Assessing the Impact of Genotype Imputation on Meta-analysis of Genetic Association Studies

by

Emmanuel Kevin Omondi

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg

Abstract

In this thesis, we study how a meta-analysis of genetic association studies is influenced by the degree of genotype imputation uncertainty in the studies combined and the size of meta-analysis. We consider the fixed effect meta-analysis model to evaluate the accuracy and efficiency of imputation-based meta-analysis results under different levels of imputation accuracy. We also examine the impact of genotype imputation on the between-study heterogeneity and type 1 error in the random effects meta-analysis model. Simulation results reaffirm that meta-analysis boosts the power of detecting genetic associations compared to individual study results. However, the power deteriorates with increasing uncertainty in imputed genotypes. Genotype imputation affects a random effects meta-analysis in a non-obvious way as estimation of between-study heterogeneity and interpretation of association results depend heavily on the number of studies combined. We propose an adjusted fixed effect meta-analysis approach for adding imputation-based studies to a meta-analysis of existing typed studies in a controlled way to improve precision and reliability. The proposed method should help in designing an effective meta-analysis study.

KEY WORDS: Between-study heterogeneity; Dosage test; Effect size; Genotype imputation; Inverse-variance weighting method

Acknowledgment

There are people without whom this thesis might not have been written, and to whom I am greatly indebted.

To Siegfried Grafe, together with Joan and Randy Porcher, you have supported my academic, social and spiritual growth through all means and have ensured that I reach the greatest academic level possible. Your support has been the pillar of every facet of my academic life for the last 13 years. In your quest to have me settled and going in Canada, you connected me with very helpful individuals in Darryl Bawdon and Jenna Silva, who ensured that I am always in time for my classes and any related activities. Their services impacted on me greatly and made my movement between Stonewall and the university easier than expected.

To my advisor, Dr. Elif F. Acar, you have been inspiration to me. In your hands I have learnt and achieved a lot. The nights we spent skyping and emailing back and forth have finally bore a worthwhile fruit. All the work in this thesis would not have been possible had I not seen the drive you had. You were always understanding, full of words of encouragement and always pushing me to the limits I thought I could never get to. You have proven to be a very valuable asset in my life and I am glad to the Almighty that I chose you as my advisor.

To the examining committee, Drs. Michelle Liu and Saman Muthukumarana, I am extremely grateful for the time you set aside to read my thesis and the revisions you proposed were worthwhile.

To my other half, Rebecca, through your patience, love and understanding, I got the peace to put this work together.

Dedication

I dedicate this thesis to my beloved late mother.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Genome-wide association studies (GWAS) are large population based studies that analyze a large number of genetic variants in the human genome for their relationship with a common disease or related traits. Prior to inception of GWAS, the major method of association testing was through genetic linkage in the family based studies. The linkage-based methods were highly successful with single-gene (monogenic) diseases, however, the same could not be replicated for common and complex multi-gene (polygenic) diseases. A proposed solution was GWAS, which boost the detection power, hence attain better performance than the linkage studies in finding genetic asociations for complex diseases (Altmüller et al., 2001).

GWAS have successfully identified many genetic variants underlying many diseases. Examples include the very first GWAS that assessed age-related macular degeneration in patients (Haines et al., 2005) and the largest GWAS so far that investigated multiple diseases such as the coronary heart disease, type 1 diabetes, type 2 diabetes, rheumatoid arthritis, Crohn's disease, bipolar disorder, and hypertension (WTCCC, 2007). A

simple Google Scholar search for "Genome-wide Association Study" yields about 52,000 scholarly published articles in all times, with 3,100 articles published in 2014 alone.

Despite the breakthrough of GWAS, the success in finding locations of genetic variations for traits of interest have been moderate. This is because the effect sizes of common variants are not large enough, and as a result, GWAS often have inflated type 1 error and low power (Li et al., 2012; Begum et al., 2012). Thus, GWAS must have relatively larger number of individuals for a desired detection power in discovering weaker genetic variants (Altshuler and Daly, 2007). Obtaining tens of thousands of individuals who are to be genotyped for hundreds of thousands to millions of genetic variants may not be feasible because of the laborious work of genoyping as well as high costs involved in the process. Genotype imputation offers one way of maintaining a large study size at a much lower cost.

Genotype imputation is a method of estimating variants, thereby allowing researchers to assess markers that are not directly genotyped in a study, or have missing values on some study individuals. It increases the number of variants to be tested for association while maintaining the sample size, hence boosts the detection power of GWAS in comparison to the case of no imputation. However, due to uncertainty arising from the probabilistic design of imputation methods, analysis based on imputed data fails to be as powerful as the ideal one that uses directly genotyped data. Thus, genotype imputation alone is not sufficient to address the problem of detecting risk variants with small effect sizes and/or lower risk allele frequencies.

A more powerful and cost-effective solution is to perform *meta-analysis*, a statistical technique that allows combining information across several independent studies. In the recent past, there have been considerable joint research efforts to amalgamate results from multiple GWAS into an extensive genome-wide meta-analysis which results in a larger sample size and an improved detection power (de Bakker et al., 2008; Begum et al.,

2012). While it has been proven successful in finding variants associated with common diseases such as Crohn's disease (Franke et al., 2010), rheumatoid arthritis (Stahl et al., 2010) and Parkinson's disease (Nalls et al., 2011), genome-wide meta-analysis has its own limitations. A major one is that because different GWAS use different genotyping array platforms (e.g., Affymetrix, Illumina), their set of genetic markers under investigation may not coincide with one another. Hence, even though these studies address the same research problem and employ same or similar methodologies, their results would not be combinable in a genome-wide level. Genotype imputation offers a natural strategy to overcome this challenge by imputing the variants that are genotyped in different platforms.

As noted earlier, imputation introduces uncertainty in genotypes, which results in a remarkable reduction of detection power of GWAS. This subsequently reduces the detection power of a genome-wide meta-analysis. By the fact that including more studies in meta-analysis also boosts power brings two conflicting effects into the whole meta-analysis process whose overall effect is not clearly understood. This has motivated the research in this thesis.

In recent years, there have been a number of publications on meta-analysis of genetic studies and various factors surrounding it (e.g., Sterne et al., 2000; Higgins et al., 2003), yet only few have addressed the impact of imputation. Zaitlen and Eskin, 2010 proposed the so-called *imputation aware meta-analysis* approach for case-control studies. Under the assumption of fixed effects, their method uses adjusted weights in the classical weighted sum of z-scores statistic (see Section 1.2) to control the contribution of imputation-based studies. The weight adjustment they suggest is based on the Pearson's correlation coefficient between the true and imputed genotypes, which quantifies the imputation quality. In as much as it improves the detection power in comparison to the standard approach, this method is yet to be implemented for quantitative traits,

studies with varying true effects sizes, as well as being replicated in other meta-analysis approaches.

Another work that has investigated the impact of imputation on meta-analysis was Li et al., 2012. Through simulations based on data from Framingham Heart Study, they found that for studies with a common true effect size, imputation induces considerable between-study heterogeneity, which in turn results in power reduction. They further considered a setting where studies have different true effect sizes, and concluded that the random effects model provides a better strategy for imputation-based meta-analysis as it accounts for between-study heterogeneity, irrespective of whether it is structural or induced.

In both papers, the empirical findings were based on meta-analysis of limited size; three studies in Li et al., 2012 and only two in Zaitlen and Eskin, 2010. Since the size of meta-analysis drastically affects both the detection power and the estimation of between-study heterogeneity, settings with more studies are needed for conclusive results. We, therefore, extensively explore the impact of imputation on genome-wide meta-analysis using a relatively large number of studies in our simulation experiments. A further inclusion in our experiments is the consideration of the degree of imputation accuracy, which neither work carefully assessed. As expected, the detection power of meta-analysis deteriorates with increasing uncertainty in the imputed data, suggesting that an adjustment strategy is necessary to reach valid conclusions.

Li et al., 2012 studied imputation-based meta-analysis for quantitative traits under both fixed effect and random effects models, but did not consider any adjustment on imputation quality. Zaitlen and Eskin, 2010, on the other hand, proposed a re-weighting scheme in their imputation-aware meta-analysis approach, but only for case-control studies with a common true effect size. In this thesis, we adapt the latter approach in the fixed effect meta-analysis, with some modifications using studies with quantitative

traits.

Our new re-weighting approach achieves a better detection power relative to the traditional approaches. Although the improvement is slight in the cases studied, our work provides a better understanding of, and treatment for, the impact of genotype imputation on meta-analysis of quantitative traits. We also assess the impact of genotype uncertainty on the type 1 error in the random effects meta-analysis. We observe that genotype uncertainty reduces the between-study heterogeneity as a result of reduced individual studies' effect sizes. Moreover, uncertainty affects the type 1 error of a random effects meta-analysis in a way that has no clear pattern. We focus on single marker tests in our analyses. However, the methods can be extended, with additional work, to GWAS settings where several markers are jointly tested.

In the remaining of this chapter, we first review statistical methods employed in meta-analysis, and then explain some basic genetics concepts together with common assumptions and genetic models in GWAS setting.

1.2 An Overview of Meta-analysis Methods

Meta-analysis refers to quantitative methods used to combine summary results of independent studies and to investigate between-study heterogeneity. Meta-analysis applications have evolved over a period of time. While some of the early statistical work goes back to Pearson, 1904, the framework for a formal anaysis is mainly due to Cochran, 1937 and Fisher, 1932. A relatively recent influential work is DerSimonian and Laird, 1986, which is credited with popularizing the use of meta-analysis in clinical trials. Besides their wide use in clinical practice, meta-analysis methods have been applied to agricultural experiments (e.g., Cochran, 1937), observational studies (e.g., Stroup et al., 2000), and lately, genetic studies (e.g., Ioannidis et al., 2001).

Meta-analysis uses a common measure of strength among studies, known as *effect size*, for amalgamation of study results. Effect sizes are obtained from a study's data using inferential statistics such as z-scores, p-values, log odds ratio, mean difference and model coefficients. A weighted mean of the effect sizes is often the outcome of interest in a meta-analysis (DerSimonian and Laird, 1986; Thompson et al., 2011), which can be inferred using different methods as detailed in Section 1.2.2.

1.2.1 Strengths and Limitations

Meta-analysis has an obvious comparative advantage over an analysis based on a single study. It aggregates results from many studies, hence, in essence, uses more information to address the problem of interest. This makes meta-analysis particularly attractive for discerning small effects, which would often be missed by individual studies, especially when conducted on a limited budget. Meta-analysis also allows researchers to investigate existence of structural and/or variational differences between studies. Once detected, such differences can usually be incorporated into meta-analysis in a flexible way. Furthermore, most meta-analysis approaches can easily handle addition of newly available study results, which essentially improves the accuracy and precision of overall summary estimates.

Despite its proven strengths, meta-analysis has several limitations. A major issue is the selection of studies to be used for a meta-anlysis, since the results highly depend on the studies selected. Meta-analysis often relies on results from successful studies of published literature. However, there exist many studies that are never published largely because of insignificant results, language bias, familiarity bias or due to inconclusive results by the authors (Rothstein et al., 2005; Kavvoura and Ioannidis, 2008). This selection of studies for publication is known as *publication bias*.

Publication bias may result in false positives, therefore it requires some serious

attention in a meta-analysis. Most available tools to detect publication bias are graphical, such as funnel plot (Light, 1984; Egger et al., 1997) and Galbraith plot (Galbraith, 1988). To counter the effect of publication bias, there have been serious efforts to include unpublished studies in a meta-analysis. To some extent, this has been enabled through collaborations in consortia and global networking among researchers (Walker et al., 2008). See, for instance, the Cochrane Collaboration (CC). Other common meta-analysis issues include validity of individual study data and results, population stratification and conflicts of interest of the studies underlying a meta-analysis study (Kavvoura and Ioannidis, 2008; Roseman et al., 2011). These aspects need careful examination prior to any meta-analysis.

1.2.2 Methods

Depending on the type of available study summaries and the model assumptions, meta-analysis methods can be classified into three major categories: p-value based approaches, fixed effect meta-analysis and random-effects meta-analysis.

p-value based approaches The Fisher's p-value approach (Fisher, 1932) is the simplest method of aggregating results across several studies. It uses the individual study p-values to combine the underlying studies, hence can be used when the effect sizes are not readily available. Given k independent studies, the Fisher approach uses the test statistic

$$x^2 = -2\sum_{j=1}^{k} \ln(p_j),$$

where p_j is the p-value of the jth study. Under the null hypothesis of "no effect", the test statistic follows a χ^2 distribution with $2k$ degrees of freedom. The p-value of the meta-analysis is given by $\Pr(X^2 > x^2)$.

The major pitfall of this method is that it treats the p-values unevenly. To clearly understand, we use the example outlined in Rice, 1990. Consider two studies with p-values 0.001 and 0.999. In the studies we reject the null hypothesis in the first one and fail to reject in the other. By p-value approach, we obtain a meta-analysis p-value of 0.008, which favours rejection of the null hypothesis, yet on average there is no significant effect. This shows that the method favours studies with smaller p-values, and this asymmetry introduces an undesired bias for the combined results.

Furthermore, the method assumes that all the studies involved are equally weighted, which is usually not the case as some studies are larger than others. Another downfall of this procedure is that the direction of the effect of each study is not taken into account. Hence, studies in opposite directions tend to strengthen each other instead of cancelling out (Begum et al., 2012).

Alternatives to the p-value approach that handle these problems include Stouffer's method (Stouffer et al., 1949). This strategy tackles the asymmetry problem by converting p-values to z-scores using $z_j = \phi^{-1}(1 - p_j)$ for a one-sided (right-tailed) test, where ϕ^{-1} is the quantile function of the standard normal distribution. For a two sided-test, the conversion requires knowledge of the effect direction, and defined as $z_j = \phi^{-1}(1 - p_j/2)$ for a positive effect and as $z_j = \phi^{-1}(p_j/2)$ for a negative one. Once the z-scores are obtained for each of the k studies, the test statistic is defined

$$z_S = \frac{\sum_{j=1}^{k} z_j}{\sqrt{k}}.$$

Under the null hypothesis, each z_j, the z-score of the jth study, is normally distributed with mean 0 and variance 1, so is the Stouffer's test statistic z_S. Hence, for a two-sided test, the p-value of meta-analysis is given by $2(1 - \phi(|z_S|))$, where ϕ is the cumulative distribution function of the standard normal.

However, Stouffer's procedure still equally weighs the studies. An improvement to the method was proposed independently by Mosteller and Bush, 1954 and Liptak, 1958. This development is a more powerful weighted z-score method that weighs studies differently as well as taking into account different directions of studies for a two-sided test. The test statistic is given by

$$z = \frac{\sum_{j=1}^{k} w_j z_j}{\sqrt{w_j^2}}.$$

The weights, w_j can be chosen as the inverses of the study variances or the square root of the study sizes. In studies that use t distribution, the weights are equivalent to the degrees of freedom (Begum et al., 2012).

The meta-analysis methods discussed so far use p-values or (converted) z-scores, from individual studies. More powerful alternatives use directly the observed effect sizes. These methods require the effect sizes to be measured with high precision on the same scale and units in all the underlying studies (Begum et al., 2012). The two popular methods are fixed effect meta-analysis (FEM) and random effects meta-analysis (REM) (Borenstein et al., 2010). These methods use the inverse-variance weighting approach to give weights to the individual studies (DerSimonian and Laird, 1986).

Fixed effect meta-analysis Fixed effect model assumes that a single true effect size θ underlies all the studies in the meta-analysis and variations in the observed effects $\hat{\theta}$ are as a result of sampling error ε. Hence, for a given jth study in a meta-analysis, the observed effect is defined as $\hat{\theta}_j = \theta + \varepsilon_j$, where ε_j is typically assumed to be normally distributed with mean 0 and variance σ^2. As a result, $\hat{\theta}_j$ is normally distributed with mean θ and variance σ^2/n_j, where n_j is the study size. The within study variance of the jth study is estimated by replacing the unknown σ^2 by the sampling error variance

estimate $\hat{\sigma}^2$, and is given by $v_j = \hat{\sigma}^2/n_j$. In the inverse-variance method, the weight of the jth study is defined as $w_j = 1/v_j$. This ensures that larger studies are given more weights since their inverse variances become larger. The overall weigthed observed effect size is given by

$$\hat{\theta}_{FE} = \frac{\sum_{j=1}^{k} w_j \hat{\theta}_j}{\sum_{j=1}^{k} w_j},$$

which is an unbiased estimate of θ, with standard error given as $\sqrt{1/\sum_{j=1}^{k} w_j}$. To test whether θ is different from a specific value θ_0, the test statistic takes the form

$$z_{FE} = \frac{\hat{\theta}_{FE} - \theta_0}{\sqrt{1/\sum_{j=1}^{k} w_j}}.$$

The test is significant when the p-value, $2(1 - \Phi(|z_{FE}|))$ is less than the significance level α.

Random effects meta-analysis In the random effects model, the true effect sizes θ_j's that underlie each study are allowed to vary from one to another. This brings in two sources of variation, (1) variation due to sampling error ε_j and (2) between-study variation ξ_j, which accounts for the deviation of true effect size θ_j from the true grand mean effect θ. Thus, the observed effect of the jth study is given by $\hat{\theta}_j = \theta_j + \varepsilon_j = \theta + \xi_j + \varepsilon_j$, where ξ_j and ε_j assumed to be independent and have normal distributions with mean zero and variances τ^2 and σ_j^2, respectively. Consequently, $\hat{\theta}_j$ has normal distribution with mean θ and variance $\sigma_j^2 + \tau^2$.

As in the fixed effect meta-analysis, the weight w_j^* of the jth study is defined using the inverse variance method, where within and between-study variances are replaced by

their estimates v_j and $\hat{\tau}^2$, i.e.,

$$\bar{w}_j = \frac{1}{v_j + \hat{\tau}^2}.$$

The estimate $\hat{\tau}^2$ is obtained from the Cochran Q-statistic,

$$Q = \sum_{j=1}^{k} w_j \left(\hat{\theta}_j - \hat{\theta}_{FE}\right)^2.$$

This statistic is also used in testing heterogeneity among studies. The τ^2 estimator is therefore given by

$$\frac{Q - df}{C},$$

where

$$C = \sum_{j=1}^{k} w_j - \frac{\sum_{j=1}^{k} w_j^2}{\sum_{j=1}^{k} w_j}.$$

Once between-study heterogeneity is estimated, the weighted observed effect $\hat{\theta}_{RE}$ is computed as

$$\hat{\theta}_{RE} = \frac{\sum_{j=1}^{k} \bar{w}_j \hat{\theta}_j}{\sum_{j=1}^{k} \bar{w}_j}.$$

The subsequent procedure is similar to the FEM case.

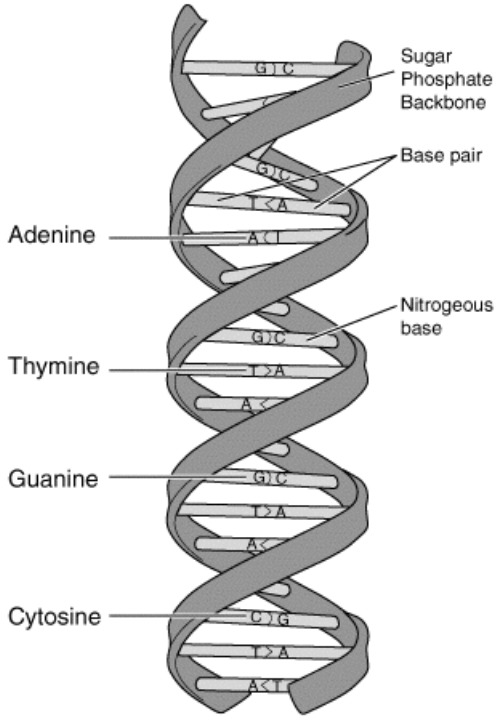
There exists other meta-analysis approaches such as Bayesian methods that utilize information from the prior assessment of the effect sizes (Sutton and Abrams, 2001) and multivariate approaches that include possible correlation between outcomes and effects (Zheng et al., 2012; Evangelou and Ioannidis, 2013).

1.3 Genetic Terminology

The purpose of this section is to familiarize the reader with basic genetic concepts. The field of genetics is very wide, therefore, we will limit our focus on concepts that are related to GWAS and this thesis.

Genetic research concerns the study of *human genome*, i.e. the entire collection of human genetic information. This information is encoded as a *DNA sequence*, which is a series of DNA bases, A, T, C and G, showing the order of nucleotides in the DNA. See Figure 1.1 (NHGRI) for a visual illustration.

Figure 1.1: Visual illustration of a DNA sequence.



Any two individuals have over 99% identical DNA sequence (IHGSC, 2001). However, the variations in remaining less than 1% may considerably cause modifications to a person's traits resulting into risks such as diseases or physical effects such as skin

complexion. The most common type of genetic variant (*polymorphism*) is a *single nucleotide polymorphism* (SNP), which occurs at a single nucleotide position in the DNA sequence. GWAS typically infer these type of variants. Other types include deletions, insertions, inversions and copy-number variations. These structural variations are more difficult to characterize (Raphael, 2012).

International research efforts to map and catalog genetic variations across human (and other species) genome resulted in a large collection of databases. In particular, the International HapMap Project (HapMap, 2005, 2007, 2010) and the 1000 Genome Project (T1000GPC, 2010, 2012) are rich sources that subsequently assist in identifying patterns in human genome map. To date over 38 million SNPs have been identified in the human genome from around 100,000 SNPs in 2005 (Zheng et al., 2012; Bush and Moore, 2012; T1000GPC, 2012).

Most human SNPs are biallelic, that is they consist of two alleles. We denote A as the major allele and a as the minor allele. Allele a usually is the *risk allele* if it is a direct cause of a trait. Its frequency in a population is referred to as *minor allele frequency* (MAF) only if the frequecy is < 0.5. We denote MAF as q while that for the major allele is denoted as p, where $p = 1 - q$. A MAF of 0.3 suggests that 30% of a population has allele a and 70% has allele A. The combinations of alleles, as a result of mating, yield *genotypes*. As illustrated in Table 1.1, biallelic SNPs have three possible genotypes AA, Aa and aa. Note that Aa and aA are treated the same, i.e., the order is not relevant.

A commonly made assumption in genetic studies is the Hardy–Weinberg equilibrium (HWE), also known as the Hardy–Weinberg principle, which allows predicting genotype frequencies from allele frequencies. The principle describes the genetic variation in a population and states that both allele and genotype frequencies in a population remain

Table 1.1: Genotype possibilities at a biallelic SNP given parental alleles.

		Father	
	Allele	A	a
Mother	A	AA	Aa
	a	aA	aa

constant from one generation to the other, provided there is random mating within the population and non existence of destabilizing factors (Edwards, 2008).

The genotype frequencies, shown in Table 1.2 are represented as probabilities of the genotypes where, $p^2 = Pr(AA)$, $q^2 = Pr(aa)$ and $2pq = Pr(Aa)$. The sum of all these proportions is a binomial expansion of the square of sum of allele frequencies, therefore, also equals to 1. The expansion, $p^2 + 2pq + q^2 = 1$, is called the Hardy-Weinberg equation. Zheng et al., 2012 show that the subsequent generations' genotype frequencies are constant given HWE.

Table 1.2: Genotype frequencies given parental alleles.

		Father		
	Allele		A	a
		Freq	p	q
Mother	A	p	p^2	pq
	a	q	qp	q^2

However, for the HWE to hold more assumptions are required. Therefore, in addition to random mating, the population size should be infinite, male and females should have identical allele frequencies, no migration and mutation and the absence of natural selection. Note that these restrictions may not be realistic in practical settings; one has to test whether HWE holds (Balding, 2006).

An observable characteristic of an individual is known as *phenotype* or *trait*. Traits

such as height, learning ability, blood pressure and weight are continuous and are often controlled by several genes (polygenic). Other traits like albinism, Huntington's disease and cystic fibrosis are discrete and are usually as a result of a single gene (monogenic) (Zheng et al., 2012). Continuous traits can also be expressed as discrete traits leading, for instance, to a case-control study setting, where cases are instances of extreme manifestation of a trait and controls otherwise.

A genetic model is a probabilistic description of inheritance. Dominant, reccessive, multiplicative and additive (dosage) models are the four most commonly analyzed models in literature as explained in Zheng et al., 2012. Table 1.3 shows a summary of these models.

Table 1.3: Summary of the genetic models.

Model	AA	Aa	aa
Dominant	0	r	r
Recessive	0	0	r
Multiplicative	0	r	r^2
Additive	0	r	$2r$

We assume that allele a is the risk allele, therefore, its presence in a genotype increases risk rth fold. In a dominant model genotypes Aa and aa have the same effect, hence having at least one copy of a results in risk, r. As for the recessive model, genotypes AA and Aa gives no risk meaning both copies of a are necessary in a genotype, i.e., aa, for risk r to be registered. The multiplicative model assumes no risk for AA, risk r for Aa and risk r^2 for aa. Additive model is linear, thus for each copy of a the risk increases by r.

In practice it is not possible to know the underlying model, therefore, some studies use a combination of models though a suitable correction for multiple testing is necessary. In GWAS, additive model is the most widely used model because it gives a reasonable

detection power for both the additive and dominant models though it is less powered for the other two (Bush and Moore, 2012). Throughout the thesis, we focus our attention to the additive model.

GWAS investigate a complete set of genetic information in many individuals in a bid to find association between genetic variants (SNPs) and heritable traits of interest (Clarke et al., 2011; Li et al., 2012). GWAS testing strategies include testing of association of multiple SNPs with a single trait, multiple SNPs with multiple traits as well as testing for interactions between genetic and environmental factors causing occurrence of a trait (Balding, 2006).

Most GWAS screen hundreds of thousands of SNPs for their associations with a large number of traits. Data on environmental variables are also often collected. Given such data structure, one may want to test genetic associations jointly on multiple SNPs and/or multivariate traits, possibly including gene-environment interactions. Since SNPs, as well as traits, are correlated, this task can be quite challenging. For computational feasibility and statistical tractability, initial SNP screening is almost always performed one-at-a-time considering one marker and one trait. Such procedures are referred to as *single-SNP tests*.

Statistical procedures for testing single-SNP associations differ according to the type of traits, the coding of genotype variable at the SNP, as well as the underlying genetic model. As has been discussed earlier, there are two types of traits; categorical and quantitative. Generalized linear models can be used to test for association between a SNP and a categorical trait, where logistic regression model is a specific case for binary traits. Other methods for binary traits include contingency tests such as Pearson χ^2-square test (Pearson, 1900), Fisher exact test (Fisher, 1922), likelihood-ratio test (Neyman and Pearson, 1933) or Cochran–Armitage trend test (Cochran, 1954b). For quantitative traits, the simple linear regression (SLR) model and the Analysis of

Variance (ANOVA) are the most widely used procedures for testing genetic associations.

There are also parametric methods of testing genetic association such as family based designs for linkage-based studies, score test, ordered categorical outcomes and bayesian alternatives (Zheng et al., 2012; Balding, 2006). Non-parametric methods include the rank-based Kruskal–Wallis one-way ANOVA. Like the ANOVA method, this method determines whether genotype groups orginate from the same distribution. When at least one group is different from the others, then significant result is assumed (Kruskal and Wallis, 1952; Acar and Sun, 2013). A comprehensive list of test procedures for both case-control data and quantitative traits can be found in Balding, 2006. In this thesis, we focus on single-SNP association tests for a quantitative trait using the SLR model. The details of the test procedure are in Chapter 2.

Before any association test is carried out on the collected data, quality of the data has to be ascertained. Therefore, testing deviations from the HWE, verifying the population homogeneity assumption and checking for missing data are mandatory data cleaning procedures for the genotype data. One also has to test any distributional assumptions on the quantitive traits.

As we have mentioned earlier, departure from the HWE can be as a result of one or a combination of inbreeding, mutation, population stratification and selection. HWE testing helps select a subset of existing SNPs that are good for association tests. See Zheng et al., 2012 for a review of statistical procedures used in testing HWE.

In large association studies, significant results for a single-SNP association test can be as a result of an underlying population structure and not necessarily because of the association between the SNP and trait. A population structure is the existence of sub-populations characterized by unique allele frequencies majorly due to non-random mating and disappearance of particular alleles as individuals migrate, die or do not

reproduce. Violation of population homogeneity can lead to both type 1 and type 2 errors, hence needs to be carefully examined in genomic control.

Another important issue in genetic studies is missing or low quality data. Missingness in phenotype data is seldom, and is typically handled by excluding the corresponding subjects from the study. Genotype data, on the other hand, are never perfect due to heavy data processing, referred to as genotyping, hence contain a lot of missing values. Imputation methods are widely used to treat the latter issue, as we detail in Chapter 2.

The outline of the thesis is as follows. Chapter 2 reviews statistical analysis of GWAS for typed and imputed SNPs, and evaluates the impact of imputation on association test results of a single study. Chapter 3 and Chapter 4 outline fixed effect and random effects meta-analysis methods, respectively. Chapter 3 contains our main contributions; namely (1) the assessment of the impact of genotype uncertainty on meta-analysis results, (2) the evaluation of the proposed re-weighting schemes for meta-analysis type 1 error and power gain, (3) an empirical judgement on the required number of imputation-based studies to achieve a certain power level in a meta-analysis. In Chapter 4, we assessed the impact of genotype uncertainty on between-study heterogeneity as well as the type 1 error in the null hypothesis case. We provide our conclusions in Chapter 5. In the appendix, we have provided part of the *r-program* codes that we used for data simulation and analyses.

Chapter 2

Association Testing in Genome-wide Association Studies

Over the last decade, GWAS have transformed into a dominant tool for exploring genetic construction of human traits (Bush and Moore, 2012), and have been successful in identifying genetic variants underlying many diseases. While there are recently emerged new technologies in genetic research, such as whole-genome and exome next generation sequencing (Hall, 2007; Zhang et al., 2011), GWAS still constitute a popular cost-effective tool for a first step in unearthing genetic locations contributing towards human susceptibility to complex diseases. The complex nature of GWAS invites a rich collection of statistical problems, concerning aspects ranging from genomic control to multiple comparison adjustments.

In this chapter we study the impact of genotype imputation on single-SNP association test results for a quantitative trait in the case of a single GWAS. This consideration provides foundation for our investigations in Chapter 3 and Chapter 4 where we address the impact of genotype imputation on meta-analysis of multiple GWAS.

In Section 2.1, we describe the problem of genotype uncertainty and briefly review some genotype imputation algorithms. The statistical framework and methods for

testing associations are outlined in Section 2.2, in the cases without and with genotype uncertainty. Section 2.3 contains simulation experiments to evaluate the impact of genotype uncertainty on test results. The chapter ends with a summary of the major conclusions.

2.1 Genotype Uncertainty

Quality checks during GWAS data cleaning result in discarding poor quality samples (individuals) or SNPs. Hence, in practice, GWAS data are almost always incomplete. SNPs violating the HWE assumption, or having a small minor allele frequency ($<$ 0.05 or 0.1) or a low genotype call accuracy ($<$0.9 or 0.99) are typically removed from the anaysis. Furthermore, individuals having many untyped genotypes (e.g., missing rate $>$5%) would be excluded from the study. Such data cleaning would improve the reliability of GWAS, but at a cost of losing information.

Genotype imputation is an appealing strategy to recover missing and/or untyped genotype information, thereby increasing power of an association study, especially for SNPs that are difficult to identify (Marchini et al., 2007; Li et al., 2009; Spencer et al., 2009). Imputation also helps in cost reduction in large studies where a subset of SNPs is typed and the remaining set is imputed (Zhang et al., 2013). More importantly, genotype imputation makes it possible to combine studies that use different genotyping platforms, hence allow meta-analysis.

Most imputation algorithms employ a combination of statistical techniques including hidden Markov models, expectation maximization algorithm, maximum likelihood estimation and random selection from a probability distribution, as well as indirect inference of neighbouring SNPs in the study and of reference panels of extensively genotyped SNPs from the International HapMap Project and 1000 Genomes Project.

There exist many software tools such as MACH (Li and Abecasis, 2006), fast-PHASE (Scheet and Stephens, 2006), P-LINK (Sham et al., 2007), IMPUTE (Marchini et al., 2007; Howie et al., 2009), BEAGLE (Browning and Browning, 2009) that perform genotype imputation. While they use different algorithms for imputation and in quantifying its accuracy, they often yield similar results (Marchini and Howie, 2010).

Imputation methods often give a probabilistic estimation of an imputed genotype as vector of group probabilities $\mathbf{p}_i = (p_{i0}, p_{i1}, p_{i2})$, with $\sum_{j=0}^{2} p_{ij} = 1$. Here, $pij = Prob(G_i = j)$ is the probability of the jth genotype on the ith individual, G_i is the genotype variable for the ith individual at a particular SNP. Table 2.1 illustrates a typical genotype imputation output for a fully imputed SNP.

Table 2.1: Imputed genotype probabilities for a fully imputed SNP.

	Genotype		
Individual	0	1	2
1	0.958	0.011	0.031
2	0.316	0.248	0.436
⋮	⋮	⋮	
n	0.000	0.621	0.379

Various strategies can be used to combine the imputed genotype probabilities to estimate the missing genotypes. The most common approaches are the best guess (BG) approach and the dosage approach. The BG approach selects the genotype value with the highest probability, that is,

$$\tilde{G}_i = \{j \ : \ \max(p_{i0}, p_{i1}, p_{i2})\},$$

while the dosage approach uses the *dosage*, the expected number of copies of the minor allele, given by

$$\tilde{G}_i = 0\ p_{i0} + 1\ p_{i1} + 2\ p_{i2}.$$

The dosage approach almost always performs better than the BG method since it loses less information. There also exist other approaches for combining the genotype probabilities that are either model-based, such as posterior probability method (Marchini et al., 2007; Aulchenko et al., 2010; Zheng et al., 2011), or non-parametric (Acar and Sun, 2013).

2.2 Association Testing

As have been discussed earlier, strategies for testing genetic associations differ for categorical and quantitative traits. In this section, we look at association testing of a quantitative trait, first at a fully typed SNP and then at an imputed SNP.

2.2.1 Association Testing at a Fully Typed SNP

The most commonly used test approaches for quantitative traits are the analysis of variance (ANOVA) and simple linear regression (SLR) models. Suppose we have a sample of n unrelated individuals. Denote the trait response by Y_i and the genotype by G_i for the ith individual. In ANOVA, the genotype is taken to be a categorical covariate. On the other hand, if we consider the genotype as numeric, the association test is addressed within the SLR model, given by

$$Y_i = \beta_0 + \beta_1 \, G_i + \varepsilon_i, \qquad i = 1, \ldots, n \quad G_i \in \{0, 1, 2\}. \tag{2.1}$$

The error terms ε_i's are assumed to be independently and identically distributed with mean 0 and variance σ^2. If the error distribution is assumed normal, then the phenotype variable Y_i follows a normal distribution with mean $\beta_0 + \beta_1 G_i$ and variance σ^2. Note that the SLR model in (2.1) accommodates the ANOVA model, when G_i's are treated categorical.

GWAS association testing under model (2.1) concerns the testing problem,

$$H_0 : \beta_1 = 0 \qquad vs. \qquad H_A : \beta_1 \neq 0, \tag{2.2}$$

where the null hypothesis specifies *no genetic association* and the alternative indicates a *genetic association* in either direction, i.e., two sided. Under the SLR model, the parameter of interest β_1 is estimated using the least-squares method, which yields

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(G_i - \bar{G})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(G_i - \bar{G})^2} = \frac{\sum_{i=1}^{n}(G_i - \bar{G})Y_i}{\sum_{i=1}^{n}(G_i - \bar{G})^2},$$

with standard error,

$$s_{\hat{\beta}_1} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{(n-2)\sum_{i=1}^{n}(G_i - \overline{G})^2}} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(G_i - \overline{G})^2}},$$

The estimator $\hat{\beta}_1$ is consistent for β_1 and have the distribution,

$$\hat{\beta}_1 \sim N\left(\beta_1,\ \frac{\sigma^2}{\sum_{i=1}^{n}(G_i - \bar{G})^2}\right).$$

Thus, the testing problem in (2.2) is addressed using the statistic,

$$z = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}.$$

Since the study sizes of GWAS are often large, the Central Limit Theorem together with Slutsky's lemma, yield an asymptotic normal distribution for this statistic, i.e.,

$$z \overset{\text{d}}{=} \frac{\hat{\beta}_1}{\sigma/\sqrt{\sum_{i=1}^{n}(G_i - \bar{G})^2}} \sim N\left(\frac{\sqrt{\sum_{i=1}^{n}(G_i - \bar{G})^2}}{\sigma}\beta_1, 1\right), \tag{2.3}$$

which corresponds to the standard normal distribution under the null hypothesis. Thus, we reject the null hypothesis in (2.2) when the p-value, $2\Pr(Z > |z|)$ is less than the level of significance α.

This test procedure is conducted for each of the SNPs in a GWAS. It is preferred due to its simplicity, as it allows fast screening of a large number of SNPs.

2.2.2 Association Testing at an Imputed SNP

Testing association for a quantitative trait at an imputed SNP is more or less similar to testing at a fully genotyped SNP. The difference only comes in with the type of approach employed in estimating genotypes from imputed genotype probabilities. We denote the imputed genotypes as $\tilde{G}$. In the BG approach, $\tilde{G}$ can be taken as categorical or numerical. This allows the use of either SLR model or ANOVA model in association testing as explained in the previous part. However, if dosage is used, only the SLR model is suitable since $\widetilde{G}$ is continuous.

Replacing G in (2.1) with $\tilde{G}$ yields

$$Y_i = \beta_0 + \beta_1 \widetilde{G}_i + \varepsilon_i, \qquad i = 1, \ldots, n, \quad \widetilde{G}_i \in [0, 2].$$

The least squares estimator of β_1 is given by

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2} = \frac{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})Y_i}{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2},$$

with the standard error,

$$s_{\tilde{\beta}_1} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \widetilde{Y}_i)^2}{(n-2)\sum_{i=1}^{n}(\tilde{G}_i - \overline{\tilde{G}})^2}} = \sqrt{\frac{\tilde{\sigma}^2}{\sum_{i=1}^{n}(\tilde{G}_i - \overline{\tilde{G}})^2}},$$

where $\widetilde{Y}_i$ is the predicted phenotype value obtained from the fitted regression line. Once the estimates $\tilde{\beta}_1$ and $s_{\tilde{\beta}_1}$ are obtained, the hypothesis in (2.2) is tested using the same procedure as in Section 2.2.1.

However, it is important to note that $\tilde{\beta}_1$ is not an unbiased estimator of β_1. Its expectation is

$$\begin{aligned} \mathrm{E}(\tilde{\beta}_1) &= \frac{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})\mathrm{E}(Y_i - \bar{Y})}{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2} \\ &= \frac{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})(G_i - \bar{G})}{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2}\,\beta_1 = \gamma\,\beta_1, \end{aligned} \tag{2.4}$$

where γ is the shift factor. Similarly, the variance of $\tilde{\beta}_1$ can be obtained as

$$\mathrm{var}(\tilde{\beta}_1) = \frac{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2\,\mathrm{var}(Y_i)}{\left(\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2}.$$

Hence, even though the distribution of $\tilde{\beta}_1$ is normal, its parameters are different from those of the distribution of $\hat{\beta}_1$, i.e.,

$$\tilde{\beta}_1 \sim N\left(\gamma\beta_1,\ \ \frac{\sigma^2}{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2}\right).$$

Provided that $\tilde{\sigma}^2$ consistent for σ^2, we obtain the following asymptotic distribution for the test statistic

$$\widetilde{z} = \frac{\tilde{\beta}_1}{s_{\tilde{\beta}_1}} \overset{\mathrm{d}}{=} \frac{\tilde{\beta}_1}{\sigma/\sqrt{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2}} \sim N\left(\frac{\gamma\,\beta_1}{\sigma/\sqrt{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2}},\ 1\right).$$

Our empirical findings in Section 2.3.3 suggest an almost similar performance of $\hat{\sigma}^2$ and $\tilde{\sigma}^2$.

Note that when the null hypothesis holds, i.e., $\beta_1 = 0$, $\tilde{z}$ has standard normal distribution. However, when $\beta_1 \neq 0$, the mean of the distribution is a scale multiple of that of $z = \beta_1 / s_{\hat{\beta}_1}$. The scale factor is shown to be

$$\frac{\mathrm{E}(\tilde{z})}{\mathrm{E}(z)} = \frac{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})(G_i - \bar{G})}{\sigma\sqrt{\sum_{i=1}^{n}(\tilde{G}_i - \bar{\tilde{G}})^2}} \frac{\sigma}{\sqrt{\sum_{i=1}^{n}(G_i - \bar{G})^2}} = r_{G,\tilde{G}},$$

where $r_{G,\tilde{G}}$ is the Pearson's correlation coefficient between the true genotypes G and the imputed ones $\tilde{G}$.

Having discussed the test procedures at a typed and imputed SNP, we will examine their empirical performance in the next section.

2.3 Simulation Study

We conduct extensive simulations majorly to assess how the degree of genotype imputation uncertainty affects the association test results based on dosage. We also compare the performance of the dosage approach with that of the BG method.

2.3.1 Simulation Design

We focus our attention to the SLR model, hence consider a quantitative trait that is normally distributed and treat G as numeric. Table 2.2 summarizes the parameter values describing our simulation scenarios. These values yield 72 different settings, for which we generate data as follows.

Since phenotype is conditional on genotype, we first generate the genotype data. Genotype generation was performed in two steps. In the first step, we obtain the genotype group sizes randomly from a trinomial distribution whose parameters are determined by the MAF and sample size under the HWE assumption. Hence, in a

Table 2.2: Parameter values describing the simulation scenarios.

Parameter	Values
MAF	0.1 0.2 0.3
n	200 500 1000
β_0	30
β_1	0 0.25 0.5 1
σ	1 3

GWAS of size n with a MAF of q, the expected group sizes for genotypes 0,1, and 2 are $n_0 = n\ (1-q)^2, n_1 = n\ 2(1-q)q$ and $n_2 = n\ q^2$, respectively, assuming HWE holds. Once the group sizes are determined randomly, we code the true genotypes accordingly. For each MAF and n combination, we generate one set of genotypes, yielding to 9 different genotype sets in total.

In the second step we induce uncertainity into the genotypes using the *Dirichlet* distribution since it is a distribution over multinomials. In Dirichlet distribution, the parameter vector, $\psi = (\psi_1, \psi_2, \psi_3)$ with $\psi_1 + \psi_2 + \psi_3 = 1$ is specified to account for the genotype imputation accuracy. The resultant output is a random vector of probabilities, $\mathbf{P} = (P_0, P_1, P_2)$, where $P_0 + P_1 + P_2 = 1$ and each of the vector component corresponds to the genotypes 0, 1 and 2 respectively.

When generating these probabilities from the Dirichlet distribution, we consider $\psi < 1$ for the true genotype group, and $(1-\psi)/2$ for the other two. The parameter values are chosen as $\psi = 0.9, 0.8$, and 0.7 representing 10%, 20% and 30% uncertainty levels, respectively. Note that the true genotypes have 0% uncertainty level reflecting the case where SNP is typed. We assume an additive model and generate the phenotype data conditional on genotypes using the SLR model in (2.1). For a given MAF and sample size, we use the true genotypes in the corresponding genotype data, and generate 10,000 Monte-Carlo samples of phenotype data, each from a normal distribution with

mean $\beta_0 + \beta_1 G$, where $G \in \{0, 1, 2\}$ and standard deviation σ. Hence, for each of genotype data set, we obtain 8 different sets of phenotype data, replicated 10,000 times.

Note that the parameter values for β_1 are chosen to reflect the cases with small, medium and large effect sizes. The σ values are selected under the alternative hypothesis case such that we obtain empirical power of at least 0.3 at 0% level of uncertainty with $n = 500$. MAFs are chosen to reflect typical settings for common markers, usually $>$0.05.

2.3.2 Association Test Results

For each genotype dataset, we obtain the dosage and BG genotype values at each uncertainty level. We then perform association tests as explained in Section 2.2.2. Our main interest is to understand the impact of genotype uncertainty on the test results as we vary the simulation parameters. In the following, we first check the accuracy of the tests under the null hypothesis and then do a power analysis.

Evaluation of Empirical Type 1 Error Rates

Under the settings with $\beta_1 = 0$, which correspond to the null hypothesis of no genetic association, we obtain the empirical type 1 error rates at $\alpha = 0.005, 0.01$ and 0.05. The results are presented in Table 2.3 for each MAF. Based on these results, we see that the empirical type 1 error rates are not significantly different from the corresponding nominal α levels for both BG and dosage approaches. Thus, neither the approach in imputing genotypes nor the degree of genotype uncertainty has a significant impact on the type 1 error rate.

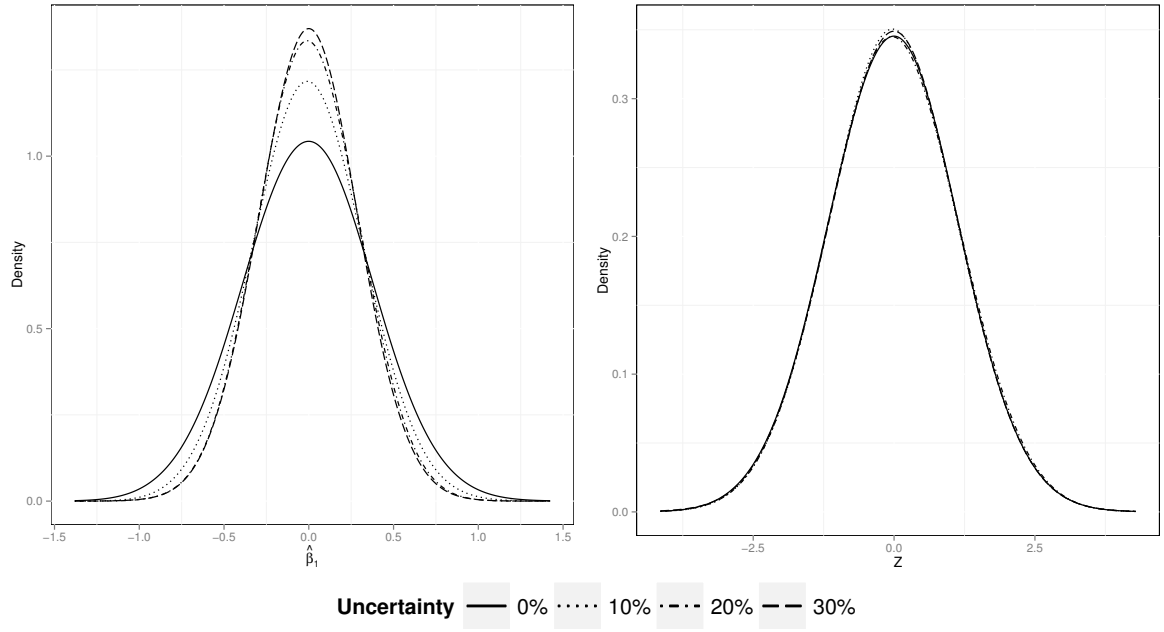
For the case with MAF = 0.1, $n = 500$, and $\sigma = 3$, we obtain the density plots of the β_1 estimates and the test statistic z under the dosage approach across different uncertainty levels. In the left panel of Figure 2.1, we see that the means of the β_1

Table 2.3: Empirical type 1 error rates of the association tests at $\alpha = 0.005, 0.01, 0.05$ for the dosage and BG procedures when $n = 500$.

			Dosage			BG		
MAF	σ		$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.05$
0.1	1	0%	0.0051	0.0097	0.0467	0.0051	0.0097	0.0467
		10%	0.0057	0.0104	0.0498	0.0050	0.0106	0.0494
		20%	0.0052	0.0103	0.0505	0.0052	0.0094	0.0487
		30%	0.0046	0.0107	0.0527	0.0051	0.0105	0.0516
	3	0%	0.0041	0.0092	0.0514	0.0041	0.0092	0.0514
		10%	0.0045	0.0090	0.0486	0.0055	0.0107	0.0466
		20%	0.0044	0.0087	0.0498	0.0062	0.0103	0.0537
		30%	0.0052	0.0099	0.0507	0.0065	0.0102	0.0469
0.2	1	0%	0.0050	0.0089	0.0474	0.0050	0.0089	0.0474
		10%	0.0044	0.0088	0.0461	0.0042	0.0085	0.0465
		20%	0.0048	0.0102	0.0498	0.0050	0.0105	0.0511
		30%	0.0051	0.0102	0.0524	0.0041	0.0085	0.0522
	3	0%	0.0057	0.0107	0.0506	0.0057	0.0107	0.0506
		10%	0.0062	0.0106	0.0520	0.0060	0.0113	0.0524
		20%	0.0061	0.0122	0.0528	0.0064	0.0127	0.0526
		30%	0.0047	0.0088	0.0500	0.0049	0.0096	0.0495
0.3	1	0%	0.0058	0.0118	0.0527	0.0058	0.0118	0.0527
		10%	0.0045	0.0109	0.0526	0.0056	0.0121	0.0520
		20%	0.0069	0.0126	0.0521	0.0051	0.0105	0.0510
		30%	0.0048	0.0103	0.0515	0.0052	0.0102	0.0522
	3	0%	0.0050	0.0092	0.0522	0.0050	0.0092	0.0522
		10%	0.0048	0.0100	0.0515	0.0054	0.0099	0.0517
		20%	0.0068	0.0120	0.0546	0.0060	0.0107	0.0531
		30%	0.0042	0.0102	0.0541	0.0051	0.0100	0.0500

estimates are centered around the true value $\beta_1 = 0$ for all uncertainty levels. However, their standard errors tend to decrease as uncertainty increases, as depicted by the

Figure 2.1: Density plots of the β_1 estimates (left) and the test statistic z (right) under the null hypothesis.



shrinking density tails. On the other hand, the density plots of z-scores in the right panel of Figure 2.1 indicate no significant difference across different levels of imputation accuracy. This supports the result in (2.3) and explain why the empirical type 1 error rates in Table 2.3 are not significantly affected by genotype uncertainty.

Evaluation of Empirical Power

For the settings with $\beta = 0.25, 0.5$ and 1, we calculate the empirical power using the nominal level $\alpha = 0.01$. The results for the dosage and BG approaches are summarized in Table 2.4 and Table 2.5, respectively. In both tables we see that the power drops significantly as uncertainty level increases. We also observe that as n, MAF and the true effect size β_1 increase, the empirical power within a specific case increases as expected. The reverse is observed when σ increases.

A cross-comparison of these results indicate that the BG approach may not be a good choice for testing genetic associations at an imputed SNP, especially when the

Table 2.4: Empirical power at $\alpha = 0.01$ using the dosage procedure.

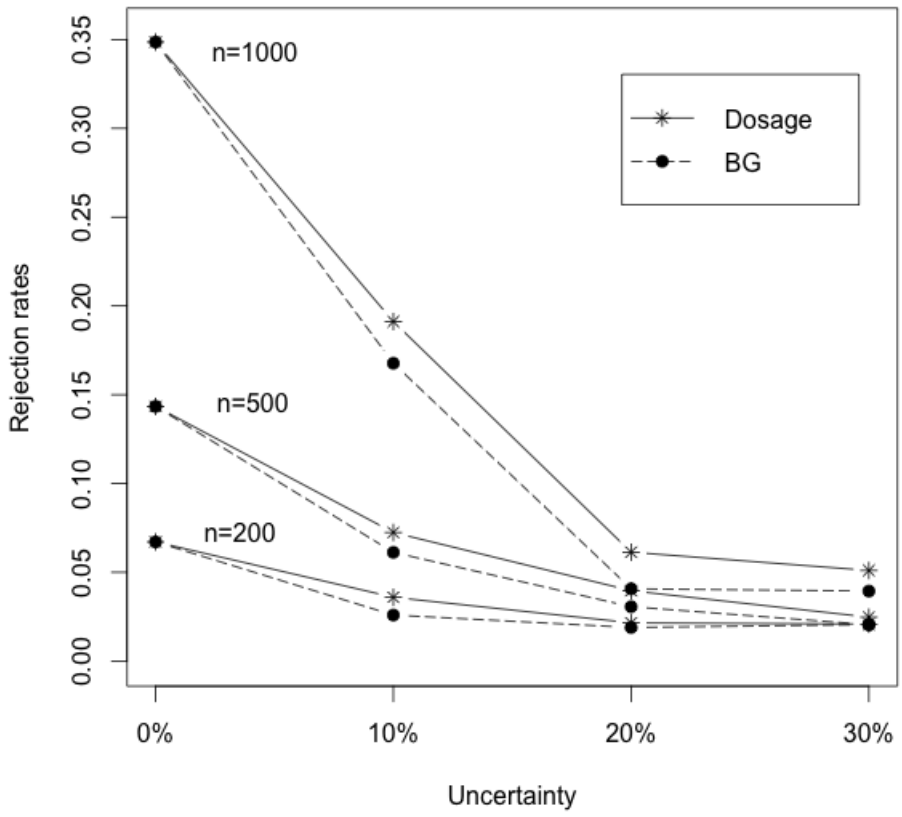
			MAF = 0.1			MAF = 0.2			MAF = 0.3		
			$n = 200$	$n = 500$	$n = 1000$	$n = 200$	$n = 500$	$n = 1000$	$n = 200$	$n = 500$	$n = 1000$
$\sigma = 1$	$\beta_1 = 0.25$	0%	0.1497	0.3696	0.7526	0.2635	0.7387	0.9674	0.3385	0.8289	0.9938
		10%	0.0717	0.1850	0.4794	0.1683	0.5747	0.8830	0.2437	0.6906	0.9749
		20%	0.0305	0.0921	0.1604	0.1198	0.3531	0.6525	0.1875	0.5163	0.9184
		30%	0.0372	0.0418	0.1251	0.0842	0.1777	0.3668	0.0739	0.2836	0.6655
	$\beta_1 = 0.5$	0%	0.6980	0.9707	1.0000	0.9128	1.0000	1.0000	0.9636	1.0000	1.0000
		10%	0.3610	0.7706	0.9913	0.7407	0.9984	1.0000	0.8863	0.9998	1.0000
		20%	0.1222	0.4547	0.7284	0.5728	0.9647	0.9997	0.7852	0.9940	1.0000
		30%	0.1490	0.2192	0.5979	0.4166	0.7557	0.9719	0.3535	0.9197	0.9997
	$\beta_1 = 1$	0%	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10%	0.9640	0.9999	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
		20%	0.5856	0.9900	0.9998	0.9977	1.0000	1.0000	0.9999	1.0000	1.0000
		30%	0.6492	0.8253	0.9982	0.9784	0.9999	1.0000	0.9525	1.0000	1.0000
$\sigma = 3$	$\beta_1 = 0.25$	0%	0.0183	0.0334	0.0640	0.0266	0.0654	0.1365	0.0322	0.0848	0.1990
		10%	0.0150	0.0229	0.0434	0.0206	0.0434	0.0976	0.0256	0.0619	0.1516
		20%	0.0118	0.0165	0.0197	0.0175	0.0364	0.0575	0.0252	0.0469	0.1076
		30%	0.0112	0.0135	0.0224	0.0167	0.0236	0.0358	0.0133	0.0279	0.0607
	$\beta_1 = 0.5$	0%	0.0671	0.1434	0.3486	0.0998	0.3393	0.6454	0.1314	0.4094	0.8006
		10%	0.0359	0.0724	0.1912	0.0690	0.2336	0.4765	0.0970	0.3001	0.6721
		20%	0.0215	0.0396	0.0613	0.0481	0.1408	0.2740	0.0801	0.2087	0.5328
		30%	0.0208	0.0247	0.0510	0.0392	0.0735	0.1387	0.0330	0.1076	0.2869
	$\beta_1 = 1$	0%	0.3116	0.6666	0.9592	0.5252	0.9563	0.9994	0.6277	0.9817	1.0000
		10%	0.1362	0.3584	0.7866	0.3396	0.8652	0.9929	0.4801	0.9349	0.9997
		20%	0.0558	0.1811	0.3296	0.2432	0.6448	0.9159	0.3772	0.8204	0.9965
		30%	0.0613	0.0850	0.2500	0.1637	0.3479	0.6474	0.1383	0.5290	0.9258

Table 2.5: Empirical power at $\alpha = 0.01$ using the best guess procedure.

			MAF = 0.1			MAF = 0.2			MAF = 0.3		
			$n = 200$	$n = 500$	$n = 1000$	$n = 200$	$n = 500$	$n = 1000$	$n = 200$	$n = 500$	$n = 1000$
$\sigma = 1$	$\beta_1 = 0.25$	0%	0.1497	0.3696	0.7526	0.2635	0.7387	0.9674	0.3385	0.8289	0.9938
		10%	0.0533	0.1589	0.4091	0.1652	0.5251	0.8355	0.2277	0.6751	0.9640
		20%	0.0258	0.0683	0.0963	0.0945	0.2649	0.5132	0.1390	0.3885	0.8708
		30%	0.0343	0.0356	0.0843	0.0613	0.1357	0.2399	0.0574	0.1747	0.5049
	$\beta_1 = 0.5$	0%	0.6980	0.9707	1.0000	0.9128	1.0000	1.0000	0.9636	1.0000	1.0000
		10%	0.2431	0.7089	0.9820	0.7351	0.9968	1.0000	0.8602	0.9997	1.0000
		20%	0.0896	0.3132	0.4901	0.4721	0.9034	0.9963	0.6531	0.9727	1.0000
		30%	0.1293	0.1492	0.4417	0.2802	0.6465	0.8818	0.2645	0.7386	0.9953
	$\beta_1 = 1$	0%	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		10%	0.8752	0.9999	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
		20%	0.4199	0.9382	0.9930	0.9903	1.0000	1.0000	0.9995	1.0000	1.0000
		30%	0.5702	0.6650	0.9869	0.8896	0.9996	1.0000	0.8747	1.0000	1.0000
$\sigma = 3$	$\beta_1 = 0.25$	0%	0.0183	0.0334	0.0640	0.0266	0.0654	0.1365	0.0322	0.0848	0.1990
		10%	0.0143	0.0214	0.0370	0.0199	0.0420	0.0850	0.0243	0.0618	0.1381
		20%	0.0129	0.0138	0.0166	0.0177	0.0291	0.0439	0.0220	0.0368	0.0970
		30%	0.0127	0.0131	0.0171	0.0142	0.0202	0.0267	0.0129	0.0233	0.0478
	$\beta_1 = 0.5$	0%	0.0671	0.1434	0.3486	0.0998	0.3393	0.6454	0.1314	0.4094	0.8006
		10%	0.0259	0.0612	0.1677	0.0680	0.2059	0.4164	0.0903	0.2875	0.6311
		20%	0.0189	0.0306	0.0407	0.0423	0.1027	0.2020	0.0641	0.1550	0.4577
		30%	0.0205	0.0205	0.0394	0.0302	0.0594	0.0938	0.0294	0.0704	0.1989
	$\beta_1 = 1$	0%	0.3116	0.6666	0.9592	0.5252	0.9563	0.9994	0.6277	0.9817	1.0000
		10%	0.0946	0.3171	0.7190	0.3320	0.8202	0.9849	0.4522	0.9241	0.9994
		20%	0.0395	0.1197	0.1964	0.1896	0.5110	0.8092	0.2820	0.6787	0.9908
		30%	0.0573	0.0662	0.1730	0.1081	0.2763	0.4690	0.1013	0.3373	0.8105

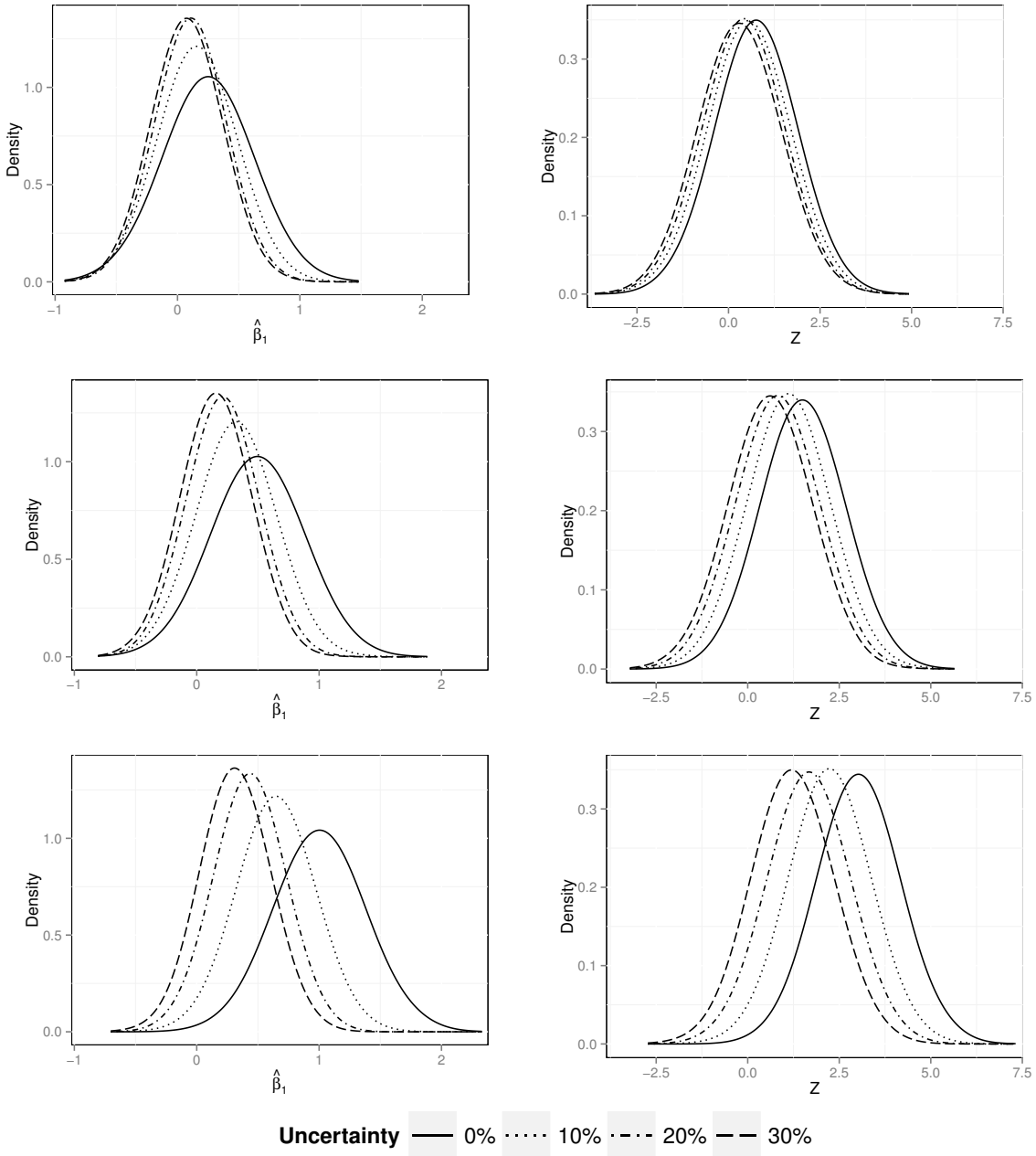
underlying model is additive. This conclusion is driven by the lower empirical power of the BG approach under all the alternative hypothesis scenarios, in comparison to the dosage approach. The performances of the two approaches are also compared graphically in Figure 2.2 for the case with MAF $= 0.1$, $n = 500$, and $\sigma = 3$. We see that the dosage approach slightly outperforms the BG procedure at all uncertainty levels. We also observe that, besides an increasing degree of uncertainty, smaller sample sizes also lower the detection power significantly.

Figure 2.2: Comparison of performances of the dosage and best guess procedures across different uncertainty levels and study sizes, when MAF $= 0.1$, $\beta_1 = 0.5$ and $\sigma = 3$.



For a better understanding of the power decrease due to imputation uncertainty, we look at the density plots of β_1 estimates and the test statistic z for the dosage

Figure 2.3: Density plots of the β_1 estimates (left) and the test statistic z (right) under the alternative hypothesis scenarios with $\beta_1 = 0.25$ (top), $\beta_1 = 0.5$ (middle), $\beta_1 = 1$ (bottom).



approach under all alternative scenarios, focusing on the case with MAF $= 0.1$, $n = 500$, and $\sigma = 3$. The top, middle and bottom panels of Figure 2.3 show the results for the scenarios with $\beta_1 = 0.25$, 0.5 and 1, respectively. The density plots for the β_1

estimates, shown in the left panels, indicate a shift from the true value towards zero, increasing in magnitude with the uncertainty level. We also observe lower variability in the distribution of the β_1 estimates at higher uncertainty levels. While this feature disappears upon standardization to z-scores (right panels), we still observe a shift towards zero, proportional to the true β_1 value in the distribution of the test statistic.

As β_1 increases, the shift from the true effect size widens with an increase in uncertainty. The effect size does not seem to significantly affect the standard error of the estimates.

2.3.3 A Closer Look at Genotype Uncertainty

We have seen that uncertainty in the genotypes reduces the detection power of GWAS. We now take a closer look at how this occurs, focusing on the cases with MAF = 0.1, $n = 500$ and $\sigma = 3$, as presented in Figure 2.2.

We first look at how the two imputation approaches differ in terms of the genotype values they produce. In Table 2.6, we report the standard deviations $s_{\tilde{G}}$ of the imputed genotypes under both approaches at each uncertainty level, where the value at 0% gives the standard deviation s_G of the true genotypes. We also include the Pearson correlation coefficient between G and $\tilde{G}$, and the shift factor $\gamma = r_{G,\tilde{G}}(s_G/s_{\tilde{G}})$ in (2.4) under each approach.

Table 2.6: The genotype standard deviations, correlation values and shift factors under the dosage and BG procedures at different uncertainty levels for the case with MAF = 0.1 and n =500.

	Dosage				BG			
	0%	10%	20%	30%	0%	10%	20%	30%
$s_{\tilde{G}}$	0.405	0.466	0.519	0.526	0.405	0.524	0.646	0.700
$r_{G,\tilde{G}}$	1.000	0.743	0.554	0.409	1.000	0.701	0.472	0.355
γ	1.000	0.646	0.432	0.315	1.000	0.541	0.296	0.205

Table 2.6 suggests that the imputed genotypes have larger variation compared to the true ones, and the variability increases with the degree of uncertainty. As expected, the correlation between the imputed and true genotypes decreases as uncertainty increases. The shift factor quantifies the amount of bias, which is $(\gamma - 1)\beta_1$, in the estimates of genetic effect. Hence, the smaller γ values at higher uncertainty levels indicate a larger bias in the corresponding estimates. Overall, we observe that BG approach produces more variation, lesser correlation and larger bias than the dosage approach at all uncertainty levels, which explains its relatively poor performance in Section 2.3.2.

Table 2.7: The mean and standard deviation (given in parentheses) of $\tilde{\sigma}^2$ calculated from 10000 Monte-Carlo samples, for the cases with MAF $= 0.1$, $n = 500$ and $\sigma = 3$ under the dosage procedure at difference uncertainty levels.

	Uncertainty Level			
	0%	10%	20%	30%
$\beta_1 = 0$	8.99670 (0.57167)	8.99689 (0.57151)	8.99668 (0.57121)	8.99671 (0.57136)
$\beta_1 = 0.25$	9.00634 (0.57431)	9.01092 (0.57488)	9.01354 (0.57488)	9.01463 (0.57474)
$\beta_1 = 0.5$	9.00468 (0.56881)	9.02325 (0.57023)	9.03370 (0.57061)	9.03886 (0.57145)
$\beta_1 = 1$	8.99889 (0.57481)	9.07340 (0.57992)	9.11334 (0.58302)	9.13656 (0.58467)

Next, we check whether genotype uncertainty has any impact on the error variance estimates $\tilde{\sigma}^2$ under the models with $\beta_1 = 0, 0.25, 0.5$ and 1, focusing on the dosage approach. For this, we look at the mean and standard deviation of $\tilde{\sigma}^2$ calculated from 10000 Monte-Carlo samples, and compare the values with those at 0% uncertainty level, which correspond to the classical mean square error $\hat{\sigma}^2$. Table 2.7 verifies that the error variance estimates $\tilde{\sigma}^2$ are equally consistent for $\sigma^2 = 9$ as the estimates $\hat{\sigma}$. This result indicate that the smaller standard error of $\tilde{\beta}_1$ is mainly due to the larger variation in the imputed genotypes at higher uncertainty levels.

Finally, we verify the results in Section 2.2 with empirical evidence from our

simulations. In Table 2.8, we report the estimated bias and variance for $\tilde{\beta}_1$ based on the results presented in Figures 2.1 and 2.3. We also include the corresponding theoretical values in parentheses for comparison purposes. The values match closely, hence our empirical findings support the theoretical results.

Table 2.8: Bias and variance of $\tilde{\beta}_1$ estimated from 10000 Monte-Carlo samples for the cases with MAF = 0.1, $n = 500$ and $\sigma = 3$ under the dosage procedure. Given in parentheses are the corresponding theoretical values based on the results in Section 2.2.

		Uncertainty Level			
	$\tilde{\beta}_1$	0%	10%	20%	30%
$\beta_1 = 0$	Bias	-0.0051 (0)	-0.0031 (0)	-0.0007 (0)	-0.0039 (0)
	Variance	0.1098 (0.1100)	0.0821 (0.0831)	0.0670 (0.0670)	0.0649 (0.0651)
$\beta_1 = 0.25$	Bias	-0.0001 (0)	-0.0892 (-0.0885)	-0.1437 (-0.1420)	-0.1726 (-0.1713)
	Variance	0.1085 (0.1100)	0.0822 (0.0831)	0.0661 (0.0670)	0.0653 (0.0651)
$\beta_1 = 0.5$	Bias	-0.0019 (0)	-0.1786 (-0.1769)	-0.2881 (-0.2840)	-0.3424 (-0.3426)
	Variance	0.1131 (0.1100)	0.0841 (0.0831)	0.0675 (0.0670)	0.0658 (0.0651)
$\beta_1 = 1$	Bias	-0.0026 (0)	-0.3534 (-0.3539)	-0.5669 (-0.5680)	-0.6845 (-0.6852)
	Variance	0.1108 (0.1100)	0.0821 (0.0831)	0.0676 (0.0670)	0.0648 (0.0651)

2.4 Summary

In this chapter, we have reviewed the association test procedures for imputation-based GWAS, and investigated the impact of genotype uncertainty on the test results via a simulation study. We have also assessed the effect of other parameters on the study results.

From the simulation results, we conclude that the larger MAF, the bigger study size, the bigger magnitude of genetic effect and the smaller the error variance, the higher the

detection power of GWAS. On the other hand, uncertainty in imputed genotypes may significantly lower the detection power. In such cases, the dosage approach offers a better imputation method since it does not loose as much information as BG method. These findings are not new, and have been established by many (see, for instance, Spencer et al., 2009 and Acar and Sun, 2013. However, to our knowledge, the mechanism behind the power decrease due to genotype uncertainty has not been addressed in such detail. We have observed that the imputation uncertainty increases the genotype variation, which consequently, reduces the standard error of the genetic effect estimates. We have also provided distributional results for the slope estimates and the z-scores under genotype uncertainty, and shown that genotype uncertainty results in underestimation of the magnitude of the genetic effect size. Since the reduced standard errors do not fully compensate for the underestimated effect sizes, we obtain smaller z-scores and thus lower detection power in imputation based tests. These findings are insightful not only for imputation-based GWAS analyses, but also for imputation-based meta-analyses, which we address in Chapter 3 and Chapter 4.

Chapter 3

Fixed Effect Meta-analysis

GWAS have been used to find SNPs that underlie common complex diseases. These studies often require large sample sizes to detect moderate effect sizes of SNPs. In our simulation study in Chapter 2, we illustrated that despite a modest study size (e.g., $n = 500$), the detection power of a study (e.g., with a true effect size of 0.5), can still be underpowered. Most genetic association studies have adopted aggregation of GWAS summaries to boost power, a statistical procedure known as GWAS meta-analysis.

The two most common genome-wide meta-analysis methods are the fixed effect meta-analysis (FEM) and random effects meta-analysis (REM) though p-value based appraoches are also not uncommon in GWAS applications. Given k studies with sample sizes $n_1, n_2, \ldots, n_k$, the general meta-analysis model for the studies is a modified form of the single GWAS model in Equation (2.1) and is given by

$$Y_i^{(j)} = \beta_0^{(j)} + \beta_1^{(j)} G_i^{(j)} + \varepsilon_i^{(j)}, \qquad i = 1, \ldots, n_j, \quad j = 1, \ldots, k. \tag{3.1}$$

The model represents the phenotype of the i^{th} individual in the j^{th} study. In its most general form, the model in (3.1) allows the intercept and slope coefficients to vary across studies. In all our discussions, we assume that the studies share a common intercept, i.e., $\beta_0^j = \beta_0$. The slope coefficient is of the main interest for testing genetic associations

and considered as the effect size in GWAS meta-analysis approaches for quantitative traits. Note that for case-control studies a commonly used effect size is the log odds ratio, which requires a different approach than considered here (Haddock et al., 1998). As in Chapter 2, we are interested in testing

$$H_0 : \beta_1 = 0 \qquad vs. \qquad H_A : \beta_1 \neq 0.$$

We discuss the FEM methodolgy for both typed and imputed genotyes in Section 3.2.1. Therein, we propose two re-weighting schemes for imputation-based meta-analysis. In Section 3.2, we provide simulation results on the performance of FEM methods under genotype uncertainty. We summarize our findings in Section 3.3.

3.1 Methods

In FEM, the underlying assumption is that all studies in the meta-analysis have a common true genetic effect size β_1 and that the studies only vary because of sampling error within each study. Therefore, the model in (3.1) becomes

$$Y_i^{(j)} = \beta_0 + \beta_1 \, G_i^{(j)} + \varepsilon_i^{(j)}, \qquad i = 1, \ldots, n_j, \quad j = 1, \ldots, k.$$

As in the single GWAS case, preliminary analyses are necesary before meta-analysis. In FEM, the aim is to make sure that the studies to be combined are as homogenous as possible (Thompson et al., 2011). This simply means ensuring that the SNP are similar, MAFs are more or less similar, and the HWE holds for all the studies. Any study that either deviates from the HWE, has significantly different MAF or shows any other sign of originating from a totally different population is either removed or its summaries adjusted appropriately. This is done to avoid potential between-study variance.

In the following, we consider a setting with k studies all investigating the same quantitative trait. Our interest is to test a particular SNP for its association with this

trait. Among many possible scenarios, we focus on the two extreme ones, (1) all the k studies have the SNP genotyped, and (2) all the k studies have the SNP imputed. Note that in practice, one would typically encounter a situation where a SNP is genotyped in some studies and imputed in others. Understanding the two extreme cases is essential to address such in-between scenarios.

3.1.1 FEM at a Fully Typed SNP

Suppose all k studies have the SNP genotyped and report a fixed effect size estimate together with its standard error. These summaries can be combined in FEM using a weighted average of the estimates with properly chosen weights. To estimate the fixed effect β_1 from the reported estimates $\hat{\beta}_1^{(j)}$, $j = 1, \ldots, k$, we use the inverse-variance weighting technique of DerSimonian and Laird, 1986, where the study weights w_j's are inversely proportional to the square of the reported standard errors s_j^2. The weighted mean genetic effect size of k studies, $\hat{\beta}_{1FE}$, therefore given by

$$\hat{\beta}_{1FE} = \frac{\sum_{j=1}^{k} w_j \hat{\beta}_1^{(j)}}{\sum_{j=1}^{k} w_j},$$

where $w_j = 1/s_j^2$ and $\hat{\beta}_1^{(j)}$ is the effect size of the jth study. Since $\hat{\beta}_{1FE}$ is a weighted mean of unbiased estimators, it is also unbiased with mean

$$\mathrm{E}\left(\hat{\beta}_{1FE}\right) = \frac{\sum_{j=1}^{k} w_j \, \mathrm{E}\left(\hat{\beta}_1^{(j)}\right)}{\sum_{j=1}^{k} w_j} = \frac{\beta_1 \sum_{j=1}^{k} w_j}{\sum_{j=1}^{k} w_j} = \beta_1.$$

Although the quantities s_j's are estimated from data, they are customarily treated as known variances of the effect size estimates. Hence, the variance of the mean effect

size is the second root of the inverse of the sum of the weights i.e.,

$$\text{var}\left(\hat{\beta}_{1FE}\right) = \frac{\sum_{j=1}^{k} w_j^2 \text{ var}\left(\hat{\beta}_1^{(j)}\right)}{\left(\sum_{j=1}^{k} w_j\right)^2} = \frac{\sum_{j=1}^{k} w_j^2 \, 1/w_j}{\left(\sum_{j=1}^{k} w_j\right)^2}$$

$$= \frac{1}{\sum_{j=1}^{k} w_j}.$$

Thus, the standard error of $\hat{\beta}_{1FE}$ is

$$s_{FE} = \frac{1}{\sqrt{\sum_{j=1}^{k} w_j}}.$$

Note that $\hat{\beta}_1^{(j)}$'s are normally distributed either exactly because the phenotype is normally distributed, or approximately via CLT since sample sizes for GWAS are large. Hence, $\hat{\beta}_{1FE}$ also has a normal distribution,

$$\hat{\beta}_{1FE} \sim N\left(\beta_1, \ \frac{1}{\sum_{j=1}^{k} w_j}\right).$$

The hypothesis in (2.2) is tested using the statistic

$$z_{FE} = \frac{\hat{\beta}_{FE}}{s_{FE}}$$

If each study reports a z-score (or a p-value together with the direction of the effect size) instead of the effect size estimates, the statistic z_{FE} can be written in terms of the

individual studies' z-scores,

$$z_{FE} = \frac{\frac{\sum_{j=1}^{k} w_j \hat{\beta}_1^{(j)}}{\sum_{j=1}^{k} w_j}}{\left(\sum_{j=1}^{k} w_j\right)^{-0.5}} = \frac{\sum_{j=1}^{k} w_j \hat{\beta}_1^{(j)}}{\sqrt{\sum_{j=1}^{k} w_j}} = \frac{\sum_{j=1}^{k} \sqrt{w_j}\ \hat{\beta}_1^{(j)}/s_j}{\sqrt{\sum_{i=1}^{k} w_j}}$$
$$= \frac{\sum_{j=1}^{k} \sqrt{w_j}\ z_j}{\sqrt{\sum_{i=1}^{k} w_j}}. \tag{3.2}$$

Thus, from the distribution in (2.3) the distribution of z_{FE} becomes

$$z_{FE} \sim N\left(\beta_1 \sqrt{\sum_{j=1}^{k} w_j},\ 1\right).$$

Under null hypothesis z_{FE} has a standard normal distribution. We reject the null hypothesis when the p-value, $2\ \Pr(Z > |z_{FE}|)$, is less than the chosen level of significance, α.

3.1.2 FEM at an Imputed SNP

Use of imputed genotypes in the underlying studies of a genome-wide meta-analysis study affects the distribution of the effect size estimate of each study, which consequently affects the result of meta-analysis. The revised FEM model at an imputed SNP is

$$Y_i^{(j)} = \beta_0 + \beta_1\ \tilde{G}_i^{(j)} + \varepsilon_i^{(j)}, \qquad i = 1, \ldots, n_j,\ \ j = 1, \ldots, k.$$

Suppose these k studies report the estimates $\tilde{\beta}_1^{(j)}$'s along with their standard errors $\tilde{s}_j$'s, $j = 1, \ldots, k$, obtained as discussed in Section 2.2.2.

$$\widetilde{\beta}_{1FE} = \frac{\sum_{j=1}^{k} \tilde{w}_j \tilde{\beta}_1^{(j)}}{\sum_{j=1}^{k} \tilde{w}_j},$$

where $\tilde{w}_j = 1/\tilde{s}_j^2$. Its standard error is

$$\tilde{s}_{FE} = \frac{1}{\sqrt{\sum_{i=1}^{k} \tilde{w}_i}}.$$

Given that $\tilde{\beta}_1^{(j)}$'s are not unbiased for β_1, $\widetilde{\beta}_{1FE}$ is also not unbiased with expectation

$$\mathrm{E}\left(\widetilde{\beta}_{1FE}\right) = \frac{\sum_{j=1}^{k} \tilde{w}_j \, \mathrm{E}\left(\tilde{\beta}_1^{(j)}\right)}{\sum_{j=1}^{k} \tilde{w}_j} = \frac{\sum_{j=1}^{k} \tilde{w}_j \, \gamma_j}{\sum_{j=1}^{k} \tilde{w}_j} \, \beta_1,$$

where γ_j is obtained from (2.4). Assuming the reported values $\tilde{s}_j$'s are the true standard errors, the variance of $\tilde{\beta}_{1FE}$ is obtained as

$$\mathrm{var}\left(\widetilde{\beta}_{1FE}\right) = \frac{\sum_{j=1}^{k} \tilde{w}_j^2 \, \mathrm{var}\left(\tilde{\beta}_1^{(j)}\right)}{\left(\sum_{j=1}^{k} \tilde{w}_j\right)^2} = \frac{1}{\sum_{j=1}^{k} \tilde{w}_j}.$$

Hence, the distribution of $\widetilde{\beta}_{1FE}$ becomes

$$\widetilde{\beta}_{1FE} \sim N\left(\frac{\sum_{j=1}^{k} \tilde{w}_j \, \gamma_j}{\sum_{j=1}^{k} \tilde{w}_j} \, \beta_1, \; \frac{1}{\sum_{j=1}^{k} \tilde{w}_j}\right).$$

Once we obtain $\widetilde{\beta}_{1FE}$ and $\tilde{s}_{FE}$, the testing of the problem in (2.2) follows the same procedure in Section 3.1.1. The test statistic for the hypothesis (2.2) is

$$\tilde{z}_{FE} = \frac{\widetilde{\beta}_{1FE}}{\tilde{s}_{FE}},$$

which can be written in terms of the study z-scores, as in (3.2), as

$$\tilde{z}_{FE} = \frac{\sum_{j=1}^{k} \sqrt{\tilde{w}_j} \, \tilde{z}_j}{\sqrt{\sum_{i=1}^{k} \tilde{w}_j}},$$

The distribution of $\tilde{z}_{FE}$ is therefore,

$$\tilde{z}_{FE} \sim N\left(\frac{\sum_{i=1}^{k} \tilde{w}_j \, \gamma_j}{\sum_{i=1}^{k} \tilde{w}_j} \, \beta_1, \; 1\right),$$

which yields a standard normal distribution under the null hypothesis. Hence, we reject the null hypothesis if $2 \Pr(Z > |\tilde{z}_{FE}|) > \alpha$.

Adjusted Weighted Effect Size Estimator

In Chapter 2 we have seen that genotype uncertainty causes underestimation of the genetic effect size, resulting in an overall decrease in the detection power of a GWAS. Since meta-analysis is a summary of several GWAS, some of which have imputed genotypes, we expected in a similar observation because it is the individual studies' effect size estimates and their respective standard errors that are used to estimate the overall effect size. Therefore, the imputation effect of the individual studies is carried over to the meta-analysis.

Since the standard error of the effect size estimate drops with increasing uncertainty, the studies with poorly imputed genotypes are given bigger weights. Given that these studies are underpowered due to genotype uncertainty and have greater weights, they underpower a meta-analysis. Thus, an imputation-aware approach becomes essential to give the poorly imputed studies lesser weights in comparison to well imputed or fully typed studies. Motivated by the work of Zaitlen and Eskin, 2010, we propose using an imputation quality score (IQS) to adjust the study weights. An IQS measures the accuracy of the imputed genotypes in relation to the true ones.

In practice, the true genotype data are not available, therefore, calculation of the imputation accuracy has to rely solely on the imputed genotypes. Most genotype imputation software report an IQS that quantifies the imputation accuracy. Different sofwares employ varied proposed methodologies. One such methodology proposed by

Li and Abecasis, 2006 calculates an estimate of the Pearson correlation coefficient (r) between true genotypes and imputed ones using

$$\hat{r} = \frac{\hat{\sigma}^2_{\tilde{G}}}{2\ \tilde{q}(1-\tilde{q})},$$

where $\hat{\sigma}^2_{\tilde{G}}$ is the empirical variance of the imputed genotypes and $\tilde{q}$ is the MAF estimate. For well imputed genotypes $\hat{r}$ would be close to 1 and close to 0 for genotypes that are poorly imputed.

In our adjusted meta-analysis experiments, we directly use the Pearson correlation coefficient r to down weigh the imputed studies. We propose two approaches, one re-weighing the effect size estimates and another method which adjusts the z-scores.

Adjustment 1 For the jth study, let r_j be the correlation between the true genotypes and the imputed genotypes. The new weight for the study takes the form

$$w_j^* = r_j\ \tilde{w}_j.$$

This is a product of the correlation coefficient and the traditional weight thereby giving well imputed studies better weights. For k studies, the adjusted overall effect size estimate is

$$\beta^*_{1FE} = \frac{\sum_{j=1}^k w_j^* \tilde{\beta}_1^{(j)}}{\sum_{j=1}^k w_j^*}.$$

The variance of β^*_{1FE},

$$s^*_{FE}{}^2 = \frac{1}{\sum_{j=1}^k w_j^*}.$$

Though a better estimator of β_1, β^*_{1FE} is still not unbiased. Its expectation is

$$\mathrm{E}\left(\beta^*_{1FE}\right) = \frac{\sum_{j=1}^k w_j^*\ \mathrm{E}\left(\tilde{\beta}_1^{(j)}\right)}{\sum_{j=1}^k w_j^*} = \frac{\sum_{j=1}^k w_j^*\ \gamma_j}{\sum_{j=1}^k w_j^*}\ \beta_1$$

and the variance is

$$\text{var}\left(\beta_{1FE}^{*}\right) = \frac{\sum_{j=1}^{k} w_j^{*2}\ \text{var}\left(\tilde{\beta}_1^{(j)}\right)}{\left(\sum_{j=1}^{k} w_j^{*}\right)^2} = \frac{\sum_{j=1}^{k} r_j^2\ \tilde{w}_j^2/\tilde{w}_j}{\left(\sum_{j=1}^{k} r_j \tilde{w}_j\right)^2}$$

$$= \frac{\sum_{j=1}^{k} r_j^2\ \tilde{w}_j}{\left(\sum_{j=1}^{k} r_j \tilde{w}_j\right)^2}.$$

Therefore, β_{1FE}^{*} is normally distributed as shown

$$\beta_{1FE}^{*} \sim N\left(\frac{\sum_{j=1}^{k} r_j \tilde{w}_j\ \gamma_j}{\sum_{j=1}^{k} r_j \tilde{w}_j}\ \beta_1,\ \frac{\sum_{j=1}^{k} r_j^2\ \tilde{w}_j}{\left(\sum_{j=1}^{k} r_j \tilde{w}_j\right)^2}\right).$$

The new test statistic for the hypothesis (2.2) is

$$z_{FE}^{*} = \frac{\beta_{1FE}^{*}}{s_{FE}^{*}}$$

We reject the null hypothesis if $2\ \Pr(Z > |z_{FE}^{*}|) > \alpha$.

Adjusment 2 Alternatively, in our second adjustment, we re-weigh the individual z-scores in

$$\tilde{z}_{FE} = \frac{\sum_{j=1}^{k} \sqrt{\tilde{w}_j}\ \tilde{z}_j}{\sum_{i=1}^{k} \tilde{w}_j} = \frac{\sum_{j=1}^{k} \check{w}_j\ \tilde{z}_j}{\sqrt{\sum_{i=1}^{k} \check{w}_j^2}},$$

where $\check{w}_j = 1/\tilde{s}_j$. The new weights take the form

$$\check{w}_j^{*} = r_j\ \check{w}_j$$

Therefore, the new FEM statistic becomes

$$\check{z}_{FE} = \frac{\sum_{j=1}^{k} \check{w}_j^{*}\ \tilde{z}_j}{\sqrt{\sum_{i=1}^{k} \check{w}_j^{*2}}},$$

We reject the null hypothesis in (2.2) when $2\ \Pr(Z > |\check{z}_{FE}^{*}|) > \alpha$.

3.2 Simulation Study

In this section, we performed a simulation study to assess the impact of genotype imputation on FEM via simulations. We consider two scenarios: (1) all studies in meta-analysis have equal study sizes where $n = n_j = 500$, (2) studies in meta-analysis have unequal study sizes where $n_j \in \{200, 300, 500, 800, 1000\}$. In each scenario, we consider four genotype uncertainty levels per study. The levels are assumed to take 0%, 10%, 20% and 30% genotype uncertainties.

Under both scenarios, we generate three sets of meta-analysis, each with five studies. For each of these 15 studies, the genotype and phenotype data are simulated independently using the steps outlined in Section 2.2. The underlying model parameters are chosen as MAF $= 0.1$, $\beta_0 = 30$, $\sigma = 3$, and for the true effect size we set $\beta_1 = 0, 0.25$ and 0.5, the first reflecting the null hypothesis, and the other two the alternative. To generate phenotype data within a study, we use the same genotype data for the null and alternative models. For each setting, we obtain 10000 Monte-Carlo replicates by generating only the phenotype data.

Within each case of a particular uncertainty level we obtain association test results of the 15 studies using the dosage approach, and record the summary data consisting of effect size estimate, standard error, z-score and p-value.

3.2.1 Meta-analysis Results

Given the summary data for each particular scenario, we perform meta-analysis using the traditional FEM method as well as the two adjusted versions. The initial meta-analysis size is set to $k = 5$ studies, where we use the first set of five studies. We then increase the meta-analysis size to $k = 10$ and $k = 15$, including the second set to first

and the third set to the first two, respectively. This consideration enables us to properly track the impact of number of studies on the meta-analysis results.

Evaluation of Empirical Type 1 error rates

We obtain the empirical type 1 error rates at $\alpha = 0.01$ under the null hypothesis of no genetic effect, i.e., $\beta_1 = 0$. Table 3.1 presents the results for both FEM methods.

Table 3.1: Empirical type 1 error rates of the three FEM methods at $\alpha = 0.01$.

		Equal study sizes			Unequal study sizes		
		5	10	15	5	10	15
Traditional FEM	0%	0.0098	0.0089	0.0101	0.0096	0.0110	0.0104
	10%	0.0103	0.0093	0.0087	0.0108	0.0090	0.0106
	20%	0.0105	0.0098	0.0098	0.0101	0.0087	0.0116
	30%	0.0098	0.0097	0.0089	0.0087	0.0105	0.0097
Adjustment 1	0%	0.0098	0.0089	0.0101	0.0096	0.0110	0.0104
	10%	0.0029	0.0026	0.0019	0.0023	0.0031	0.0022
	20%	0.0010	0.0007	0.0005	0.0004	0.0005	0.0015
	30%	0.0001	0.0000	0.0000	0.0000	0.0001	0.0001
Adjustment 2	0%	0.0098	0.0089	0.0101	0.0096	0.0110	0.0104
	10%	0.0100	0.0092	0.0088	0.0108	0.0093	0.0105
	20%	0.0100	0.0099	0.0100	0.0105	0.0092	0.0119
	30%	0.0096	0.0101	0.0088	0.0091	0.0104	0.0098

As expected, at 0% uncertainty level, the adjusted versions under each case give the same value as the traditional FEM method, which does not deviate much from the nominal level $\alpha = 0.01$. However, as uncertainty increases, only the traditional FEM and the Adjustment 2 achieve the nominal type 1 error value of 0.01. For these two methods, we observe that genotype uncertainty does not significantly affect the rates.

Adjustment 1, on the other hand, consistently gives lower empirical type 1 error rates, which tend towards zero as uncertainty increases. This seems to be mainly due to the increased standard error of the adjusted overall effect estimate. In all three methods, we observe fairly similar results for the cases with equal and unequal study sizes considered in the simulation study.

Evaluation of Empirical Power

In power assessment, we considered two setups, one with $\beta_1 = 0.25$ and the other $\beta_1 = 0.5$. The results are summarized in Table 3.2 where both the traditional and adjustment 2 FEM approaches have been used. These summaries are for the cases where meta-analysis is performed on studies that fall with a specific genotype uncetainty level.

In all cases, we see that the empirical power increases as we increase the number of studies in a meta-analysis. Use of bigger true effect size in simulation also has a significant impact on the power since it results in larger overall effect size estimates which consequently give significant results. Meta-analysis with unequal study sizes has better empirical power compared to the one with equal study sizes. This is attributed to the larger average sample size of 560 individuals in the unequal study sizes compared to the 500 individuals in the equal study sizes case. Most importantly, for a fixed true effect size and fixed number of studies in a meta-analysis, we observe that power decreases as uncertainty increases. From Table 3.2, we see that our adjusted FEM method performs slightly better than the traditional FEM in all cases.

The two FEM methods considered are not significantly different as can be observed from the table, however, our adjusted FEM method performs better in most of the cases. The empirical power gain due to the adjustment is not significantly large, therefore, we can say that both methods perform equally when studies in a meta-analysis have a uniform level of uncertainty.

Table 3.2: Empirical power of the traditional and adjusted FEM methods at at $\alpha = 0.01$.

	FEM		Equal study sizes			Unequal study sizes		
			5	10	15	5	10	15
$\beta_1 = 0.25$	Traditional	0%	0.2051	0.4683	0.5966	0.2640	0.5515	0.6765
		10%	0.0976	0.2940	0.4187	0.1254	0.3490	0.4773
		20%	0.0566	0.2115	0.3208	0.0829	0.2608	0.3821
		30%	0.0322	0.1381	0.2279	0.0409	0.1640	0.2710
	Adjusted	0%	0.2051	0.4683	0.5966	0.2640	0.5515	0.6765
		10%	0.0985	0.2947	0.4185	0.1251	0.3491	0.4771
		20%	0.0570	0.2104	0.3205	0.0825	0.2613	0.3815
		30%	0.0326	0.1376	0.2280	0.0421	0.1645	0.2709
$\beta_1 = 0.5$	Traditional	0%	0.8154	0.9896	0.9995	0.9087	0.9983	1.0000
		10%	0.4910	0.8580	0.9699	0.6071	0.9317	0.9902
		20%	0.2860	0.6010	0.8236	0.4108	0.7161	0.9016
		30%	0.1168	0.2874	0.4767	0.1803	0.3906	0.5932
	Adjusted	0%	0.8154	0.9896	0.9995	0.9087	0.9983	1.0000
		10%	0.4947	0.8600	0.9704	0.6070	0.9318	0.9907
		20%	0.2864	0.6042	0.8250	0.4125	0.7194	0.9032
		30%	0.1169	0.2879	0.4797	0.1839	0.3951	0.5971

In practice the studies involved often have varying genotype uncertainty levels. To simulate this case, we take the first five studies at 0% uncertainty level and combine them with the studies in the other two sets at 0%, 10%, 20% and 30% uncertainty levels. This way, we consider the impact of adding imputed study results to meta-analysis when we initially have $k = 5$ typed study results. We include 0% uncertainty level to reflect the ideal cases with $k = 10$ and $k = 15$ typed studies. Table 3.3 is a summary of the empirical power for both traditional FEM method and the adjusted FEM for mixed uncertainty levels. We observe that empirical power of our adjusted meta-analysis method is better than the traditional FEM in all the cases that we considered.

Table 3.3: Empirical power of the traditional and adjusted FEM methods at $\alpha = 0.01$.

	FEM		Equal study sizes		Unequal study sizes	
			10	15	10	15
$\beta_1 = 0.25$	Traditional	10%	0.3609	0.4695	0.4361	0.5401
		20%	0.2850	0.3802	0.3716	0.4655
		30%	0.2045	0.2773	0.2756	0.3504
	Adjusted	10%	0.3680	0.4777	0.4447	0.5477
		20%	0.3080	0.4013	0.3949	0.4850
		30%	0.2491	0.3150	0.3260	0.3913
$\beta_1 = 0.5$	Traditional	10%	0.9508	0.9915	0.9851	0.9975
		20%	0.8767	0.9473	0.9369	0.9785
		30%	0.7586	0.8220	0.8596	0.9063
	Adjusted	10%	0.9571	0.9930	0.9866	0.9980
		20%	0.9051	0.9633	0.9588	0.9871
		30%	0.8430	0.8951	0.9258	0.9542

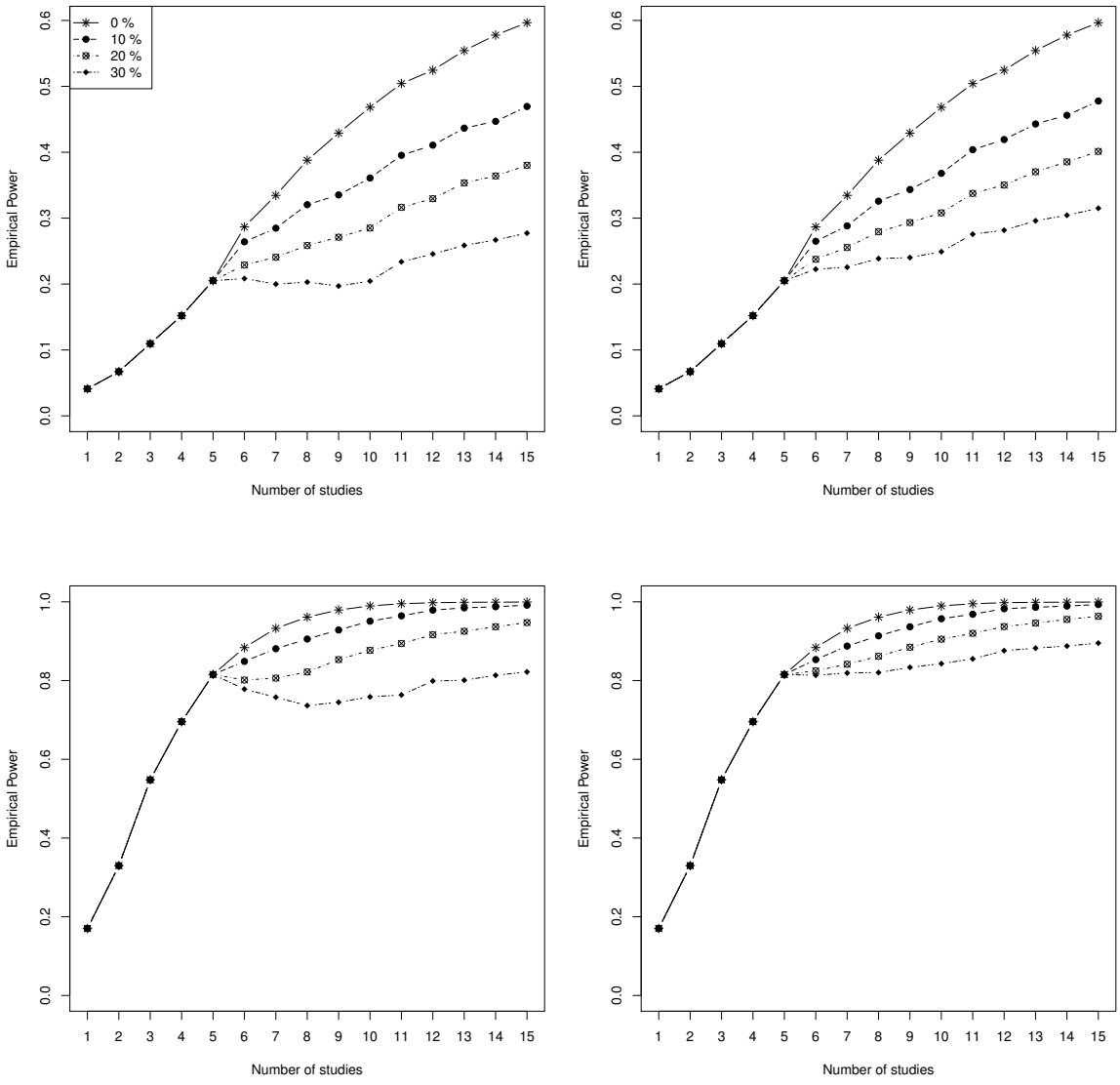
3.2.2 Adding Imputed Studies

In this section, we investigate the gradual increase in the meta-analysis power as we add studies one at a time for the case with equal study sizes. The first five studies have fully typed studies and from the the sixth study, we have four different cases through adding studies from the four uncertainty levels independently. We have considered both the standard FEM method and our adjusted FEM method.

Figure 3.1 shows empirical power progression as we add one study at a time in the equal sized studies case. For the first 5 studies where there is no genotype uncertainty, we obtain the empirical power for both FEM methods. With no genotype uncertainty, there is a steady increase of power as studies in the meta-analysis increase until the 15th study. With the other uncertainty levels, the increase is relatively lower. As uncertainty level increases, the rate of empirical power increment reduces relative to the uncertainty

level as shown by the branches where 0% is the top branch and 30% is the bottom branch. This observation is same for the two β_1's and for both FEM approaches.

Figure 3.1: Plots showing progression of empirical power with increase in number of studies using old FEM method (left panel) and adjusted FEM (right panel) for $\beta_1 = 0.25$ (top panel) and 0.5 (bottom panel).



However, for the uncertain meta-analysis cases where the adjusted FEM model has been used, the branching of the line plots is narrowing towards the line plot for

0% uncertainty. This indicates that the adjusted method performs better than the traditional FEM method. It is also important to note that, sometimes when an extra study is added in a meta-analysis, increase in power is not always guaranteed. This is quite clear on the plots for the traditional FEM (see Figure 3.1), where at 30% uncertainty level. For example, where $\beta_1 = 0.5$, we see a decrease in empirical power when the 6th, 7th and 8th studies are added, thereafter, there is a marginal increase in power.

To achieve empirical power of about 0.95 for the equal sized studies meta-analysis, where $\beta_1 = 0.5$ and the first five studies are fully typed, we need three more studies at 0% uncertainty level or five more studies at 10% uncertainty level for both the traditional FEM and adjusted FEM. However, at 20% uncertainty level, nine more studies are required for the adjusted FEM. The traditional FEM does not achieve this power even with 10 extra studies. Therefore, for a desired power of a meta-analysis study at an imputed SNP, more studies may be required for the traditional FEM in comparison to the adjusted FEM.

3.3 Summary

In this chapter, we have evaluated the impact of genotype uncertainty on FEM results, compared the accuracy of the traditional FEM and the adjusted FEM, and provided an empirical judgement on the number of imputation-based studies required to achieve a desired power level under both methods.

Initially, we proposed two adjustment methods, adjustment 1 and adjustment 2. From the simulation study results, we observed that adjustment 1 gives inaccurate type 1 error rates at 0.01 level of significance for both equal and unequal study meta-analyses. Therefore, we settled on the adjustment 2, which we refer to as adjusted FEM, in our subsequent analyses. From the results, we have verified that in both the traditional FEM

and adjusted FEM, type 1 error is not significantly affected by any or a combination of the simulation parameters. We have also ascertained that a bigger effect size and addition of more studies result in better detection power in a meta-analysis. We noted that a meta-analysis study that has a superior average study size of the individual studies give better power.

On comparing the two meta-analysis methods, in the case where the individual studies have varied uncertainty levels, we have observed that the adjusted FEM gives slighlty better power in comparison to the traditional FEM. In addition, the adjusted FEM often requires fewer imputation-based studies to achieve a certain power, especially when there is siginificant genotype uncertainty. However, when the uncertainty in the genotypes is high, a desired power may never be achieved because some of the added studies may decrease the meta-analysis power. This is more pronounced in the traditional FEM.

Chapter 4

Random Effects Meta-analysis

In Chapter 3, we have discussed the genome-wide meta-analysis model that assumes that the underlying studies have the same true effect size. Therefore, the studies estimate a common effect and the differences are as a result of sampling error. Occasionally, the estimated effect sizes may vary in the individual studies because of both the differences in the underlying true effects sizes of the studies and sampling error. The differences in the true effect sizes maybe a result of geographical differences of the studies or differences in studies populations. The estimated individual studies' effect sizes may also show these variances because of the different statistical methodologies employed in the studies. When the observed effect sizes of individual studies are different because of not only the sampling error but also other unexplained factors, we assume that *between-study variance/heterogeneity* is present in the studies.

When between-study variance is proved to be present among studies then random effects meta-analysis (REM) is advisable to be used over FEM model. Like FEM, REM also uses a weighted mean of estimated effect sizes of the primary studies to estimate the overall genetic effect size. However, in REM the weighting is determined in two steps, (1) first using the FEM weighting method to obtain the FEM overall effect size and (2) adding the between-study variance to the within study variance to obtain the

total variance, which is then used in the inverse variance weighting scheme to estimate the overall effect size.

In this chapter, we explore the findings of Li et al., 2012. Using more studies, we investigate the effect of genotype uncertainty on between-study heterogeneity and type 1 error in the test of association under REM. REM methodolgy for both typed and imputed genotypes is discussed in Section 4.1. Therein, we discuss methods for quantifying and testing the between-study heterogeneity as well as testing for association. In Section 4.2, we provide simulation results on the accuracy of REM method under genotype uncertainty. We summarize our findings in Section 4.3.

4.1 Methods

In the preliminary analysis of a meta-analysis study, if we discover between-study heterogeneity, we suspect that the individual studies' effect size estimates are from different distributions and hence prefer REM model for genome-wide meta-analysis. In REM, the model in (3.1) takes the form

$$Y_i^{(j)} = \beta_0 + \beta_1^{(j)} G_i^{(j)} + \varepsilon_i^{(j)}, \qquad i = 1, \ldots, n_j, \;\; j = 1, \ldots, k,$$

where $\beta_1^{(j)} \sim N(\beta_1, \tau^2)$. β_1 is the grand mean of the individual study true effect sizes and τ^2 is their variance.

However, before any meta-analysis is undertaken, we have to quantify and test for heterogeneity. For k studies addressing the same problem on the same quantitative trait on a particular SNP, we explore heterogeneity for both fully typed studies on one hand and purely imputed studies on the other. We are particularly interested in how meta-analysis study conclusions are affected in these two cases in the prescence of heterogeneity.

4.1.1 REM at a Fully Typed SNP

The first step in REM is to test for heterogeneity for the studies involved in a meta-analysis. We are therefore interested in testing

$$H_0 : \tau^2 = 0 \qquad vs. \qquad H_A : \tau^2 > 0, \tag{4.1}$$

The null hypothesis defines *no heterogeneity* and the alternative states some *heterogeneity*. In meta-analysis, a common way of testing the problem in 4.1 is to use *Cochrane Q test* (Cochran, 1950, 1954a).

Consider k studies such that $\hat{\beta}_1^{(1)}$, $\hat{\beta}_1^{(2)}$, $\ldots$, $\hat{\beta}_1^{(k)}$ are independently, normally distributed effect size estimators with mean β_1 and variance $\sigma_j^2 + \tau^2$, where σ_j^2 is the within study variance for jth study. The test statistic for testing heterogeneity is

$$Q = \sum_{j=1}^{k} w_j \left(\hat{\beta}_j - \hat{\beta}_{1FE}\right)^2,$$

where w_j and $\hat{\beta}_{1FE}$ are obtained as outlined in Section 3.1.1 and Q is χ^2 distributed with $k-1$ degrees of freedom. To obtain the mean of Q, we first need know the distribution of $\hat{\beta}_{1FE}$ in the presence of heterogeneity. Since each $\hat{\beta}_1^{(j)}$ are normally distributed with mean β_1, $\hat{\beta}_{FE}$ is also normally distributed with mean β_1 and variance,

$$\begin{aligned}
\text{var}\left(\hat{\beta}_{1FE}\right) &= \text{var}\left(\frac{\sum_{j=1}^{k} w_j \hat{\beta}_1^{(j)}}{\sum_{j=1}^{k} w_j}\right) = \frac{\sum_{j=1}^{k} w_j^2 \text{var}\left(\hat{\beta}_1^{(j)}\right)}{\left(\sum_{j=1}^{k} w_j\right)^2} \\
&= \frac{\sum_{j=1}^{k} w_j^2 \left(\sigma_j^2 + \tau^2\right)}{\left(\sum_{j=1}^{k} w_j\right)^2} = \frac{\sum_{j=1}^{k} w_j \sigma_j^2 + \tau^2 \sum_{j=1}^{k} w_i^2}{\left(\sum_{j=1}^{k} w_j\right)^2}.
\end{aligned}$$

Therefore, the mean of Q is

$$\mathrm{E}\left(Q\right) = \mathrm{E}\left(\sum_{j=1}^{k} w_j \left(\hat{\beta}_1^{(j)} - \hat{\beta}_{1FE}\right)^2\right) = \sum_{j=1}^{k} w_j \, \mathrm{E}\left(\left(\hat{\beta}_1^{(j)} - \beta + \beta - \hat{\beta}_{1FE}\right)^2\right)$$

$$= \sum_{j=1}^{k} w_j \left(\mathrm{E}\left(\hat{\beta}_1^{(j)} - \beta\right)^2 + \mathrm{E}\left(\beta - \hat{\beta}_{1FE}\right)^2 + 2\mathrm{E}\left(\hat{\beta}_1^{(j)} - \beta\right)\left(\beta - \hat{\beta}_{1FE}\right)\right).$$

Solving each part of the equation, we obtain

$$\mathrm{E}\left(Q\right) = k - 1 + \tau^2 \left(\sum_{j=1}^{k} w_j - \frac{\sum_{j=1}^{k} w_j^2}{\sum_{j=1}^{k} w_j}\right). \tag{4.2}$$

Under null hypothesis, Q is χ^2 distributed with degrees of freedom $k-1$ while under the alternative hypothesis it has a non-centrality parameter given by the last term in (4.2). We reject the null hypothesis in the problem (4.1) when $\Pr(\chi^2 > Q) > \alpha$.

Q statistic only helps in testing heterogeneity but can not quantify the severity of heterogeneity among the studies on a common scale. The degree of heterogeneity is often quantified by I^2 (Higgins et al., 2003) and is given by

$$I^2 = \frac{Q - (k-1)}{Q} \, 100\%.$$

I^2 ranges from 0% to 100% where 100% denotes extreme heterogeneity. Heterogeneity has been naively categorized into low, moderate and high for I^2 values 25%, 50% and 75%, respectively.

If we obtain significant results in the test of heterogeneity, we perform meta-analysis using the REM model. The overall observed effect size is estimated using the inverse-variance weighting technique where the variance is the sum of the between-study and within study variance estimates. Within study variance is reported in study summaries

and treated as the true value as in Section 3.1.1 while the between-study variance is estimated using expectation of Q-statistic in (4.2) and is given by

$$\hat{\tau}^2 = \frac{Q - (k-1)}{C},$$

where

$$C = \left(\sum_{j=1}^{k} w_j - \frac{\sum_{j=1}^{k} w_j^2}{\sum_{j=1}^{k} w_j} \right).$$

Therefore, the weight of the jth study in REM is

$$\bar{w}_j = \frac{1}{s_j^2 + \hat{\tau}^2},$$

which yields the overall effect size estimate

$$\hat{\beta}_{1RE} = \frac{\sum_{j=1}^{k} \bar{w}_j \hat{\beta}_1^{(j)}}{\sum_{j=1}^{k} \bar{w}_j}.$$

Given that $\hat{\beta}_{1RE}$ is a weighted mean of unbiased estimators, it is also unbiased.The derivation of its distribution follows the same procedure in the FEM model and is given by

$$\hat{\beta}_{1RE} \sim N \left(\beta_1, \ \frac{1}{\sum_{j=1}^{k} \bar{w}_j} \right).$$

The test statistic for the hypothesis in (2.2) is

$$z_{RE} = \frac{\hat{\beta}_{SE}}{s_{RE}},$$

where $s_{RE} = \sqrt{1/\sum_{j=1}^{k} \bar{w}_j}$. In the case where each study reports only the z-score the statistic z_{RE} can be written in terms of the z-scores,

$$z_{RE} = \frac{\sum_{j=1}^{k} \sqrt{\bar{w}_j}\, z_j}{\sqrt{\sum_{i=1}^{k} \bar{w}_j}},$$

with distribution

$$z_{RE} \sim N\left(\beta_1 \sqrt{\sum_{j=1}^{k} \bar{w}_j},\ 1\right).$$

Therefore, under the null hypothesis z_{RE} has a standard normal distribution. We reject the null hypothesis when the p-value, $2\ \Pr(Z > |z_{RE}|)$, is less than the significance level, α.

4.1.2 REM at an Imputed SNP

If all the studies in the genome-wide meta-analysis have imputed genotypes, the distribution of the overall effect size changes because the distribution of the individual studies' effect size estimators are affected as we had seen in Chapter 2. In the case of REM the model in (3.1) becomes

$$Y_i^{(j)} = \beta_0 + \beta_1^{(j)}\, \tilde{G}_i^{(j)} + \varepsilon_i^{(j)}, \qquad i = 1, \ldots, n_j,\ \ j = 1, \ldots, k.$$

Assume that all the k studies in REM report the effect size estimates $\tilde{\beta}_1^{(j)}$'s along with their standard errors $\tilde{s}_j$'s, $j = 1, \ldots, k$. We test the between-study heterogeneity as shown in Section 4.1.1, with the test statistic is given as

$$\tilde{Q} = \sum_{j=1}^{k} \tilde{w}_j \left(\tilde{\beta}_j - \tilde{\beta}_{1FE}\right)^2,$$

where $\tilde{w}_j = 1/\tilde{s}_j^2$ and $\tilde{\beta}_{1FE}$ is estimated as shown in Section 3.1.2. The I^2 value for imputed SNP based studies is

$$\tilde{I}^2 = \frac{\tilde{Q} - (k-1)}{\tilde{Q}} \, 100\%.$$

For significant heterogeneity results, the between-study variance is estimated in a similar way as in Section 4.1.1 and is given by

$$\hat{\tilde{\tau}}^2 = \frac{\tilde{Q} - (k-1)}{\tilde{C}},$$

where

$$\tilde{C} = \left(\sum_{j=1}^{k} \tilde{w}_j - \frac{\sum_{j=1}^{k} \tilde{w}_j^2}{\sum_{j=1}^{k} \tilde{w}_j} \right).$$

By inverse variance method the weights of the j imputed study becomes

$$\tilde{\tilde{w}}_j = \frac{1}{\tilde{s}_j^2 + \hat{\tilde{\tau}}^2}.$$

Therefore, the mean effect size estimate is

$$\hat{\tilde{\beta}}_{1RE} = \frac{\sum_{j=1}^{k} \tilde{\tilde{w}}_j \hat{\beta}_1^{(j)}}{\sum_{j=1}^{k} \tilde{\tilde{w}}_j},$$

and its standard error is

$$\tilde{s}_{RE} = \frac{1}{\sqrt{\sum_{i=1}^{k} \tilde{\tilde{w}}_i}}.$$

Since $\tilde{\beta}_1^{(j)}$'s are not unbiased for β_1, $\widetilde{\beta}_{1RE}$ is not unbiased as well. Its distribution is

derived the same way as that of FEM mean effect size in Section 3.1.2 and is given by

$$\widetilde{\beta}_{1RE} \sim N\left(\frac{\sum_{j=1}^{k} \tilde{\tilde{w}}_j \, \gamma_j}{\sum_{j=1}^{k} \tilde{\tilde{w}}_j} \, \beta_1, \; \frac{1}{\sum_{j=1}^{k} \tilde{\tilde{w}}_j}\right).$$

Once we obtain $\widetilde{\beta}_{1RE}$ and $\tilde{s}_{RE}$, similar procedure outlined in Section 3.1.1 is followed for testing of the problem in (2.2). The test statistic becomes

$$\tilde{z}_{RE} = \frac{\widetilde{\beta}_{1RE}}{\tilde{s}_{RE}},$$

which can also be written in terms of the individual studies' z-scores, as in (3.2),

$$\tilde{z}_{RE} = \frac{\sum_{j=1}^{k} \sqrt{\tilde{\tilde{w}}_j} \, \tilde{z}_j}{\sqrt{\sum_{i=1}^{k} \tilde{\tilde{w}}_j}}.$$

The distribution of $\tilde{z}_{RE}$ is,

$$\tilde{z}_{RE} \sim N\left(\frac{\sum_{i=1}^{k} \tilde{\tilde{w}}_j \, \gamma_j}{\sum_{i=1}^{k} \tilde{\tilde{w}}_j} \, \beta_1, \; 1\right),$$

Under the null hypothesis the distribution of $\tilde{z}_{FE}$ is standard normal distribution. We reject the null hypothesis if $2 \Pr(Z > |\tilde{z}_{RE}|) > \alpha$.

4.2 Simulation Study

In this section, we assess the impact of uncertainty on REM results via simulations. We particularly focus on the assessment of heterogeneity and the accuracy of REM results under the null hypothesis. As in the FEM simulations, we consider both equal and unequal study sizes, and form three sets, each consisting of 5 studies. Four genotype uncertainty levels, 0%, 10%, 20% and 30%, are included for each study. Within each

setting we choose τ^2 values that result in approximately, low, moderate and and high heterogeneity levels. The respective τ^2 values considered are, 0.04, 0.16 and 1 for low, moderate and high heterogeneity for a meta-analysis that consist of 5 equally sized studies.

Data are generated similar to as outlined in 3.2. The underlying model parameters are chosen as MAF $= 0.1$, $\beta_0 = 30$, $\sigma = 3$, and β_1 is set at 0 for the null hypothesis. The underlying true efffect size of each study is randomly obtained from normal distribution where mean is β_1 and variance τ^2. These true β_1 values are only generated once for studies with both equal and unequal study sizes. The generated true effect sizes at each heterogeneity level are shown in Table 4.1 for the case with $k = 5$ studies.

Table 4.1: True effect sizes for different τ^2 values in a five study genome-wide meta-analysis using the REM model under null hypothesis.

Parameter	Effect sizes				
$\tau^2 = 0.04$	0.0928	-0.0207	-0.0017	0.0633	0.2158
$\tau^2 = 0.16$	0.0598	-0.6365	-0.4337	0.6796	-0.4667
$\tau^2 = 1$	-0.2357	1.9711	0.2288	-0.6493	-0.1134

To generate phenotype data within a study, we use the same genotype data for all the different heterogeneity levels. For each setting, we obtain 10000 Monte-Carlo replicates by generating only the phenotype data. Within each case of a particular uncertainty level we obtain association test results of the 15 studies using the dosage approach, and record the summary data consisting of effect size estimate, standard error, z-score and p-value.

4.2.1 Meta-analysis Results

Having obtained the individual studies' summary results we perform test for heterogeneity, quantify the heterogeneity and perform meta-analysis as outlined in Section 4.2 within a specific scenerio using the REM method.

Evaluation of Heterogeneity

We investigate the impact of genotype uncertainty on the power and the degree of between-study heterogeneity under the alternative hypothesis of the problem in 4.1. We also compare our results to Li et al., 2012. Table 4.2 is a summary of the empirical power of the test of heterogeneity and average I^2 values under the null hypothesis.

We observe that for a fixed number of either equal or unequal sized studies, there is a consistent drop in the rejection rates and I^2 values as uncertainty increases at moderate and high heteregoneity levels. This is illustrated in Figure 4.1, under the scenario with equal sized studies. At low heterogeneity, power for the test of heterogeneity is very low and is around the nominal level $\alpha = 0.01$. The mean values of I^2 confirms this observation.

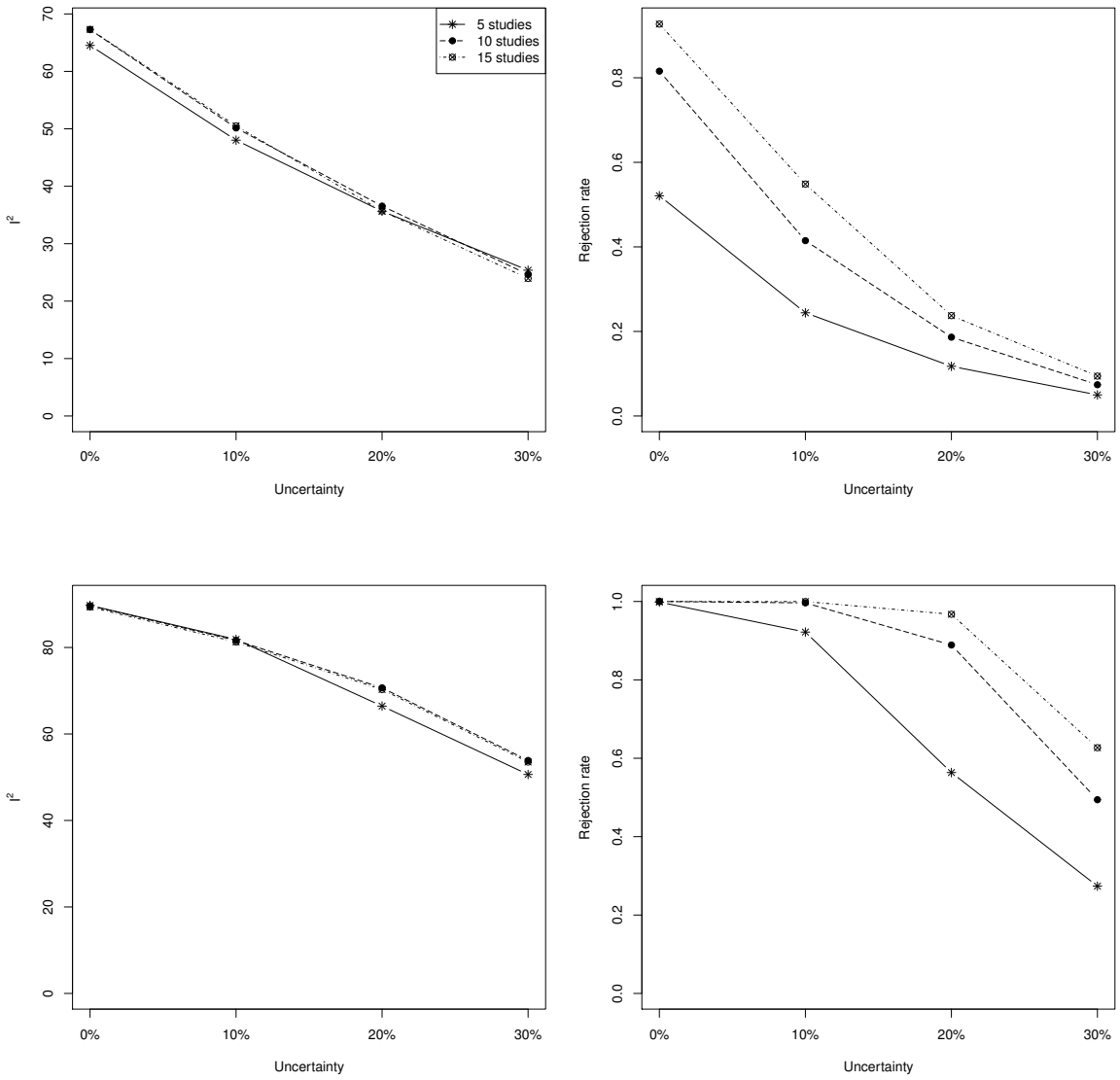
In the individual studies, the true effect sizes are necessarily not zero (see Table 4.1), hence as uncertainty increases, the estimated effect sizes tend towards zero and become smaller as explained in Chapter 2. Since they are converging to a common value as uncertainty increases, the size of Q-statistic diminishes and as consequence,empirical power and I^2 reduces. Therefore, contrary to Li et al., 2012 conclusion, we can say that uncertainty reduces the observed between-study heterogeneity.

Moreover, for moderate and high heterogeneity, there is an increase in significant results as the number of studies increases. Interestingly, in most cases, the degree of uncertainty decreases, which indicates that increasing the number of studies in REM may result in less degree of between-study heterogeneity. At low heterogeneity, there is

Table 4.2: Empirical power and mean I^2 for the three heterogeneity levels.

			Equal Study sizes			Unequal Study sizes		
			5	10	15	5	10	15
Empirical power	Low	0%	0.0150	0.0172	0.0194	0.0121	0.0157	0.0178
		10%	0.0133	0.0152	0.0130	0.0133	0.0147	0.0155
		20%	0.0117	0.0112	0.0137	0.0116	0.0117	0.0134
		30%	0.0117	0.0127	0.0128	0.0128	0.0129	0.0124
	Moderate	0%	0.5210	0.8157	0.9271	0.4919	0.7422	0.8876
		10%	0.2441	0.4148	0.5486	0.2091	0.3834	0.5089
		20%	0.1175	0.1864	0.2374	0.0986	0.1651	0.2165
		30%	0.0499	0.0739	0.0943	0.0536	0.0680	0.0962
	High	0%	0.9984	1.0000	1.0000	1.0000	1.0000	1.0000
		10%	0.9219	0.9964	0.9997	0.9651	0.9996	1.0000
		20%	0.5635	0.8889	0.9673	0.7355	0.9668	0.9972
		30%	0.2737	0.4942	0.6269	0.4306	0.6948	0.8584
I^2	Low	0%	16.0727	13.9683	12.9916	15.5243	13.2332	12.0453
		10%	14.8967	13.0009	11.6278	14.9602	12.7738	11.6033
		20%	14.5008	12.4415	11.0200	14.3414	12.1528	10.7104
		30%	13.8290	11.7150	10.4121	14.0167	11.8725	10.4809
	Moderate	0%	64.5148	67.2644	67.3141	63.1034	64.1932	64.8441
		10%	48.0203	50.1470	50.5094	45.0656	48.4187	48.8730
		20%	35.6481	36.5462	35.6875	33.4479	34.2283	34.2184
		30%	25.3793	24.6590	23.9501	26.2304	24.5396	24.5531
	High	0%	89.7298	89.6461	89.3477	92.4364	91.9183	91.6832
		10%	81.8684	81.6372	81.2655	84.5482	85.0946	84.34841
		20%	66.4446	70.6618	70.2822	73.7772	76.0299	76.0126
		30%	50.6289	53.8755	53.5116	60.2275	62.1787	63.2575

Figure 4.1: Plots showing impact of genotype uncertainty on between-study heterogeneity for the equal sized studies. The top panel plots illustrate the case of moderate heterogeneity and the bottom panel ones illustrate high heterogeneity. The left panel plots show average I^2 values and the right panel plots are for rejection rates at $\alpha = 0.01$.



an unclear behaviour of the rejection rates as the number of studies increases, this is because the studies have almost similar true effect sizes.

Evaluation of Empirical Type 1 error rates

For a nominal type 1 error of $\alpha = 0.01$, we obtain the empirical type 1 error rates under the null hypothesis of no genetic effect for low, moderate and high heterogeneity levels as shown in Table 4.3.

Table 4.3: Empirical type 1 error rates of the three heterogeneity levels under the REM method at $\alpha = 0.01$.

		Equal Study sizes			Unequal Study sizes		
		5	10	15	5	10	15
Low	0%	0.0153	0.0210	0.0282	0.0109	0.0160	0.0203
	10%	0.0117	0.0142	0.0173	0.0101	0.0116	0.0144
	20%	0.0095	0.0133	0.0142	0.0089	0.0098	0.0109
	30%	0.0080	0.0100	0.0119	0.0076	0.0078	0.0101
Moderate	0%	0.0062	0.0073	0.0123	0.0094	0.0163	0.0707
	10%	0.0098	0.0120	0.0270	0.0140	0.0223	0.0764
	20%	0.0088	0.0142	0.0232	0.0130	0.0207	0.0733
	30%	0.0073	0.0135	0.0189	0.0070	0.0107	0.0158
High	0%	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	10%	0.0006	0.0005	0.0008	0.0007	0.0001	0.0004
	20%	0.0034	0.0013	0.0028	0.0021	0.0018	0.0044
	30%	0.0043	0.0062	0.0096	0.0090	0.0113	0.0141

For a fixed number of studies, the meta-analysis empirical type 1 error rate decreases as uncertainty increases at low heterogeneity but increases at high heterogeneity. At moderate heterogeneity, type 1 error rate pattern is not quite clear, as in some cases, it is higher and in others lower than the nominal significance level. At low and moderate heterogeneity, we also observe a steady increase in type 1 error as the number of studies increases in meta-analysis. The number of false positives are higher at 0% levels of uncertainty. On the other hand, at high heterogeneity there is almost no rejection, especially at low uncertainty levels. These varying patterns make it quite difficult

to reach a common conclusion on the impact of genotype uncertainty. However, it is obvious that the DerSimonian and Laird, 1986 approach fails to produce accurate results even when there is no uncertainty.

4.3 Summary

In this chapter, we have assessed the impact of genotype imputation on REM results using a simulation study. First, we have investigated the heterogeneity then type 1 error and lastly compared our results to Li et al., 2012.

From the results, we can conclude that genotype uncertainty reduces the degree of between-study heterogeneity. Our findings are contrary to Li et al., 2012, who deduced that "imputation may cause between-study heterogeneity, especially when imputation was performed in some but not all of the sub-samples". We have also observed that as the number of studies increases, more significant results are observed, however, the degree of heterogeneity drops. This may explain Li et al., 2012 observation since there were only three studies and as we have seen, fewer studies often show higher degree of heterogeneity.

The simulation results have also shown that in association testing, type 1 error of REM is either underestimated or overestimated depending on the number of studies in the meta-analysis or genotype uncertainty level. As a result, the test results obtained may be questionable.

Chapter 5

Conclusion

Our study was set out to assess the impact of genotype uncertainty on the genome-wide meta-analysis. In Chapter 2 we have given the empirical findings for a single GWAS while in Chapter 3 and Chapter 4, the empirical findings are for FEM and REM methods respectively. In FEM we have investigated how uncertainty affects both type 1 error and power of a meta-analysis and in REM, we have evaluated the impact of uncertainty on between-study heterogeneity and type 1 error.

From our empirical results, we have observed that larger the sample size, bigger effect size and higher MAF give higher detection power for a single-SNP GWAS. In contrast, more variation in the trait lowers power. Also, increase in genotype uncertainty results in underestimation of the effect size and at the same time reduces the within study variance of a GWAS. Overall, this reduces the detection power of the test of association, since the reduced within study variance does not fully compensate for the reduced estimated effect size. In GWAS we also noted that dosage method of imputation performs better than the BG method.

In FEM, we have observed that increasing number of studies and having a bigger true effect size in a meta-analysis increase the detection power. Like GWAS, genotype uncertainty reduces power. Our adjusted FEM performs almost equally to the traditional

FEM in cases where the studies to be combined have a uniform uncertainty level. The adjusted FEM, however, gives a slighlty better power than the traditional FEM where the first five studies are fully typed and the added studies have a uniform uncertainty level. To achieve a desired power, more studies are required for an imputation-based meta-analysis where the first five studies are fully typed and the added studies have imputed genotypes. This is achievable up to a certain uncertainty level because studies with higher levels of uncertainty tend to reduce the meta-analysis study power (see Figure 3.1).

In REM, we observed that type 1 error rates are not accurately estimated due to the extra between-study heterogeneity factor. There have been previous attempts to solve this problem on type 1 error (See Hartung and Knapp, 2001, Sidik and Jonkman, 2002, Sidik and Jonkman, 2003, Han and Eskin, 2011 and IntHout et al., 2014). Through simulation study, we also observed that in the test of heterogeneity, uncertainty reduces the number of significant results as well as the magnitude of heterogeneity in instances where heterogeneity is moderate or high. In addition, more studies give more significant results with lower degree of heterogeneity.

In the FEM model, our method has not fully compensated for the power lost due to uncertainty. Although, our simulations results have shown that it is better than the traditional method, we recommend that a further research work to be done to improve this method especially where studies in the meta-analysis have more than two levels of heterogeneity. REM is more challenging due to the extra factor, therefore, before conducting a power analysis, we recommend further investigation on the type 1 error. Sound conclusions under REM would be possible only after an accurate procedure is obtained.

Appendix

The following is part of the r-code used in data simulation and analyses.

```
#########################################################################
## FUNCTIONS
#########################################################################
rdirichlet <- function(n.d, alpha1) {
#########################################################################
# This function generates data under dirichlet distribution with
# parameter vector alpha1
#########################################################################
    l <- length(alpha1)
    x <- matrix(rgamma(l*n.d,alpha1),ncol=l,byrow=TRUE)
    sm <- x%*%rep(1,l)
    return(x/as.vector(sm))
  }

###############################################
# simple check:
###############################################
# p= rdirichlet(100, c(0.1,0.1,0.8))
# colSums(p)/100
###############################################

#########################################################################
add.func <- function(p.vec){
#########################################################################
# calculates dosage value for a vector of genotype probabilities
#########################################################################
```

```
if(length(p.vec)!=3){stop("Parameter vector should be 3-dimensional")}

G.dosage <- sum(p.vec*c(0, 1, 2))
return(G.dosage)
}

##############################################
# simple check:
##############################################
# add.func(p) # gives error
# d = apply(p, 1, "add.func")  # OK
##############################################

#############################################################################
Geno.data <- function(n, q, unc.lev = 0){
#############################################################################
# simulates genotype data at a given MAF and with different uncertainty
# levels
#############################################################################

# number of uncertainity levels
L.unc <- length(unc.lev)

# population genotype probabilities assuming H-W equilibrium.
p <- 1 - q
pg.p <- c(p^2, 2*p*q, q^2)

# number of individual studies in a meta-analysis
k <- length(n)
```

```
# Defining the empty objects for mega-probabilities and mega-dosage
# values
mega.genoprob <- vector("list",k)
mega.dosage   <- vector("list",k)

for(i in 1:k){

# genotype group sizes for the ith study
  geno.size <- rmultinom(1, n[i], pg.p)

# Defining the empty objects for the ith study
  genoprob <- array(NA, c(n[i], 3, L.unc),
                  dimnames = list(1:n[i], c("G0", "G1", "G2" ),
                      paste(unc.lev*100, "%", "uncert", sep = " ")))

  dosage <- matrix(data = NA, nrow = n[i], ncol = length(unc.lev),
                   dimnames = list(1:n[i], paste(unc.lev*100, "%",
                                         "uncert", sep = " ")))

for(j in 1:L.unc){
genoprob[,,j] <- round(rbind(rdirichlet(geno.size[1],
c(1 - unc.lev[j],(unc.lev[j])/2, (unc.lev[j])/2)),
rdirichlet(geno.size[2], c((unc.lev[j])/2, 1 - unc.lev[j],
(unc.lev[j])/2)), rdirichlet(geno.size[3], c((unc.lev[j])/2,
(unc.lev[j])/2, 1 - unc.lev[j]))), 3)
dosage[,j] <- apply(genoprob[,,j] , 1, add.func)
}

mega.genoprob[[i]] = genoprob
```

```
mega.dosage[[i]] = dosage
}

# Returning results
all.data <- list("Dosage"= mega.dosage, "GenoProb" = mega.genoprob)
return(all.data)
}

##########################################
# simple check:
##########################################
# gdat = Geno.data(n=c(10,15), q=0.2, unc.lev = c(0, 0.1))
##########################################

#########################################################################
Pheno.data_FE <- function(n, param, G.size){
#########################################################################
# simulates phenotype data under Fixed effect model (FEM) given the
 # true genotype vector
#########################################################################
# n: vector of sample sizes
# param: c(Beta.0, Beta.1, sigma)
# G.size: matrix of genotype group sizes for all studies
#########################################################################

if(length(param)!=3){stop("Parameter vector should have dimension 3!")}

# Renaming the parameters
Beta.0 = param[1]
Beta.1 = param[2]
```

```
Sigma = param[3]

# number of studies in the meta-analysis
k <- length(n)
if(k==1){G.size= t(as.matrix(G.size)) }

# Defining an empty list

mega.pheno <-vector("list",k)

  for(i in 1:k){
  y0<-rnorm(G.size[i,1], mean= Beta.0, sd=Sigma)
  y1<-rnorm(G.size[i,2], mean= Beta.0+  Beta.1, sd=Sigma)
  y2<-rnorm(G.size[i,3], mean= Beta.0+2*Beta.1, sd=Sigma)
  mega.pheno[[i]] <- c(y0, y1, y2)
}

return(list("Phenotypes"= mega.pheno))
}

############################################
# simple check:
############################################
# G.size = matrix(NA, nrow=k,ncol=3)
# for(i in 1:k){G.size[i,] = colSums(gdat$GenoProb[[i]][,,1])}
# Pheno.data_FE(n=c(10,15), param=c(100,2,5), G.size)
# Pheno.data_FE(n=10, param=c(100,2,5), G.size[1,])
############################################

#########################################################################
```

```
Pheno.data_RE <- function(n, Beta.0, Beta.1, Sigma, G.size){
#########################################################################
# simulates phenotype data under Random effects model (REM) given the
# true genotype vector
#########################################################################
# n: vector of sample sizes
# Beta.0: fixed intercept
# Beta.1: random slope values from N(mean= B1, sd=tau)
# Sigma: error variance
# G.size: matrix of genotype group sizes for all studies
#########################################################################

# number of studies in the meta-analysis
k <- length(n)

if(length(Beta.1)!=k){stop("Provide a vector of slope coefficients")}
if(k==1){G.size= t(as.matrix(G.size)) }
# Defining an empty list

mega.pheno <-vector("list",k)

for(i in 1:k){
  y0<-rnorm(G.size[i,1], mean= Beta.0, sd=Sigma)
  y1<-rnorm(G.size[i,2], mean= Beta.0+  Beta.1[i], sd=Sigma)
  y2<-rnorm(G.size[i,3], mean= Beta.0+2*Beta.1[i], sd=Sigma)
  mega.pheno[[i]] <- c(y0, y1, y2)
}

return(list("Phenotypes"= mega.pheno))
```

```
}

##########################################
# simple check:
##########################################
# bvals = rnorm(k, mean=2, sd= 3)
# Pheno.data_RE(n=c(10,15), Beta.0=100, Beta.1=bvals, Sigma=5,
# G.size)
##########################################

#############################################################################
test.stats <- function(Phenotypes, Genotypes){
#############################################################################
# performs the dosage test for a single study
#################################################################

Genotypes= as.matrix(Genotypes)
Dimen= ncol(Genotypes)

test.results <- matrix(data=NA, nrow = 4, ncol = Dimen, dimnames =
    list(c("Beta1.estimate", "Beta1.error", "t-statistic", "p-value"),
           paste("Level", 1:Dimen, sep = " ")))

for(i in 1:Dimen){
  test.result <- summary(lm(Phenotypes~Genotypes[,i]))
  test.results[1,i] <-test.result$coefficients[2,1]
  test.results[2,i] <-test.result$coefficients[2,2]
  test.results[3,i] <-test.result$coefficients[2,3]
  test.results[4,i] <-test.result$coefficients[2,4]
}
```

```
return(test.results)
}

##########################################
# simple check:
##########################################
# G = gdat$Dosage[[1]]
# Y = Pheno.data_FE(n=10, param=c(100,2,5), G.size[1,])$Phenotypes[[1]]
# test.stats(Y,G)
##########################################

#########################################################################
meta_loop.test <- function(Phenotypes, Genotypes){
#########################################################################
# performs the dosage test for each study in a meta-analysis
#################################################################

k <- length(Phenotypes)
Dimen <- dim(Genotypes[[1]])[2]
all.study_sum <- vector("list", Dimen)

for(j in 1:Dimen){
    all.study_sum[[j]] <- matrix(data = NA, nrow = 4, ncol = k,
                    dimnames=list(c("Beta1.estimate", "Beta1.error",
                                    "t-statistic", "p-value")))
    colnames(all.study_sum[[j]]) <- paste("Study", 1:k, sep = " ")
}
```

```
for(i in 1:k){
# Summaries for the i-th study of k-studies
    test.res <- test.stats(Phenotypes[[i]], Genotypes[[i]])

# Loop for within study summaries for all the uncertainity levels
for(j in 1:Dimen){
    all.study_sum[[j]][,i] <- test.res[,j]
    }
   }

return(all.study_sum)
}

############################################
# simple check:
############################################
# G = gdat$Dosage
# Y = Pheno.data_FE(n=c(10,15), param=c(100,2,5), G.size)$Phenotypes
# meta_loop.test(Y,G)
############################################

#########################################################################
cochran.q_test <- function(b, se){
#########################################################################
# calculates the Cochran's Q-test statistic and reports p-value and
# I.square for heterogeneity test
#################################################################
# b: vector of slope estimates for study samples
# se: vector of standard errors of slope estimates
```

```
#################################################################

if(length(b) < 2){stop("Length of b must be greater than 1")}
  else if(length(se) < 2){stop("Length of se must be greater than 1")}
  else if(length(b)!=length(se) ){warning("Length of se must be equal
                                          to length of b")}

# Obtain weights, length of w/se and the degrees of freedom
  w<-1/(se^2)
  k<-length(w)
  df<-k-1

# Weighted mean
  b_w<- sum(w*b)/sum(w)

# Q-statistics
  q.stat<-sum(w*(b-b_w)^2)

# p-value
  value <-  1 - pchisq(q.stat, df, ncp = 0)

  # I^2 calculation
  I.sq <- ((q.stat - df)/q.stat)*100
  if(I.sq <0) I.sq  = 0

  results<-list("Weighted.beta_est"= b_w, "Q_statistic"= q.stat,
                "P_value"= value, "I_square" = I.sq )

  return(results)
  }
```

```
##########################################
# simple check:
##########################################
# b = meta_loop.test(Y,G)[[1]][1,]
# se = meta_loop.test(Y,G)[[1]][2,]
# cochran.q_test(b,se)
##########################################

#########################################################################
meta_func <- function(b, se, model, Qstat = FALSE){
#########################################################################
# performs meta-analysis for FEM and/or REM
#################################################################
# b: vector of slope estimates for study samples
# se: vector of standard errors of slope estimates
# model: Fixed Effect Model or Random Effect Model ("FE" or "RE")
# Qstat: heterogeneity test (TRUE or FALSE)
#################################################################

# reports results under FE and RE models
if(missing(model)){model= "BOTH"}

# Results from the test of heterogeneity.
Cochran.Q = cochran.q_test(b, se)
pval_Q <-  Cochran.Q$P_value # Cochran Q-test p-value

k<-length(b) # Number of studies in the meta-analysis
w <- (se)^(-2)   # Study weights
```

```
#################################################################
# First part: Fixed effect model
#################################################################
  if(model == "FE" | model=="BOTH"){

          b.fe <-  Cochran.Q$Weighted.beta_est    # Weighted beta estimate
          V.fe <-  1/sum(w)    # Variance of weighted beta estimate

    # Test statistic
    z.fe <- b.fe/sqrt(V.fe)    # Z-score statistic

    # p-values
    p.val_fe <- 2*(1-pnorm(z.fe,mean=0,sd=1))   # Two sided test

    # Where p.val_fe > 1
     if(p.val_fe > 1) p.val_fe = 2 - p.val_fe

    # Results for fixed effects

    results.fe<- list("Weighted_beta.estim" = b.fe,
    "Error" = sqrt(V.fe), "Z_Score" = z.fe, "Beta.P_value" = p.val_fe)
  }

#################################################################
# Second part: Random effects model
#################################################################

  if(model == "RE" | model=="BOTH"){

   # Calculating tau.sq_stat:
    q.stat<- Cochran.Q$Q_statistic # Q-statistics
```

```
c<-sum(w)-(sum(w^2))/sum(w)  # Weighting factor
df <- k-1 # degrees of freedom

tau.sq_stat <- (q.stat-df)/c
# In case tau.sq_stat <0, return 0
if(tau.sq_stat<0) tau.sq_stat = 0

# Total variance
var.t<- 1/w + tau.sq_stat

# RE weights
w.re <- 1/var.t

# Weighted beta under random effects
b.re<- sum(w.re*b)/sum(w.re)

# Variance of weighted mean
V.re <- 1/sum(w.re)

# Test statistic
z.re <- b.re/sqrt(V.re) # Z-score statistic

# p-values
p.val_re <- 2*(1-pnorm(z.re,mean=0,sd=1)) # Two sided test
# Where p.val_re > 1
if(p.val_re > 1) p.val_re = 2 - p.val_re

# Results for random effects
results.re<- list("Weighted_beta.estim" = b.re,
"Error" = sqrt(V.re), "Z_Score" = z.re, "Beta.P_value" = p.val_re,
```

```
                    "Tau.Sq"= tau.sq_stat)
  }

#################################################################
#  Heterogeneity results
#################################################################
 if (Qstat == TRUE){
         results.q<- list("Q_statistics" = Cochran.Q$Q_statistic,
    "Q_stat.P_value" = pval_Q, "I_square" = Cochran.Q$I_square)
    }

#################################################################
# Results to return:
#################################################################
# Returns fixed effect results
#################################################################

if (model == "FE"){
    if(Qstat == FALSE){ # Qstat is not required
return(results.fe)
    }else{  # Qstat is required
      results.nfe<-list("FE_results" = results.fe,
                        "Q.statistics_results" = results.q)
      return(results.nfe)
    }}

# Returns random effect results
#################################################################

if (model == "RE"){
```

```
    if(Qstat == FALSE){ # Qstat is not required
      return(results.re)
    }else{ # Qstat is required
      results.nre<-list("RE_results" = results.re,
                        "Q.statistics_results" = results.q)
      return(results.nre)
    }}

# Returns both FE and RE results
#################################################################

 if (model=="BOTH"){
    if(Qstat == FALSE){ # Qstat is not required
      results.nullF <- list("FE_results" = results.fe,
                            "RE_results" = results.re )
      # output results
      return(results.nullF)
      }else{ # Qstat is required
      results.nullT <- list("FE_results" = results.fe,
      "RE_results" = results.re,
      "Q.statistics_results" = results.q )

      return(results.nullT)
    }}}
###############################################
# simple check:
###############################################
# meta_func(b, se, Qstat = TRUE)
###############################################
```

Bibliography

Acar, Elif F. and Sun, Lei. A generalized kruskal-wallis test incorporating group uncertainity with application to genetic association studies. *Biometrics*, 69:427 – 435, 2013.

Altmüller, Janine, Palmer, Lyle J., and Wjst, Matthias. Genomewide scans of complex human diseases: True linkage is hard to find. *American Journal of Human Genetics*, 69(6):936 – 950, December 2001.

Altshuler, David and Daly, Mark. Guilt beyond a reasonable doubt. *Nature Genetics*, 39(7):813 – 815, 2007.

Aulchenko, Yurii S, Struchalin, Maksim V, and van Duijn, Cornelia M. Probabel package for genome-wide association analysis of imputed data. *BMC Bioinformatics*, 11:1 – 10, 2010.

Balding, David J. A tutorial on statistical methods for population association studies. *Nature Review Genetics*, 7(10):781 –791, 10 2006.

Begum, Ferdouse, Ghosh, Debashis, Tseng, George C., and Feingold, Eleanor. Comprehensive literature review and statistical considerations for gwas meta-analysis. *Nucleic Acids Research*, 40(9):3777 – 3784, 2012.

Borenstein, Michael, Hedges, Larry V., Higgins, Julian P.T., and Rothstein, Hannah R. A basic introduction to fixed-effect and a basic introduction to fixed-effect and

random-effects models for meta-analysis. *Research Synthesis Methods*, 1:97 – 111, 2010.

Browning, B. and Browning, S. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American Journal of Human Genetics*, 84:210–223, 2009.

Bush, W. S and Moore, J. H. Chapter 11: Genome-wide association studies. *PLoS Computational Biology*, 8(12), 2012.

CC, Cochrane Collaboration. URL http://www.cochrane.org.

Clarke, Geraldine M, Anderson, Carl A, Petterson, Fredrik H, Cardon, Lon R, Morris, Andrew P, and Zondervan, Krina T. Basic statistical analysis in genetic case-control studies. *Nature Protocols*, 6(2):121 – 133, 2011.

Cochran, William G. Problems arising in the analysis of a series of similar experi-ments. *Journal of the Royal Statistical Society*, 4:102–118, 1937.

Cochran, William G. The comparison of percentages in matched samples. *Biometrika*, 37:256–266, 1950.

Cochran, William G. The combination of estimates from different experiments. *Biometrics*, 10:101 – 129, 1954a.

Cochran, William G. Some methods for strengthening the common chi-squared tests. *Biometrics*, 10(4):417–451, 1954b.

de Bakker, Paul I.W., Ferreira, Manuel A.R., Jia, Xiaoming, Neale, Benjamin M., Raychaudhuri, Soumya, and Voight, Benjamin F. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Human Molecular Genetics*, R7(R2):R122–R128, 2008.

DerSimonian, Rebecca and Laird, Nan. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7:177 – 188, 1986.

Edwards, A. W. F. G. h. hardy (1908) and hardy–weinberg equilibrium. *Genetics*, 179 (3):1143 – 1150, 2008.

Egger, Matthias, Smith, G. Davey, Schneider, M., and Minder, C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 1997.

Evangelou, Evangelos and Ioannidis, John P. A. Meta-analysis methods for genome-wide association studies and beyond. *Nature Review Genetics*, 14:379–389, 2013.

Fisher, Ronald Aylmer. On the interpretation of χ^2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, 85(1):87–94, 1922.

Fisher, Ronald Aylmer. *Statistical Methods for Research Workers*. Oliver Oliver and Boyd, 1932.

Franke, Andre, McGovern, Dermot P B, and et. al. Genome-wide meta-analysis increases to 71 the number of confirmed crohn's disease susceptibility loci. *Nature Genetics*, 45:1118–1125, 2010.

Galbraith, R.F. Graphical display of estimates having differing standard errors. *Technometrics*, 30(3):271–288, 1988.

Haddock, C. Keith, Rindskopf, David, and Shadish, William R. Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods*, 3(3):339 – 353, 1998.

Haines, Jonathan L., Hauser, Michael A., Schmidt, Silke, Scott, William K., Olson, Lana M., Gallins, Paul, Spencer, Kylee L., Kwan, Shu Ying, Noureddine, Maher,

Gilbert, John R., Schnetz-Boutaud, Nathalie, Agarwal, Anita, Postel, Eric A., and Pericak-Vance, Margaret A. Complement factor h variant increases the risk of age-related macular degeneration. *Science*, 308(5720):419 – 421, 15 April 2005.

Hall, Neil. Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology*, 210(9):1518–1525, 2007.

Han, Buhm and Eskin, Eleazar. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *The American Society of Human Genetics*, 88(586–598), 2011.

HapMap, The International HapMap Consortium. A haplotype map of the human genome. *Nature*, 437(7063):1299–320, 2005.

HapMap, The International HapMap Consortium. A second generation human haplotype map of over 3.1 million snps. *Nature*, 449(7164):851 – 861, 2007.

HapMap, The International HapMap Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311):52–58, 2010.

Hartung, Joachim and Knapp, G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine*, 20(12):1771–1782, 2001.

Higgins, Julian P T, Thompson, Simon G, Deeks, Jonathan J, and Altman, Douglas G. Measuring inconsistency in meta-analyses. *British Medical Journal*, 2003.

Howie, B. N., Donnelly, P., and Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*, 2009.

IHGSC, International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 2001.

IntHout, Joanna, Ioannidis, JP, and Borm, George F. The hartung-knapp-sidik-jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard dersimonian-laird method. *BMC Medical Research Methodology*, 14(25), 2014.

Ioannidis, JP, Ntzani, EE, Trikalinos, TA, and Contopoulos-Ioannidis, DG. Replication validity of genetic association studies. *Nature Genetics*, 29(3):306 109, 2001.

Kavvoura, Fotini K. and Ioannidis, John P. A. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Human Genetics*, 123(1 - 14), 2008.

Kruskal, W. H. and Wallis, W. A. Use of ranks in one-criterion variance analysis. *Journal of American Statistical Association*, 47(583 - 621), 1952.

Li, Jian, fang Guo, Yan, Pei, Yufang, and Deng, Hong-Wen. The impact of imputation on meta-analysis of genome-wide association studies. *PLoS ONE*, 7, 2012.

Li, Y and Abecasis, G. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *American Journal of Human Genetics*, 2006.

Li, Yun, Willer, Cristen, and Abecasis, Gonçalo. Genotype imputation. *Annual Review of Genomics and Human Genetics*, 10:387 – 406, 2009.

Light, Richard J. *Summing up : the science of reviewing research.* Cambridge, Mass. : Harvard University Press, 1984.

Liptak, T. On the combination of independent tests. *Magyar Tud. Akad. Mat. Kutato Int.*, 3:171 – 197, 1958.

Marchini, Jonathan and Howie, Bryan. Genotype imputation for genome-wide association studies. *NATURE REVIEWS*, 11(499 - 511), 2010.

Marchini, Jonathan, Howie, Bryan, Myers, Simon, McVean, Gil, and Donnelly, Peter. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics*, 39:906 – 913, 2007.

Mosteller, F. and Bush, R. R. Selected quantitative techniques. In *Handbook of Social Psychology*, volume 1. Addison-Wesley, Cambridge, Mass., 1954.

Nalls, MA, Plagnol, V, Hernandez, DG, Sharma, M, Sheerin, UM, Saad, M, Simón-Sánchez, J, Schulte, C, Lesage, S, Sveinbjörnsdóttir, S, Stefánsson, K, Martinez, M, Hardy, J, Heutink, P, Brice, A, Gasser, T, Singleton, AB, and Wood, NW. Imputation of sequence variants for identification of genetic risks for parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*, 377(9766): 641–649, 2011.

Neyman, Jerzy and Pearson, Egon S. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 231:289–337, 1933.

NHGRI, National Human Genome Research Institute. URL http://www.genome.gov/Pages/Hyperion//DIR/VIP/Glossary/Illustration/Images/dna.gif.

Pearson, Karl. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302): 157–175, 1900.

Pearson, Karl. Certain enteric fever inoculation statistics. *British Medical Journal*, 2 (2288):1243–1246, 1904.

Raphael, Benjamin J. Chapter 6: Structural variation and medical genomics. *PLoS Comput Biol*, 8(12), 2012.

Rice, William R. A consensus combined p value test and the family wide significance of component tests. *Biometrics*, 6:303 – 308, 1990.

Roseman, Michelle, Milette, Katherine, Bero, Lisa A, Coyne, James C, Lexchin, Joel, Turner, Erick H, and Thombs, Brett D. Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments. *JAMA*, 305(10):1008–1017, 2011.

Rothstein, Hannah R., Sutton, Alexander J., and Borenstein, Michael. *Publication bias in meta-analysis : prevention, assessment and adjustments*. Chichester, England ; Hoboken, NJ : Wiley, 2005.

Scheet, P. and Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, 78:629–644, 2006.

Sham, Pc, Purcell, S, and et. al. Plink: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3): 559–575, 2007.

Sidik, K and Jonkman, JN. A simple confidence interval for meta-analysis. *Statistics in Medicine*, 21(21):3153–3159, 2002.

Sidik, K and Jonkman, JN. On constructing confidence intervals for a standardized mean difference in meta-analysis. *Communications in Statistics - Simulation and Computation*, 32(4):1191–1203, 2003.

Spencer, Chris C. A., Su, Zhan, Donnelly, Peter, and Marchini, Jonathan. Designing

genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. *PLoS genetics*, 5(5), 2009.

Stahl, Eli A, Raychaudhuri, Soumya, and et. al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genetics*, pages 508–514, 2010.

Sterne, Jonathan A.C, Gavaghan, David, and Egger, Matthias. Publication and related bias in meta-analysis. *Journal of Clinical Epidemiology*, 53(11):1119–1129, 2000.

Stouffer, Samuel A., Suchman, Edward A., DeVinney, Leland C., Star, Shirley A., and Williams, Robin M. The american soldier: Adjustment during army life. *Studies in Social Psychology in World War II*, 1949.

Stroup, DF, Berlin, JA, Morton, SC, Olkin, I, Williamson, GD, Rennie, D, Moher, D, Becker, BJ, Sipe, TA, and Thacker, SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. meta-analysis of observational studies in epidemiology (moose) group. *JAMA*, 283(15):2008 – 2012, 2000.

Sutton, Alexander J. and Abrams, Keith R. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res*, 10:277–303, 2001.

T1000GPC, The 1000 Genomes Project Consortium. A map of human genome variation from population scale sequencing. *Nature*, 467(7319), 2010.

T1000GPC, The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491:56–65, 2012.

Thompson, John R, Attia, John, and Minelli, Cosetta. The meta-analysis of genome-wide association studies. *Brief Bioinform*, 12(3):259 – 269, 2011.

Walker, Esteban, Hernandez, Adrian V., and Kattan, Michael W. Meta-analysis: Meta-analysis:its strengths and limitations. *Cleveland Clinic Journal of Medicine*, 75(6): 431–439, 2008.

WTCCC, The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661 – 678, 06 2007.

Zaitlen, Noah and Eskin, Eskin. Imputation aware meta-analysis of genome-wide association studies. *Genet Epidemiol*, 34(6):537–42, 2010.

Zhang, Jun, Chiodini, Rod, Badr, Ahmed, and Zhang, Genfa. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 38(3):95–109, 2011.

Zhang, Peng, Zhan, Xiaowei, Rosenberg, Noah A., and Zöllner, Sebastian. Genotype imputation reference panel selection using maximal phylogenetic diversity. *Genetics*, 195(5), 2013.

Zheng, Gang, Yang, Yaning, Zhu, Xiaofeng, and Elston, Robert C. *Analysis of Genetic Association Studies*. Springer New York Dordrecht Heidelberg London, 2012.

Zheng, Jin, Li, Yun, Abecasis, Gonçalo R., , and Schee, Paul. A comparison of approaches to account for uncertainty in analysis of imputed genotypes. *Genet Epidemiol*, 35(2):102 – 110, 2011.