Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes

Running title: Linear models in binary phenotype meta-analysis

James P Cook[1,*], Anubha Mahajan[2,*], and Andrew P Morris[1,2]

[1]Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK

[2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK

[*]Equal contribution

Address for correspondence:

Prof Andrew Morris

Department of Biostatistics, University of Liverpool

Block F, Waterhouse Building, 1-5 Brownlow Street, Liverpool L69 3GA, UK

E-mail: a.p.morris@liverpool.ac.uk

Telephone: 0151 794 49756

Conflict of interest. The authors declare no conflict of interest.

ABSTRACT

Linear mixed models are increasingly utilised for the analysis of genome-wide association studies (GWAS) of binary phenotypes because they can efficiently and robustly account for population stratification and relatedness through inclusion of random effects for a genetic relationship matrix. However, the utility of linear (mixed) models in the context of meta-analysis of GWAS of binary phenotypes has not been previously explored. In this investigation, we present simulations to compare the performance of linear and logistic regression models under alternative weighting schemes in a fixed-effects meta-analysis framework, considering designs that incorporate variable case-control imbalance, confounding factors and population stratification. Our results demonstrate that linear models can be used for meta-analysis of GWAS of binary phenotypes, without loss of power, even in the presence of extreme case-control imbalance, provided that one of the following schemes is used: (i) effective sample size weighting of *Z*-scores; or (ii) inverse-variance weighting of allelic effect sizes after conversion onto the log-odds scale. Our conclusions thus provide essential recommendations for the development of robust protocols for meta-analysis of binary phenotypes with linear models.

Keywords: genome-wide association study; binary phenotype; meta-analysis; linear regression model; logistic regression model; mixed model.

INTRODUCTION

Linear mixed models (LMMs) have received increasing prominence in the analysis of genome-wide association studies (GWAS) of complex human traits because they account for genetic structure, across participants, which arises from population stratification, cryptic relatedness or close familial relationships[1-7]. In this framework, structure is modelled by means of a genetic relationship matrix (GRM), constructed from genome-wide SNP genotype data across study participants (or from known familial relationships). A random effects model is then used to evaluate the evidence of association for a SNP by accounting for the contribution of the GRM to the overall variance of the trait. This flexible modelling framework can incorporate fixed effects to account for covariates, and can be used to estimate components of heritability that are explained by (subsets of) genotyped SNPs[8,9].

Linear models assume that the outcome of interest is a quantitative trait with a Gaussian distribution. However, it has become increasingly common to employ LMM approaches in population- and family-based GWAS of binary phenotypes because of their flexibility in accounting for structure, and their computational tractability in comparison to logistic mixed models. Linear models have the disadvantage that allelic effect estimates cannot be interpreted, directly, in terms of the odds ratio (OR), although approximations on the log-odds scale can be obtained[10]. Recent studies have also demonstrated that LMMs have less power than traditional logistic regression modelling techniques in GWAS of case-control phenotypes unless ascertainment is adequately accounted for[11,12].

Whilst the properties of linear (mixed) models in the analysis of GWAS of binary phenotypes *at the cohort level* have been previously explored[10], their utility in the context of *meta-analysis* has not been investigated. In this study, therefore, we present simulations

to compare the type I error rates and power of generalised linear (mixed) models under alternative weighting schemes in a fixed-effects meta-analysis framework. We consider a range of study designs that incorporate variable case-control imbalance across GWAS to reflect the increasing use of large-scale, population-based biobanks, and investigate the impact of confounders and population stratification on the properties of the analytical strategies. We conclude by making recommendations for the development of robust protocols for meta-analysis of GWAS of binary phenotypes with linear (mixed) models, which will be highly relevant in the era of large-scale consortium efforts to unravel the genetic basis of complex human diseases.

MATERIALS AND METHODS

Consider a GWAS of n participants, with binary phenotypes, genome-wide genotypes and additional covariates denoted by $\mathbf{y}$, $\mathbf{G}$ and $\mathbf{x}$, respectively. We denote the phenotype of the ith participant by $y_i \in \{0,1\}$, and their genotype at the jth SNP by $G_{ij} \in [0,2]$, coded under a dosage model in the number of minor alleles. In a generalised linear mixed modelling framework,

$$g(E[\mathbf{y}]) = \alpha + \beta \mathbf{G}_j + \boldsymbol{\gamma}\mathbf{x} + \mathbf{u} \quad (1)$$

where $g(.)$ is the link function, β is the allelic effect of the jth SNP on the phenotype, and $\boldsymbol{\gamma}$ is a vector of covariate regression parameters. In this expression, $\mathbf{u}$ is a vector of random effects, defined by $\mathbf{u} \sim \mathrm{MVN}(0,\lambda\mathbf{K})$, for the variance component λ and GRM $\mathbf{K}$, derived from genome-wide SNP data (or known familial relationships) to account for population

structure. A likelihood ratio test with one degree of freedom is then formed by comparing the maximised log likelihood of the unconstrained model (1), with that obtained under the null hypothesis of no association, $\beta = 0$. Note that model (1) reduces to a generalised linear model (no random effects) for $\lambda = 0$, which is appropriate in the absence of structure due to population stratification and/or familial relationships.

Under a logistic regression model, for the logit link function, the maximum likelihood estimate of the allelic effect, $\hat{\beta}_{LOG}$, can be interpreted directly as the log-OR of the *j*th SNP. However, under a linear regression model, for the identity link function, the maximum likelihood estimate of the allelic effect, $\hat{\beta}_{LIN}$, is measured on the wrong scale. Nevertheless, we can obtain an approximation of the allelic log-OR and corresponding variance from the linear model[10], given by

$$\hat{\beta}'_{LOG} \approx \frac{\hat{\beta}_{LIN}}{\hat{\alpha}_{LIN}(1-\hat{\alpha}_{LIN})}$$

and

$$\mathrm{Var}(\hat{\beta}'_{LOG}) = \frac{Var(\hat{\beta}_{LIN})}{[\hat{\alpha}_{LIN}(1-\hat{\alpha}_{LIN})]^2}$$

where $\hat{\alpha}_{LIN}$ is the maximum likelihood estimate of the intercept. In practice, $\hat{\alpha}_{LIN}$ is usually obtained from the null model for which $\beta_{LIN} = 0$, because the effect of any SNP on the phenotype is expected to be small. Here, we estimate $\hat{\alpha}$ by the proportion of participants that are cases, for which the correction factor $[\hat{\alpha}_{LIN}(1-\hat{\alpha}_{LIN})]^{-1}$ is minimised when the numbers of cases and controls in the study are equal (i.e. no imbalance). This

transformation of parameter estimates from the linear regression model has been demonstrated to provide an accurate approximation of the allelic log-OR provided that genetic effects are small, the case-control ratio is well balanced, and the SNP is common[10].

Fixed-effects meta-analysis. Consider *N* GWAS, for which we have tested for association of the phenotype with the *j*th SNP under a generalised linear model (1). We denote the *effective* sample size of the *k*th GWAS by n_k, given by

$$\frac{4n_{0k}n_{1k}}{n_{0k}+n_{1k}}$$

where n_{0k} and n_{1k} denote the numbers of controls and cases, respectively. In the kth GWAS, we also denote the *p*-value obtained from the regression model by p_k, and the estimated allelic effect from the regression model by $\hat{\beta}_k$.

Under an effective sample size weighting scheme, we obtain a combined Z-score for association of the *j*th SNP across GWAS by

$$Z^{(SS)} = \frac{\sum_k \phi^{-1}\left(\frac{p_k}{2}\right)\left(\frac{\hat{\beta}_k}{|\hat{\beta}_k|}\right)\sqrt{n_k}}{\sum_k \sqrt{n_k}}$$

where ϕ^{-1} is the inverse normal distribution function. Alternatively, under an inverse variance weighting scheme, we obtain an estimate of the allelic effect of the *j*th SNP on the phenotype, and the corresponding variance, across GWAS by

$$B = \mathrm{Var}(B)\left(\sum_k \hat{\beta}_k[\mathrm{Var}(\hat{\beta}_k)]^{-1}\right)$$

where

$$\mathrm{Var}(B) = \left(\sum_k [\mathrm{Var}(\hat{\beta}_k)]^{-1}\right)^{-1}$$

We then obtain a combined *Z*-score for association of the *j*th SNP across GWAS by

$$Z^{(IV)} = \frac{B}{\sqrt{Var(B)}}$$

Simulation study. We have performed a series of detailed simulations to investigate the type I error rates and power of alternative approaches to study-level association testing of a binary phenotype (linear and logistic regression modelling) in the context of fixed-effects meta-analysis (with effective sample size or inverse variance weighting schemes), summarised in **Table 1**.

Our first study design consisted of ten cohorts of a binary phenotype, ascertained from the same population, each comprising of 2,000 participants. We considered three scenarios for case-control imbalance, described in **Table 2**, such that the meta-analysis comprised a total of 10,000 cases and 10,000 population controls: (i) no imbalance (1:1 ratio in each cohort); (ii) moderate imbalance (variable ratio of 3:1 to 1:3 across cohorts); and (iii) extreme imbalance (variable ratio of 19:1 to 1:19 across cohorts). For each scenario, we

investigated models of association parameterised according to: (i) the risk allele frequency (RAF) of the causal SNP, denoted q; and (ii) the allelic OR for the risk allele, denoted ψ.

For each model, we generated 10,000 replicates of genotype data for the causal SNP in the study participants. For each replicate, genotypes were simulated in the required numbers of cases and controls in each cohort, according to the causal SNP RAF and allelic OR, and assuming Hardy-Weinberg equilibrium. Specifically, genotypes in cases and controls were simulated from a multinomial distribution, with probabilities given by

$$P(RR|\text{case}) = \psi^2 q^2 / T$$

$$P(Rr|\text{case}) = 2\psi\, q(1-q)/T$$

$$P(rr|\text{case}) = (1-q)^2/T$$

$$P(RR|\text{control}) = q^2$$

$$P(Rr|\text{control}) = 2q(1-q)$$

$$P(rr|\text{control}) = (1-q)^2$$

where R denotes the risk allele and $T = \psi^2 q^2 + 2\psi q(1-q) + (1-q)^2$.

To assess the impact of confounders on the alternative analysis strategies, we also simulated a binary covariate for each individual from a Bernoulli distribution, taking the value 1 in cases with probability $\omega/(1-\omega)$ and 0 otherwise, and taking the value 1 in controls with probability $1/(1-\omega)$ and 0 otherwise.

We also investigated the impact of population stratification on the alternative analysis strategies. Within each cohort, cases and controls were ascertained from subpopulation A with probabilities θ and (1-θ), respectively, and were otherwise ascertained from subpopulation B. The RAFs in subpopulations A and B were assumed to be 0.4 and 0.6,

respectively, and used to generate genotypes at the causal SNP under Hardy-Weinberg equilibrium, from a multinomial distribution, as defined above. For each individual, we then simulated genotype data for 1,000 additional uncorrelated SNPs, assuming Hardy-Weinberg equilibrium, and independent of case-control status, from a multinomial distribution. For each SNP, we assumed minor allele frequencies of 0.2 and 0.8, respectively, in subpopulations A and B. Genotypes at the 1,000 SNPs were then used to construct the GRM within each cohort.

Our second study design consisted of two cohorts of a binary phenotype, ascertained from the same population. The first cohort consisted of 1,000 cases and 1,000 controls. The second cohort represented a large biobank of 100,000 individuals, within which we investigated the impact of the extent of case-control imbalance on the meta-analysis. For each scenario, we assumed a causal SNP RAF of 0.5 and an allelic OR of 1.25, and generated 10,000 replicates of genotype data for the causal SNP in the study participants. For each replicate, genotypes were simulated in the required numbers of cases and controls in the two cohorts, assuming Hardy-Weinberg equilibrium, from a multinomial distribution, as described above.

For both study designs, we utilised a linear Wald test, implemented in EPACTS, to obtain parameter estimates and association *p*-values under a linear regression model (no random effects) within each cohort for each replicate. To obtain parameter estimates under a logistic regression model (no random effects) within each cohort, we utilised a Firth bias-corrected likelihood ratio test, also implemented in EPACTS, which has been demonstrated to be more robust to case-control imbalance than Wald or score statistics for binary outcomes[13]. To obtain parameter estimates under a linear mixed model (random effects for GRM) within each cohort, we utilised EMMAX[1], also implemented in EPACTS. We

combined summary statistics through fixed-effects meta-analysis with effective sample size and inverse variance weighting using METAL[14] and GWAMA[15], respectively.

Across all scenarios, each test of association, after meta-analysis, was evaluated at nominal significance thresholds of $p<0.05$ and $p<0.01$, and at the traditional genome-wide standard of $p<5\times10^{-8}$. For estimated allelic effect sizes on the log-odds scale (from the logistic regression model and after conversion from the linear regression model), we also evaluated bias and mean square error (MSE).

RESULTS

No population stratification or confounders. We first considered the properties of fixed-effects meta-analysis of association summary statistics obtained from linear and logistic regression models *without* random effects for the GRM, for simulations generated in the absence of structure or confounders. **Supplementary Figure S1** presents the type I error rate (at a nominal 5% significance threshold) of each of the analytical strategies considered (**Table 1**) for a SNP with RAF in the range of 1-50%. For all frequencies investigated, the type I error rate was consistent with the nominal significance threshold of $p<0.05$, irrespective of the analytical approach and the extent of case-control imbalance.

Figure 1 presents the power (at genome-wide significance) of each of the analytical strategies considered (**Table 1**), as a function of the allelic OR, for a SNP with RAF in the range of 1-50%. There is no appreciable difference in power between the five approaches unless there is extreme case-control imbalance. In this extreme imbalance setting, the power of the meta-analysis under inverse-variance weighting of effect sizes from the linear model (without conversion to the log-odds scale) is substantially lower than for the other

approaches. However, we also observe a loss in power of the meta-analysis under inverse-variance weighting of effect sizes from the logistic regression model for rare SNPs (RAF 1%), irrespective of the extent of case-control imbalance, which has not been reported previously. We observe the same pattern of results at less stringent significance levels (**Supplementary Figure S2**), with the inverse-variance weighting of effect sizes from the linear model (without conversion to the log-odds scale) being substantially less powerful when there is extreme case-control imbalance.

Supplementary Figures S3 and S4 present the bias and MSE of the estimated allelic OR after meta-analysis under the inverse-variance weighting of effect sizes from the logistic regression model and the linear regression model after conversion to the log-odds scale. Results are presented as a function of the allelic OR. There is minimal difference in both metrics between the two meta-analysis strategies. However, for rare SNPs (RAF 1%), the meta-analysis under inverse variance weighting of effect sizes from the logistic regression model underestimates the allelic OR, irrespective of case-control imbalance, explaining the reduction in power of this strategy that was observed above.

Impact of a confounding variable in the absence of population stratification. We next considered the properties of fixed-effects meta-analysis of association summary statistics obtained from linear and logistic regression models *without* random effects for the GRM, for simulations generated in the absence of structure, but where the binary phenotype was also correlated with a confounding variable. We assumed a causal SNP with RAF 50% and an allelic OR of 1.15 for the binary phenotype. **Supplementary Figure S5** presents the power (at genome-wide significance) of each of the five analytical strategies considered (**Table 1**), as a function of the relative risk of the confounding variable, defined by $\omega/(1-\omega)$. As

expected, there is a general decline in power to detect association across analytical strategies as the relative risk of the confounder of the binary phenotype increases. However, as demonstrated by the simulations in the absence of confounders, the inverse-variance weighting of effect sizes from the linear model (without conversion to the log-odds scale) was less powerful when there is extreme case-control imbalance.

Supplementary Figure S5 also presents the bias and MSE of the estimated allelic OR after meta-analysis under the inverse-variance weighting of effect sizes from the logistic regression model and the linear regression model after conversion to the log-odds scale. Results are presented as a function of the relative risk of the confounding variable. Irrespective of the case-control imbalance, the estimated allelic OR after conversion to the log-scale becomes increasingly biased (underestimated) as the relative risk of the confounding variable increases, although power is not affected.

Impact of population stratification. We then considered the properties of fixed-effects meta-analysis of association summary statistics obtained from linear regression models, with and without random effects for the GRM, for simulations generated in the presence of population stratification (cases and controls ascertained from subpopulations A and B). **Supplementary Figure S6** presents the type I error rate (at a nominal 5% significance threshold) of each analytical strategy considered (**Table 1**) as a function of the probability, θ, that a case is ascertained from subpopulation A. Irrespective of the extent of population stratification, the type I error rate was consistent with the nominal significance threshold of $p<0.05$ for any fixed-effects meta-analysis strategy using the linear model with random effects for the GRM. However, as expected, type I error rates became increasingly inflated

as the extent of population stratification was elevated for all fixed-effects meta-analysis strategies using the linear model without a random effect for the GRM.

Figure 2 presents the power (at genome-wide significance) of the three fixed-effects meta-analysis strategies that aggregate association summary statistics from the linear model with random effects for the GRM, for a causal SNP with allelic OR of 1.15 for the binary phenotype. There is no appreciable difference in power between the analytical strategies, unless there is extreme case-control imbalance. In this extreme imbalance setting, the power of the meta-analysis under inverse-variance weighting of effect sizes from the linear model (without conversion to the log-odds scale) is substantially lower than for the other approaches. The difference in power between these approaches is consistent, irrespective of the extent of population stratification.

Impact of inclusion of a population biobank with extreme case-control imbalance. Finally, we considered the properties of fixed-effects meta-analysis of association summary statistics obtained from linear and logistic regression models without random effects for the GRM, for simulations generated in the absence of structure. In these simulations, association summary statistics were aggregated from a population biobank of 100,000 participants with extreme case-control imbalance and a balanced case-control study of 2,000 participants. **Figure 3** presents the power (at genome-wide significance) of each of the analytical strategies considered (**Table 1**), for a causal SNP with RAF 50% and an allelic OR of 1.25, as a function of the number of cases in the population biobank. As reported above, in this extreme imbalance setting, the power of the meta-analysis under inverse-variance weighting of effect sizes from the linear model (without conversion to the log-odds scale) is substantially lower than for the other approaches. The difference in power reduces

as the extent of the imbalance in the biobank decreases (i.e. the proportion of cases increases), and thus has most detrimental impact for rare diseases.

DISCUSSION

We have presented simulations to evaluate the utility of linear models in the context of meta-analysis of GWAS of binary phenotypes. Our results highlight that the extent of case-control imbalance across studies can have a major impact on the performance of a linear regression model. We have demonstrated that, for extreme imbalance, meta-analysis under inverse-variance weighting of allelic effect estimates from a linear regression model results in a substantial reduction in power, unless they are first converted onto the log-odds scale. This is of particular importance because existing, widely-used software[16] for the meta-analysis of association summary statistics from LMMs implements inverse-variance weighting of allelic effect estimates *without* conversion to the log-odds scale.

For a binary phenotype, under a linear regression model, the standard error of an allelic effect estimate is dependent on multiple factors, including allele frequency, total sample size, OR and variance of the trait. For a fixed total sample size, the variance of the trait (and thus standard error of the allelic effect estimate) decreases as the case-control imbalance becomes more extreme. However, the power to detect association with the binary phenotype is less in imbalanced studies, and they should, in fact, be given less weight in any meta-analysis. Correction of allelic effect estimates from the linear regression model onto the log-odds scale circumvents this issue by inflating the corresponding standard error by a factor that is inversely proportional to the case-control imbalance.

Case-control imbalance is becoming increasingly widespread in GWAS of binary phenotypes, particularly with the availability of large-scale, extensively-studied, population-based biobanks, often with linkage to electronic medical records[17-20]. The utility of linear models in these extremely imbalanced case-control designs has not been previously studied in the context of meta-analysis. Crucially, our investigation highlights that linear models can be used for meta-analysis of GWAS of binary phenotypes, without loss of power, even in the presence of extreme case-control imbalance, provided that one of the following schemes is used: (i) effective sample size weighting of Z-scores; or (ii) inverse-variance weighting of allelic effect sizes after conversion onto the log-odds scale.

Our simulations demonstrate that meta-analysis of association summary statistics for binary phenotypes from LMMs is robust to population stratification, even in the presence of extreme case-control imbalance. However, it is important to note that this conclusion is valid only when population stratification does not lead to violation of the LMM assumption of homoscedasticity, for which residual variances are constant, irrespective of covariates[21,22]. Heteroscedasticity can occur in the presence of population stratification, for example, when strata have variable case-control imbalance or heterogeneous disease risk. Under these circumstances, LMMs are valid only for variants that have similar RAFs across strata, such that there is only weak confounding due to structure. Otherwise, computationally efficient software will be required to implement logistic mixed models on the scale of the whole-genome.

Conflict of interest. The authors declare no conflict of interest.

Acknowledgements. Andrew P. Morris is a Wellcome Trust Senior Fellow in Basic Biomedical Science (under award WT098017).

Supplementary Information. Supplementary Information accompanies this paper on the European Journal of Human Genetics website (http://www.nature.com/ejhg).

REFERENCES

1. Kang HM, Sul JH, Service SK, *et al*. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 2010; **42**: 348-354.
2. Zhang Z, Ersoz E, Lai CQ, *et al*. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 2010; **42**: 355-360.
3. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* 2010; **11**: 459-463.
4. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods* 2011; **8**: 833-835.
5. Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E, Heckerman D. Improved linear mixed models for genome-wide association studies. *Nat Methods* 2012; **9**: 525-526.
6. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012; **44**: 821-824.
7. Svishcheva GR, Axenovich TI, Belonogova NM, van Duijn CM, Aulchenko YS. Rapid variance components-based method for whole-genome association analysis. *Nat Genet* 2012; **44**: 1166-1170.

8. Yang J, Benyamin B, McEvoy BP, *et al*. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42**: 565-569.
9. Zaitlen N, Kraft P. Heritability in the genome-wide association era. *Hum Genet* 2012; **131**: 1655-1664.
10. Pirinen M, Donnelly P, Spencer CCA. Efficient computation with a linear mixed model on large-scale genetic data sets with applications to genetic studies. *Ann App Stat* 2013; **7**: 369-390.
11. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application mixed-model association methods. *Nat Genet* 2014; **46**: 100-106.
12. Hayeck TJ, Zaitlen NA, Loh P-R, *et al*. Mixed model with correction for case-control ascertainment increases association power. *Am J Hum Genet* 2015; **96**: 720-730.
13. Ma C, Blackwell T, Boehnke M, Scott LJ, GoT2D Investigators. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genet Epidemiol* 2013; **37**: 539-550.
14. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinf* 2010; **26**: 2190-2191.
15. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinf* 2010; **11**: 288.
16. Liu DJ, Peloso GM, Zhan X, *et al*. Meta-analysis of gene-level tests for rare variant association. *Nat Genet* 2014; **46**: 200-204.
17. Roden DM, Pulley JM, Basford MA, *et al*. Development of a large-scale de-identified DNA biobank to enable personalised medicine. *Clin Pharmacol Ther* 2008; **84**: 362-369.

18. Tayo BO, Teil M, Tong L, *et al*. Genetic background of patients from a university medical center in Manhattan: implications for personalized medicine. *PLoS One* 2011; **6**: e19166.

19. Leitsalu L, Haller T, Esko T, *et al*. Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol* 2015; **44**: 1137-1147.

20. Sudlow C, Gallacher J, Allen N, *et al*. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**: e1001779.

21. Conomos MP, Laurie CA, Stilp AM, *et al*. Genetic diversity and association studies in US Hispanic/Latino populations: applications in the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet* 2016; **92**: 165-184.

22. Chen H, Wang C, Conomos MP, *et al*. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am J Hum Genet* 2016; **98**: 653-666.

FIGURE LEGENDS

Figure 1. Power to detect association (at genome-wide significance, $p<5x10^{-8}$) of a binary phenotype with a causal SNP, in the absence of population stratification or confounders, using alternative meta-analysis strategies for summary statistics obtained from linear and logistic regression models without random effects for the GRM (**Table 1**). Results are presented as a function of the allelic OR, for a causal SNP with RAF in the range of 1-50%, and for variable extent of case-control imbalance (defined in **Table 2**).

Figure 2. Power to detect association (at genome-wide significance, $p<5x10^{-8}$) of a binary phenotype with a causal SNP, in the presence of population stratification (cases and controls ascertained from subpopulations A and B), using alternative meta-analysis strategies for summary statistics obtained from linear regression models with random effects for the GRM (**Table 1**). Results are presented as a function of the probability that a case is ascertained from subpopulation A, for a causal SNP with allelic OR of 1.15 for the binary phenotype, and for variable extent of case-control imbalance (defined in **Table 2**).

Figure 3. Power to detect association (at genome-wide significance, $p<5x10^{-8}$) of a binary phenotype with a causal SNP, in the absence of population stratification or confounders, using alternative meta-analysis strategies for summary statistics obtained from linear and logistic regression models without random effects for the GRM (**Table 1**). Association summary statistics were aggregated from a population biobank of 100,000 participants with extreme case-control imbalance and a balanced case-control study of 2,000 participants.

Results are presented for a causal SNP with RAF 50% and an allelic OR of 1.25, as a function of the number of cases in the population biobank.

Table 1. Summary of approaches to study-level association testing of a binary phenotype (linear and logistic regression modelling) in the context of fixed-effects meta-analysis (with effective sample size or inverse variance weighting schemes).

Study-level analysis	Random effects?	Summary statistic	Meta-analysis weighting	Meta-analysis summary statistic(s)
logistic regression	No	*p*-value	effective sample size	*p*-value
logistic regression	No	allelic effect on log-odds scale	inverse variance	*p*-value and effect size on log-odds scale
linear regression	No	*p*-value	effective sample size	*p*-value
linear regression	No	allelic effect on linear scale	inverse variance	*p*-value and effect size on linear scale
linear regression	No	allelic effect converted to log-odds scale	inverse variance	*p*-value and effect size on log-odds scale
linear regression	GRM	*p*-value	effective sample size	*p*-value
linear regression	GRM	allelic effect on linear scale	inverse variance	*p*-value and effect size on linear scale
linear regression	GRM	allelic effect converted to log-odds scale	inverse variance	*p*-value and effect size on log-odds scale

Table 2. Summary of case-control counts in each cohort for alternative imbalance scenarios considered in the simulation study.

Cohort	No imbalance		Moderate imbalance		Extreme imbalance	
	Cases	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**
1	1,000	1,000	600	1,400	100	1,900
2	1,000	1,000	700	1,300	300	1,700
3	1,000	1,000	800	1,200	500	1,500
4	1,000	1,000	900	1,100	700	1,300
5	1,000	1,000	1,000	1,000	900	1,100
6	1,000	1,000	1,000	1,000	1,100	900
7	1,000	1,000	1,100	900	1,300	700
8	1,000	1,000	1,200	800	1,500	500
9	1,000	1,000	1,300	700	1,700	300
10	1,000	1,000	1,400	600	1,900	100
Total	**10,000**	**10,000**	**10,000**	**10,000**	**10,000**	**10,000**