Meta-Analysis of Genome-Wide Association Studies: No Efficiency Gain in Using Individual Participant Data

D. Y. Lin* and D. Zeng

Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina

To identify genetic variants with modest effects on complex human diseases, a growing number of networks or consortia are created for sharing data from multiple genome-wide association studies on the same disease or related disorders. A central question in this enterprise is whether to obtain summary results or individual participant data from relevant studies. We show theoretically and numerically that meta-analysis of summary results is statistically as efficient as joint analysis of individual participant data (provided that both analyses are performed properly under the same modeling assumptions). We illustrate this equivalence with case-control data from the Finland-United States Investigation of NIDDM Genetics (FUSION) study. Collating only summary results will increase the number and representativeness of available studies, simplify data collection and analysis, reduce resource utilization, and accelerate discovery. *Genet. Epidemiol.* 34:60–66, 2010. © 2009 Wiley-Liss, Inc.

Key words: complex diseases; GWAS consortia; joint analysis; mega analysis; SNPs; summary results

Contract grant sponsor: National Institutes of Health.
*Correspondence to: Danyu Lin, Ph.D., Department of Biostatistics, University of North Carolina, McGavran-Greenberg Hall, CB #7420, Chapel Hill, NC 27599-7420. E-mail: lin@bios.unc.edu
Received 26 January 2009; Revised 8 April 2009; Accepted 17 April 2009
Published online 21 October 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20435

INTRODUCTION

Genome-wide association studies (GWAS) have yielded new findings for many complex human diseases. Because complex diseases are influenced by an array of genetic variants mostly with small to moderate effects, it is difficult for one GWAS to provide unequivocal findings. Indeed, the odds ratios of disease with SNPs that have been observed in GWAS thus far are typically less than 1.5, and the majority of positive findings have emerged only after aggressive data sharing across multiple studies. For example, the initial findings from individual type 2 diabetes GWAS were ambiguous, but a number of disease loci with odds ratios of 1.1–1.4 were identified conclusively after combining results from several studies [Saxena et al., 2007; Scott et al., 2007; Zeggini et al., 2007, 2008].

Recognizing the need and benefits of data sharing, GWAS investigators have formed various networks or consortia to share data on the same disease or related disorders [Kavvoural and Ioannidis, 2008]. For example, the Psychiatric GWAS Consortium we are involved with has enrolled 47 studies in five major disorders [The Psychiatric GWAS Consortium Steering Committee, 2009]. Some of these consortia have attempted to obtain raw data on individual participants, as opposed to summary results that are used in traditional meta-analysis. The raw data from all available studies can then be analyzed simultaneously. Such analysis is commonly called joint analysis or mega-analysis. We will use the term mega-analysis and refer to the traditional method of combining summary results as meta-analysis.

A major motivation for obtaining raw, individual-level data is the general perception that mega-analysis is statistically more efficient than meta-analysis since it utilizes much more detailed information. However, obtaining raw data is difficult, costly, and time-consuming. Some investigators are unwilling or unable to share raw data. For the Tobacco and Genetics Consortium we are involved with, the majority of the investigators were unable to provide raw data due to IRB issues and/or study policies that prohibit the sharing of raw data. Excluding studies that do not contribute raw data will reduce statistical power and limit the generalizability of the findings. Furthermore, the sheer scale of GWAS data poses significant practical challenges in storing and analyzing raw data from a large number of studies.

We show in this article that meta-analysis (when performed properly) is as efficient as mega-analysis in that the estimates of any genetic effect produced by the two methods have approximately the same variance. Thus, there is no need to obtain raw data. Even if raw data are available, one can analyze the data for each study separately and then combine the summary results through meta-analysis. This will greatly facilitate the analysis, especially if raw data are available only on a subset of studies.

METHODS

We wish to combine results from K studies with n_k participants in the kth study. For the analysis of each SNP, the data consist of (Y_{ki}, X_{ki}), where Y_{ki} is the disease status (1 = disease, 0 = no disease) for the ith participant of the kth study, and X_{ki} is the corresponding genotype score. (Under the additive mode of inheritance, the genotype score is the number of minor alleles; under the dominant

© 2009 Wiley-Liss, Inc.

model, the genotype score indicates, by the values 1 vs. 0, whether or not the individual has at least one minor allele; under the recessive model, the genotype score indicates, by the values 1 vs. 0, whether or not the individual has two minor alleles. For an untyped SNP, the unknown genotype score may be imputed by the expected genotype score.) We assume the following logistic regression model:

$$\Pr(Y_{ki}=1)=\frac{e^{\alpha_k+\beta X_{ki}}}{1+e^{\alpha_k+\beta X_{ki}}}, \quad (1)$$

where the α_k's are study-specific intercepts, and β is the log odds ratio representing a common genetic effect across studies.

Let $\widehat{\beta}_k$ be the maximum likelihood estimate of β by maximizing the likelihood function of the kth study

$$L(\alpha_k,\beta)=\prod_{i=1}^{n_k}\frac{e^{Y_{ki}(\alpha_k+\beta X_{ki})}}{1+e^{\alpha_k+\beta X_{ki}}},$$

and let V_k be the variance estimate of $\widehat{\beta}_k$. Then the inverse-variance meta-analysis estimate of β is

$$\left(\sum_{k=1}^{K}V_k^{-1}\right)^{-1}\sum_{k=1}^{K}V_k^{-1}\widehat{\beta}_k,$$

and its variance is estimated by

$$\left(\sum_{k=1}^{K}V_k^{-1}\right)^{-1}.$$

To perform mega-analysis, we obtain the maximum likelihood estimate of β and its variance estimate by maximizing the joint likelihood function

$$\prod_{k=1}^{K}L(\alpha_k,\beta).$$

We show in the Appendix that the meta-analysis and mega-analysis estimates of β have approximately the same variance, so the two methods have approximately the same efficiency.

We can add covariates to model (1) in both meta-analysis and mega-analysis. The covariates may include environmental factors or principal components [Price et al., 2006] used to adjust for population stratification. The numbers and types of covariates need not be the same across studies. Meta-analysis of covariate-adjusted genetic effects is approximately as efficient as mega-analysis using individual-level covariate data (see the Appendix for details).

If the effects of some covariates are the same across studies, then one can improve the efficiency of mega-analysis by incorporating this restriction into the joint likelihood function and thus estimating fewer parameters. However, the efficiency gain is usually minimal because the number of covariates is much smaller than the sample sizes of typical GWAS. Interestingly, one can achieve the same efficiency gain by performing a multivariate version of meta-analysis (see the Appendix for details). The multivariate version of meta-analysis is not generally recommended because it requires additional summary results and the assumption of common covariate effects may not be appropriate.

Both meta-analysis and mega-analysis assume a common genetic effect across studies. This assumption does not affect the validity of association testing since the genetic effects are all zero under the null hypothesis of no association. However, it is important to determine whether meta-analysis or mega-analysis is more powerful when the effect sizes are unequal among studies. We show in the Appendix that the estimates produced by meta-analysis and mega-analysis are approximately the same and their variance estimates are also approximately the same when the genetic effects are unequal across studies, so that the two methods have similar statistical powers.

RESULTS

SIMULATION STUDIES

To demonstrate the equivalence between meta-analysis and mega-analysis, we present here some simulation results on combining two case-control studies. We

TABLE I. Mean effect estimates, standard errors, and powers at the 10^{-7} significance level for meta-analysis and mega-analysis of case-control data

Study 1 (MAF= 0.3)			Study 2 (MAF = 0.2)			Meta-analysis			Mega-analysis		
OR	Cases	Controls	OR	Cases	Controls	Mean	SE	Power	Mean	SE	Power
1.4	1,000	1,000	1.4	1,000	1,000	1.402	0.076	0.812	1.402	0.076	0.814
	1,500	1,500		500	500	1.402	0.074	0.865	1.402	0.074	0.866
	500	500		1,500	1,500	1.402	0.079	0.745	1.402	0.079	0.747
	750	1,500		1,500	750	1.402	0.076	0.814	1.402	0.076	0.815
	1,500	750		750	1,500	1.402	0.076	0.812	1.402	0.076	0.814
1.5	1,000	1,000	1.3	1,000	1,000	1.411	0.077	0.840	1.411	0.077	0.843
	1,500	1,500		500	500	1.459	0.077	0.967	1.459	0.077	0.967
	500	500		1,500	1,500	1.359	0.076	0.543	1.360	0.076	0.550
	750	1,500		1,500	750	1.408	0.076	0.830	1.408	0.076	0.841
	1,500	750		750	1,500	1.413	0.077	0.850	1.414	0.078	0.847
1.3	1,000	1,000	1.5	1,000	1,000	1.383	0.075	0.736	1.383	0.075	0.741
	1,500	1,500		500	500	1.338	0.070	0.594	1.339	0.070	0.599
	500	500		1,500	1,500	1.436	0.081	0.858	1.437	0.081	0.861
	750	1,500		1,500	750	1.386	0.075	0.755	1.386	0.076	0.748
	1,500	750		750	1,500	1.380	0.074	0.720	1.381	0.074	0.737

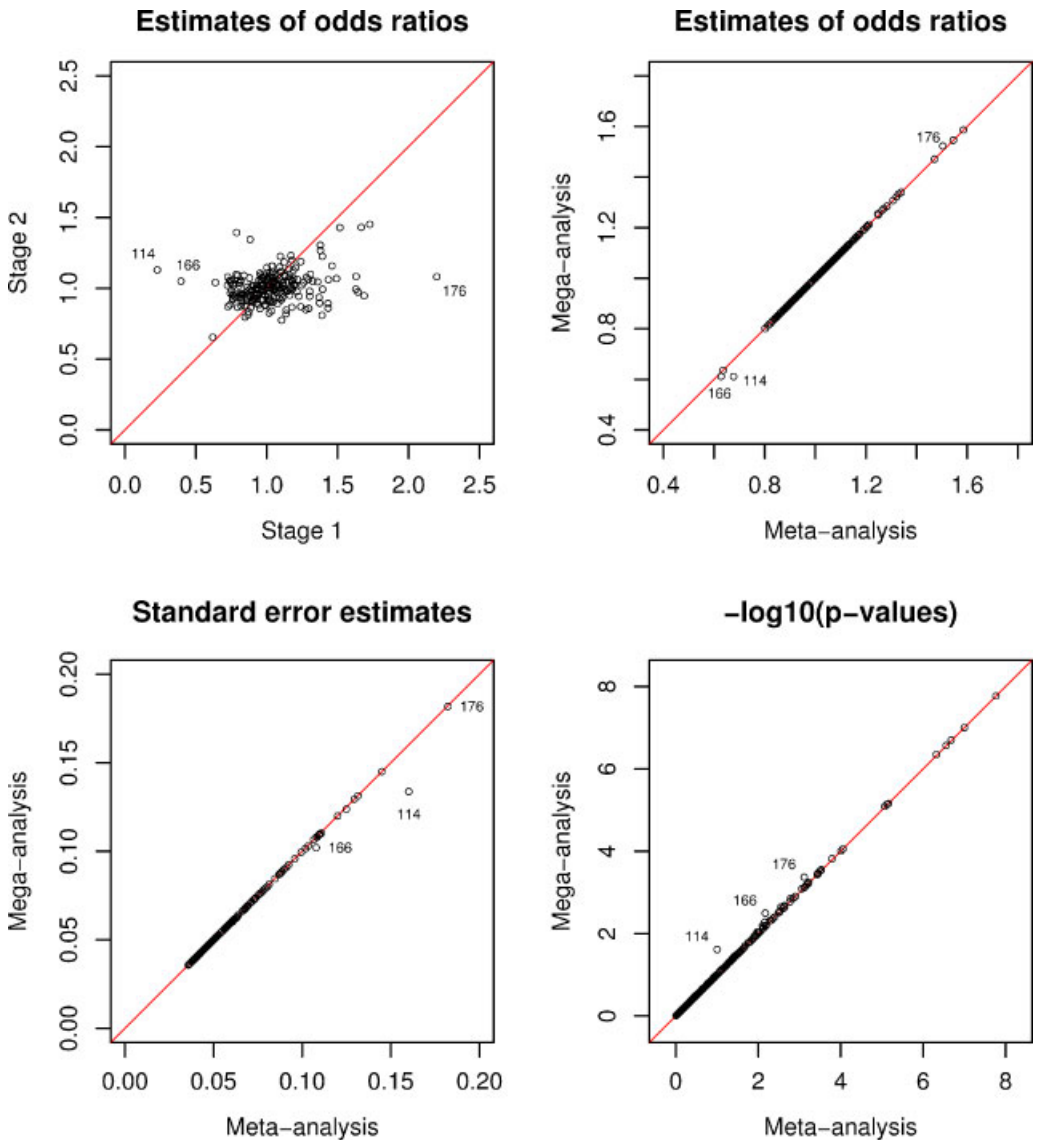

Fig. 1. Analysis of stages 1 and 2 data from the FUSION study. The top left panel compares the individual estimates of odds ratios between stages 1 and 2; the top right panel compares the combined estimates of odds ratios between meta-analysis and mega-analysis; the bottom left panel compares the standard error estimates between the two methods; and the bottom right panel compares the $-\log_{10}$ (*P*-values) between the two methods. In each panel, the red line indicates where the values on the two axes are equal.

simulated data from model (1), in which the SNP of interest had population minor allele frequencies (MAFs) of 0.3 and 0.2 in studies 1 and 2, respectively, and X_{ki} was the number of minor alleles. We set $\alpha_1 = -3$, $\alpha_2 = -2.2$, and $\beta = \log 1.4$. We also considered unequal values of β for the two studies. Note that e^{β} pertains to the odds ratio (OR) of disease with the SNP under the additive mode of inheritance. We obtained various combinations of the numbers of cases and controls for the two studies. For each combination of the simulation parameters, we generated 10 million data sets and performed meta-analysis and mega-analysis of each data set under model (1). The results are summarized in Table I.

When the SNP effects are the same between the two studies, the mean estimates of the SNP effects and the standard errors are identical up to the third decimal point between meta-analysis and mega-analysis, and the powers are identical up to the second decimal point. When the SNP effects are different between the two studies, there are some slight differences between the two methods, and either method can be slightly more powerful than the other.

FUSION DATA

For illustration with empirical data, we considered the Finland-United States Investigation of NIDDM Genetics (FUSION) study [Scott et al., 2007]. The FUSION study genotyped 1,161 Finnish type 2 diabetes (T2D) cases and 1,174 Finnish normal glucose-tolerant (NGT) controls on 317,503 SNPs on the Illumina HumanHap300 BeadChip in

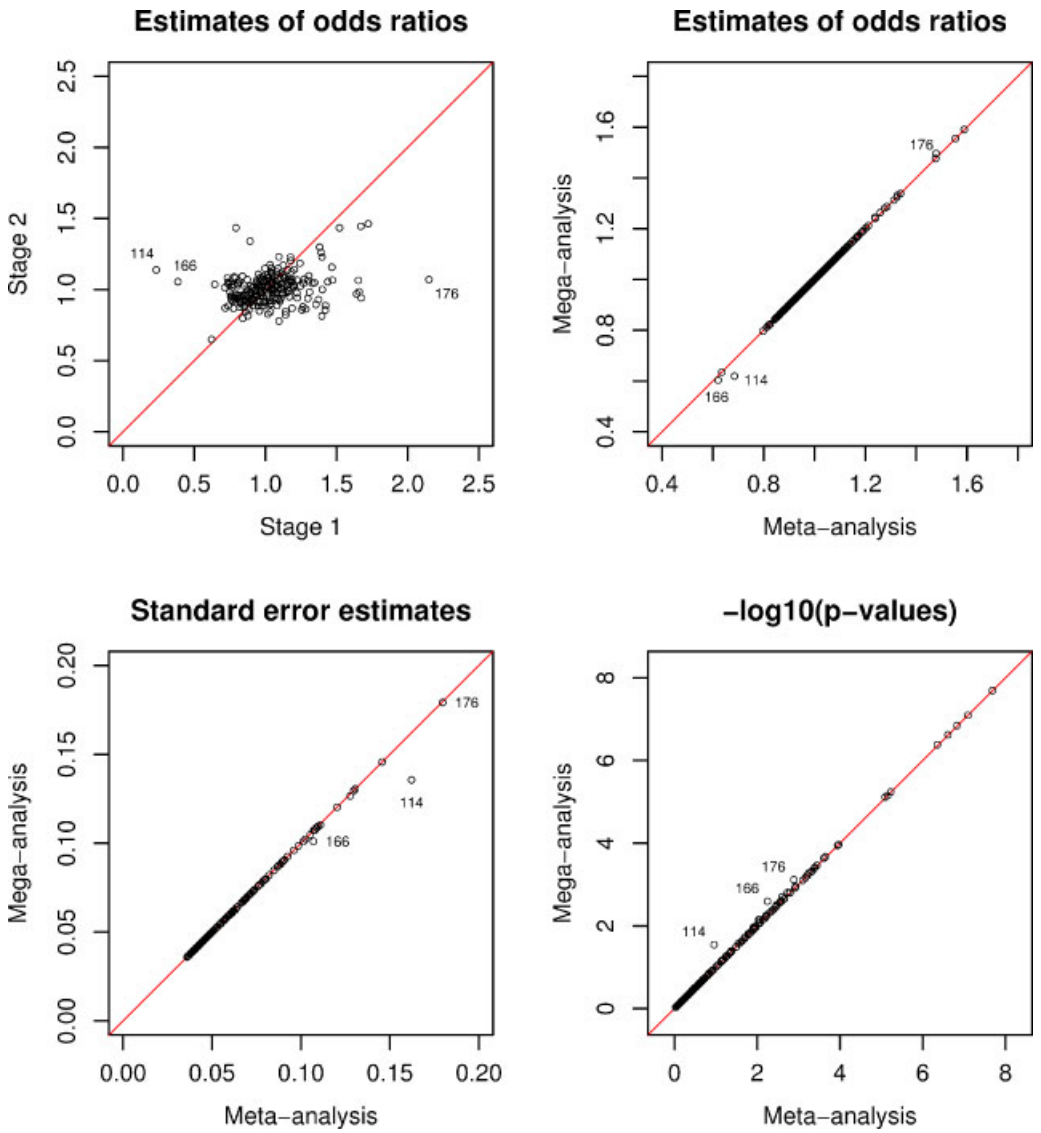

Fig. 2. Analysis of stages 1 and 2 data from the FUSION study adjusted for age and sex. The top left panel compares the individual estimates of odds ratios between stages 1 and 2; the top right panel compares the combined estimates of odds ratios between meta-analysis and mega-analysis; the bottom left panel compares the standard error estimates between the two methods; and the bottom right panel compares the $-\log_{10}$ (*P* values) between the two methods. Both meta-analysis and mega-analysis allow age and sex effects to be different between stages 1 and 2. In each panel, the red line indicates where the values on the two axes are equal.

stage 1 of a two-stage design. Based on the stage-1 results and the findings of other studies, the study genotyped 224 SNPs in an additional 1,204 Finnish T2D cases and 1,253 Finnish NGT controls. The subjects with missing genotypes on a particular SNP were excluded from the analysis of that SNP. All subjects have age and sex information.

We performed meta-analysis and mega-analysis of T2D status on the 224 SNPs that were genotyped in both stage 1 and stage 2 of the FUSION study. The results under the additive mode of inheritance are displayed in Figure 1. The individual estimates of odds ratios vary considerably between stages 1 and 2. The combined estimates of odds ratios and the corresponding standard error estimates are virtually identical between meta-analysis and mega-analysis, and consequently the two sets of *P*-values are virtually identical. The only noticeable differences lie in SNPs 114, 166, and 176, which have observed MAFs of approximately 0.9, 1.6, and 3.1%. For SNPs with low MAFs, the individual estimates of genetic effects may be unstable, which may cause the combined estimates to be different between meta-analysis and mega-analysis. Such differences are unlikely to alter the rankings of the top SNPs because the *P*-values associated with rare SNPs tend to be non-significant.

For further illustration, we included age and sex as covariates in the logistic regression model. When age and sex are allowed to have different effects between stages 1 and 2, meta-analysis and mega-analysis again produce virtually identical results (see Fig. 2). When age and sex are assumed to have common effects between stages

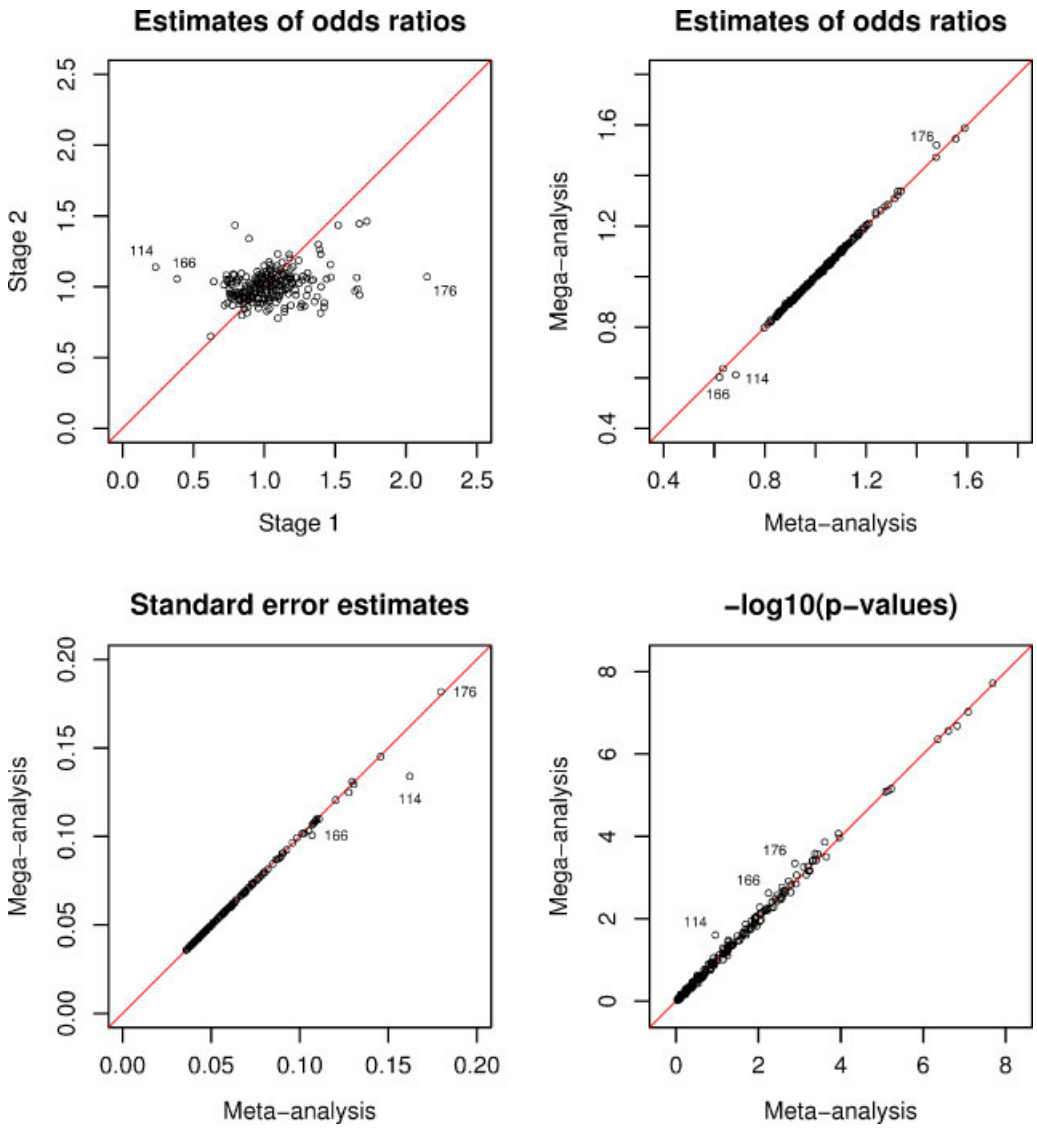

Fig. 3. Analysis of stages 1 and 2 data from the FUSION study adjusted for age and sex. The top left panel compares the individual estimates of odds ratios between stages 1 and 2; the top right panel compares the combined estimates of odds ratios between meta-analysis and mega-analysis; the bottom left panel compares the standard error estimates between the two methods; and the bottom right panel compares the $-\log_{10}$ (*P*-values) between the two methods. Mega-analysis assumes age and sex effects to be the same between stages 1 and 2 whereas meta-analysis does not. In each panel, the red line indicates where the values on the two axes are equal.

1 and 2 in mega-analysis, the results between the two methods are slightly more different (see Fig. 3).

DISCUSSION

Publication bias is a major concern in meta-analysis of literature results. One may reduce or avoid this kind of bias by planning GWAS meta-analysis prospectively to take advantage of all available studies and all available SNPs. By using summary results rather than raw data, one can increase the number of available studies and thus enhance the power of the analysis and the generalizability of the findings.

In many applications, it is desirable to adjust for participant-level covariates, such as principal components and environmental exposures. Such data are not available in published reports. In a consortium setting, the covariate adjustments can be made within each study and the covariate-adjusted estimates of genetic effects can then be combined through meta-analysis. It is logistically much simpler to provide such adjusted estimates than to transfer raw data. Indeed, this is the strategy adopted by the Tobacco and Genetics Consortium and many other consortia. If the covariate effects are the same across studies, then the mega-analysis that incorporates that restriction tends to be more efficient than the traditional meta-analysis. However, the efficiency gain is generally minimal and the same efficiency gain can be achieved by using a multivariate version of meta-analysis (see the Appendix for details).

We have focused on binary traits. In a related paper, Olkin and Sampson [1998] showed that, for comparing treatments with respect to a continuous outcome in clinical trials,

meta-analysis is equivalent to mega-analysis if the treatment effects and error variances are constant across trials. It follows from the arguments of the Appendix that all the conclusions of this article hold for quantitative traits and indeed for any traits under any study designs; the details are given in Lin and Zeng [2009].

By working with raw data, one can ensure that all studies use the same quality-control criteria and estimate the same quantities. However, such standardization and harmonization of information can be achieved by requiring all participating investigators to follow a common set of guidelines on quality control and statistical analysis so that the data are filtered and analyzed in the same way across studies before summary results are submitted.

ACKNOWLEDGMENTS

The authors are grateful to Drs. Michael Boehnke and Heather Stringham and other FUSION investigators for providing the data used in this article. They are also grateful to Dr. Kuo-Ping Li for his programming assistance.

REFERENCES

Cox DR, Hinkley DV. 1979. Theoretical Statistics. London: Chapman and Hall.

Kavvoura1 FK, Ioannidis JPA. 2008. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. Hum Genet 123:1–14.

Lin VT, Zeng V. 2009. On the relative efficiency of using summary statistics versus indivudual level data in meta-analysis. Biometrika, in press.

Olkin I, Sampson A. 1998. Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics 54: 317–322.

Price AL, Patterson NJ, Plenge, RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38:904–909.

Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PIW, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Boström KB, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen M-R, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjögren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, DeFelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn G-W, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336.

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding C-J, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li X-Y, Conneely KN, Riebow NL, Sprau AQ, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345.

The Psychiatric GWAS Consortium Steering Committee. 2009. A framework for interpreting genome-wide association studies of psychiatric disorders. Molecular Psychiatry 14:10–17.

Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JRB, Rayner, NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney ASF, The Wellcome Trust Case Control Consortium (WTCCC), McCarthy MI, Andrew T, Hattersley AT. 2007. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316:1336–1341.

Zeggini E, Scott LJ, Saxena R, Voight BF, for the Diabetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. 2008. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nature Genetics 40:638–645.

APPENDIX: TECHNICAL DETAILS

We adopt the notation of the Methods section. Let $\widehat{\alpha}_k$ and $\widehat{\beta}_k$ be the maximum likelihood estimates (MLEs) of α_k and β based on the likelihood function of the kth study, and let $\widetilde{\alpha}_k$ and $\widetilde{\beta}$ be the MLEs of α_k and β based on the joint likelihood function. Note that $\widetilde{\beta}$ is the mega-analysis estimate of β. Write $\theta_k = (\alpha_k, \beta)$, $\widehat{\theta}_k = (\widehat{\alpha}_k, \widehat{\beta}_k)$, and $\widetilde{\theta}_k = (\widetilde{\alpha}_k, \widetilde{\beta})$. Also, define

$$I_k(\theta_k) = \sum_{i=1}^{n_k} v_{ki}(\theta_k) X_{ki}^2 - \left\{ \sum_{i=1}^{n_k} v_{ki}(\theta_k) X_{ki} \right\}^2 \Bigg/ \sum_{i=1}^{n_k} v_{ki}(\theta_k),$$

where $v_{ki}(\theta_k) = e^{\alpha_k + \beta X_{ki}}/(1 + e^{\alpha_k + \beta X_{ki}})^2$. According to the MLE theory [Cox and Hinkley, 1979], the variances of $\widehat{\beta}_k$ and $\widetilde{\beta}$ are estimated by $V_k = I_k^{-1}(\widehat{\theta}_k)$ and

$$\mathrm{Var}(\widetilde{\beta}) = \left\{ \sum_{k=1}^{K} I_k(\widetilde{\theta}_k) \right\}^{-1},$$

respectively. The inverse-variance meta-analysis estimate of β is

$$\widehat{\beta} = \left\{ \sum_{k=1}^{K} I_k(\widehat{\theta}_k) \right\}^{-1} \sum_{k=1}^{K} I_k(\widehat{\theta}_k)\, \widehat{\beta}_k, \tag{A1}$$

and its variance is estimated by

$$\mathrm{Var}(\widehat{\beta}) = \left\{ \sum_{k=1}^{K} I_k(\widehat{\theta}_k) \right\}^{-1}.$$

Note that $\mathrm{Var}(\widehat{\beta})$ takes the same form as $\mathrm{Var}(\widetilde{\beta})$: the only difference is that I_k is evaluated at $\widehat{\theta}_k$ in the former and at $\widetilde{\theta}_k$ in the latter. Denote $n = \sum_{k=1}^{K} n_k$. Under model (1) of the Methods section, $\widehat{\alpha}_k$ and $\widetilde{\alpha}_k$ converge to α_k while $\widehat{\beta}_k$ and $\widetilde{\beta}$ converge to β (as sample sizes n_k increase), so that $\widehat{\beta}$ also converges to β while $\mathrm{Var}(n^{1/2}\widehat{\beta})$ and $\mathrm{Var}(n^{1/2}\widetilde{\beta})$ converge to a common constant. Thus, $n^{1/2}(\widehat{\beta} - \beta)$ and $n^{1/2}(\widetilde{\beta} - \beta)$ are asymptotically normal with mean 0 and with a common variance, which implies that meta-analysis and mega-analysis are asymptotically equivalent.

To accommodate covariates, we extend Equation (1) of the Methods section as follows:

$$\Pr(Y_{ki}=1)=\frac{e^{\alpha_k+\beta X_{ki}+\gamma_k^{\mathrm{T}}Z_{ki}}}{1+e^{\alpha_k+\beta X_{ki}+\gamma_k^{\mathrm{T}}Z_{ki}}}, \tag{A2}$$

where Z_{ki} is the vector of covariates for the ith participant of the kth study, and γ_k is the corresponding vector of log odds ratios. By incorporating the unit component into Z_{ki} and the intercept α_k into γ_k, Equation (A2) can be written in a more compact form

$$\Pr(Y_{ki}=1)=\frac{e^{\beta X_{ki}+\gamma_k^{\mathrm{T}}Z_{ki}}}{1+e^{\beta X_{ki}+\gamma_k^{\mathrm{T}}Z_{ki}}}.$$

The likelihood functions given in the Methods section are modified to reflect the inclusion of covariates in the model. Write $\theta_k=(\beta,\gamma_k)$. Let $\widehat{\theta}_k$ and $\widetilde{\theta}_k$ denote the MLEs of θ_k based on the likelihood function of the kth study and the joint likelihood function, respectively. Then all the results of the previous paragraph hold with the redefinition of

$$I_k(\theta_k)=\sum_{i=1}^{n_k}v_{ki}(\theta_k)X_{ki}^2-\left\{\sum_{i=1}^{n_k}v_{ki}(\theta_k)X_{ki}Z_{ki}^{\mathrm{T}}\right\}\times\left\{\sum_{i=1}^{n_k}v_{ki}(\theta_k)Z_{ki}Z_{ki}^{\mathrm{T}}\right\}^{-1}\left\{\sum_{i=1}^{n_k}v_{ki}(\theta_k)X_{ki}Z_{ki}\right\},$$

where $v_{ki}(\theta_k)=e^{\beta X_{ki}+\gamma_k^{\mathrm{T}}Z_{ki}}/(1+e^{\beta X_{ki}+\gamma_k^{\mathrm{T}}Z_{ki}})^2$.

If the effects of covariates are the same across studies, then Equation (A2) becomes

$$\Pr(Y_{ki}=1)=\frac{e^{\alpha_k+\beta X_{ki}+\gamma^{\mathrm{T}}Z_{ki}}}{1+e^{\alpha_k+\beta X_{ki}+\gamma^{\mathrm{T}}Z_{ki}}}. \tag{A3}$$

By expanding X_{ki} to include Z_{ki}, Equation (A3) can be written as

$$\Pr(Y_{ki}=1)=\frac{e^{\alpha_k+\beta^{\mathrm{T}}X_{ki}}}{1+e^{\alpha_k+\beta^{\mathrm{T}}X_{ki}}},$$

in which the vector β represents both the genetic effect and the covariate effects. Redefine

$$I_k(\theta_k)=\sum_{i=1}^{n_k}v_{ki}(\theta_k)X_{ki}X_{ki}^{\mathrm{T}}-\left\{\sum_{i=1}^{n_k}v_{ki}(\theta_k)X_{ki}\right\}\times\left\{\sum_{i=1}^{n_k}v_{ki}(\theta_k)X_{ki}^{\mathrm{T}}\right\}\Big/\sum_{i=1}^{n_k}v_{ki}(\theta_k),$$

where $v_{ki}(\theta_k)=e^{\alpha_k+\beta^{\mathrm{T}}X_{ki}}/(1+e^{\alpha_k+\beta^{\mathrm{T}}X_{ki}})^2$. By the arguments of the first paragraph, $\widehat{\beta}$ and $\widetilde{\beta}$ are asymptotically normal with mean β and with a common covariance matrix. Thus, performing the multivariate version of meta-analysis on the vector of parameters β yields an estimate of the genetic effect that is asymptotically as efficient as the mega-analysis estimate when covariate effects are the same across studies.

Because model (A2) has K sets of covariate effects whereas model (A3) only has one set, mega-analysis is generally more efficient under model (A3) than under model (A2). Thus, univariate meta-analysis, which is asymptotically equivalent to mega-analysis under model (A2), is generally less efficient than mega-analysis under model (A3). However, the efficiency loss is minimal in large samples. Although one can avoid the efficiency loss by performing multivariate meta-analysis, it is more difficult to obtain multivariate than univariate summary statistics.

All the above results assume that the genetic effects are the same across studies. This assumption does not affect the type I error of association testing since all genetic effects are zero under the null hypothesis of no association. Nevertheless, it is of practical importance to determine the relative power of meta-analysis vs. mega-analysis when genetic effects are unequal. By taking the differences between the score functions of $L_k(\alpha_k,\beta)$ and $\prod_{k=1}^{K}L_k(\alpha_k,\beta)$ and applying the mean-value theorem, we can show that

$$\widetilde{\beta}=\left\{\sum_{k=1}^{K}I_k(\theta_k^*)\right\}^{-1}\sum_{k=1}^{K}I_k(\theta_k^*)\,\widehat{\beta}_k,$$

where θ_k^* lies between $\widehat{\theta}_k$ and $\widetilde{\theta}_k$. Thus, $\widetilde{\beta}$ takes the same form as $\widehat{\beta}$ shown in Equation (A1), the difference being that I_k is evaluated at θ_k^* in the former and at $\widehat{\theta}_k$ in the latter. As indicated before, the only difference between $\mathrm{Var}(\widetilde{\beta})$ and $\mathrm{Var}(\widehat{\beta})$ is that I_k is evaluated at $\widetilde{\theta}_k$ in the former and at $\widehat{\theta}_k$ in the latter. Note that I_k depends on θ_k through $v_{ki}(\theta_k)$ only. It can be shown that $v_{ki}(\theta_k)$ does not change its values drastically when θ_k varies between $\widehat{\theta}_k$ and $\widetilde{\theta}_k$ in case-control studies with modest genetic effects. Thus, $\widehat{\beta}$ and $\widetilde{\beta}$ are approximately the same, and so are $\mathrm{Var}(\widehat{\beta})$ and $\mathrm{Var}(\widetilde{\beta})$. Consequently, the power of meta-analysis is similar to that of mega-analysis even when genetic effects are unequal across studies.