Science, medicine, and the future: Genetic epidemiology

Jaakko Kaprio

BMJ 2000;320:1257-1259
doi:10.1136/bmj.320.7244.1257

Updated information and services can be found at:
http://bmj.com/cgi/content/full/320/7244/1257

These include:

References
This article cites 19 articles, 6 of which can be accessed free at:
http://bmj.com/cgi/content/full/320/7244/1257#BIBL

2 online articles that cite this article can be accessed at:
http://bmj.com/cgi/content/full/320/7244/1257#otherarticles

Rapid responses
You can respond to this article at:
http://bmj.com/cgi/eletter-submit/320/7244/1257

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Topic collections
Articles on similar topics can be found in the following collections

Genetics (3926 articles)

Notes

To order reprints of this article go to:
http://www.bmjjournals.com/cgi/reprintform

To subscribe to BMJ go to:
http://bmj.bmjjournals.com/subscriptions/subscribe.shtml
Clinical review

Science, medicine, and the future
Genetic epidemiology
Jaakko Kaprio

Research in disease aetiology has shifted towards investigating genetic causes, powered by the human genome project. Successful identification of genes for monogenic disease has led to interest in investigating the genetic component of diseases that are often termed complex—that is, they are known to aggregate in families but do not segregate in a mendelian fashion. Genetic epidemiology has permitted identification of genes affecting people’s susceptibility to disease, although progress has been much slower than many people expected. While the role of genetic factors in diseases such as hypertension, asthma, and depression is being intensively studied, family studies and the large geographical and temporal variation in the occurrence of many diseases indicate a major role of the environment. Thus, it is necessary to consider findings about susceptibility genes in the context of a population and evaluate the role of genetic factors in relation to other aetiological factors. This article discusses some approaches used to resolve the genetic architecture of disease and to study the relation of genes to environmental factors in the population.

Methods
I used peer reviewed publications and selected reviews as the main information sources for this article.

What is genetic epidemiology?
Genetic epidemiology is the study of the aetiology, distribution, and control of disease in groups of relatives and of inherited causes of disease in populations. From its parent disciplines of genetics and epidemiology, it has inherited the key elements of studying defined populations while investigating the roles of genes and the environment in relation to each other and endeavouring to account for the known biology of diseases. Quantifying the risk associated with genetic variation is a prerequisite for assessing the use of this new knowledge in medicine.

Is there a genetic component to disease?
The primary goal of genetic epidemiology is resolving the genetic architecture of a disease—that is, establishing whether it has a genetic component and the relative size of that genetic effect in relation to environmental effects. In this context the environment is understood to encompass everything non-genetic, from the intrauterine environment to physical and chemical effects and to behavioural and social aspects. Effects from different environmental categories are insufficiently taken into account in most genetic epidemiological studies.

The estimation of the genetic component comes from family studies, in which the disease risk in relatives of a patient is compared with the general risk of disease in the population (see box). Naturally, it is important that the patients studied are representative of the population. However, an increased risk in family members does not necessarily indicate that the disease has an inherited component accounted for by genetic variation. Familial aggregation can be due to non-genetic factors in the family environment, such as the physical environment of the home and the family’s socioeconomic status. Interindividual differences in religiosity, a protective factor against alcohol misuse in many societies, are largely accounted for by non-genetic familial factors. Stratification of risk by degree of relatedness (first degree versus second degree relatives) and comparisons with unrelated individuals living in the same household (typically spouses) can help distinguish between genetic and non-genetic familial effects. A thorough family history provides excellent information about possible genetic risk in families.

Families with several diseased family members, particular those with large pedigrees, are particularly informative, both for establishing that genes matter.
and for identifying the specific genes. Such families are rare for the common diseases now at the centre of genetic epidemiological research. Other traditional designs for distinguishing non-genetic family effects from genetic effects have been studies of twins and adoptees (see box), but study of half-sibs, who are increasingly common with higher divorce rates, is also valuable. Combinations of designs, such as the inclusion of parents and sibs in twin studies, can permit more incisive estimation of the role of genetic factors and account for assortative mating and transmission of non-genetic effects from parents to offspring.

A model of complex disease

After the size of a genetic component has been established, we seek to establish how many genes are contributing to the disease. In complex diseases (see figure) many genes act through several intermediate phenotypes to increase disease risk, but the same genes can also influence other diseases. Environmental factors can independently affect the risk of disease, but also act through the intermediate phenotypes. For diseases such as coronary heart disease, we know something of the genetics of such intermediate phenotypes such as blood lipids, haemostatic factors, and blood pressure.

In any single gene conferring disease susceptibility there are generally multiple alleles that affect disease risk to different degrees. For example, the cystic fibrosis gene has over 800 mutations associated with the disease. A decade of research has indicated that the genotype poorly predicts phenotype,10 bringing new complexity to the diagnosis of cystic fibrosis while permitting identification of carriers of the cystic fibrosis gene. Such multiplicity of mutations and disease associated alleles is more the rule than the exception. Also, mutations in the cystic fibrosis gene are associated with other phenotypes such as male infertility and allergic bronchopulmonary aspergillosis.10 For other diseases, multiple genes are known to be involved. Migraine has been shown to have a genetic component in family and twin studies,12 but identification of migraine genes has so far been restricted to a rare subtype of migraine, familial hemiplegic migraine. Calcium channel genes on chromosomes 1 and 19 account for many but not all cases of familial hemiplegic migraine. To complicate matters further, not even all family members with a mutation have manifest migraine, and these disease mutations have not been convincingly associated with the common forms of migraine.12

While the findings in subtypes of common disease are invaluable as clues to the molecular pathology of the disease, the complexity of single gene disorders reminds us that progress in identifying susceptibility genes will probably be slow. Some single gene effects have been found for common diseases, and others will probably be found in families with multiple cases. However, findings in hypertension,13 breast cancer,14 and colorectal cancer15 suggest that known genes account for only a fraction of the estimated genetic component. This may be because there are genes of large effect yet to be identified, but it is more likely to be because genetic susceptibility is due to multiple genes of small effects, gene-gene interactions, and gene-environment interactions of complex nature that are difficult to assess, at least in humans, with current study designs. Also many risk factors for disease—such as smoking, alcohol consumption, obesity, and physical inactivity—aggregate in families, and genetic factors are partly responsible for that familial aggregation.16 17

Complex interactions are probably important in explaining differences in disease prevalence among populations. For example, population based twin studies in the Nordic countries in the 1990s suggest that the heritability of asthma is about 70%.18 Strictly comparable earlier studies are not available, but twin studies from the 1970s suggested that the heritability was under 50%, and at the same time asthma has increased in prevalence. While susceptibility genes for asthma cannot have changed in the population during one generation, their expression and interaction with environmental factors may have changed, and may be reversed if the appropriate environmental factors can be identified and eliminated.

Clinical and public health implications of identifying disease genes

The frequency of the alleles of a disease gene and the allele effect size are informative about the impact of a specific allele on disease risk in an individual and in a population. Most allelic variation has relatively small effects on disease risk and is thus of little use clinically by itself.

For example, in the 1980s it was established that a person with the epsilon 4 allele of the apo E gene has,
on average, slightly higher serum cholesterol concentrations than people without the allele. At the population level, variability in apo E accounts for a substantial proportion (about 7%) of the variability in cholesterol levels. In contrast, mutations in the low density lipoprotein receptor have large and clinically important effects on individual cholesterol levels even though their impact on population variability is small because they are rare. In the 1990s people with the apoE e4 allele were also found to have a higher risk of Alzheimer’s disease, but not sufficiently so to be of diagnostic value, and a recent study reported that recurrence of intracerebral haemorrhage is higher in subjects with the e2 or e4 allele compared with those with the e3/e3 genotype. In contrast, the frequency of e4 allele was lower than expected in patients with neoplasia of the proximal colon.

The factor V Leiden mutation increases risk of venous thrombosis about seven or eight times, yet it is only one factor among many genetic and acquired causes, and there is evidence for both gene-gene and gene-environment interactions. Also, screening for the factor V Leiden mutation before prescribing oral contraceptives to prevent venous thrombosis is not cost effective.

Many reported associations between a gene and disease are not consistently replicated, and then both meta-analyses and large studies are needed to establish their true existence (for reviews see the Human Genome Epidemiology Network, www.cdc.gov/genetics/hugenet/). The association of the I/D polymorphism of the gene for angiotensin converting enzyme with myocardial infarction has provided conflicting results from mostly small studies. Now, a study of 4629 cases of myocardial infarction and 5034 controls (data from the ISIS-3 trial) has provided a relative risk of 1.10 (95% confidence interval 1.00 to 1.21) for the DD genotype.

As a person’s genotype cannot (yet) be changed, primary and secondary prevention strategies to reduce risk would have to be based on influencing modifiable risk factors. This presupposes that the relation of other risk factors to disease risk is the same among people with and without the susceptibility genotype. But this may not always be the case—as suggested by the finding that smoking may reduce breast cancer risk among carriers of the BRCA1 and BRCA2 gene mutations and indicating the need for careful epidemiological studies. However, knowledge of specific genes for some forms of common disease has brought many new patients and families for genetic counselling.

Future

Studies of large populations, especially of genetically and culturally homogeneous ones, have been proposed because we are still far from identifying susceptibility genes for common diseases such as migraine, depression, asthma, schizophrenia, and coronary heart disease. The effort by deCODE Genetics to combine information about genealogy, medical records, and genetic information on the entire Icelandic population is being carefully followed, as the usefulness of genetic isolates for studying common diseases is not as clearly established as for rare diseases.

Investigators in clinical trials and epidemiological studies should endeavour to store DNA to permit evaluation of possible gene-disease associations for their clinical impact. Genetic information will be useful if it provides additional information about aetiology, diagnosis, or prognosis compared with what is currently available. For many diseases, this is likely to be the case, and it will lead to greater integration of genetic information into clinical practice and public health. The use of genetic information will become routine in many fields of medicine, possibly through genotyping profiling on gene chips. On the other hand, because both genes and environment are involved in complex diseases, environmental causes and gene-environment interactions should continue to be carefully assessed.

Competing interests: JK is codirector of the Finnish twin cohort study and secretary general of the International Society for Twin Studies.