EPIDEMIOLOGIC REVIEWS

VOLUME 2 · 1980

Edited by Philip E. Sartwell, M.D.
and Neal Nathanson, M.D.

EPIDEMIOLOGIC REVIEWS

Volume 2
1980

Edited by Philip E. Sartwell
and Neal Nathanson

THE JOHNS HOPKINS UNIVERSITY PRESS
BALTIMORE AND LONDON

Printed in the United States of America

The Johns Hopkins University Press, Baltimore, Maryland 21218
The Johns Hopkins Press Ltd., London

Library of Congress Catalog Card Number 79–7564
ISBN 0–8018–2404–4 hardcover
ISBN 0–8018–2405–2 paperback
ISSN 0193–936X

EPIDEMIOLOGIC REVIEWS
Volume 2, 1980

Epidemiologic Reviews is published under the editorial management of the *American Journal of Epidemiology* and is sponsored by the Society for Epidemiologic Research and the International Epidemiological Association.

Assistance in the review of some of the papers in this volume was kindly provided by George Baer, Francis L. Black, Genrose Copley, Ellen B. Gold, Franz Matschinsky and Stanley A. Plotkin.

For more information about this review series, write to the *American Journal of Epidemiology,* 550 N. Broadway, Suite 201, Baltimore, Maryland 21205, U.S.A.

INTRODUCTION

Volume 2 of *Epidemiologic Reviews,* like its predecessor published in 1979, is produced under the editorial management of the *American Journal of Epidemiology.* In light of the favorable reception given to that volume, the same general policies have been followed.

In these volumes we have relied mainly on reviews solicited by the editors, but have included a few that were submitted to us for consideration. The subjects chosen are mostly ones on which recent work has substantially increased knowledge of the topics, and which are perceived as important to the public health. In the current volume, of the 11 papers, two are on methodological subjects and the others are divided between infectious and noninfectious conditions. However, four of the papers—those on juvenile diabetes, sudden infant death syndrome, Creutzfeldt-Jakob disease, and rubella vaccination—serve to remind us that infectious and noninfectious categories of etiology do not always remain hard and fast.

It is not necessary to defend the production of review volumes in the various medical specialties, or in scientific fields other than medicine. Not only are collections of reviews useful, as indicated by their proliferation; each author of a scientific paper or thesis rightly feels obliged to present a review of the status of information on his or her topic. These range from the briefest summaries to the most exhaustive treatises. The aim of our volumes is to produce documents falling between these extremes. We hope that our authors will cite enough of the early work to lay a foundation for the more recent work but not write a total history of the topic.

The limits of epidemiology are debatable, especially with regard to the inclusion of evaluation of health services and health planning. At present, we tend to adhere to the limits suggested by a definition of epidemiology as the study of the distribution of disease in human populations, with particular emphasis on those determinants of disease which offer clues to etiology or control. Both volumes thus far prepared, however, do include topics not specifically encompassed by this definition.

Epidemiology has had a more rapid growth than most other disciplines of the sciences allied to medicine. It is increasingly recognized as important in all branches of clinical medicine, preventive medicine, and public health. The increased concern over environmental, pharmacologic, and iatrogenic hazards has led to recognition that only through epidemiologic methods can many of these hazards be established. The need for review volumes in this field is fully justified by these trends. There is no other review series in the English language devoted wholly to epidemiology.

The editors wish to express their appreciation to the staff of the *American Journal of Epidemiology,* and particularly to Martha Myricks, Christopher T. George and Frances Stark, for their efficient and dedicated work in the preparation of this volume.

PHILIP E. SARTWELL
NEAL NATHANSON

CONTENTS

EPIDEMIOLOGIC REVIEWS
Vol. 2, 1980

Printed in U.S.A.

GENETIC EPIDEMIOLOGY: FOUR STRATEGIES

WILLIAM J. SCHULL[1] AND KENNETH M. WEISS[2]

What is genetic epidemiology? Morton (1) has defined it as the "science that deals with the etiology, distribution and control of disease in groups of relatives or with the genetic causes of disease in populations." Sing and Moll (2), however, contend that it is the study of the interaction of environmental and genetic determinants in common diseases. They would agree that its methods are different from those of "traditional" epidemiology. Still others fail to recognize it as either an appropriate or even a necessary conjunction of nouns. If there is no consensus as to what genetic epidemiology may be, it is clearly difficult to trace its historical development. We shall, therefore, limit our attention to the modern era, that is, to the period which has followed the rediscovery of Mendel's experiments and conjectures.

Garrod (103), with the aid of Bateson, can be credited with the first unambiguous effort to determine the incidence and prevalence of a simple, recessively inherited disease, namely, alcaptonuria, and to understand the factors which contributed thereto. He saw the importance of consanguineous unions in the occurrence of this and similarly inherited disorders and thereby identified one of the cardinal criteria we now associate with recessive inheritance. In 1902, with uncommon prescience, he also wrote, "If it is, indeed, the case that in alkaptonuria and the other conditions mentioned we are dealing with individualities of metabolism and not with the results of morbid processes the thought naturally presents itself that these are merely extreme examples of chemical behaviour which are probably elsewhere present in minor degrees and that just as no two individuals of a species are absolutely identical in bodily structure neither are their chemical processes carried out on exactly the same lines." Thus, he recognized too a heterogeneity in the risk of disease albeit in a biochemical context. Discovery of the inherited antigenic differences in erythrocytes at approximately the same time led ultimately to the search for disease associations or the identification of risk factors, if you will. This search gained momentum in the forties and early fifties, but then waned to return full cry with the discovery of the histocompatibility antigens. Others, notably George Hardy, Wilhelm Weinberg, Sewall Wright, J. B. S. Haldane and Ronald Fisher, to mention but a few, began in the early decades of this century to lay the framework for a mathematical theory to describe the factors which lead to the persistence, loss or spread of genes in populations. This was not, however, done in a conspicuously public health oriented manner; in fact, rarely was any allusion made to the public health implications of these notions. But the theory which was aborning was to provide the basis for segregation analysis and our understanding of quantitative inheritance in man, limited though it may still be.

As far as we know, the first unequivocal attempts to relate this theory to epidemi-

[1] Radiation Effects Research Foundation, Hiroshima, Japan (on leave, Center for Demographic and Population Genetics, University of Texas Health Science Center, Houston, TX 77030).

[2] Center for Demographic and Population Genetics, University of Texas Health Science Center, Houston, TX 77030.

This work has been supported by National Institute of Aging Grant No. AG 01028, National Cancer Institute Grant No. CA 19311, National Institute of General Medical Sciences Grant Nos. GM 19513 and KO4 GM 00230, and National Heart and Lung Institute Grant No. HL 15614.

ology were made by Neel and Schull (3) in a chapter entitled "Genetics and Epidemiology" in a 1954 textbook. They attempted to identify simple criteria of use to the epidemiologist from which he or she might assess the influence of genetic factors in the etiology of a given disorder, disease susceptibility and the like. They argued that the geneticist should function not as a competitor but as a member of an epidemiologic team. Writing on genetics and epidemiology, not genetic epidemiology, they advocated a partnership, not a disciplinary hybrid. Subsequently, other attempts to wed these perspectives occurred. One of the more calculated of these took place some 15 years ago when a conference of epidemiologists and geneticists was held at the University of Michigan, the proceedings of which were subsequently published under the title, *Genetics and the Epidemiology of Chronic Diseases* (4). Somewhat earlier efforts, of course, exist which focus on the geographic differences in the distribution of genetic disease (e.g., see references 5 and 6) but these were basically instances of geneticists examining the geographic distribution of genetic disease. No appeal was made to the methodologies of the epidemiologist. Morton (1) has noted that by 1967 when he and his colleagues examined the consequences of racial crossing in Hawaii (7) "there was general agreement that synthesis of goals and methods from epidemiology and genetics was inevitable and desirable." Unfortunately, they contributed little to the synthesis he acknowledged as needed. Their techniques were still those of the geneticist prone to accept epidemiology only on his or her terms. However, within the past several years the expression genetic epidemiology has not only achieved some currency in the medical and genetic literature but there exists a potentially constructive search for an identity.

It is our contention that if genetic epidemiology warrants existence as a discipline, it should be syncretic. It must in some way meld the problems and techniques of the epidemiologist and the geneticist. It should not be merely a fathoming of pedigrees, however large, nor solely a search for major genes or heritability estimates. It must recognize the inherent connectedness of genetic diversity with such general problems as aging, carcinogenesis, evolutionary biology, demographic change, natural stages of growth and development (and decay), environmental pollution and so on. It should involve a broader, more basic biology, one which recognizes, for example, changing orders of importance of diseases as a result of the removal of competing causes of death and disability or that the somatic accumulation of mutational events is a phenomenon of inheritance.

What, now, can the geneticist bring to this wedding? We describe four investigative strategies or techniques: segregation analysis, path analysis, fixed clusters of relatives, and a cohort of genealogies. In our estimation, these can contribute to the advance, ineluctable surely, in our understanding of the interaction of nature and nurture which subtends our individual differences in risk of disease and disability. They are neither exhaustive nor mutually exclusive nor will they be set forth in all of their complexity. Our aim is to titillate not to titivate. They have been selected to illustrate a methodology which we maintain is different from that of more conventional areas of epidemiology.

Segregation analysis

One of the intellectual appeals of Mendelian inheritance is the preciseness of the predictions which follow from the segregation of genes. Each of the simple modes of inheritance leads to a testable statement about the frequency of a specific phenotype in a sibship, the so-called segregation ratio; the value varies, of course, as a function of the type of mating

and the mode of inheritance postulated. Precise as the predictions of genotypes and phenotypes may be, however, given the small size of human families, chance can obscure the true mode of inheritance in any given nuclear family. Consequently, to test the hypothesis of an autosomal dominant or an autosomal recessive mode of inheritance, as examples, we must be able to combine data derived from a number of different families. The greater the number combined, the less likely are random departures from expectation to obscure the true situation. These families may vary not only as to the type of mating and the number of children, but also in the manner through which they come to the investigator's attention, that is, are ascertained.

Simply and traditionally viewed, family data are collected in only two different ways, namely, we either select families (parents) at random without reference to the phenotypes of the offspring they contain, or we select families which contain at least one individual with a specified phenotype, commonly a disease. Intuitively, it is clear that if the phenotype of interest is rare, the random selection of families, without reference to the phenotypes of the children, will result in few affected individuals; whereas if only families with one or more affected children are selected, families capable genetically of producing an affected child who fail to do so by chance will be systematically excluded. Similarly, if an effort is made to identify independently all affected individuals in a given population, all families with affected children irrespective of their number will be selected, that is, ascertainment will be complete. If the selection process focuses upon a particular sex or age in addition to the occurrence of the phenotype of interest, the likelihood that a given family will be included will depend upon the number of affected persons in the family. Ascertainment will be incomplete. Clearly, then, the observed segregation ratio will depend not only upon chance but the method of ascertainment as well, that is, the manner of family selection. The need to account for the systematic biases introduced through our methods of collection of data was recognized early in the century; indeed, Wilhelm Weinberg and Ronald Fisher, notably, developed a corpus of statistical techniques to be used to analyze family data and gave us such expressions as proband or propositus, complete and single selection, the probability of ascertainment, and the like. Their methods continued in vogue and largely unchanged from the thirties through the mid-fifties.

It remained for Newton Morton to revitalize these methods and to call attention forcibly to the new opportunities and more complex models which the advent of computers had made possible. One need no longer be restricted to those simple models or analytic alternatives where algebraically closed solutions were possible. Computer-oriented numerical methods opened new vistas. He and his colleagues (see reference 8 for a bibliography and a fuller discussion) soon had extended segregation analysis, as he termed these new developments, to include the estimation of other parameters, such as prevalence and inbreeding, and the search for etiologic heterogeneity. They applied these methods with profit and insight into a variety of diseases (e.g., (9)). But untempered enthusiasms also took segregation analysis into the morass of lethal and detrimental equivalents and fractioned genes, issues now little discussed, indeed, little remembered. The emphasis was, however, on single-locus models and the analysis of independently sampled sibships. It was assumed the latter were identified through specific affected individuals, so-called probands or propositi. The likelihood functions which formed the bases of the equations of estimation were conditioned on the parental phe-

notypes and derived under somewhat restrictive sampling assumptions. Stene (10) has since shown that the equations for complete and single ascertainment, the commonly used ones, hold under a more general set of assumptions, that is, whenever the phenotypes can be clearly distinguished; the probability of an affected child depends upon the mating type; births are independent trials, and the birth of an affected child does not lead to family limitation or overcompensation. Some of these concerns have Swiftian overtones, however, for few genetic surveys in the past have adhered carefully to rigorous sampling procedures. They involved what Thompson and Cannings (11) call "registry sampling," but even this suggests a systematization of sampling which is dubious. Too frequently, the sampling frame or registry was undefined, indeed, undefinable, and no pretense at enumeration of the frame was made. Thompson and Cannings (11) and Elston (12) have begun a long overdue reanalysis of the statistical issues posed by different sampling schemes.

Concurrent with the developments we have just described, an interest began to grow in the analysis of the role of genetic factors in those many, seemingly discrete traits which cluster in families but do not segregate simply. Numerous congenital malformations are notable examples. Edwards (13, 14) examined the incidence of a particular trait among specific classes of relatives of probands under several genetic hypotheses. Of particular interest, as a model for the interpretation of the kinds of traits to which we allude, was the emerging notion of an underlying, continuously distributed liability to the trait coupled with a "threshold," that is, some minimal liability necessary to bring about the postulated pathologic process. Falconer (15, 16) used this notion and the concept of heritability in quantitative genetics to develop a model which he and his colleagues (17, 18) then applied to diabetes mellitus. His model is essentially descriptive rather than analytic and as others have noted may mask rather than elucidate genetic heterogeneity. It may, nonetheless, under restrictive circumstances be possible to deduce the action of a single locus or small clusters of loci from the heritabilities from different kinds of relatives. Complex segregation analysis (19, 20) can incorporate these notions, for it recognizes recurrence risks as variable among families with a given parental phenotype. Simple segregation analysis, of course, posits a constant recurrence risk among families of a particular mating type; it is assumed that the genotype-phenotype relationship is regular in the sense used by Cotterman (21).

It became increasingly self-evident that there was a need to distinguish or to discriminate between the existence of 1) a single major locus to which a substantial amount of interindividual variability could be assigned, 2) numerous genes (so-called polygenes) with individually small effects, and 3) environmental correlations between relatives (cultural inheritance). Because of this need, a search was begun for still more general segregation models. Elston and Stewart (22) described such a model for the analysis of extended families or pedigrees which was not restricted to dichotomous traits. It could accommodate an arbitrarily large pedigree so long as it was derived from a single pair of ancestors and included no consanguineous unions. It assumed the families studied to be a random sample from the population (i.e., not ascertained through probands) and stated the distribution of genotypes given the parents' genotypes, i.e., conditional on the latter. Campbell and Elston (23) then showed how the method of stochastic matrices, introduced by Li and Sacks (24) to derive the frequency and correlation between relatives, could be extended to the analysis of summary data. Elston (25) soon extended the model he and Stewart had earlier devel-

oped to embrace a variable age of onset and ascertainment through probands. Lange and Elston (26) were able to remove the restrictions in the Elston-Stewart model. They introduced a graph-theoretic definition of a pedigree and the notions of simple and complex pedigrees. The former is any pedigree in which the chain of descent is necessarily downwards or horizontal through the pedigree; all other pedigrees are viewed as complex. Their approach encompassed an infinite set of loci but assumed no environmental correlations, and nonassortative mating.

The more complex the models have grown, the less and less accessible they have become to the average investigator. Unlike their simple predecessors of Fisher and Weinberg, the calculations required are formidable and time-consuming. Computer programs obviously exist for a wide variety of segregation models (27), but many investigators may find them more intimidating than helpful. All of the models depend upon the solution of a series of likelihood equations and, as previously stated, the solutions are reached by iterative methods, commonly through the use of the Newton-Raphson procedure. As the dimensionality of the likelihood surface grew, concern mounted over the robustness of the models and the numerical methods used. The Newton-Raphson method is known to fix occasionally upon a local rather than the global maximum, and the assumptions inherent in the models became more difficult to identify explicitly. MacLean et al. (28) have addressed some of these issues through the use of simulated data. They contend that complex segregation analysis can discriminate between effects due to a major locus and those due to polygenes and cultural inheritance. They further show that the procedure is reasonably robust but if and only if the model simultaneously includes a major locus, polygenic inheritance and environmental effects common to siblings. Even then, however, it is essential that heterogeneity among mating types be tested. They conclude that most of the information on the contribution of a major locus to a quantitatively varying trait is lost if the trait is reduced to a dichotomy. Thus, for example, if blood pressure variability is reduced to the two states, normotensive and hypertensive, most of the effects of a major locus, if one exists, will be lost.

Go et al. (29) have approached these matters somewhat differently but again through the use of simulated data. They examine robustness and efficiency under unconditional as well as conditional likelihoods. They too conclude that the complex model is robust insofar as a major locus is concerned, and that variation in family size has little effect on this robustness. They show that skewness and polygenic inheritance do not lead to the spurious detection of a major locus, as has frequently been conjectured, unless they occur together and in the presence of a moderate environmental correlation between siblings (greater than zero but less than 0.4). Finally, they find that unconditional likelihoods are more efficient than conditional ones. The latter finding has prompted Elston and Sobel (30) to further examine unconditional likelihoods and the effects of ascertainment. They find that most of the problems associated with earlier efforts to modify the unconditional likelihood of a pedigree to allow for independent ascertainments can be remedied through the use of a rigidly defined sampling frame.

Today, though these methods have seen limited application, and this only in the hands of their advocates, they have produced some interesting findings. Elston et al. (31) have examined the distributions of cholesterol and triglycerides in a large kindred and have identified a major locus in the transmission of hypercholesterolemia and hypertriglyceridemia. Similarly, Gerrard et al. (32) have identified a major locus in the control of immunoglob-

ulin E levels. Morton and Rao (33) have chastised human geneticists who "tend to hug the coasts of Mendelian traits without venturing into the deep waters of complex inheritance." But in a world increasingly devoid of biologic benchmarks, it seems to us that investigators can be forgiven if they are unable to distinguish between satori and vertigo.

Path analysis

Segregation analysis, as we have just seen, is applicable to the analysis of data on nuclear and extended families. Path analysis, to which we now turn, is a method for decomposing correlations between quantitative phenotypes of pairs of related individuals into their genetic and environmental causal components. It is applicable whenever 1) a reasonable causal model among certain measurable variables can be constructed, 2) the relative effects of these causal elements are the characteristics of moment, 3) a specific genetic model is not known, 4) the basis of analysis is pairs of related individuals, 5) pre-disease states, symptoms, or measurements are the phenotypes to be analyzed, and 6) the inclusion of a complex of environmental cause factors is of primary interest. The basis of path analysis is a causal diagram, which consists of the elements in a hierarchical structure connected by arrows indicating the posited directions of causation. Path analysis was devised by Sewall Wright (34, 35) who early saw the special power of the method to compute genotypic correlations among relatives in any patterned system of mating (36). Although he also applied his method to the analysis of the causation of variation in IQ (37), its use was basically restricted to agricultural genetics and kinship computations until fairly recently.

The best systematic account of the statistical basis and properties of the path method is due to Li (38), but the reader might also consult Kempthorne (39). Discussions of the method in terms of its applicability, and special characteristics, are given by Wright (40–43) and by Tukey (44) who evaluates the various methods of regression, correlation, and path analysis. There are discussions of genetic epidemiologic applications in Li (45), Elston and Rao (46), Rao and Morton (47), and Morton and Rao (33).

It is only possible to indicate the basic nature of the method here. Figure 1 is a typical (schematic) path diagram for a trait with mixed genetic and environmental causes. Unobserved elements are circles; observed ones squares. All elements which have a causal relation to another element are connected by a single-headed arrow. Relationships between elements which are due to ultimate, but unknown, other factors are represented by double-headed arrows, and the coefficients on such arrows are the correlation coefficients between the variables. The other coefficients are "path coefficients," that is, regression coefficients standardized by the ratio of the standard deviations of the connected variates. Certain path coefficients will, in genetic applications, be known a priori from genetic theory, such as the genotypic correlation between parents and offspring. Correlations between all pairs of observed variables can be computed (the data consist of sets of related individuals and measurements on them and their environments). The object is to find correlations between the unobserved variables or to evaluate the contributions such variables make to the variation in the phenotype, acting singly or jointly with other variates in the causal scheme.

There are two basic properties of a path diagram of interest to the estimation of parameters. First, the correlation between any two variates in the diagram is determined by summing the multiplied path coefficients from all paths connecting them, following certain rules for tracing out paths. Second, the variance in any

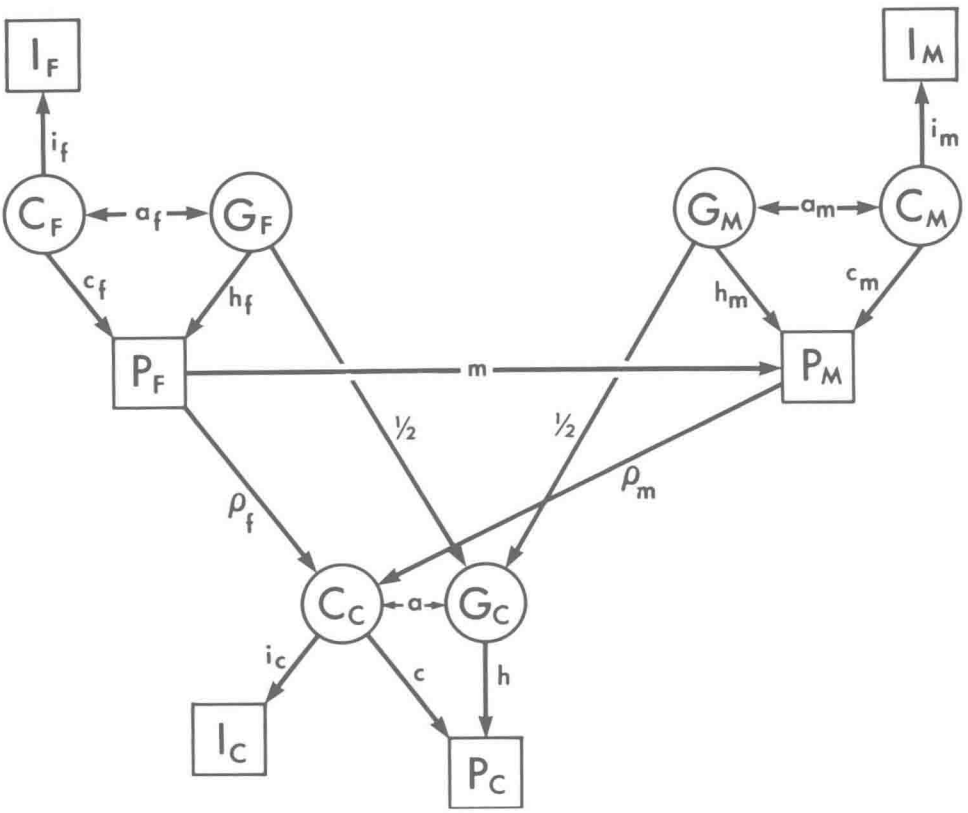

FIGURE 1. Schematic path diagram. □ = observed value, ○ = unobserved. Double-headed arrows are correlation coefficients; single-headed arrows are standardized regression coefficients. I = index, P = phenotype of parent or child, G = genotype, C = common environment; all of these for mother, father, and child. Paths are labeled so that they may be specified in the appropriate equations. Example of a path equation: the correlation between father's phenotype and child's environment, $r_{C_c P_f} = \rho_f + a\frac{1}{2}h_f + \rho_m M$.

variable may be decomposed into a component equal to the square of the path coefficients of factors acting independently on it, plus components of joint action of correlated causal elements in the diagram. In practice, one writes down relations equating observed correlations to their contributory path coefficients and then computes the best fitting values of the parameters. If statistical solutions are to be found, the diagram must have enough coefficients and paths, of course, that the entire system is overdetermined relative to the observed correlations. It cannot be overly stressed that the investigator bears the burden of constructing a reasonable path diagram. *Any* properly constructed diagram will lead to a set of complete, consistent parameter values.

Path diagrams are particularly useful for incorporating several elements of importance and interest in the study of chronic disease phenotypes. First is the ability to include the effects of assortative mating among the parents. Second, the model permits consideration of many different ways in which culture or "environment" influences the phenotypes of interest. Not only can environmental factors be considered which, in one's view, directly lead to changed phenotypes (e.g., cholesterol intake and hypertension), but individuals of the same nuclear family may be viewed as similar because they

share environments. Parent-offspring correlations may appear to have higher genetic components than are really correct if some cultural traits are passed on directly from parent to offspring (e.g., relevant attitudes, behavior, diet). The problem of "cultural inheritance" has been considered at length by Cavalli-Sforza and Feldman (e.g., (48) and many other papers). Cloninger et al. (49) have looked at this problem too and have concluded that under most conditions data from separation "experiments" (e.g., twins reared apart, broken homes) will be needed to separate genetic from cultural inheritance.

Rao et al. (50, 51, see also Morton (52)) advocate the extensive use of environmental "indices" to increase the number of available correlations and the power of path methods to separate out important causes. There are two ways that environmental elements may be constructed as "causes" of the phenotypic variance. One approach is to develop an index in a way thought to be reasonable from environmental factors but not considering phenotypes. Another is to regress phenotype on a series of environmental factors and to assign to each phenotype the regressed value of the index. This runs the obvious risk of considering things which are related to (produced by) genotype as environments. Rice et al. (53) and Cloninger et al. (54) note that such indices 1) may really measure some genetic/phenotypic value, such as midparent phenotype, and 2) must be checked independently of the path analysis to determine if they are well constructed.

The use of path diagrams should not be restricted to estimation of parameters alone since any proper diagram will yield such estimates. It is of primary interest to check for goodness-of-fit to determine whether the diagram so analyzed may be rejected as an explanation of the observed data (45, 50, 51). Rao et al. (51) advocate the use of likelihood tests to determine the most reasonable diagram among several plausible ones. Such an approach may, for example, rule out common environments as important factors, or assortative mating. Where this approach is merited, the diagram may be built up until the degrees of freedom to estimate parameters are exhausted.

Although much has been written recently about path analysis, its actual application to epidemiologic problems has not been extensive. In fact, the most active use of the method has been to decompose variability in IQ scores into genetic and various environmental components. Although somewhat feebly rationalized as epidemiology because individuals with low IQs are medical or at least social problems (33), this application is more methodological than epidemiologic. Morton and Rao (33) claim to have found, among other things, an additive genetic contribution ("heritability") of about 70 per cent of the variation in IQ. Both their methodology and algebraic accuracy have been energetically challenged by Goldberger (55, 56). Cloninger et al. (54) find "significant and substantial" contributions to IQ of both cultural and genetic factors. With all that is known about the potential misleading effects of gene-environment interactions, dominance (or other nonlinear effects), and all that is not known about whether such effects are important, path analysis should be applied to other more straightforward problems.

Some applications in such directions exist. Morton and Rao (33) review some traits which they have studied by various quantitative genetic models including path analysis: these include lipoprotein levels (57, 58), immunoglobulin E levels and disease (32), periodontal disease (59) in which they find no important genetic effects, and birth weight (33).

The major application, to date, of path methods has been to the study of various cardiovascular disease-related measurements. Several investigators (57, 58,

60–63) have looked at the genetic and environmental sources of variation in serum cholesterol levels, in particular, dividing the environmental components into those common to individuals in the same household (related to cultural inheritance when parent and offspring are involved) and those special to individuals. Path analysis was but one method they used. Sing and Orr (60) found that about 20 per cent of the full sib correlation in cholesterol level is due to common environment, genes being responsible for about 58 per cent. Sing et al. (61) deal with this further and note findings very similar to those of Morton et al. (57) and Rao et al. (58) in Hawaii, in terms of the relative contributions of the various components.

Weinberg et al. (62) studied blood pressure by path analysis among Bogalusa, Louisiana, schoolchildren as part of a SCOR-A program study. The goal, again, was to determine the relative importance of additive genetic factors and common environments in sib-sib correlations. Only one type of relative pair (sibs) was available and thus only one parameter (heritability or environment) could be estimated at a time. Neither genetic factors nor common environments could be rejected as explaining the observed familial similarity of blood pressure. Though the power of their test is low, it indicates that common environments can in a sense be separated from common genes, and that both are important factors.

The effects of environment and other health-related variables (e.g., weight, age, etc.) on blood pressure have been studied by Ward et al. (63) in groups of migrants and nonmigrants in the Micronesian Tokelau Islands. Path analysis has demonstrated the importance of the genetic component in risk and has shown the way in which risk factors in different cultural environments contribute to hypertension.

We must continue to stress, as have others, that path analysis is used specifically because one wants to partition causal variance in a specific population. It is not known, and is a matter of controversy, how widely any such variance partitioning will apply (38, 44, 45, 64, 65). The fact that the results of the partitioning are population specific and dependent on the assumption of only linear effects is, in practice, usually overlooked totally or nearly so. It can be said, however, that since intervention studies for chronic disease risk factors usually have population-specific public health objectives, the path method has potential. Thus, in conditions which are truly polygenic, that is, in which one locus does not predominate in producing risk heterogeneity, contemporary chronic disease epidemiology and its preventive aims probably can make use of the path method or its derivatives.

Fixed clusters of relatives

Segregation and path analysis commonly proceed from a "sampling" of individuals with some one or combination of attributes to the relatives of those individuals. The families so studied may differ substantially in size as well as other characteristics; they share in common only the presence of one or more persons ostensibly chosen through the process we term ascertainment and take to be synonymous with probability sampling. The result is a series of clusters of individuals of unequal size. Other strategies of sampling are clearly possible. We shall be concerned in this section primarily with the sampling of clusters of individuals of equal size, and we begin with undoubtedly the oldest such fixed cluster, namely, twins.

Over a century ago, Späth (66) and Galton (67) called attention to the uniqueness of twins and their usefulness in the appraisal of the nature-nurture problem. Although the notions of inheritance prevailing then were not the ones to which we presently subscribe, twins and twin

studies have enjoyed and continue to enjoy a certain standing in genetic research. Numerous registries—national or statewide, large and small, complete and incomplete—exist; prominent among these are the Scandinavian Twin Registries (68, 69), and the National Academy of Sciences-National Research Council Twin Registry (70). The importance of twins comes, of course, from the fact that members of a monozygotic pair (MZ) have identical genotypes, barring mutation, and thus any dissimilarity which may exist between them must be due to the action of environmental agents either postnatally or *in utero*. Interest in dizygous (DZ) twins stems from the fact that, while differing genetically (to the same degree as a pair of siblings), they enjoy certain environmental similarities, e.g., birth rank and maternal age, not enjoyed by single-born offspring, and thus afford a measure of environmental effects not otherwise possible.

Assessment of the contributions of nature and nurture rests either upon the frequencies of concordance, that is, the occurrence of the same phenotype, in identical (MZ) and fraternal (DZ) twin pairs, if the trait of interest is dichotomous or at least discrete, or the proportion of variation common to the two types of twins, if the trait is continuous. Heritability, the effect ascribed to genetic factors, has commonly been defined in the first instance as

$$H = \frac{\text{CMZ} - \text{CDZ}}{100 - \text{CDZ}}$$

where CMZ and CDZ are the percentages of concordant monozygotic and dizygotic twins, respectively, and in the second instance as (71)

$$H = \frac{r_{MZ} - r_{DZ}}{1 - r_{DZ}} = \frac{V_{DZ} - V_{MZ}}{V_{DZ}}$$

where r_{MZ} and r_{DZ} are the intraclass correlation coefficients with respect to monozygotic and dizygotic twins, respectively, and V_{MZ} and V_{DZ} are the respective mean square deviations for the two twin types. *It should be noted that the utility of this definition has been repeatedly challenged (e.g., 33, 72–74), and the two definitions of heritability bear little relationship to one another.* Smith (73) urges the use of $2(r_{MZ} - r_{DZ})$ in the second instance since it eliminates nongenetic familial effects on twins and provides a more readily interpretable estimate of genetic determination. A better concordance model is to be found in Allen and Hrubec (74). A fuller exposition of these and other methods of assessment of twin data will be found in Cavalli-Sforza and Bodmer (75). It should be noted that analysis of twin data must also take into account possible differences in rates of diagnosis, ascertainment and other factors, e.g., differences in frequencies by sex and age of onset.

An extension of the method so briefly adverted to above involves the study of identical twins reared apart (51, 76). Theoretically, such studies can give a direct estimate of genetic heritability, but in practice this seems unlikely for random placement of the twins, an essential, seems most improbable. Moreover, identical twins reared apart are difficult to identify and increasingly difficult to collect. Appeal to this type of study has been common, however, in the acrimonious and unproductive debate over the significance of the alleged racial differences in intelligence.

The limitations of the conventional twin method have been repeatedly pointed out (e.g., 33, 77). These include not only unequal probabilities of ascertainment but a variety of biases of both pre- and postnatal origin. Among these are the so-called natal factors (dissimilarities in the positions of the fetuses in the uterus, *in utero* crowding, the special conditions of implantation and ultimately delivery), lateral inversions (differing de-