Comment

Gene environment interaction

H Campbell

The development of powerful new techniques for locating and identifying human genes is creating new opportunities for advances in our understanding of human disease. The discovery of specific genetic factors which confer susceptibility to common diseases such as cancer, diabetes, and asthma will greatly increase our ability to study the interplay of genetic and environmental factors in causing disease.

Genetic and environmental causes of disease

Very few diseases can be considered to be purely genetic or environmental in origin. The role of environmental factors in precipitating certain classical genetic conditions is well recognised¹⁻³ and genetic risk factors have been identified in diseases such as squamous cell lung cancer that are considered to be caused largely by exposure to environmental agents.⁴⁻⁶ A distinction is often made between genetic disorders with 'simple' Mendelian inheritance and those with more complex inheritance patterns, for example due to reduced penetrance or gene environment interactions.⁷ However, it is becoming apparent that many of these 'simple' Mendelian disorders are in fact exceedingly complex⁸⁻⁹ and that interactions between environmental factors and genetic susceptibility are important in determining disease risk.¹⁰⁻¹¹

In 1967, Motulsky drew attention to the numerous single gene mutations which were recognised to affect both the nature and rate of drug metabolism¹ and pointed out that this provided a model for genetically determined differential susceptibility to disease.¹² A differential ability to detoxify phenytoin may help to explain the occurrence of congenital malformations in infants of some women on antiepileptic treatment but not in others;¹³ and acetylator status has been linked to incidence of drug side effects.¹⁴⁻¹⁵

Traditional epidemiological investigation has focused on environmental factors in large part due to the desire to identify disease determinants which have the potential to be modified. They have not been designed to identify the portion of the variability in the incidence of a disease which can be assigned to genetic causes,¹⁶ and there has been little attempt to detect subgroups of individuals who

because of their genotype are more susceptible to certain diseases.¹⁷ Most genetic studies on the other hand have concentrated on underlying biology and not on environmental determinants.¹⁶⁻¹⁸ Gene mutations causing disease and genetic risk factors (allelic variants in many genes that may not cause disease but interact with other genes and environmental factors in causing disease) have been identified through family based linkage analysis both in pedigrees and by non-parametric methods in sibling pairs, and haplotype sharing studies.¹⁹⁻²¹ However, classic methods of genetic analysis such as linkage and segregation analysis have mostly been carried out in highly selected pedigrees with multiple affected members and this marked selection bias means that findings cannot be extrapolated to the population. Analysis of heritability through variance component analysis or path analysis has been used to compare the relative contributions of genetic and environmental factors to observed phenotypic variance. However, this form of analysis has often assumed that no gene environment interactions are present or relegates their contribution to residual effects.²² More recently, population based association studies of candidate genes have been used to identify genetic risk factors and this approach can be used to explore gene environment interactions (see below).²³⁻²⁶

Importance of gene environment interactions

Gene environment interaction is best conceived of as the genetic control of sensitivity to the environment.²⁷⁻²⁸ Experimental studies in animals have shown that these interactions are fundamental properties of most genetic systems.²⁷ As noted above, pharmacogenetic studies have confirmed these findings by showing that different genotypes respond differently to exposure to certain environmental risk factors.²⁹ There is increasing interest in genetic susceptibility in the aetiology and pathogenesis of common diseases such as coronary heart disease, cancer, diabetes, and congenital malformations. This is important not only in order to document genetically determined differences in susceptibility to the actions of environmental agents but also in order to more precisely understand the association between

Department of Public Health Sciences,
University of Edinburgh, Edinburgh
EH8 9AG.
H Campbell

Correspondence to:
Dr H Campbell

Accepted for publication
May 1996

Table 1 Approaches to the investigation of gene environment interactions

Population studies	Family studies
Case-control study	Twin studies
Cohort study (family history is the exposure)	Relative study (proband is the exposure)
Population based case-control family study	

environmental factors and disease.³⁰ If gene environment interactions are important in the causation of a disease then failure to consider genetic susceptibility may result in biased risk estimates for environmental exposures and may lead to incorrect inferences being made.³¹ Epidemiological investigations are therefore increasingly incorporating the collection of detailed family information, or laboratory measurements of specific genetically determined enzyme systems, proteins, or DNA markers that may be implicated in disease pathogenesis.

Study of gene - environment interactions

McMichael has highlighted the need to, ‘incorporate the emerging array of molecular biological measurements into epidemiological research and thus broaden its scope’.³² He has drawn attention to the fact that molecular techniques can be used to measure susceptibility, internal dose, biologically effective dose, or biological effect and that such detection of early biological responses may offer added insight in disease causation.

Advances in molecular technology are making it increasingly feasible to search for disease susceptibility genes in both population and family settings. Polymorphic markers acting as reference points are being used to localise genes by systematic investigation of the genome.³³ As information accumulates on confirmed associations between common diseases and specific alleles and on allele distributions in populations, the study of gene environment interactions will be facilitated.

To discriminate among the five identified models of gene environment interaction, individuals must be classified according to the presence or absence of the predisposing genotype and the environmental risk factor.³⁴ Data on disease status can then be stratified according to observed levels of the genetic and environmental factors and interpreted according to these models. This approach can be applied with genetic risk factors which have been fully identified or, in family studies only, with genetic markers with demonstrated genetic linkage to the disease. When genetic marker data are not available, family history data can be used as rough indicator of genetic susceptibility but there is potential for significant misclassification.³⁵

The main approaches to the investigation of gene environment interactions are given in table 1 and are discussed below. Inbreeding, admixture, migration, and adoption studies which can all give general information on genetic and environmental effects are not discussed as the contribution and design of these studies are all covered in standard texts.²²

Approaches to the investigation of gene environment interactions

CASE - CONTROL STUDIES

The availability of highly polymorphic markers close to candidate genes (which may be involved in the aetiology and pathogenesis of common diseases) has meant that standard case control designs can be adapted to study differences in genetic susceptibility and gene environment interactions in addition to the classical study of environmental exposure-disease associations.³⁶ Environmental risk factors are examined together with relevant genetic markers as independent predictors of disease and as interacting factors with the genetic marker.^{37 38} When the genetic marker/polymorphism is common and the frequency of exposure to the environmental agent is in the range of 30 to 70% then the required sample sizes to detect gene environment interactions are modest.³⁶

There may be problems with selection bias if participation rates differ with marker levels. However, genetic markers are generally stable thus reducing the likelihood of misclassification from marker instability or from the presence of the disease affecting the marker. Interpretation of marker-disease associations has to consider the possibility of linkage disequilibrium to a nearby gene, and confounding when (unrecognised) population subgroups differ both in marker and disease frequencies. These stratification effects can be avoided by careful matching on ethnic, geographical and ethnic background or by adopting internal family controls.³⁹ Observed associations between genetic risk factors and a disease should be checked by investigating the presence of linkage if the mode of inheritance known.³⁶

COHORT STUDIES

This approach follows the classic cohort study design with identification of environmental exposure and genetic risk factor status followed by a period of follow up during which new episodes of disease are registered.⁴⁰

TWIN STUDIES

Monozygotic (MZ) twins have an identical genetic complement whereas dizygotic (DZ) twins share only 50% and are in this respect no different to full siblings. Genetic and environmental effects can be studied by looking at MZ and DZ twin concordance for qualitative traits and correlation for quantitative traits. When MZ twin concordance is greater than that found among DZ twins, genetic effects are inferred, although differences between these two groups in terms of implantation patterns, intrauterine positions, prenatal circulation, and labour and delivery experience can also account for this finding. Similarly MZ twins may share more similar environments and this can be accentuated by self selection bias in voluntary twin registries. Lack of MZ twin concordance and DZ twin concordance greater than concordance in full siblings suggests environmental effects. Studies of discordant twin pairs have been particularly

fruitful.^{41 42} Gene environment interactions have recently been studied in the context of the above twin study designs⁴³⁻⁴⁵ or by applying interventions to MZ pairs.⁴⁶ International collaboration will be required to achieve sufficient power in future twin studies which consider gene environment interactions.

THE POPULATION BASED CASE-CONTROL FAMILY STUDY

The population based case-control family study approach has been proposed as one method of measuring the population burden associated with cancer genes and studying the roles of genetic and environmental risk factors in cancer susceptibility.^{47 48} In these studies investigators can evaluate the role of environmental factors, specific gene loci, and gene environment interactions in the aetiology and pathogenesis of diseases.³⁵ A specified group of relatives of both cases and population based controls are studied by collecting information on environmental risk factor exposure and taking blood or buccal smears as a source of DNA to establish the presence or absence of genetic risk factors. It has been suggested that this type of study design can be widely applied and will become a standard tool in epidemiology in the future.⁴⁸

FAMILY STUDIES INVESTIGATING CLUSTERS OF RELATIVES

Family studies can be used to explore gene environment interactions particularly when disease susceptibility genes have not been well characterised.⁴⁹ The distinctive feature of family data is the statistical dependence arising from the biological relationships and shared environment. Inferences from these data depend on developing models that represent this dependence by a limited number of parameters.⁵⁰⁻⁵⁴ However familial aggregation of a disease can be due to recall bias with cases being more likely to report a family history of disease than controls, or to shared environmental exposure. Lack of familial aggregation can be due to very low susceptibility in the majority of individuals or susceptibility which is unexpressed unless appropriate environmental exposure takes place.⁵⁵

There are problems of misclassification when family history is used for specification of a genetic trait.^{55 56 57} In addition, this approach is problematic when dealing with very common diseases. The relative risk of a disease in relatives of patients compared with the general population is usually more than an order of magnitude lower than the ratio of risk in susceptibles to that in non-susceptibles.⁵⁸ Even a weak familial aggregation in disease, as measured by an odds ratio in first degree relatives of as low as 1.5-2, is consistent with there being one or more strong familial risk factors.^{55 58}

Where genetic risk factors have been fully identified, disease risks among the relatives of cases (proband) can be stratified by the number of alleles shared ibd (identical by descent) with the proband and by the presence or absence of specific environmental exposures.^{22 38 59}

The statistical analysis of gene environment interactions in family studies has been carried out through analysis of variance or regression analysis and by nonlinear factor analysis.^{60 61}

Conclusion

Concern has been expressed that refocusing on genetic susceptibility of individuals could detract from the important public health goal of creating a less hazardous environment and might lead to genetic reductionism. This view, however, reinforces the polarised traditional debate of nature versus nurture. We need to reinterpret this issue not as one of genes versus environment but as the need for an active search for interactions between genes and the environment if we are to better understand the aetiology of common diseases.

Epidemiological study of disease aetiology needs to combine accepted methods of collecting data on environmental exposure with the study of specific genetic markers / polymorphisms. Epidemiologists should not see their role uniquely with population based data but recognise, in addition, the value of collecting and analysing family based data.³⁰ The incorporation of DNA data into epidemiological studies and the investigation of gene environment interactions holds the promise of being able to determine the association between risk factors and disease more accurately, clarify causal pathways and contribute to disease prevention.

- 1 Evans DAP. Pharmacogenetics. In: Emery AEH, Rimoin DL eds. *Principles and practice of medical genetics*. Vol 2. New York: Churchill, Livingstone, 1983; 1389-400.
- 2 Lenke RR, Levy HL. Maternal phenylketonuria and hyperphenylalaninemia. *N Engl J Med* 1980;303: 1202-8.
- 3 Khoury MJ, Beaty TH, Cohen BH. The interface of genetics and epidemiology. *J Chron Dis* 1986;39: 963-78.
- 4 Tokuhata GK, Lilienfeld AM. Familial aggregation of lung cancer in humans. *J Natl Cancer Inst* 1963;30: 289-312.
- 5 Caporaso NF, Tucker MA, Hoover RN. Lung cancer and the debrisoquine metabolic phenotype. *J Natl Cancer Inst* 1990;82: 1264-72.
- 6 Ooi WJ, Elston RC, Chen VW. Increased familial risk for lung cancer. *J Natl Cancer Inst* 1986;76: 217-22.
- 7 Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994;265: 2037-48.
- 8 Alper JS. Genetic complexity in single gene diseases. *BMJ* 1996;312: 196-7.
- 9 Wolf U. The genetic contribution to the phenotype. *Hum Genet* 1995;95: 127-48.
- 10 Ottman R. An epidemiologic approach to gene-environment interaction. *Genet Epidemiol* 1990;7: 177-85.
- 11 Ottman R, Susser E, Meisner M. Control for environmental risk factors in assessing genetic effects on disease familial aggregation. *Am J Epidemiol* 1991;134: 298-309.
- 12 Motulsky AG, Campbell-Kratt JM. Population genetics of glucose-6 phosphate dehydrogenase deficiency of the red cell. In: *Proceedings of the Conference on Genetic Polymorphisms and Geographic Variation in Disease*. New York: Grune and Stratton, 1962; 159-91.
- 13 Strickler S, Dansky L, Anderman E, Spielberg S. Pharmacogenetic predisposition to phenytoin-induced birth defects. *Pediatr Res* 1985;19: 180A.
- 14 Woosley RL, Drayer DE, Reidenberg MM. Effect of acetylator phenotype on the rate at which procainamide induces antinuclear antibodies and lupus syndrome. *N Engl J Med* 1978;208: 1157-9.
- 15 Eichelbaum M, Kroemer HK, Mikus G. Genetically determined differences in drug metabolism as a risk factor in drug toxicity. *Toxicol Lett* 1992;64: 15-22.
- 16 Sellers TA, Anderson VE, Potter JD, et al. Epidemiologic and genetic follow-up study of 544 Minnesota breast cancer families: design and methods. *Genet Epidemiol* 1995;12: 417-29.
- 17 Motulsky AG (editorial). Genetic epidemiology. *Genet Epidemiol* 1984;1: 143-4.
- 18 Khoury MJ, Beaty TH, Cohen BH. Interaction of genetics and epidemiology in the literature. *Genet Epidemiol* 1986;3: 269-77.
- 19 Blackwelder WC, Elston RC. A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet Epidemiol* 1985;2: 85-97.

- 20 Knapp M, Seuchter SA, Baur MP. The haplotype-relative-risk (HRR) method for analysis of association in nuclear families. *Am J Hum Genet* 1993;52: 1085-93.
- 21 Plomin R, Owen MJ, McGuffin P. The genetic basis of complex human behaviors. *Science* 1994;264: 1733-46.
- 22 Khoury MJ, Beaty TH, Cohen BH. *Fundamentals of genetic epidemiology*. New York: Oxford University Press, 1993.
- 23 Hwang SJ, Beaty TH, Panny SR, et al. Association study of transforming growth factor alpha (TGFA) TaqI polymorphism and oral clefts: indication of gene-environment interaction in a population-based sample of infants with birth defects. *Am J Epidemiol* 1995;141: 629-36.
- 24 Duijn CM, Haverkes LM, Van Broeckhoven C, de Knijck P, Hofman A. Apolipoprotein E genotype and association between smoking and early onset Alzheimer's disease. *BMJ* 1995;310: 627-31.
- 25 Hunt SC, Hasstedt SJ, Wu LL, Williams RR. A Gene-environment interaction between inferred kallikrein genotype and potassium. *Hypertension* 1993;22: 160-68.
- 26 Clerget-Darpoux F, Bonaiti-Pellie C. Strategies based on marker information for the study of human diseases. *Ann Hum Genet* 1992;56: 145-53.
- 27 Eaves LJ. The resolution of genotype environment interaction in segregation analysis of nuclear families. *Genet Epidemiol* 1984;1 215-28.
- 28 Mather K, Jinks JL. Biometrical genetics. *The study of continuous variation*. London: Chapman and Hall, 1982.
- 29 Vesell ES. Genetic and environmental factors causing variation in drug response. *Mutat Res* 1991;247: 241-57.
- 30 Khoury MJ, Stewart W, Beaty TH. Effect of genetic susceptibility on causal inference in epidemiologic studies. *Am J Epidemiol* 1987;126: 561-7.
- 31 Phillips PH, Linet MS, Harris EL. Assessment of family history information in case-control cancer studies. *Am J Epidemiol* 1991;133: 757-65.
- 32 McMichael AJ. Molecular epidemiology: new pathway or new travelling companion? *Amer J Epidemiol* 1994;140: 1-11.
- 33 Cooper DN, Clatton JF. DNA polymorphisms and the study of disease associations. *Hum Genet* 1988;78: 299-312.
- 34 Ottman R. An epidemiologic approach to gene-environment interaction. *Genet Epidemiol* 1990;7: 177-85.
- 35 Khoury MJ, Beaty TH. Applications of the case-control method in genetic epidemiology. *Am J Epidemiol* 1994;139: 134-49.
- 36 Hwang SJ, Beaty TH, Liang KY, Coresh J. Minimum sample size estimation to detect gene-environment interaction in case-control designs. *Am J Epidemiol* 1994;140: 1029-37.
- 37 Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, Liu ET. The Carolina breast cancer study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat* 1995;35: 51-60.
- 38 Khoury MJ, Flanders WD, Lipton RB. The affected sib-pair method in the context of an epidemiologic study design. *Genet Epidemiol* 1991;8: 277-82.
- 39 Falk CT, Rubinstein P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* 1987;51: 227-33.
- 40 Cohen BH. Chronic obstructive pulmonary disease: A challenge in genetic epidemiology. *Am J Epidemiol* 1980;112: 274-88.
- 41 Floderus B, Cederlof R, Friberg L. Smoking and mortality: A 21-year follow-up based on the Swedish Twin Registry. *Int J Epidemiol* 1988;17: 332-40.
- 42 Kaprio J, Koskenvuo M. Cigarette smoking as a cause of lung cancer and coronary heart disease. A study of smoking-discordant twin pairs. *Acta Genet Med Gemellol* 1990;39: 25-34.
- 43 Heller DA, De Faire U, Pedersen NL, Dahlen G, McClearn GE. Genetic and environmental influences on serum lipid levels in twins. *N Engl J Med* 1994;328: 1150-6.
- 44 Braun MM, Caporaso NE, Page WF, Hoover RN. Genetic component of lung cancer: cohort study of twins. *Lancet* 1994;344: 440-3.
- 45 Brennan P, Silman AJ. An investigation of gene environment interaction in the etiology of rheumatoid arthritis. *Am J Epidemiol* 1994;140: 453-60.
- 46 Bouchard C, Tremblay A, Despres JP, et al. The response of long-term overfeeding in identical twins. *N Engl J Med* 1990;322: 1477-82.
- 47 Hopper JL, Giles GG. A study of breast cancer in Victorian families. *Proceedings of the Cold Spring Harbor conference on cancer and cell biology* (Abstract). 1996.
- 48 Hopper JL. Epidemiology of genetic epidemiology. *Acta Genet Med Gemello* 1992;41: 261-73.
- 49 Khoury MJ, James LM. Population and familial relative risks of disease associated with environmental factors in the presence of environment interaction. *Am J Epidemiol* 1993;137: 1241-50.
- 50 Demenais F, Elston RC. A general transmission probability model for pedigree data. *Hum Hered* 1987;31: 93-9.
- 51 Bonney GE. Regressive logistic models for familial disease and other binary traits. *Biometrics* 1986;42: 611-25.
- 52 Morton NE, MacLean CJ. Analysis of family resemblance. III. Complex segregation analysis of quantitative traits. *Am J Hum Genet* 1974;26: 489-503.
- 53 Lalouel JM, Rao DC, Morton NE, Elston RC. A unified model for complex segregation analysis. *Am J Hum Genet* 1983;35: 816-26.
- 54 Diepgen TL, Blettner M. Analyses of familial aggregation of atopic diseases by using odds-ratio regression models. *Am J Epidemiol* 1995;141: S31(123A).
- 55 Hopper JL, Carlin JB. Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. *Am J Epidemiol* 1992;136: 1138-47.
- 56 Hunt SC, Williams RR, Barlow GK. A comparison of positive family history definitions for defining risk of future disease. *J Chronic Dis* 1986;39: 809-21.
- 57 Williams RW. Understanding genetic and environmental risk factors in susceptible persons. *West J Med* 1985;141: 99-806.
- 58 Peto J. Genetic predisposition to cancer. In: Cairns J, Lyon JL, Skolnick M, eds. *Cancer incidence in defined populations* (Banbury report no 4). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1980; 203-13.
- 59 Shah S, Green JR. Disease susceptibility genes and the sib-pair method: a review of recent methodology. *Ann Hum Genet* 1994;8: 381-95.
- 60 Freeman GH. Statistical methods for the analysis of genotype-environment interactions. *Heredity* 1973;31: 339-54.
- 61 Molenaar PCM, Boomsma DI. Application of nonlinear factor analysis to genotype-environment interaction. *Behav Genet* 1987;17: 71-80.