American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2014. This work is written by (a) US Government employee(s) and is in the public domain in the US.

Vol. 180, No. 6
DOI: 10.1093/aje/kwu196
Advance Access publication:
August 26, 2014

Editorial

Editorial: Updated Guidance on Human Genome Epidemiology (HuGE) Reviews and Meta-Analyses of Genetic Associations

Marta Gwinn, John P. Ioannidis, Julian Little, and Muin J. Khoury*

* Correspondence to Dr. Muin J. Khoury, Office of Public Health Genomics, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop E-61, Atlanta, GA 30333 (e-mail: MKhoury@cdc.gov).

Abbreviations: *AJE*, *American Journal of Epidemiology*; GWAS, genome-wide association study; HuGE, human genome epidemiology.

Even before the human genome was sequenced, Khoury and Dorman published an editorial in the *American Journal of Epidemiology* (*AJE*) calling for a population-based approach to health-related discoveries that stemmed from the Human Genome Project (1). The approach was termed "human genome epidemiology" (HuGE for short) and led to the formation of an informal collaboration (HuGENet) (2) to explore the systematic use of epidemiologic methods to investigate the role of genetic variation in health and disease. That inaugural editorial outlined the goals of human genome epidemiology, which ranged from estimating the population prevalence of gene variants to evaluating genetic tests and services.

In 2000, the *AJE* proposed a schema for publishing HuGE reviews, which were envisioned as systematic summaries of population-based data on gene-disease associations for use by researchers, health officials, and policymakers (3). Each review would address a particular combination of 1 or more genetic variants and health outcomes, summarizing the available data on genotype prevalence, gene-disease associations, and gene-environment interactions and commenting on its implications for population health. The *AJE* offered to consider HuGE reviews for publication, and 10 more journals followed suit. From 2000 through 2013, the *AJE* published 65 HuGE reviews, and 58 more reviews appeared in other journals. Together, these HuGE reviews assessed the relationships of variants of 195 genes with specific outcomes, including inherited disorders (e.g., sickle-cell anemia), common diseases (e.g., coronary artery disease), several cancers, birth defects, and other conditions.

The surge in genetic association studies during the last decade has been well documented in online databases, including the HuGE Navigator and the GWAS Catalog, which captures genome-wide association studies (GWAS) (4, 5). The number of genetic association studies published annually grew from 2,514 in 2001 to 10,940 in 2013 (http://www.hugenavigator.net). From the beginning, it was clear that many reported genetic associations with common diseases were spurious; furthermore, even consistently replicated associations had mostly small effects (6). Meta-analysis thus emerged as an important tool for assessing gene-disease associations, both for neutralizing reports of spurious associations and for revealing subtle associations. In 2009, HuGENet authors recommended reporting the results of primary genetic association studies in sufficient detail to allow their evaluation for quality and inclusion in systematic reviews (7, 8). Updated HuGE review guidelines recommended the use of meta-analysis to arrive at summary estimates of association, as well as a heuristic to help gauge their reliability (the "Venice criteria") (9).

In 2009, Minelli et al. (10) published a systematic review of meta-analyses of published genetic association studies. They found that general methodological problems—such as failure to document search strategy, account for publication bias, or test for heterogeneity—were common, although no more common than among meta-analyses in other fields. However, methodological considerations specific to genetics—such as Hardy-Weinberg equilibrium, genotype frequency, choice of genetic model, and population stratification—were often poorly addressed or completely ignored. The authors concluded with a set of practical recommendations for the conduct and reporting of such meta-analyses.

Although meta-analyses of genetic associations account for only a small proportion of publications on gene-disease associations, they have proliferated rapidly from only 29 in 2001 to 1,606 in 2013. During that time, the genetic association research strategy shifted focus from candidate gene studies that evaluated 1 or a few polymorphisms to GWAS for the discovery of new candidate genes; recent years have seen an increase in studies of rare variants in both common and

rare diseases. Meta-analyses remain relevant to all of these approaches; however, ongoing concerns about their quality raise questions about their contribution to the field (11, 12). The editors of *PLOS ONE* recently cited these concerns when outlining a new policy regarding meta-analyses of genetic association studies submitted for publication (13). Authors of such manuscripts must provide a rationale for the meta-analysis, including relevant context; refer to relevant GWAS and indicate whether they are included in the analysis; and document that methodological points have been addressed, using a checklist based in part on earlier guidelines (7, 8, 10).

The publication of duplicate meta-analyses has become an issue in human genome epidemiology, as well as in other fields (14). Although some independent replication is useful, unnecessary redundancy is wasteful. A 2013 editorial in the *BMJ* stated that systematic reviews "should identify existing reviews as a compulsory first step" (15, p. 1), adding further clarity to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, which asks authors of systematic reviews and meta-analyses to "describe the rationale for the review in the context of what is already known" (16, p. 4).

For meta-analyses of genetic associations, this context should include both candidate gene analyses and relevant GWAS. Integrating results from these 2 approaches is not always straightforward because their goals are somewhat different: GWAS search for new associations (assuming uniformly low prior probability), whereas candidate gene analyses examine associations suggested by knowledge of biological pathways or other prior information. The single nucleotide polymorphism markers examined in GWAS, particularly in earlier studies, may not include candidate functional variants, and imputation is not always possible (17). Because GWAS publications tend to report only statistically significant associations, meta-analyses limited to published data may omit relevant negative GWAS findings.

Few studies have systematically compared the results of GWAS and candidate gene association studies of a specific phenotype. One systematic review of genetic associations with cancer calculated the false positive reporting probability for each association, given either of 2 levels of prior probability (0.001 or 0.000001) and odds ratio (1.2 or 1.5) (18). This analysis applied a threshold of false positive reporting probability of 0.2 or less to identify 163 unique associations reported from GWAS and 66 reported from candidate gene analyses; only 27 (13%) of all associations were reported in both groups, usually with similar effect sizes. The increasing use of high-resolution microarrays and sequencing approaches in GWAS should enable more direct comparisons with studies of candidate gene variants.

Methods for integrating results from different types of studies and for establishing appropriate significance thresholds are needed to support the use of meta-analysis in human genome epidemiology (19). Sound meta-analyses of genetic associations are fundamental to higher-level epidemiologic synthesis, such as in field synopses that summarize all well-established genetic associations with a specific disease (20). Poorly conducted, incomplete, or redundant meta-analyses create noise, undermining rather than improving our ability to detect genetic signals in epidemiologic studies.

The *AJE* endorses and extends the *PLOS ONE* approach to setting basic standards for meta-analyses of genetic associations, including HuGE reviews. Any such manuscript submitted to the *Journal* must describe the rationale for conducting the review in the context of related publications. The manuscript should also provide full methodological details, including whether and how GWAS or other agnostic (e.g., sequencing) data have been incorporated (see Appendix). We ask authors who submit a meta-analysis of genetic associations to document that they have supplied all the necessary methodological information by completing the checklist that is available online in the Web Appendix (available at http://aje.oxfordjournals.org/). Editors of the *Journal* will evaluate the information supplied by the authors when deciding whether to forward HuGE review manuscripts for peer review. Reviewers of meta-analyses will be encouraged to use information supplied in the checklist to evaluate methodological quality.

As advances in genomic and other omic technologies continue to reveal the complexity of human genetic variation in increasing detail, epidemiologists can help researchers, health-care providers, and policymakers keep their eye on the ball. Without population-based evidence—summarized in efficient, high-quality meta-analyses—we won't know whether, how, or how much genomic information has the potential to improve population health.

ACKNOWLEDGMENTS

Author affiliations: McKing Consulting Corporation, Atlanta, Georgia (Marta Gwinn); Stanford Prevention Research Center, Stanford University, Palo Alto, California (John P. Ioannidis); Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada (Julian Little); and Office of Public Health Genomics/CSELS/OPHSS, Centers for Disease Control and Prevention, Atlanta, Georgia (Muin J. Khoury).

Conflict of interest: none declared.

REFERENCES

1. Khoury MJ, Dorman JS. The human genome epidemiology network. *Am J Epidemiol*. 1998;148(1):1–3.
2. Centers for Disease Control and Prevention. Human Genome Epidemiology Network (HuGENet). http://www.cdc.gov/genomics/hugenet/. Updated January 29, 2013. Accessed April 18, 2014.
3. Khoury MJ, Little J. Human genome epidemiologic reviews: the beginning of something HuGE. *Am J Epidemiol*. 2000;151(1):2–3.
4. Yu W, Gwinn M, Clyne M, et al. A navigator for human genome epidemiology. *Nat Genet*. 2008;40(2):124–125.
5. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(Database issue):D1001–D1006.
6. Ioannidis JP, Ntzani EE, Trikalinos TA, et al. Replication validity of genetic association studies. *Nat Genet*. 2001;29(3):306–309.

7. Sagoo GS, Little J, Higgins JP. Systematic reviews of genetic association studies. Human Genome Epidemiology Network. *PLOS Med*. 2009;6(3):e28.
8. Little J, Higgins JPT, Ioannidis JPA, et al. STrengthening the REporting of Genetic Association Studies (STREGA)—an Extension of the STROBE Statement. *PLOS Med*. 2009;6(2): e1000022.
9. Ioannidis JP, Boffetta P, Little J, et al. Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol*. 2008;37(1):120–132.
10. Minelli C, Thompson JR, Abrams KR, et al. The quality of meta-analyses of genetic association studies: a review with recommendations. *Am J Epidemiol*. 2009;170(11):1333–1343.
11. Ioannidis JP, Chang CQ, Lam TK, et al. The geometric increase in meta-analyses from China in the genomic era. *PLOS One*. 2013;8(6):e65602.
12. Osnabrugge RL, Head SJ, Zijlstra F, et al. A systematic review and critical assessment of 11 discordant meta-analyses on reduced-function CYP2C19 genotype and risk of adverse clinical outcomes in clopidogrel users [published online ahead of print]. *Genet Med*. (doi:10.1038/gim.2014.76).
13. PLOS ONE Editors. Meta-analyses of genetic association studies – PLOS ONE's approach. The PLOS ONE Community Blog. http://blogs.PLOS.org/everyone/2014/04/04/meta-analyses-genetic-association-studies-PLOS-ones-approach/. Published April 4, 2014. Accessed June 16, 2014.
14. Siontis KC, Hernandez-Boussard T, Ioannidis JP. Overlapping meta-analyses on the same topic: survey of published studies. *BMJ*. 2013;347:f4501.
15. Moher D. The problem of duplicate systematic reviews. *BMJ*. 2013;347:f5040.
16. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Med*. 2009;6(7):e1000097.
17. Duan Q, Liu EY, Croteau-Chonka DC, et al. A comprehensive SNP and indel imputability database. *Bioinformatics*. 2013; 29(4):528–531.
18. Chang CQ, Yesupriya A, Rowell JL, et al. A systematic review of cancer GWAS and candidate gene meta-analyses reveals limited overlap but similar effect sizes. *Eur J Hum Genet*. 2014; 22(3):402–408.
19. Broer L, Lill CM, Schuur M. Distinguishing true from false positives in genomic studies: p values. *Eur J Epidemiol*. 2013; 28(2):131–138.
20. Theodoratou E, Montazeri Z, Hawken S, et al. Systematic meta-analyses and field synopsis of genetic association studies in colorectal cancer. *J Natl Cancer Inst*. 2012;104(19):1433–1457.

APPENDIX

Author guidelines for meta-analyses and HuGE Reviews of genetic associations submitted to the *American Journal of Epidemiology*

The following must be addressed within the text of the manuscript:

1. The rationale for the meta-analysis.
   - What is the scope of the meta-analysis? What genes and variants are included? What range of phenotypes is considered?
   - How is meta-analysis of this genetic association relevant to epidemiologic research?
2. The contribution that the meta-analysis makes to existing knowledge.
   - What other relevant meta-analyses or systematic reviews have been published? How were these identified? What meta-analyses have considered other variants in the same gene? What meta-analyses have considered related phenotypes?
   - Why is (another) meta-analysis needed? Have important data (published or unpublished) become newly available? Is there a priori interest in analysis of specific population subgroups?
   - Were previous meta-analyses methodologically flawed?
3. Whether GWAS (or other agnostic platform data, e.g., sequencing) relevant to the meta-analysis have been published and whether and how these were included in the analysis.
   - What GWAS or other agnostic data have examined the phenotype of interest?
   - Have any of these GWAS or other agnostic data reported associations with the same genetic variant? How are these data included in the meta-analysis?
   - Are unpublished data available from relevant GWAS and other agnostic data? Have these been requested and included in the meta-analysis? If specific genotypic data on the variant of interest are not available, can data be inferred from GWAS variants in linkage disequilibrium with the variant of interest?