Genetic Epidemiologic Studies on Age-specified Traits

NIA Aging and Genetic Epidemiology Working Group

This commentary calls attention to the value of combining genetic and epidemiologic methods in studies to understand the determinants of two crucial aspects of aging: ages at which certain outcomes (e.g., disease, mortality) occur and rates of change with age of individual's characteristics (e.g., physiologic functions, disease risk factors). Inclusion of age in the specification of traits in genetic epidemiologic studies could lead to improved strategies to increase healthy life expectancy and evaluate individuals' risk for age-related morbidity. Special issues that make genetic epidemiologic approaches important for studies of age-specific phenomena as well as opportunities and challenges for such studies are discussed, including study designs, sampling frames, databases, analytic tools, and related methodological issues. This commentary is based on a report prepared by the Aging and Genetic Epidemiology Working Group, convened by the National Institute on Aging to review opportunities for research on the genetic epidemiology of aging-related outcomes. The report, which contains more extensive discussion, literature review, and references, is available on the World Wide Web at http://www.nih.gov/nia/conferences/GeneticReport111199.htm. Am J Epidemiol 2000;152:1003–8.

aging; disease; mortality; survival

This commentary calls attention to the value of combining genetic and epidemiologic methods in studies to understand the determinants of two crucial aspects of aging: ages at which certain outcomes (e.g., diseases and mortality) occur and rates of change with age of individuals' characteristics (e.g., physiologic functions, disease risk factors). The benefits of integrating genetic and epidemiologic methods for studies on a broad range of topics have recently been thoroughly reviewed (1). Here, the focus is confined to special issues that make genetic epidemiologic approaches especially important for studies of age-specific phenomena and related methodological issues and research opportunities. We address opportunities for studies to identify regions of the genome that contain genes affecting a trait (genome scans), as well as effects of specific alleles on a trait. (In this commentary, genome refers to all of an individual's genetic material, genotype refers to an individual's specific genetic characteristics, allele refers to one of two or more alternative forms of a given gene, and polymorphism refers to a relatively common allele, in contrast to rare mutations.)

This commentary is based on a report prepared by the Aging and Genetic Epidemiology Working Group convened by the National Institute on Aging to review opportunities for research on the genetic epidemiology of aging-related outcomes. This report, which contains more extensive discussion, literature review, and references, is available on the World Wide Web at http://www.nih.gov/nia/conferences/GeneticReport111199.htm.

AGE-SPECIFIED TRAITS

Although genetic epidemiologic studies on diseases of aging are increasing rapidly (2), most have defined the trait of interest simply as the presence or absence of a disease (or occurrence of an adverse event) or the level of a given risk factor at one point in time. However, it is a familiar epidemiologic concept that individuals vary not only in whether they develop a disease but also in the ages at which
disease occurs and in the rates at which premorbid changes progress to disease. Further, success in avoiding a disease cannot be fully characterized without indicating the age range over which individuals are at risk. These issues can be addressed by including age in the specification of the traits for which genetic and other risk factors are to be studied.

Two types of age-specified traits are **survival traits** (the age at which a specified outcome has occurred or has not occurred) and **rate-of-change traits** (the rate at which individuals' characteristics change over a specified age interval).

**Survival traits**

Inclusion of age of onset in definitions of traits can allow factors that affect risk for a disease at different ages to be distinguished. A survival trait, for example, can be defined in terms of the age range over which an individual is at risk for an outcome such as disease onset or a morbidity event. In this case, survivornship could be defined as a continuous variable (e.g., years surviving from birth, or from a given age, until the outcome) or as a dichotomous variable (e.g., occurrence or nonoccurrence of the outcome before a given age or between one age and another).

Studies of age-related diseases have found relations between age of onset and the magnitude of genetic contributions to risk (3) and contributions of specific genes (4, 5). This approach has commonly used two age ranges (e.g., early- and late-onset of a disease). However, adequate sample sizes could allow several age ranges to be compared. Differences between genetic risk factors contributing to early-onset familial cases of conditions such as Alzheimer’s disease and those contributing to late-onset nonfamilial cases are well appreciated, suggesting that analogous differences could exist among age ranges for nonfamilial forms of other age-related diseases.

Distinguishing survival traits by the age intervals over which survival is measured has other potential advantages. Many age-related conditions are likely to be highly polygenic, posing difficulties for identifying genes with effects. Distinguishing among age intervals could diminish these difficulties, since the number of genes affecting risk within any one age interval could be fewer than the number affecting risk over the entire life span. Study of age-specific survival effects may also avoid confusion about genes that have beneficial effects in one age range and harmful ones in another. The importance of such genes in aging has been predicted on the basis of evolutionary biological considerations (6).

The value of differentiating among age intervals in searches for genetic risk factors for disease in old age is reinforced by the substantial mortality over this age range. To be at risk for a disease in very old age, one must first survive to old age. As discussed below, the likelihood of survival to advanced age is influenced by genetic factors. This implies genetic differences among successively older age cohorts. In view of probable gene-gene interactions affecting disease risks, these cohort differences could influence the risks or protection conferred by any one allele.

**Broad survival traits.** Identification of the effects of a genotype on time to onset of a particular condition does not determine its effects on overall survival or risk for other conditions. This is important because risk factors for one age-related disease can protect against another disease (e.g., obesity predisposes to osteoarthritis but protects against osteoporosis). This may be true for many genetic risk factors. A lack of effect on life span and/or disability by a genotype that has large effects on disability and/or mortality from a common disease suggests that it may have opposite effects on disability, morbidity, or mortality from some other disease. Broad survival traits such as longevity, active life expectancy (ALE) (disability-free survival), or health expectancy (disease-free survival) can clarify the public health significance of disease-specific findings.

Longevity may be characterized simply by age of death or by a more sophisticated measure such as integrated mortality risk (7), derived from the sum of yearly mortality risks until death, which adjusts for birth cohort and other factors. European studies of twins and other populations indicate that genetic factors account for about 25–35 percent of the variance in life span (8–10). The moderate size of this genetic effect reinforces the importance of environmental and other factors. Nevertheless, in terms of attributable risk, genetic effects on longevity in the population are important. Their share of the variance in life span accounts for a number of person-years of life equivalent to that of a much higher share of the variance of any disease-specific mortality because of the breadth of the outcome (all-cause mortality). Evidence also suggests genetic effects on the likelihood of reaching very old age (11).

Using longevity as a trait, without additional information on functional status, cannot resolve whether a factor promotes healthy functional survival or survival with disability. This has stimulated interest in measures of ALE, the duration of survival without disability (12). The ratio of ALE to life span is also of interest, particularly because of its effects on resources needed to support dependent older persons, and could be addressed by measuring both ALE and longevity in the same study (13). Health expectancy (survival without any one of a specified set of diseases or disabilities) is also a trait of public health interest (14, 15).

Outcome rates for these broad survival traits generally rise exponentially with age, at least until very old age (16). The pattern of this rise in risk with age is an important aspect of these traits. For example, the effects of a genotype would have differing biologic, medical, and demographic implications, depending on whether it 1) primarily affects outcomes in middle age or early old age, but had little effect on survival to very advanced age, or 2) affects the rate of exponential increase in outcomes across all ages, thus substantially increasing the proportion surviving to very advanced age.

Study of broad survival traits could also facilitate detection of genes affecting two important characteristics whose genetic influences might be missed in studies of disease-specific survival traits:

**Delayed or accelerated risk for multiple diseases.** A genotype that delays or accelerates onset of several diseases could affect survival significantly, but be undetected in disease-specific studies if its effect on any one disease were small.
Unrecognized age-related pathologies. Many important pathologies of aging may remain to be defined. Recent findings indicate pathologic significance of age-related changes, such as vascular stiffening (17) and loss of muscle mass (18), that were considered “normal” aging until recently. Genes with effects on unrecognized pathologies might not be detected in studies of disease-specific survival but could be found by studies to identify genetic effects on broader outcomes.

Rate-of-change traits

Genetic factors may influence not only physiologic functions at one time point but also their rates of change with age, which can influence whether and when disease occurs. There is evidence that genetic factors significantly affect rates of change with age in several physiologic measures (19, 20). Similar to survival traits, there are benefits to distinguishing among different age ranges, including intervals before pathologic consequences occur. The latter are of particular interest regarding disease prevention strategies. The feasibility of using rate of change of a function as an outcome measure is affected not only by the function’s measurement properties at one time point but also by the properties of measures of its change. Methodologies to analyze such data are being refined (21, 22).

BENEFITS OF COMBINING GENETIC AND EPIDEMIOLOGIC APPROACHES TO AGE-SPECIFIED TRAITS

A major challenge to the study of age-specified traits is the fact that characterizing these traits requires information on individuals’ characteristics over long time intervals, often spanning many decades. Risk factors such as physiologic functions, environmental exposures, lifestyle, socioeconomic status, and disease history change throughout an individual’s life span. This implies the need to know the status of these characteristics at multiple time points.

Conversely, genetic characteristics are generally stable over time. Accordingly, characterization of genetic risk factors does not require sequential measurements or retrospective data and can therefore be done at any time in a study. Genotype information in studies of survival traits can thus be particularly useful when utilizing retrospective or case-control study designs because these types of studies often have very limited information on subjects’ prior status, particularly when considering survival over long intervals. Examination of potential genetic risk factors in such studies clearly does not eliminate all problems caused by incomplete data on prior status of other risk factors and exposures. However, it adds a set of potential risk factors whose status is more certain and can improve the likelihood of obtaining conclusive findings. Additionally, as more information on the effects of polymorphisms in numerous genes becomes available, it may be increasingly possible to infer the earlier status of individuals’ other risk factors by combining their genotype information with other data on prior exposures and behavior and by knowledge of gene-environment interactions.

Longitudinal studies, on the other hand, by definition collect information on individuals over time. Many existing studies have a wide range of measures of clinical, physiologic, and environmental exposure characteristics of individuals, often spanning decades. These data, when combined with genotype information, could provide exceedingly valuable opportunities to learn how age-specified traits are affected by genotype and gene-environmental interactions over time. In addition, analytic methods developed for analyzing longitudinal data from epidemiologic studies could be adapted and applied to incorporate genetic information.

Epidemiologic expertise in dealing with large populations, sampling frames, and detailed characterization of subjects is particularly needed for genetic studies on age-specified traits for several reasons: Large samples will tend to be required because of the need to distinguish among age groups and because of the genetic complexity of many broad survival traits. Further, potential birth cohort differences add a source of heterogeneity that must be reliably characterized to avoid artifactual findings. Studies of rare extreme traits, such as exceptional longevity, can benefit especially from large sampling frames, especially if identification of probands’ relatives is also needed.

OPPORTUNITIES AND METHODOLOGICAL ISSUES FOR STUDIES

There are several opportunities to increase the ability of epidemiologic studies to identify factors that affect age-specified traits. These include strategies to deal with methodological issues, as well as new resources.

Inclusion of multiple family members and generations in longitudinal studies

As noted above, longitudinal studies have unique value for genetic epidemiologic studies on age-specified traits. Including family members in such studies allows the use of powerful genetic methods based on comparisons of relatives (such as linkage analyses in families, sib-pair studies, and family-based association studies). A few existing longitudinal studies contain familial information (23–25). However, many more could benefit from the addition of family members.

Strategies to reduce population-stratification artifact

Immigration, internal migration, and intermarriage have caused great differences in the genetic makeup of different age cohorts in most areas, unrelated to their longevity. Thus, comparisons of polymorphism frequencies in the very old with those in the young or middle aged are very susceptible to artifactual results that can result from this population stratification (26). These problems probably have contributed to inconsistent results among many such studies, e.g., on associations between longevity and human leukocyte antigen polymorphisms (27–29) or genes for cardiovascular risk factors (30). However, replication of case-control study findings on survival outcomes in multiple differing populations provides confidence in their validity, as with the APOE locus and longevity (31–33). The need to consider designs based on comparisons of relatives as an
alternative to case-control comparisons of unrelated individuals is especially great for studies of survival outcomes such as longevity.

Strategies to deal with missing parental data

Family-based studies of age-specified traits expressed late in life often face methodological issues stemming from lack of data on parents' genotypes and other characteristics. (Such data can clarify the pattern of transmission of the trait and increase statistical power of inferences about genetic contributions to it.) There are design and analytic strategies to address this problem. For example, if multiple siblings are available, it is often possible to determine parental genotypes from analyses of sibling genotypes. Statistical methods are also available to cope with missing parental data (34). Longitudinal studies including multiple generations have the additional option of prospectively collecting data on, for example, elderly parents and middle-aged children and waiting until age-specified traits of interest can be ascertained in the children.

Linkages between demographic, family, and health care databases

Data on individuals' ages in national demographic and health care data sets provide opportunities to develop large sampling frames that are particularly valuable for genetic epidemiologic studies of age-specified traits. By linking data sets combining family information with other containing current vital status and health information, one could identify large numbers of potential subjects of specified ages with specified health conditions. The value of such linkages is illustrated by Scandinavian studies that use linkages of national demographic and health data sets with genealogic information (including longevity data) and family information (35, 36). The United States does not have this range of national data sets, but linkages of American data sets could yield combined demographic, health, and family data on individuals.

One such opportunity relates to databases of the Health Care Financing Administration (HCFA), the Social Security Administration (SSA), and the US Census. The age of almost all Americans aged 65 years and older can be tracked through HCFA data and linked to health data (37). However, with some exceptions (38), one cannot directly identify familial relations from these data. SSA data include information useful for identifying relatives (place of birth, parents' names, and women's maiden names) and could be linked to HCFA databases. However, although HCFA considers requests to release data with personal identifiers to and contact beneficiaries identified through such data (subject to privacy and confidentiality regulations), SSA presently will not release such data. The Census enumerates people by households, including the name, age, sex, and relation to head of household. It allows access to data on individuals from censuses taken at least 72 years ago. Linking this family information with current information on vital status and health in HCFA, SSA, and other databases is potentially feasible.

Cell repositories and databases

Cell repositories and databases would be particularly valuable for genetic epidemiologic studies on certain age-specified traits. This is true for studies of rare traits such as exceptional longevity. The time needed to recruit enough of these subjects with such traits could dissuade an investigator from studying them. However, if a repository acquired material from them as they were identified, studies could be done relatively quickly once enough were accumulated. In addition, many elderly persons on whom there are extensive, long-term longitudinal data may die before the full range of useful studies involving them can be designed. Establishment of repositories for such populations would allow future research that would otherwise be impossible. Study design and planning would also be aided by establishment of registries and databases on families who have an aggregation of traits such as extended longevity or the absence of specific chronic diseases of aging.

CONCLUSIONS

Aging is a universal phenomenon whose relations to genetic factors cannot be ignored. Besides insights to be gained into determinants of survival traits and rate-of-change traits, there are hazards posed by not using age-specified traits in genetic epidemiologic studies. The literature on genetic factors in diseases of adult life is replete with contradictory findings. Although many of the causes of discrepancies are unrelated to age factors, age specification can reduce heterogeneity and thereby remove hidden causes of discrepancy and improve the likelihood of statistically significant findings. Thus, adding the dimension of age to outcomes in genetic epidemiologic studies can provide better insights into diseases as well as aging.

From a public health perspective, inclusion of age-specified traits in genetic epidemiologic studies could lead to improved strategies to increase healthy life expectancy and evaluate individuals' risks for age-related morbidity. Such insights could come not only from identifying genotypes with significant effects on the variance of age-specified traits in populations but also from genotypes responsible for rare extreme values of these traits, which can illuminate mechanisms affecting them in greater numbers of persons. It is also worth stressing the point that studying age-specified traits can identify protective factors as well as risk factors. Discovery of genetic factors or gene-environment interactions that contribute to successful aging, such as survival to very old age with no adverse health events or lack of decline in function, could be important benefits of such research.

ACKNOWLEDGMENTS

Participating institutions and principal staff were as follows: Dr. Evan C. Hadley and Winifred K. Rossi, Geriatrics Program, National Institute on Aging, Bethesda, MD; Dr. Steven Albert, Gertrude H. Sergievsky Center, Columbia
REFERENCES


Am J Epidemiol Vol 152, No. 11, 2000


