 14. When fewer than 6 background studies have a heritability as high as 25%, the type I error rates do not t exceed the nominal 5% rate. Again, the type I error rate is in¤ated when 9 or more background studies have a heritability greater than 10%. The type I error rate is 20% when all 14 background studies have a heritability of 25%. In summary, the type I error rate of the IM-EB stays below to the nominal 5% rate when most of the background studies have a heritability less than 10%. At the same time, we did find some in¤ated type I error rates of the IM-EB method when most of background studies have a higher heritability. This is expected because the empirical Bayes based method borrows the information from the other studies. If there are a large number of studies with the large QTL effect, the empirical Bayes based method will detect a QTL even the results from the primary study shows small to no effect. However, from an EB perspective, it is debatable whether this situation is truly a ”null’ situation.

We then investigate the power of the IM-EB method. We generated 1,000 data sets with 5, 10, and 15 studies. In the primary study, the heritability of QTL was set to 15%. In the background studies, the heritability was set either 0 or some value between 5% and 25% in increments of 5%. The number of background studies with non-zero QTL effect varied and all background studies with non-zero QTL effect had the same heritability.

Figure 1.3 shows the power of the IM-EB method, which is the proportion of simulations in which the null hypothesis was rejected. In these simulations, the QTL was positioned 65cM from the p-terminus of the chromosome. In the previous subsection, we used 1.645 as the 95% cutoff value to reject the null hypothesis of no linkage between the marker and the QTL. This value is only valid for one single study. When the empirical Bayes based method was used, this cutoff value tends to be conservative. We followed the method proposed by Beasley et al. (2005) to determine the cutoff value. We simulated 1,000 data sets with 5, 10, and 15 studies. All studies had no QTL effect. For the IM-EB method, the 95% cutoff values were 1.464, 1.406, and 1.224 for 5, 10, and 15 studies. These simulated cutoff values were used as critical values to reject the null hypothesis at the nominal 5% level.

It can be seen from Figure 1.3, the power of the IM-EB estimator can be substantially increased when a majority of background studies have the same or higher QTL effect. When all 4, 9, and 14 background studies have a heritability of 15%, the power of the IM-EB estimator increases from 0.191 (the power of the IM estimator for an individual study) to 0.266, 0.343, and 0.466, respectively. When all 4, 9, and 14 background studies have a heritability of 25%, the power of the IM-EB estimator increases to 0.322, 0.525, and 0.688, respectively. The power of the IM-EB estimator also increases even when some of the background studies disagree with the primary study. For example, when about half of the background studies have no QTL effect and half of the background studies have the same heritability of 15%, the power of

the IM-EB estimator is 0.224, 0.221, and 0.345 for 4, 9 and 14 background studies, respectively. As would be expected, the increase in power is slightly less than the situation when all of the studies agreed.

To see how the existence of other QTLs along the same chromosome affects the power of the IM-EB method, we simulated data sets in which all background studies had a heritability of 15% but half of them had the QTL positioned 35cM from the p-terminus of the chromosome. The power of the IM-EB estimator at each marker locus is shown in Figure 1.4. We find that the IM-EB estimator increases the power to detect linkage near the QTL of interest at a very small cost of in¤ated type I error rates at 35cM.

1.4.2 The Accuracy of the IM-EB estimates for QTL Location and QTL Effect

To investigate the accuracy of the IM-EB estimates for QTL location, we recorded their mean value (MEAN), their standard error (STD), and the square root of the mean squared difference between the estimates and the true value (MSE). The simulations adapted here were same with those described in the previous subsection. Specifically, we generated 1,000 data sets with 5, 10, and 15 studies. In the primary study, the QTL was positioned 65cM from the p-terminus of the chromosome and the heritability of the QTL was set to 15%. In the background studies, the heritability was set either to 0 or some value between 5% and 25% in increments of 5%. The number of background studies with non-zero QTL effect varied. Under any particular condition, all background studies with non-zero QTL effect had the same heritability.

The mean and MSE of the IM and IM-EB point estimates for the QTL location and effect under several different simulation strategies are presented in Table 1.1 and 1.2. Several general conclusions emerged from these two tables. First, as expected, the empirical Bayes based method (the IM-EB method) using multiple studies estimate the QTL location and effect more precisely and supply a smaller MSE than does the IM method using an individual study in most situations we simulated. This improvement becomes more notable with more independent studies having larger QTL heritability included in the analysis. For 5 studies, the MSE of the estimates for the QTL location is reduced 5% (from 41.4 to 39.0) when all 4 background studies have a heritability of 25%. For 10 studies, the MSE of the estimates for the QTL location is reduced 18% (from 42.4 to 34.9) when all 9 background studies have a heritability of 25%. Second, the heterogeneity among background studies and the disagreement between the primary study and background studies only slightly affect the accuracy of the IM-EB estimates. In addition, we did not observe a large bias for the estimates of either the QTL location or effect in the presence of other QTLs and different QTL effects in the background studies, as observed in Zhang et al. (2005).

1.5 Discussion

With availability of multiple genome scan studies detecting the linkage between the same QTL and the marker, there is a need to develop novel methods that can borrow or combine information from all available studies. Historically, there are two kinds of methods: the meta-analysis methods and the empirical Bayes based methods. In this paper, we summarized the empirical Bayes based methods (Beasley at el., 2005; Zhang et al., 2005) and assessed their performance using extensive simulations. We found that the empirical Bayes based methods have more power to detect the QTL and provide more precise estimates of QTL location and effect than do methods using an individual study.

To assess the effect of the heterogeneity among studies, we assumed the background studies could have no QTL effect, have a non-zero QTL effect different from that of the primary study, or have the QTLs different from that of the primary study. Although the in¤uence of these factors varies, the empirical Bayes methods were generally robust under all simulated situations. That is, they had more power to detect the QTL and yielded more precise estimates for QTL location and effect, with type I error increased only under extreme situations. In simulations, we assumed that all studies had identical marker maps. This is not required by the empirical Bayes based methods. In addition, varied marker maps across studies had the slight impact on the empirical Bayes based methods and could be helpful in a few situations (Zhang et al., 2005).

We did not compare the empirical Bayes based methods with meta-analysis methods in this paper. Zhang et al. (2005) compared several empirical Bayes based methods with a weighted least-square methods developed by Etzel and Guerra (2002). Their results showed that no method was superior to any other under all simulation situations. Although it is great of interest to conduct such comparison, it is important to point out that the empirical Bayes based methods introduced here are not meta-analysis methods. In the meta-analysis methods, results from several studies of the same relationship are combined to obtain an overall inference or estimate of that relationship. In such an analysis, the results of the studies are combined with equal regard weighted by their relative precisions. In the empirical Bayes based methods, there is one study of primary interest, whereas the rest of studies are regarded as background studies. The results obtained from background studies are incorporated as prior information to improve the inference or estimate for the primary study.

In summary, we conclude that the empirical Bayes based methods can account for between-study heterogeneity. They can have more power to detect linkage between markers and QTL and provide more precise estimates for the QTL location and effect.

1.6 Acknowledgement

This paper is supported by grant R01ES09912 from National Institutes of Health.

1.7 References

D. B. Allison, M. Heo, "Meta Analysis of linkage data under worst-case conditions: a demonstration using the human OB region", Genetics, Vol 148, pp. 859–865, 1998.

D. B. Allison, J. R. Fernandez, M. Heo, T. M. Beasley, "Testing the robustness of the new Haseman-Elston quantitative-trait loci-mapping procedure", Am. J. Hum. Genet., Vol 61, pp. 249–252, 2000.

C. I. Amos, "Robust variance-components approach for assessing genetic linkage in pedigrees", Am. J. Hum. Genet., Vol 54, pp. 535–543, 1994.

T. M. Beasley, H. Wiener, K. Zhang, A. A. Bartolucci, C. I. Amos, D. B. Allison, "Empirical Bayes Method for Incorporating Data from Multiple Genome Scans", Hum. Hered., Vol 60, pp. 36–42, 2005.

D. Blacker, L. Bertram, A. Saunders, T. Moscarillo, M. Albert, et al., "Results of a high-resolution genome screen of 437 Alzheimer's disease families", Hum. Mol. Genet., Vol 12, pp. 23–32, 2003.

G. E. Bonney, K. K. Amfoth, S. L. Sherman, B. J. B. Keats, "An application of empirical Bayes methods to updating linkage information on chromosome 21", Cytogenet. Cell. Genet., Vol 59, pp. 112–113, 1992.

B. P. Carlin, T. A. Louis, "Empirical Bayes: Past, present and future", J. Am. Stat. Assoc., Vol 95, pp. 1286–1289, 2000a.

B. P. Carlin, T. A. Louis: Bayes and Empirical Bayes Methods for Data Analysis, 2nd Edition. Chapman and Hall, Boca Raton, 2000b.

E. Drigalenko, "How sib pairs reveal linkage", Am. J. Hum. Genet. Vol 63, pp. 1242–1245, 1998.

B. Efron, C. Morris, "Stein's estimator and its competitors - An empirical Bayes approach", J. Am. Stat. Assoc., Vol 68, pp. 117–130, 1973.

B. Efron, C. Morris, "Data Analysis using Stein's estimator and its generalizations", J. Am. Stat. Assoc., Vol 70, pp. 379–421, 1975.

R. C. Elston, S. Buxbaum, K. B. Jacobs, J. M. Olson, "Haseman and Elston revisited", Genet. Epidemiol., Vol 19, pp. 1–17, 2000.

C. J. Etzel, R. Guerra, "Meta-analysis of genetic-linkage analysis of quantitative-trait-loci", Am. J. Hum. Genet. Vol 71, pp. 56–65, 2002.

A. A. Fisher: Statistical methods for research workers, Oliver and Boyd, New Landon, CT, 1925.

D. W. Fulker, S. S. Cherney, C. R. London, "Multipoint interval mapping of quantitative trait loci using sib pairs", Am. J. Hum. Genet., Vol 56, pp. 1224–1333, 1995.

C. Gu, M. Province, A. Todorov, D. C. Rao, "Meta-analysis methodology for combining non-parametric sibpair linkage results: genetic homogeneity and identical markers", Genet. Epidemiol., Vol 15, pp. 609–626, 1998.

R. Guerra: Meta-analysis in human genetic studies, In: Elston RC, Olson J, Palmer L (eds) Biostatistical genetics and genetic epidemiology. John Wiley & Sons, New York, 2002.

R. Guerra, C. J. Etzel, D. R. Goldstein, S. R. Sain, "Meta-analysis by combining p-values: simulated linkage studies", Genet. Epidemiol. Vol 17, Suppl 1, pp. S605–S609, 1999.

J. K. Haseman, R. C. Elston RC, "The investigation of linkage between a quantitative trait and marker locus", Behav. Genet., Vol 2, pp. 3–19, 1972.

L. V. Hedges, I. Olkin: Statistical methods for meta-analysis, Academic Press, New York, 1985.

L. Kruglyak, E. S. Lander, "Complete multipoint sib-pair analysis for quantitative trait and quantitative traits", Am. J. Hum. Genet., Vol 57, pp. 439–454, 1995.

Z. Li, D. C. Rao, ”Random effects model for meta-analysis of multiple quantitative sib pair linkage studies”, Genet. Epidemiol., Vol 12, pp. 377–383,1996.

J. R. Lockwood, K. Roeder, B. Devlin, ”A Bayesian hierarchical model for allele frequencies”, Genet. Epidemiol., Vol 20, pp. 17–33, 2001.

C. N. Morris, ”Parametric empirical Bayes inference: theory and applications”, J. Am. Stat. Assoc., Vol 78, pp. 47-55, 1983.

J. P. Rice, ”The role of meta-analysis in linkage studies of complex traits”, Am. J. Med. Genet., Vol 74, pp. 112–114, 1997.

P. C. Sham, S. Prucell, S. S. Cherny, G. R. Abecasis, ”Powerful regression based quantitative linkage analysis of general pedigrees”, Am. J. Hum. Genet., Vol 71, pp. 238–253, 2002.

R. Tang, C. N. Morris: Gaussian Regression Interactive Multilevel Model, 2003.

L. H. Wise, J. S. Lanchbury, C. M. Lewis, ”Meta-analysis of genome scans”, Ann. Hum. Genet., Vol 63, pp. 263–272, 1999.

J. S. Witte, ”Genetic analysis with hierarchical models”, Genet. Epidemiol., Vol 14, pp. 1137–1142, 1997.

X. Xu, S. Weiss, X. Xu, J. J. Wei, ”A unified Haseman-Elston method for testing linkage with quantitative traits”, Am. J. Hum. Genet., Vol 67, pp. 1025–1028, 2000.

K. Zhang, H. Wiener, T. M. Bealey, V. George, C. I. Amos, D. B. Allison, ”An empirical bayes method for updating inferences in analysis of quantitative trait loci using information from related genome scans”, Submitted for publication, 2006.

1.8 Figures and Tables

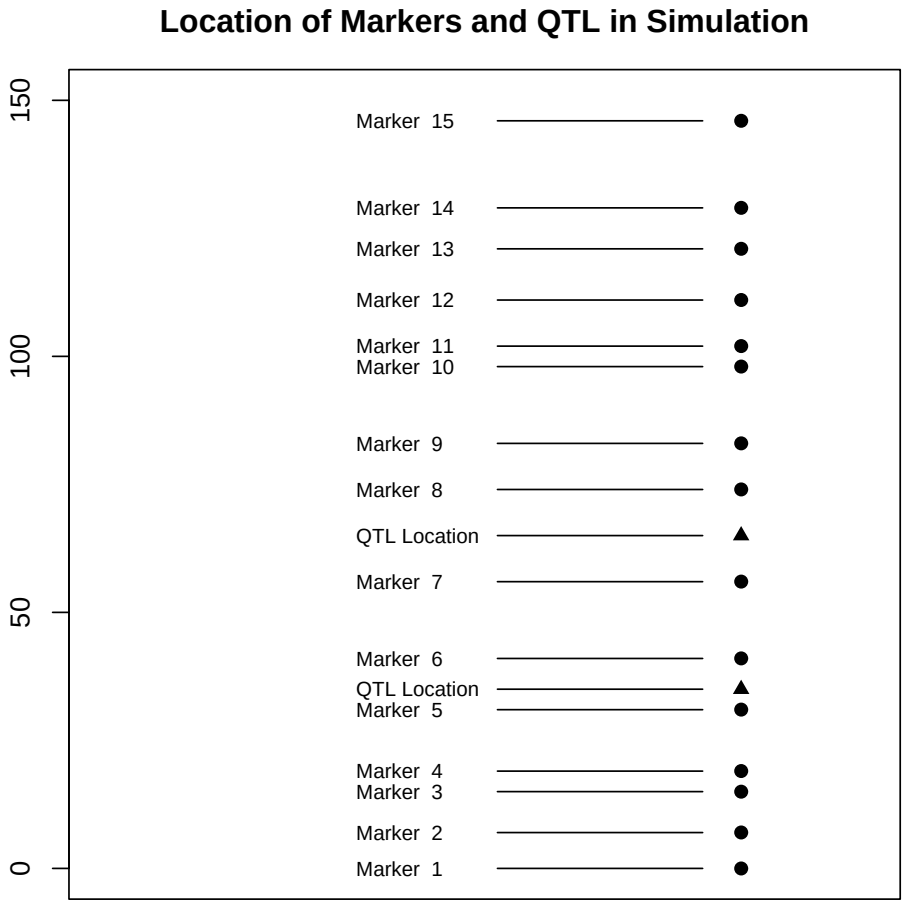

Figure 1.1 *The actual map for 15 micro-satellite markers from the National Institute of Mental Health Alzheimer's Diseases Genetics Initiative and the locations of two hypothetical QTLs used in simulations. The minimum distances between the marker and two QTLs, 65cM and 35cM from the p-terminus of the chromosome, are 9cM and 4cM, respectively.*

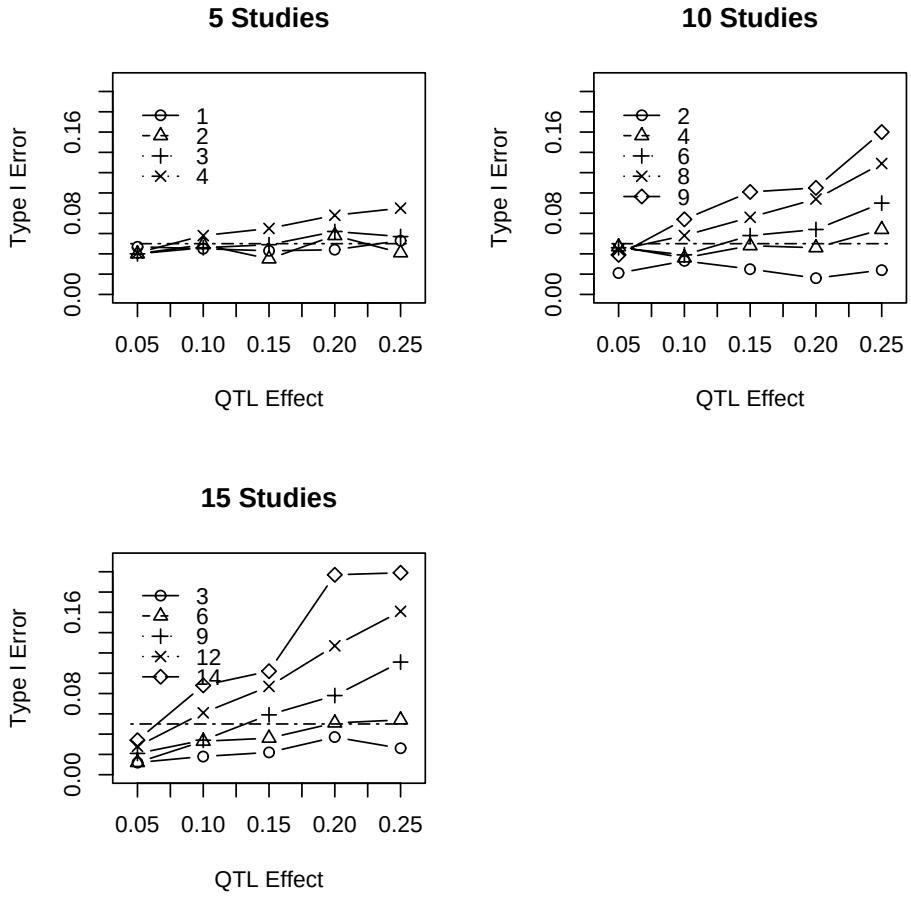

Figure 1.2 *The type I error rates of the IM-EB estimator at the 65cM from the p-terminus of the chromosome with 5, 10, and 15 studies. The QTL in all studies were positioned 65cM from the p-terminus of the chromosome. In the primary study, the heritability of QTL was set to 15% and the number of background study having non-zero QTL effect varied.*

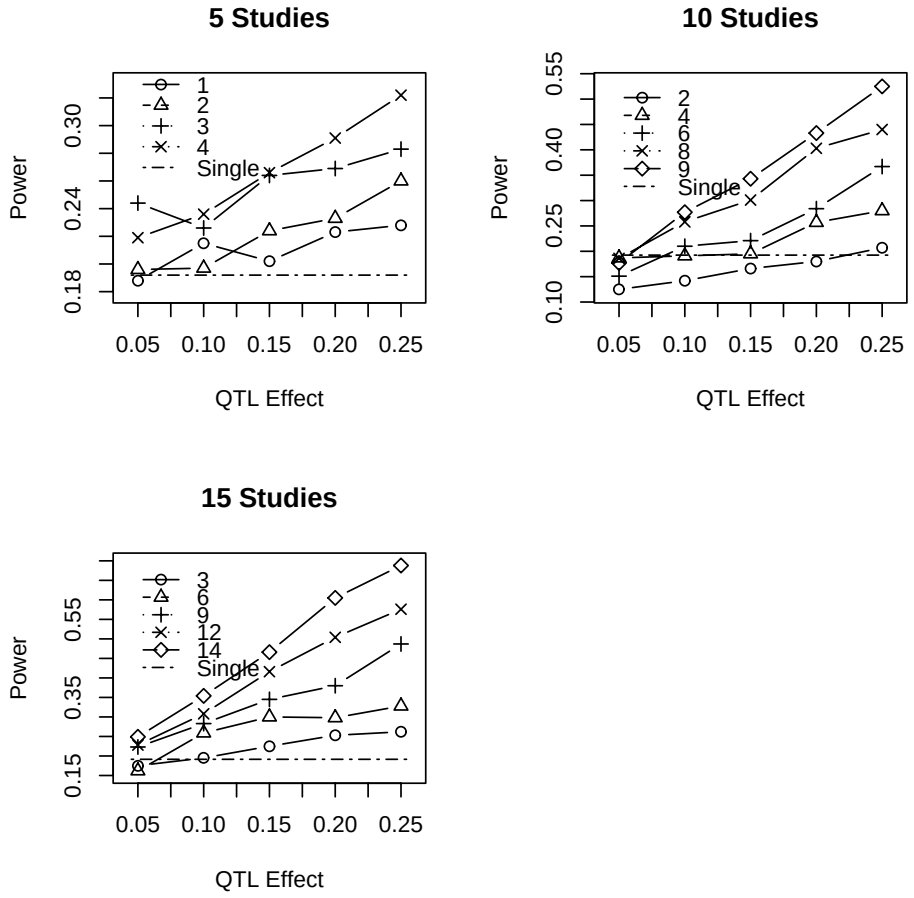

Figure 1.3 *The power of the IM estimator and IM-EB estimator with 5, 10, and 15 studies at 65cM from the p-terminus of the chromosome. The QTL in all studies was positioned 65cM from the p-terminus of the chromosome. In the primary study, the heritability of QTL was set to 15% and the number of background study having non-zero QTL effect varied.*

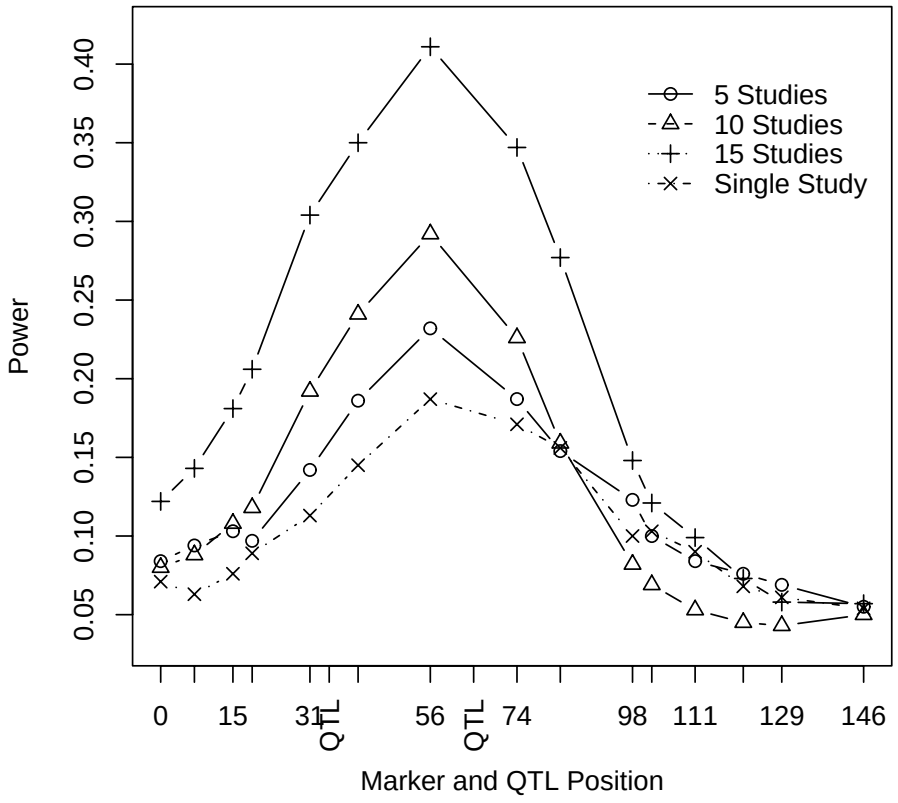

Figure 1.4 *The power of the IM estimator and IM-EB estimator with 5, 10, and 15 studies at the marker loci. In all studies, including the primary study and background studies, the heritability was set to 15%. In the primary study and half of the background studies, the QTL was positioned 65cM from the p-terminus of the chromosome. In another half of the background studies, the QTL was positioned 35cM from the p-terminus of the chromosome.*

Table 1.1 *The mean and MSE (in parentheses) for the point estimates of QTL location.*

Number of QTLs in Background Studies	Number of Studies	Number of Background Studies with non-zero QTL effect	Method	The Heritability in Background Studies		
				5%	15%	25%
One QTL at 65cM	5	2	IM	71.8(43.4)	69.2(43.6)	69.6(43.9)
	5	2	IM-EB	72.2(44.2)	70.6(42.1)	68.7(42.5)
	5	4	IM	65.4(41.7)	66.8(41.9)	69.5(41.4)
	5	4	IM-EB	64.9(40.6)	67.8(40.7)	69.3(39.0)
	10	4	IM	71.1(41.8)	68.2(42.2)	69.6(42.4)
	10	4	IM-EB	72.0(41.4)	68.1(40.2)	70.0(40.1)
	10	9	IM	70.8(43.5)	71.0(41.8)	69.9(42.2)
	10	9	IM-EB	68.9(41.9)	69.5(37.5)	67.2(34.9)
	15	6	IM	72.4(42.6)	70.4(42.6)	70.8(41.9)
	15	6	IM-EB	71.9(43.3)	68.4(40.4)	70.8(38.3)
	15	14	IM	69.8(41.2)	68.3(42.7)	71.0(43.2)
	15	14	IM-EB	69.9(39.3)	69.4(38.5)	68.7(33.2)
One QTL at 65cM One QTL 65cM at 35cM	5	2	IM	70.8(42.7)	70.8(42.6)	66.0(42.2)
	5	2	IM-EB	69.4(42.3)	70.2(42.4)	64.8(41.9)
	5	4	IM	67.8(41.5)	70.0(42.9)	70.641.9)
	5	4	IM-EB	66.3(41.1)	68.6(41.2)	65.8(40.2)
	10	4	IM	72.0(42.2)	69.1(42.5)	72.0(43.0)
	10	4	IM-EB	71.6(41.7)	68.4(42.9)	67.7(41.2)
	10	9	IM	70.2(42.6)	68.7(42.5)	68.2(42.4)
	10	9	IM-EB	65.8(42.0)	65.2(38.3)	62.7(36.8)
	15	6	IM	69.9(41.3)	69.7(41.9)	69.8(42.5)
	15	6	IM-EB	70.6(41.8)	66.0(40.7)	64.8(39.5)
	15	14	IM	69.8(43.1)	70.7(42.7)	69.7(41.4)
	15	14	IM-EB	67.4(41.5)	63.6(37.9)	60.3(34.2)

Table 1.2 *The mean and MSE (in parentheses) for the point estimates of QTL effect.*

Number of QTLs in Background Studies	Number of Studies	Number of Background Studies with non-zero QTL effect	Method	The Heritability in Background Studies		
				5%	15%	25%
One QTL at 65cM	5	2	IM	0.27(0.18)	0.27(0.18)	0.26(0.18)
	5	2	IM-EB	0.22(0.14)	0.23(0.14)	0.23(0.14)
	5	4	IM	0.27(0.18)	0.27(0.17)	0.26(0.18)
	5	4	IM-EB	0.22(0.14)	0.22(0.14)	0.23(0.14)
	10	4	IM	0.27(0.19)	0.27(0.18)	0.26(0.18)
	10	4	IM-EB	0.17(0.10)	0.17(0.10)	0.18(0.10)
	10	9	IM	0.27(0.18)	0.27(0.18)	0.27(0.18)
	10	9	IM-EB	0.17(0.10)	0.19(0.10)	0.19(0.10)
	15	6	IM	0.27(0.18)	0.27(0.18)	0.27(0.18)
	15	6	IM-EB	0.14(0.08)	0.15(0.09)	0.16(0.09)
	15	14	IM	0.27(0.18)	0.27(0.18)	0.27(0.18)
	15	14	IM-EB	0.15(0.08)	0.16(0.09)	0.17(0.08)
One QTL at 65cM One QTL 65cM at 35cM	5	2	IM	0.26(0.17)	0.26(0.17)	0.27(0.18)
	5	2	IM-EB	0.21(0.13)	0.22(0.13)	0.22(0.14)
	5	4	IM	0.27(0.18)	0.27(0.18)	0.26(0.17)
	5	4	IM-EB	0.22(0.14)	0.22(0.14)	0.23(0.14)
	10	4	IM	0.27(0.18)	0.27(0.17)	0.27(0.18)
	10	4	IM-EB	0.16(0.09)	0.17(0.09)	0.17(0.10)
	10	9	IM	0.27(0.18)	0.27(0.18)	0.27(0.18)
	10	9	IM-EB	0.17(0.09)	0.18(0.10)	0.19(0.10)
	15	6	IM	0.27(0.18)	0.27(0.18)	0.27(0.18)
	15	6	IM-EB	0.14(0.08)	0.15(0.08)	0.15(0.08)
	15	14	IM	0.27(0.17)	0.27(0.18)	0.28(0.19)
	15	14	IM-EB	0.14(0.07)	0.16(0.08)	0.17(0.08)