American Journal of Epidemiology
Vol. 156, No. 4

Printed in U.S.A.
DOI: 10.1093/aje/kwf053

HUMAN GENOME EPIDEMIOLOGY

Commentary: Epidemiology and the Continuum from Genetic Research to Genetic Testing

Muin J. Khoury

From the Office of Genomics and Disease Prevention, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA.

Received for publication September 5, 2001; accepted for publication June 13, 2002.

Abbreviation: HuGE, Human Genome Epidemiology.

In January 2001, the Centers for Disease Control and Prevention and the National Institutes of Health convened an expert panel to develop recommendations for evaluating and synthesizing data from epidemiologic studies of the human genome. Experts in medicine, genetics, epidemiology, statistics, laboratory sciences, prevention effectiveness, and the social sciences discussed examples drawn from cancer, cardiovascular disease, human immunodeficiency virus infection, and other areas. Participants discussed issues for evaluating and synthesizing data from epidemiologic and genetic test studies (table 1) relevant to three areas: 1) prevalence of gene variants and gene-disease associations, 2) gene-environment and gene-gene interactions, and 3) evaluation of genetic tests. The workshop recommendations in areas 1 and 3 are included in this issue of the *Journal* (1, 2). The recommendations of area 2 (gene-environment interaction) are in progress (D. J. Hunter, Channing Laboratory, unpublished manuscript). I summarize the meeting's background and highlight the importance of the panel's recommendations.

Many scientists believe that advances in human genetics and the Human Genome Project will play a central role in the practice of medicine and public health in the 21st century by predicting and preventing disease and promoting health (3). However, to ensure a systematic translation of genetic research into clinical practice, ongoing epidemiologic data are needed, in addition to studies of gene function and biologic pathways, to quantify the impact of gene variants on the risk of various diseases and to identify and quantify the impact of modifiable risk factors that interact with gene variants (4). So far, most studies in this area come from family-based studies or highly selected groups. Results from population-based studies will help medical and public health professionals better target medical, behavioral, and environmental interventions.

A systematic application of epidemiologic methods and approaches to the human genome—HuGE (4)—represents the continuum from gene discovery (traditional domain of genetic epidemiology) to risk characterization (domain of molecular epidemiology) and evaluation of genetic tests and services (applied epidemiology and health services research). As a multidisciplinary field, epidemiology has begun to address issues related to post-gene discovery with increasing emphasis on characterization of gene effects and genetic tests in populations (what do you do with a gene after you find one?). The continuum of studies can be divided into the three areas that are the topics of the workshop papers: 1) assessing the population prevalence of gene variants and evaluating genotype-disease associations; 2) assessing the impact of gene-environment and gene-gene interaction on disease risk; and 3) evaluating the usefulness and impact of genetic tests in populations. Because of the numerous genes that are discovered on a regular basis, an epidemiologic approach is needed for all three study domains. An analysis of the published epidemiologic literature on human genes for 2001 reveals that, of the 2,042 published articles, most reported on only the population prevalence of gene variants or simple gene-disease associations (82.0 percent), while 14.5 percent integrated the study of interactions (gene-gene and gene-environment) and only 3.5 percent dealt with evaluation of genetic tests (5). Epidemiologic studies of gene-environment interaction and genetic tests are bound to increase as more genes are discovered, characterized, and used to develop diagnostic and predictive tests.

Correspondence to Dr. Muin J. Khoury, Office of Genomics and Disease Prevention, National Center for Environmental Health, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341 (e-mail: muk1@cdc.gov).

TABLE 1. Issues to consider in evaluating and integrating epidemiologic studies of human genes and genetic test data

What are the data elements needed to evaluate and integrate data?
Prevalence and genotype-disease associations
Gene-gene and gene-environment interactions
Genetic tests
What should be the methodological standards for reporting individual studies?
Reproducibility
Objectivity
Case definition (delineation and adequacy of)
Comparison (control) group definition (delineation and adequacy of)
Quantitative summary of results (how do we deal with potential confounding from population stratification?)
What should be the recommendations for synthesis and grading evidence from published and unpublished data?
Reporting of background (objective, study design, outcome measures, etc.)
Methods (including investigating heterogeneity)
Results
Discussion
Conclusion (including identifying gaps, future research)
What should be the recommendations for summarizing and presenting data on Web sites and databases?
Guidelines/inclusion criteria for studies
Design format for individual studies
Updating with new information
Addressing unpublished data

NEED FOR STANDARDS FOR REPORTING AND SYNTHESIS OF EPIDEMIOLOGY AND GENETIC TEST DATA

Although the need for epidemiologic studies on human genes is acknowledged, many such studies are based on small numbers and use convenient control subjects. Issues of validity and reliability of genotyping methods need to be addressed. In addition, special attention needs to be paid to ethical and informed consent guidelines for the conduct of such studies (5). Several authors have conducted systematic, peer-reviewed synopses of epidemiologic aspects of human genes, prevalence of allelic variants in different populations, population-based disease risk information, gene-environment interaction, and quantitative data on genetic tests and services (6). These reviews have uncovered the need for unified guidelines that can be used to synthesize results of the increasing number of such studies (7). Although several groups have addressed guidelines for the evaluation and synthesis of a number of areas (e.g., controlled clinical trials), no such recommendations exist that cover the spectrum of epidemiologic studies of the human genome. In an analysis of the epidemiologic quality of molecular genetic research, Bogardus et al. (8) used seven methodological standards to evaluate the quality of studies in four mainstream medical journals. They found that, in spite of the major molecular genetic advances, 63 percent of the articles did not comply with two or more quality standards. This finding emphasizes the need for methodological standards in reporting such studies. Based on an expert panel workshop held in 1997, Stroup et al. (9) published a proposal for reporting results of meta-analysis of observational studies in epidemiology but did not specifically address genetic studies. Bruns et al. (10) provided a checklist for the reporting of studies on the diagnostic accuracy of medical tests. A workshop sponsored by the National Cancer Institute led to a monograph on innovative study designs and analytic approaches to the genetic epidemiology of cancer (11). This series of articles was useful in outlining the spectrum of study designs in gene discovery and characterization in relation to disease, but it does not provide concrete guidance on the evaluation and synthesis of such studies.

Moreover, there is little discussion about epidemiologic approaches to evaluating genetic tests. Many of the genetic tests that will emerge in the next decades will be used not only for diagnostic purposes but also to predict the risk of developing disease in otherwise healthy people and to make decisions about potentially preventive interventions or therapies. The use of genetic tests in this context will depend heavily on the quality of epidemiologic information that summarizes the relation between genotypes and disease and how such relation is modulated by the presence of other factors, such as drugs and environmental exposures. This is clearly illustrated in the hypothetical case scenario shown by Collins (12) of a man aged 23 years who is receiving the results of various genetic tests in the year 2010. This person's genetic test report included probabilistic information on the risk of various diseases for genotypic combinations at several loci. Such information will have to be based on properly designed epidemiologic information on genotype-disease associations and gene-gene and gene-environment interaction and on how the risk of these diseases can be reduced using different interventions.

To address the need for systematic analysis of genetic tests, the Foundation for Blood Research (13) is developing a model approach for assembling, analyzing, disseminating, and updating existing data on the safety and effectiveness of DNA-based genetic tests and testing algorithms. Over a 3-year period, up to 10 tests for different disorders will be evaluated for analytic validity, clinical validity, clinical utility, and related ethical, legal, and social issues. The goal of this effort is to design, test, and validate a working model that can be used to collect and interpret data on genetic tests to provide a basis for transition from genetic discoveries to clinical practice.

With the expected increase in the number of genetic tests in the next decade, we hope that the accompanying papers provide guidance for researchers conducting, reporting, and reviewing the results of epidemiologic studies involving human gene variants and also for producers of data on

genetic tests. Ultimately, we hope that consumers of such information (e.g., policy makers, clinicians, public health practitioners, and the general public) will become more savy in making sense of how information on genes and genetic tests can be used to improve health and prevent disease.

REFERENCES

1. Little J, Bradley L, Bray MS, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am J Epidemiol 2002;156:300–10**.**
2. Burke W, Atkins D, Gwinn M, et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. Am J Epidemiol 2002;156:311–18**.**
3. Collins FS, McKusick VM. Implications of the Human Genome Project for medical science. JAMA 2001;285:540–4.
4. Khoury MJ, Dorman JS. The Human Genome Epidemiology Network. (Editorial). Am J Epidemiol 1998;148:1–3.
5. CDC Office of Genomics and Disease Prevention. The Human Genome Epidemiology Network. Accessed April 8, 2002. (http://www.cdc.gov/genomics/hugenet/default.htm).
6. Beskow LM, Burke W, Merz JF, et al. Informed consent for population-based research involving genetics. JAMA 2001; 286:2315–21.
7. Revised guidelines for submitting HuGE reviews. Am J Epidemiol 2000:151:4–6.
8. Bogardus ST, Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: the need for methodologic standards. JAMA 1999;281:1919–26.
9. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008–12.
10. Bruns DE, Huth EJ, Magid E, et al. Toward a checklist for reporting of studies of diagnostic accuracy of medical tests. Clin Chem 2000;46:893–5.
11. Innovative study designs and analytic approaches to the genetic epidemiology of cancer. J Natl Cancer Inst Monogr 1999;26:1–105.
12. Collins FS. Shattuck lecture—medical and societal consequences of the Human Genome Project. N Engl J Med 1999; 341:28–37.
13. Centers for Disease Control and Prevention. Model system for collecting, analyzing and disseminating information on genetic tests. Accessed August 7, 2001. (http://www.cdc.gov/genomics/activities/fbr.htm).

American Journal of Epidemiology
Vol. 156, No. 4

Printed in U.S.A.
DOI: 10.1093/aje/kwf053

HUMAN GENOME EPIDEMIOLOGY

Commentary: Epidemiology and the Continuum from Genetic Research to Genetic Testing

Muin J. Khoury

From the Office of Genomics and Disease Prevention, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA.

Received for publication September 5, 2001; accepted for publication June 13, 2002.

Abbreviation: HuGE, Human Genome Epidemiology.

In January 2001, the Centers for Disease Control and Prevention and the National Institutes of Health convened an expert panel to develop recommendations for evaluating and synthesizing data from epidemiologic studies of the human genome. Experts in medicine, genetics, epidemiology, statistics, laboratory sciences, prevention effectiveness, and the social sciences discussed examples drawn from cancer, cardiovascular disease, human immunodeficiency virus infection, and other areas. Participants discussed issues for evaluating and synthesizing data from epidemiologic and genetic test studies (table 1) relevant to three areas: 1) prevalence of gene variants and gene-disease associations, 2) gene-environment and gene-gene interactions, and 3) evaluation of genetic tests. The workshop recommendations in areas 1 and 3 are included in this issue of the *Journal* (1, 2). The recommendations of area 2 (gene-environment interaction) are in progress (D. J. Hunter, Channing Laboratory, unpublished manuscript). I summarize the meeting's background and highlight the importance of the panel's recommendations.

Many scientists believe that advances in human genetics and the Human Genome Project will play a central role in the practice of medicine and public health in the 21st century by predicting and preventing disease and promoting health (3). However, to ensure a systematic translation of genetic research into clinical practice, ongoing epidemiologic data are needed, in addition to studies of gene function and biologic pathways, to quantify the impact of gene variants on the risk of various diseases and to identify and quantify the impact of modifiable risk factors that interact with gene variants (4). So far, most studies in this area come from family-based studies or highly selected groups. Results from population-based studies will help medical and public health professionals better target medical, behavioral, and environmental interventions.

A systematic application of epidemiologic methods and approaches to the human genome—HuGE (4)—represents the continuum from gene discovery (traditional domain of genetic epidemiology) to risk characterization (domain of molecular epidemiology) and evaluation of genetic tests and services (applied epidemiology and health services research). As a multidisciplinary field, epidemiology has begun to address issues related to post-gene discovery with increasing emphasis on characterization of gene effects and genetic tests in populations (what do you do with a gene after you find one?). The continuum of studies can be divided into the three areas that are the topics of the workshop papers: 1) assessing the population prevalence of gene variants and evaluating genotype-disease associations; 2) assessing the impact of gene-environment and gene-gene interaction on disease risk; and 3) evaluating the usefulness and impact of genetic tests in populations. Because of the numerous genes that are discovered on a regular basis, an epidemiologic approach is needed for all three study domains. An analysis of the published epidemiologic literature on human genes for 2001 reveals that, of the 2,042 published articles, most reported on only the population prevalence of gene variants or simple gene-disease associations (82.0 percent), while 14.5 percent integrated the study of interactions (gene-gene and gene-environment) and only 3.5 percent dealt with evaluation of genetic tests (5). Epidemiologic studies of gene-environment interaction and genetic tests are bound to increase as more genes are discovered, characterized, and used to develop diagnostic and predictive tests.

Correspondence to Dr. Muin J. Khoury, Office of Genomics and Disease Prevention, National Center for Environmental Health, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341 (e-mail: muk1@cdc.gov).

TABLE 1. Issues to consider in evaluating and integrating epidemiologic studies of human genes and genetic test data

What are the data elements needed to evaluate and integrate data?
Prevalence and genotype-disease associations
Gene-gene and gene-environment interactions
Genetic tests
What should be the methodological standards for reporting individual studies?
Reproducibility
Objectivity
Case definition (delineation and adequacy of)
Comparison (control) group definition (delineation and adequacy of)
Quantitative summary of results (how do we deal with potential confounding from population stratification?)
What should be the recommendations for synthesis and grading evidence from published and unpublished data?
Reporting of background (objective, study design, outcome measures, etc.)
Methods (including investigating heterogeneity)
Results
Discussion
Conclusion (including identifying gaps, future research)
What should be the recommendations for summarizing and presenting data on Web sites and databases?
Guidelines/inclusion criteria for studies
Design format for individual studies
Updating with new information
Addressing unpublished data

NEED FOR STANDARDS FOR REPORTING AND SYNTHESIS OF EPIDEMIOLOGY AND GENETIC TEST DATA

Although the need for epidemiologic studies on human genes is acknowledged, many such studies are based on small numbers and use convenient control subjects. Issues of validity and reliability of genotyping methods need to be addressed. In addition, special attention needs to be paid to ethical and informed consent guidelines for the conduct of such studies (5). Several authors have conducted systematic, peer-reviewed synopses of epidemiologic aspects of human genes, prevalence of allelic variants in different populations, population-based disease risk information, gene-environment interaction, and quantitative data on genetic tests and services (6). These reviews have uncovered the need for unified guidelines that can be used to synthesize results of the increasing number of such studies (7). Although several groups have addressed guidelines for the evaluation and synthesis of a number of areas (e.g., controlled clinical trials), no such recommendations exist that cover the spectrum of epidemiologic studies of the human genome. In an analysis of the epidemiologic quality of molecular genetic research, Bogardus et al. (8) used seven methodological standards to evaluate the quality of studies in four mainstream medical journals. They found that, in spite of the major molecular genetic advances, 63 percent of the articles did not comply with two or more quality standards. This finding emphasizes the need for methodological standards in reporting such studies. Based on an expert panel workshop held in 1997, Stroup et al. (9) published a proposal for reporting results of meta-analysis of observational studies in epidemiology but did not specifically address genetic studies. Bruns et al. (10) provided a checklist for the reporting of studies on the diagnostic accuracy of medical tests. A workshop sponsored by the National Cancer Institute led to a monograph on innovative study designs and analytic approaches to the genetic epidemiology of cancer (11). This series of articles was useful in outlining the spectrum of study designs in gene discovery and characterization in relation to disease, but it does not provide concrete guidance on the evaluation and synthesis of such studies.

Moreover, there is little discussion about epidemiologic approaches to evaluating genetic tests. Many of the genetic tests that will emerge in the next decades will be used not only for diagnostic purposes but also to predict the risk of developing disease in otherwise healthy people and to make decisions about potentially preventive interventions or therapies. The use of genetic tests in this context will depend heavily on the quality of epidemiologic information that summarizes the relation between genotypes and disease and how such relation is modulated by the presence of other factors, such as drugs and environmental exposures. This is clearly illustrated in the hypothetical case scenario shown by Collins (12) of a man aged 23 years who is receiving the results of various genetic tests in the year 2010. This person's genetic test report included probabilistic information on the risk of various diseases for genotypic combinations at several loci. Such information will have to be based on properly designed epidemiologic information on genotype-disease associations and gene-gene and gene-environment interaction and on how the risk of these diseases can be reduced using different interventions.

To address the need for systematic analysis of genetic tests, the Foundation for Blood Research (13) is developing a model approach for assembling, analyzing, disseminating, and updating existing data on the safety and effectiveness of DNA-based genetic tests and testing algorithms. Over a 3-year period, up to 10 tests for different disorders will be evaluated for analytic validity, clinical validity, clinical utility, and related ethical, legal, and social issues. The goal of this effort is to design, test, and validate a working model that can be used to collect and interpret data on genetic tests to provide a basis for transition from genetic discoveries to clinical practice.

With the expected increase in the number of genetic tests in the next decade, we hope that the accompanying papers provide guidance for researchers conducting, reporting, and reviewing the results of epidemiologic studies involving human gene variants and also for producers of data on

genetic tests. Ultimately, we hope that consumers of such information (e.g., policy makers, clinicians, public health practitioners, and the general public) will become more savy in making sense of how information on genes and genetic tests can be used to improve health and prevent disease.

REFERENCES

1. Little J, Bradley L, Bray MS, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am J Epidemiol 2002;156:300–10**.**
2. Burke W, Atkins D, Gwinn M, et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. Am J Epidemiol 2002;156:311–18**.**
3. Collins FS, McKusick VM. Implications of the Human Genome Project for medical science. JAMA 2001;285:540–4.
4. Khoury MJ, Dorman JS. The Human Genome Epidemiology Network. (Editorial). Am J Epidemiol 1998;148:1–3.
5. CDC Office of Genomics and Disease Prevention. The Human Genome Epidemiology Network. Accessed April 8, 2002. (http://www.cdc.gov/genomics/hugenet/default.htm).
6. Beskow LM, Burke W, Merz JF, et al. Informed consent for population-based research involving genetics. JAMA 2001; 286:2315–21.
7. Revised guidelines for submitting HuGE reviews. Am J Epidemiol 2000:151:4–6.
8. Bogardus ST, Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: the need for methodologic standards. JAMA 1999;281:1919–26.
9. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008–12.
10. Bruns DE, Huth EJ, Magid E, et al. Toward a checklist for reporting of studies of diagnostic accuracy of medical tests. Clin Chem 2000;46:893–5.
11. Innovative study designs and analytic approaches to the genetic epidemiology of cancer. J Natl Cancer Inst Monogr 1999;26:1–105.
12. Collins FS. Shattuck lecture—medical and societal consequences of the Human Genome Project. N Engl J Med 1999; 341:28–37.
13. Centers for Disease Control and Prevention. Model system for collecting, analyzing and disseminating information on genetic tests. Accessed August 7, 2001. (http://www.cdc.gov/genomics/activities/fbr.htm).