

Draft - Not for Quotation - May 31, 2000

DEMOGRAPHY IN THE AGE OF GENOMICS: A FIRST LOOK AT THE PROSPECTS

Douglas Ewbank
Population Studies Center
University of Pennsylvania

Prepared for discussion at the Meeting on bioindicators, Committee on Population, National Research Council, February 10-11, 2000.

Douglas Ewbank's research on genetics and demography is supported by NIA grant #R01-AG16683.

The popular consensus seems to be that genetics is the wave of the future. Information technology was the driving force that changed our economy and our society during the late twentieth century. Genetics is expected to have similar effects on medicine and the social sciences during coming decades. The 1980s and 1990s produced numerous developments in molecular biology, statistics, and computer technology. These developments make it easier to associate observed traits (e.g., diseases, risk factors for disease, personality traits, or differences in protein structures) to specific genes. The resulting changes in our understanding of genetics are so profound that Weiss (1996) has suggested they may amount to a paradigm shift. A few examples suggest the speed of change.

- The first *positional cloning* (identification of a gene by virtue of its location in the genome rather than by its biochemical function) occurred in 1986. By 1990, when the Human Genome Project (HGP) began, only a handful of genes had been identified this way since it was such a tedious process. The discovery of the gene for Huntington's chorea came in 1993, ten years after it was learned that it had to be near one end of chromosome 4. Improvements in molecular biology have greatly speeded up this process. By 1997 the number of genes identified by positional cloning was close to 100 (Collins, Guyer and Chakravarti, 1997).

- The development of new statistical techniques for studying complex traits was marked in 1993 by the publication of three textbooks in genetic epidemiology (Weiss, 1993; Khoury, 1993; Schulte and Perera, 1993). This development continued through the 1990s with improved computer programs and estimation procedures. Recent developments have reopened the debate about the best way to find genes associated with complex traits (Risch and Merikangas, 1996; Long, Grote and Langley, 1997; Bell and Taylor, 1997; Gambaro, Anglani, D'Angelo, 2000).

- The HGP's first five-year plan, for 1993-98, was to map the human genome using marker loci. By 1994, they had already published a map with about three times the resolution that they had planned for 1998. They have now sequenced and checked over 50% of the genome thus providing a complete description of the genome of a "consensus" individual. This was accomplished well ahead of the goal set for 2003.

The revolution in genetic epidemiology was just becoming apparent in 1989 when there was a meeting of geneticists and demographers to discuss "convergent issues in genetics and demography" (Adams, Lam, Hermalin, Smouse, 1990). Reading the resulting volume, it is clear that in 1989 there really were no issues pulling demographers and geneticists together. Genetics was just getting to the point where it could begin to address the kinds of questions that interest demographers. Now, more than ten years later, the nature of the revolution in genetics is clearer and we can begin to consider how it might affect demography.

For demographers, and for social scientists in general, there are several options for dealing with genetics. The first is simply to ignore it. Since we are primarily interested in the social and behavioral factors affecting demographic variables, there is a temptation to ignore genetic differences. This may be a reasonable option as long as ignoring genetics doesn't distort our estimates of the effects of social and behavioral factors. A second option is to use samples of twins (or other related individuals) to control for unobserved heterogeneity associated with genetics. However, samples of twins are hard to collect, especially when there is a need for twins raised apart in different environments. In addition, twin studies don't allow us to directly address questions about the importance of specific genes. This makes it difficult to understand differences between populations and to forecast the potential impact of developments in genetic medicine.

The third strategy is to include in our analyses data on the genetics of individuals or gene

frequencies for populations. Adding genetic information to our analyses could reduce the amount of unobserved heterogeneity and produce estimates of the contribution of specific genes to variations among individuals or across populations. This is not yet a real option. As long as most of the genes that have been identified are associated with rare diseases (like Huntington's chorea or sickle cell anemia) the potential impact of genetics on demographic research is very limited. However, genetic epidemiologists are now searching for genes that have large effects on common conditions. During the next ten years this might lead to discoveries that will substantially alter demographic research.

This paper examines how future research on complex traits made possible by the HGP will affect demography. There are two ways in which demographic research might change. First, research on the genetic basis for common diseases and mortality will benefit from applications of demographic multi-state modeling. Eventually, this could change epidemiology more than demography. Second, research on the determinants of health and behaviors could expand to include controls for genetic differences. As more genes are linked to common diseases and behaviors, adding genetic data into statistical analyses will become more attractive. However, it is important to be realistic about what we can expect from genetics. In particular, demographers need to think ahead about what kinds of genetic associations will be useful for our purposes.

Before turning to the implications for demography of new development in genetics, it is useful to examine the developments in genetic epidemiology during the past 15 years. This review provides a framework within which to discuss the likely developments in genetics in the next five to ten years.

An Outline of Genetic Epidemiology

The revolution in genetics has been driven largely by developments in molecular biology. However, for demographers the important changes can be more easily described through developments in genetic epidemiology.¹ *Genetic epidemiology* is study of the relationship between *genotypes* (the particular combination of genes carried by an individual) and *phenotypes* (observable traits). The choice of statistical methods depends on whether the trait is *quantitative* (continuous variables like body weight) or *qualitative* (i.e., discrete variables such as being overweight or a case of diabetes). Genetic variation results from errors in chromosome duplication which lead to different forms of a gene (termed *alleles*). The most common difference among alleles is single base-pair differences called *single-nucleotide polymorphisms* (SNPs). Some mutations render the gene completely incapable of performing its intended function, but most alleles have no noticeable effect on gene functioning. Most genes have only one allele with high frequency and many (often hundreds) of rare alleles (Weiss, 1993).

¹ Ridley's popular book, *Genome: An Autobiography of a species in 23 Chapters* (1999) provides a fascinating review of the history of genetics and an exceptionally clear discussion of the complexity of inheritance revealed by recent research. Weiss (1996) provides a summary of how the developments of the 1980s and early 1990s have moved genetics away from a simple Mendelian view. Lander and Schork (1994) provide a non-statistical discussion of the basic approaches in genetic epidemiology. Weiss (1993) presents the basic statistical methods and an excellent overview of human evolution. A brief overview of the molecular biology that makes the HGP possible can be found in a series of articles by Ellsworth and Monolio (1999a, 1999b, and 1999c) which include excellent glossaries of important terms.

An individual's genotype is defined by the particular combination of alleles he or she carries. Common alleles (1% or more of the alleles in the population) are termed *polymorphisms*. A gene is not apt to explain much of the variation in risk in a population unless it has common polymorphisms or numerous different alleles that are all associated with substantial excess risk. Genes that have at least two genotypes that are associated with very different risks or with very different distributions of quantitative traits are known as *major* genes or *oligogenes*. Major genes for quantitative traits are usually defined as those for which the mean values for two genotypes differ by at least 2.5 times the standard deviation within genotypes (Weiss, 1993).

Most variables of interest to demographers are what genetic epidemiologists call *complex traits*. They are traits that are affected by numerous genes as well as the environment and interactions between environment and genotype. Variables like mortality, "health status," and limitations of activities of daily living are extreme cases of complex traits. However, even the individual health problems that demographers consider as components of health and mortality are very complex.

The genetics revolution started with breakthroughs that increase the ability of genetic epidemiology to link specific genes with individual traits. The identification of the genes responsible for specific traits then forms the basis for all of the other aspects of the genetic revolution including the promises of medical genetics and the potential future use of genetic information in demographic research. Recent developments will greatly increase the rate of discovery of genes associated with complex traits.

It is useful to distinguish four areas of research in genetic epidemiology in humans. The first two examine the role of genetic factors without reference to specific genes. The third area involves research to actually identify the relevant genes by determining their position on individual chromosomes. The fourth area uses genetic differences between populations to study the origin of populations.²

Twin studies and family studies of the contribution of genetics to observed differences among individuals. Genetic epidemiologists have long relied on studies of twins and other related individuals to estimate the relative importance of genetics in determining various traits. They apply variance component models to decompose differences in quantitative traits (like blood pressure) into components associated with genetics, family environment, individual unshared traits, and interactions among these factors. Different sample designs give information about different factors. For example, comparisons of *monozygotic* (i.e., identical) and *dizygotic* (fraternal) twins provide estimates of the contribution of genetics. Comparisons of monozygotic twins raised apart provide estimates of the contribution of shared environment. One outcome of these studies is estimates of *heritability*, the proportion of the variation in the distribution of a quantitative trait that is explained by genetics.³ For example, a study of Danish twins estimated that about 25% of the variation in life span is genetically determined (Herskind et al., 1996). Twins studies have produced estimates of heritability for a wide range of traits. For example it has been estimated that genes explain 25% to 50% of the variation in the risk of cancer, IQ

² Genetic research in non-human populations often involves cross-breeding which enables researchers to increase the frequency of a trait (and therefore the associated genes). It can also be used to increase genetic heterogeneity.

³ The equivalent measure for qualitative traits is a relative risk.

scores, risk-taking behavior, and sexuality. Estimates of heritability are responsible for much of the excitement (and anxiety) surrounding recent developments in genetics.

Demographers and economists have occasionally applied variance component models to twin data (e.g., Behrman, Rosenzweig and Taubman, 1994). However, they have also used data on twins as controls for genetics to improve estimates of the effects of other variables. For example, they have used data on twins to control for genetic endowments and improve estimates of the economic returns to education (Miller, Mulvey, and Martin, 1995; Behrman, Rosenzweig and Taubman, 1996).

Inheritance patterns for genetically determined traits. Studies of families to determine inheritance patterns use *segregation analysis*. By examining the proportions of siblings (or more distant relatives) that exhibit a trait it is possible to distinguish various genetic patterns (e.g., a single recessive gene) and to estimate the rate of *penetrance* (the probability of developing the trait given a specific genotype). Until the late 1980s this research was primarily focused on *Mendelian models*, that is qualitative traits caused by single genes with a high rate of penetrance. This research led to an expansion of genetic counseling. It was generally most successful for diseases associated with clearly defined outcomes with high penetrance and young ages at onset. Recently, the focus has shifted to the search for genes associated with variation in quantitative traits. Quantitative traits are generally *multifactorial* and *polygenic*, that is they are determined by the interaction of several genes or between genes and environment.⁴ Individual genes contributing to a quantitative trait are called *QTLs* (quantitative trait loci).

The inheritance patterns of traits associated with multiple genes are much more difficult to discern. During the 1980s advances in statistical techniques and computer speed led to the development of QTL models. These models assume a trait is controlled by one two important genes with moderate to large effects (termed *oligogenes*) and numerous other genes with much smaller effects (jointly termed *polygenes*). These models require strong assumptions about the distributions of the relative importance of these genes including the number of oligogenes. A brief overview of segregation models is provided by Weiss (1990). His textbook on genetic epidemiology (1993) provides a more complete discussion.

Segregation studies are complicated by gene-environment interactions. For example, segregation studies in families that exhibit large variation in relevant environmental variables may fail to identify oligogenes. Also, segregation analyses performed in populations with different environments may lead to very different conclusions because the effects of some genes may be masked by the environment.

The search for genes responsible for specific traits. The study of inheritance patterns only provides evidence that there are genes associated with a given trait. The next step is to identify the specific *loci* (i.e., locations on chromosomes) that contains these genes. There are two approaches to locating the loci associated with specific traits.

The first is *association studies*. The simplest association studies compare a trait to the presence of known alleles of a *candidate gene*.⁵ For quantitative traits this involves samples of

⁴ However, see Weiss (1993) for a discussion of the quantitative variability caused by different alleles of a single gene (PAH) associated with PKU, phenylketonuria, a well-known genetic disease.

⁵ Gambaro, Anglani, and D'Angelo (2000) discuss the difficulties in selecting candidate genes. They note that association studies are actually based on candidate alleles which are even more difficult to identify than candidate genes.

cases and controls. Studies of quantitative traits test for differences in means among genotypes. Candidate genes are often associated with a known protein. For example, the vitamin D receptor gene was a logical candidate for involvement in osteoporosis (Ralston, 1997). Alternatively, genes identified through rare alleles can be tested for the effects of more common alleles. For example, rare mutations of the genes encoding type I collagen (COLIA1 and COLIA2) lead to a severe osteoporotic condition. Therefore, a more common polymorphism is a candidate for explaining the more common osteoporosis (Ralston, 1997). The list of candidate genes will probably expand rapidly once the human genome is completely sequenced (Guo and Lange, 2000).

In the absence of a candidate gene, it is possible to do a *whole-genome scan* to look for genes associated with a trait. Testing correlations at many loci raises the problem of multiple comparisons. However, recent analyses have demonstrated that whole-genome scans can be efficient methods for identifying genes associated with specific traits even after adjusting for multiple comparisons (van den Oord, 1999). However, whole-genome scans require a large number of candidate alleles or SNPs, not just candidate genes. The HGP and other groups are beginning to address this need (see below). The availability of a large number of known alleles may make association studies the method of choice for identifying the genes associated with complex traits.

Association studies are prone to two common problems that can lead to spurious correlations. First a gene may show a close correlation with the trait because it is very close to the true causal gene on the same chromosome (see the discussion of linkage below).⁶ This can lead to close associations in one population that are not replicable in other populations since the correlations among neighboring genes will differ among populations. A second problem is *population admixture*. In a population, a trait that is more common in one ethnic group will appear to be correlated with any allele that also happens to be more common in that group. Therefore, association studies should be performed in relatively homogeneous populations like Finland and Iceland and in small populations of individuals descended from a small number of ancestors.⁷

The second approach involves *linkage analysis*, *fine mapping*, and *positional cloning*. This has been the predominant method used for genetic research during the past decade. It enables researchers to identify first a region of a chromosome then a gene based solely on the position of the gene without any knowledge of its function. Linkage takes advantage of the fact that the chromosomes inherited from your parents are not always passed on to your children intact. Instead, the two copies of the chromosome sometimes exchange segments (called *recombination*). Because of recombination, it is possible to associate the inheritance of a trait to the inheritance of a segment of a chromosome. Loci that are physically close to each other on a chromosome are more apt to remain together after recombination. Loci that are very close will

⁶ With incomplete mapping of SNPs, it is also possible to find a spurious correlation with an SNP in a non-coding region, which can not affect a trait. In this way, SNPs can act like marker loci to identify the neighborhood within a gene in which a relevant allele is located (Collins, Guyer and Chakravarti, 1997).

⁷ Another solution to this problem is to use a sample of affected individuals and their parents to control for differences in allele frequencies among populations (van den Oord, 1999).

be in *linkage disequilibrium*.⁸ It is therefore possible to examine the frequency of the trait in relation to the occurrence of *genetic markers* (known sequences of nucleotides that occur at specific locations on chromosomes). The relevant gene probably lies between the two markers that are most highly correlated with the presence of the trait. The more markers that are available, the smaller the area identified by linkage analysis. The first goal of the HGP was to produce a finer genetic map to improve the precision of linkage studies.

Linkage analysis leads to a candidate region of a chromosome. For example, linkage analysis suggested that there was a gene associated with the risk of Alzheimer's disease (AD) in the long arm of chromosome 19 (labeled 19q).⁹ The gene can then be identified within this region through *fine mapping* based on *positional cloning* (Ellsworth and Monolio, 1999b) The gene for AD turned out to be the gene for apolipoprotein-E, APOE, (Corder et al., 1993) which is described below. Fine mapping is a time consuming process since there can be hundreds of genes between markers. This process will be eased by the complete sequencing of the human genome.

Linkage studies require large pedigrees (i.e., families in which the trait in question is unusually common). Linkage can be very difficult for traits that don't follow Mendelian inheritance (i.e., a single gene with few alleles). It can also be difficult in the case of common alleles. When the risky allele is common, many individuals will be homozygous for the risky allele. Since the two copies of the allele may be linked with different markers, inheritance may not always be associated with the same marker. This problem complicated early linkage studies of Alzheimer's disease and the identification of a linkage to chromosome 19 (Lander and Schork, 1994; Corder et al., 1993).

Studies of the Origins of Human Populations. A third area of research applies knowledge about the geographic distribution of a few dozen alleles (often including genes for blood type) or markers of genotype (e.g., lactose intolerance) to infer historical relationships among populations (Cavalli-Sforza, Menozzi, and Piazza, 1994). When combined with archeological and linguistic evidence, these maps provide important insights into the origin of man, ancient migration streams (Owens and King, 1999), and the role of evolution in human history. An excellent example of the use of mapping is research on the geographic distribution of lactose malabsorption (Simoons, 1978; Weinberg, 1999).

The Human Genome Project.

The HGP will significantly increase the speed of discovery of the genes associated with specific traits. Linkage analysis and genome-wide scans depend on the availability of numerous markers and maps of the genes that lie between them. The mapping of the human genome will provide a very detailed map thereby increasing the ability to narrow in on the specific loci associated with a given trait. This development, combined with improved statistical methods and expanded computer power, makes possible large-scale searches for the genes associated with

⁸ If the two loci were not linked, the inheritance of an allele of one gene would be independent of the inheritance of an allele of the other gene. Independent inheritance is associated with equilibrium when there is random mating. Therefore, correlated risks of joint inheritance of two alleles constitutes disequilibrium.

⁹ The short arm of each chromosome is labeled *p* and the long arm is *q*.

complex traits. The full sequencing will also expand the identification of candidate genes based on an understanding of the functioning of genes (Guo and Lange, 2000).

Weiss (1998) points out that the HGP was originally designed to produce a map for an “average” individual. To social scientists it is genetic diversity that is important. Heterogeneity is also central to genetic epidemiology research on humans. The only way we can study the action of a gene is by observing mutations which alter gene functioning. The study of diversity was added to the goals for HGP in 1998. The goal is mapping 100,000 polymorphisms involving SNPs by 2003 (Collins et al., 1998). The HGP search for SNPs will take advantage of the diversity of the U.S. population. Although this is a huge number, it is estimated that there are about 200,000 SNPs in protein-coding regions (cSNPs) which are apt to be most important for understanding disease (Collins, Guyer, and Chakravarti, 1997). A second project involving the Wellcome Trust and ten international pharmaceutical partners was formed in 1999 to identify 300,000 DNA variants (Cardon and Watkins, 2000).

Demography and the Genetics of Complex Traits

The developments in genetic epidemiology during the past fifteen years have greatly expanded the opportunities for identifying the genes associated with variation in complex quantitative traits. As more genes are identified, the potential gain from incorporating genetic information into demographic research will increase dramatically.

Measured genotypes associated with common traits are in some ways ideal variables for the kinds of research conducted by demographers and other social scientists. The reason is simple: genotype is fixed at birth. This has two implications for the relationship between genetics and demography. First, demographic models are ideally suited to the study of fixed traits. Second, we can add genetic information to our statistical model and improve the fit without introducing complex correlations with other independent variables.

The following sections discuss potential applications of demographic models to the study of complex traits and the use of genetic information in research on standard demographic variables. These are the two areas where developments in genetic epidemiology are apt to have the biggest impact on demography and demography is apt to have the biggest impact on epidemiology.

Demographic Models for Studying Major Genes Affecting Common Diseases

Once a gene for a common, complex condition has been identified, there will be numerous questions about its effect in populations. These problems are apparent in research on the only known gene like this, the apolipoprotein-E gene (APOE). APOE is so unique and so heavily studied that few discussions of genetics can avoid using it as an example. It is a major risk factor for both ischemic heart disease, IHD, (Wilson et al., 1996) and Alzheimer’s disease (Corder et al., 1993; Farrer et al., 1997). The APOE gene has three common polymorphisms labeled ,2, ,3, and ,4.¹⁰ Therefore, individuals have one of six possible genotypes ,2/2, ,2/3, ,2/4, ,3/3, ,3/4, or ,4/4. The ,3/3 is the most common genotype comprising about 60-70% in all populations. The ,3/4 and ,4/4 genotypes are associated with increased risk of both IHD and AD. The ,2/2 and ,2/3 genotypes are associated with reduced risk of AD.

¹⁰ Numerous very rare mutations of the APOE gene have also been discovered. However, there has not been any research examining the effects of these mutations.

One issue raised by the discovery of major genes involves differences in the amount of excess risk at different ages. For example, the effect of APOE ϵ_4 on the risk of AD increases with age up to about age 60 and declines at the oldest ages (Farrer et al., 1999). Similarly, a segregation analysis of the risk of lung cancer suggests that there is a major gene that has a very large effect on the risk under age 60, but only marginal effects after age 80 (Gauderman and Morrison, 2000). The genetic effects on breast cancer also change with age (see discussion below). These changes might be the result of unobserved heterogeneity in the risks of disease, cohort trends in risk (e.g., Colilla et al., 2000) or changes in the nature of the disease with age. For example, very early onset AD might involve a very different natural history than AD at later ages.

A second complication arises when a gene is associated with more than one disease. For example, the gene for the vitamin D receptor appears to affect bone density. Alleles that reduce bone density might increase the risk of osteoporosis but reduce the risk of osteoarthritis (Uitterlinden et al., 1997). Most epidemiologic research examines individual diseases. For example, almost all of the research on APOE examines only its relationship with AD or with IHD. One reason for this is that few studies include both a thorough examination for dementia and precise diagnoses of cardiac events or measures of serum lipids. Case-control studies in particular are designed to study one well-defined condition.

A third set of issues arises when two or more major genes are identified as being associated with the same disease. Since most epidemiologic research focuses on the effects of single genes, there may be little direct evidence of the combined effects of several genes. If the effects of different genes are not additive, estimates of the effects of alleles of one gene might differ between populations because of unobserved differences at other loci. The interactions of the two genes can be very complex especially if the gene frequencies differ across populations or the effects of each gene change with age. For example, mutations of the PS-1 and PS-2 genes are associated with very early onset AD (Lendon, Ashall and Goate, 1997). It appears that the increase in the relative risks of AD under age 60 associated with the ϵ_4 allele of APOE are due to complications introduced by PS-1 and PS-2 (Ewbank, unpublished results).

These problems complicate research on mortality differences by APOE genotype. APOE ϵ_4 allele is clearly associated with increased risk of death due to both IHD and AD, at least in males (Ewbank, 1999). There are numerous studies which suggest the importance of APOE for mortality, but few provide direct evidence of mortality differentials by genotype. No single study is large enough to provide solid evidence of the effect of APOE on mortality at various ages. Therefore, it is necessary to combine studies to understand the effects of APOE genotype on the level and age pattern of mortality at the oldest ages.

The process of comparing and combining studies is complicated by the fact that published analyses present different types of data. For example, the published analysis of data from the Kungsholmen Project in Sweden provides estimates of excess risk of death over a seven-year period by APOE genotype for a cohort (Corder et al., 1996). Eichner et al. (1993) provide data from a case-control study showing excess risk associated with the ϵ_4 allele. Stengård et al. (1995) provide data on the APOE allele frequencies (i.e., the proportion of alleles, not individuals, that are of each APOE type) for survivors and decedents over a five-year period. Several studies document that the ϵ_4 allele is less common among centenarians than among octogenarians which suggests excess mortality (Schächter et al., 1994; Asada et al., 1996; Louhija et al., 1994). Each of these studies provides evidence on excess risk of death associated with the ϵ_4 allele. However, the measures provided by the studies are not comparable.

Demographic multi-state models are ideally suited to dealing with these issues. Since genotype is a fixed trait, it is possible to develop a basic multi-state model and apply it separately to each genotype with different risks. The model can easily incorporate multiple causes of death, age patterns of onset of disease that incorporate heterogeneity, and, if necessary, duration of disease. If two or more major genes are involved, the multi-state model can be applied separately to each combination of genotypes. If there are subgroups in the population defined by non-genetic characteristics (e.g., race/ethnicity or sex) these can be incorporated as well. The effects of different combinations of risk by genotype on the total population can be studied by merely summing the l_x or ${}_nL_x$ columns for the subgroups to get a multi-state table for the whole population.

Genetic epidemiology has developed most of the tools necessary for identifying the genes associated with common complex diseases. However, once those genes have been identified demographers and epidemiologists will have to develop the tools to study their joint impact of health and mortality. Demographic multi-state models will become increasingly important for sorting out the interactions between multiple genes and diseases.

Potential Uses of Genetic Information in Demographic Research

The recent revolution in genetic epidemiology means that demographers will increasingly have the option of including measured genotypes in their data collection and analyses. Several areas of demographic research are particularly ripe for including information on the genotypes of individuals. The following sections discuss four applications. The first is the use of genes in demographic research on the social correlates of health status, mortality and other demographic events. This discussion examines the potential impact of genetic information on analyses based on linear regression, logistic regression or survival analysis.

The second section discusses the potential value to studies of the age pattern of mortality or onset of ill health at the older ages. This involves a discussion of the potential contribution of genetic information to understanding heterogeneity in frailty.

The third area is the potential value of genetic information in research on differences between populations. This includes cross-national comparisons and research on differences by race and ethnicity. The fourth section discusses how the discovery of genes associated with behavior might change demographic research. The fourth section discusses the potential importance to demography of longevity genes (or “gerontogenes”). The final section discusses a few examples of epidemiologic surveys that include genetics information. Data from these surveys might be useful for the development of methods for incorporating genetics into demographic research.

These discussions lead to the concept of “demogenes.” These genes have a sufficiently large impact that their effects are important at the population level. The number of demogenes discovered in the next decade will determine how much impact genetics has on demography during the next twenty years.

Genetic Information in Studies of Differences among Individuals.

To understand what demography would gain from genetic information, we have to consider the range of analytical approaches that constitute the bulk of current demographic research. Much demographic research is based on regression-type statistical methods. The outcome of interest, y , might be death, the onset of disability, savings behavior, or retirement.

This is generally related to a vector of social and behavioral variables, \mathbf{X} . We are now considering what would be gained by adding a vector of genetic characteristics, \mathbf{G} . The resulting generalized linear model is:

$$f(y_i) = \alpha + \beta \mathbf{X}_i + \gamma \mathbf{G}_i + \theta \mathbf{X}_i \mathbf{G}_i . \quad (1)$$

The choice of the functional form for $f(y)$ determines whether the model is a simple linear regression, a logistic, probit or Poisson regression or some other model. The following paragraphs examine how genetic information might be utilized in linear regression, logistic regression, and survival analysis to study the effect of social, economic, and behavioral characteristics on demographic variables. Each type of analysis highlights different issues.

Predicting Status or Explaining Variance. Segregation analyses of numerous traits (including longevity) suggest that we will not explain more than 60-90% of the variance in many outcomes if we ignore genetics. However, sample sizes in the largest surveys are too small to capture all of the variance attributable to genetic variation in standard regression-type analyses. We will generally have to focus on a small number of genes that are most relevant to our research. For a gene to be useful for statistical analysis it would have to have common alleles associated with large differences in the dependent variable. An allele (or a group of alleles) found in only 0.01% of the population or that increased risk by only 0.01% contributes very little to understanding the distributions of demographic variables.

An alternative is to develop ways of combining data from multiple sources to control for the effects of rare genotypes. For example, we might combine data from a demographic survey with data from case-control studies. The case-control samples would provide power for estimating the effects of rare genotypes but they include very few social and behavioral variables. Therefore, we would need methods for combining data sets that do not include all of the same variables. A simple method is to impose estimates of the effects of the genotypes from epidemiologic studies on the analysis of data from a demographic survey. For example, we could adjust observed blood pressure measurements for known differences between genotypes. Similarly, we could use offsets in Poisson regressions to incorporate the combined effect of many genotypes. This approach would be relatively easy to implement, would allow controls for even the rarest genotypes, and could make use of published estimates from the case-control studies or from meta-analyses.

However, using point estimates from case-control surveys would exaggerate the precision of the resulting estimates. Even the largest case-control studies produce estimates with large confidence intervals. Fully utilizing external sources of information to inform adjustments would require maximizing the joint likelihood of observing the data reported in different data sets. Methods for imputing missing data could be adapted to this problem.

Genes as Controls to Improve Estimates of the Effects of Other Variables. In some cases excluding genetic variables, the $\gamma \mathbf{G}_i$ term in equation (1), might lead to biased estimates of the effects of social and behavioral variables, the β . In linear regression, excluding a relevant variable from the analysis does not bias the estimates of the parameters of interest unless that variable is highly correlated with the independent variable of interest. However, this is not apt to be the case. A genotype that is associated with a health outcome (e.g., a gene associated with cancer) is not apt to also be highly correlated with behavioral risk factors (e.g., diet, smoking,

etc.).¹¹ In particular, since genotype is not determined by choice, we don't have the problem of joint determination that often causes problems in social research.¹²

The problem is somewhat different in logistic regression. Omitting a variable biases the estimates of the other coefficients even if there is no correlation between the omitted variable and the variable of interest. However, in the absence of interaction effects the magnitude of the bias is related to the magnitude of the effect of the omitted variable. To have a noticeable effect, an omitted genotype would have to be an important determinant of the probability of the event in question. Therefore, excluding genetic determinants from analyses of differences in health and behavior is not apt to be a frequent source of significant error.

Genes as Effect Modifiers. Recent developments in the methodologies for linkage analysis and association studies have increased the power of genetic epidemiology to study gene-environment interactions (Yang and Khoury, 1997; van den Oord, 1999). For demographers, controlling for genotype is potentially important when it interacts with social or behavioral variables. In equation (1) effect modifiers are shown as the interaction term $\theta X_i G_i$, which genetic epidemiologists often refer to simply as $G \times E$. The full effect of the behavioral variables is mis-specified if the interaction term is omitted. Gene-environment interactions would only cause serious biases if alleles associated with significant interactions with environmental variables were very common.

For example, there are two common alleles that probably have significant interactions with a high fat diet. Carriers of the APOE ϵ_4 allele may be more susceptible to the effects of high fat diets on the risk of death to ischemic heart disease. Similarly, the 825T allele of the gene for the G protein β_3 subunit (GNB3) appears to be associated with obesity (Siffert et al., 1999). If the relationship of the 825T allele to obesity is replicated, understanding the effect of diet on health might require interaction terms between diet and the APOE and GNB3 genotypes. Understanding these interactions might also be important for predicting future trends in mortality. In particular, African populations have high allele frequencies for the risky genotypes of both genes (allele frequencies of about 20% for the ϵ_4 allele of APOE and about 80% for the 825T allele of GNB3) and may therefore be particularly susceptible to the health and mortality risks associated with a high fat diet (Zekraou et al., 1997; Corbo et al., 1999; Siffert et al., 1999).

Genes as Instrumental Variables Research on the relationship between health and social status variables is complicated by the fact that social, economic and health variables are so intertwined. For example, we might want to control for health status when studying the economic correlates of labor force participation. However, we will never be able to measure all of the determinants of participation rates. If some of the unmeasured determinants are also

¹¹ An exception would be genes that are risk factors for behaviors. An example would be genotypes that increased the risk of addiction to alcohol or nicotine. In these cases the intermediate variables model is actually inverted. The effect of such a genotype on health would work through the intermediate behavior variable. However, if demographers are only interested in the link between the behavior and the outcome, there may be little gained by incorporating the genetic risk factors for the behavior since the selectivity associated with the genotype is not apt to be correlated with other behaviors of interest to demographers.

¹² There is a chance that some genes that cause ill health at early ages might affect social status variables determined at young ages such as education or marital status.

correlated with health status (for example, psychological variables, family history, or personal circumstances), then the regression estimates will be biased. This would affect the estimates of all of the coefficients, not just the coefficient on health status. One approach to this problem is the use of instrumental variables. This involves replacing the health status variable with variables that are highly correlated with health status but not correlated with the unobserved variables. Although this approach is good in theory, it is generally very difficult to find appropriate instrumental variables.

Genetic information might be useful instrumental variables. Since genotype is determined at birth, it is not affected by any aspects of life history. If sufficient genetic information were available to identify a substantial fraction of those with high risks of health problems, we could replace actual health status with measures of genetic risk of health problems. However, we are rarely interested in controlling for specific health conditions and overall health status is determined by a large number of genes. Therefore, the value of genes as instrumental variables for health status will depend on the number of common genotypes associated with excess risk of the most common health problems.

Genes as Sources of Heterogeneity in Survival Analysis:

In introducing a 1990 volume edited by Adams, Hermalin, Lam and Smouse entitled *Convergent Issues in Genetics and Demography*, Julian Adams suggested that

“[p]erhaps the most striking difference in approach and paradoxically the best hope for a convergence of the two fields can be seen in the way in which the fields view within-population variation.” (Pg. 10)

A third of the book was devoted to the section entitled “Heterogeneity, Phenotypic Variation, and Frailty.” Demographers are primarily interested in studying the variation associated with social and behavioral variables. However, other sources of variation cannot be safely ignored in survival analyses and multi-state models. In Trussell and Rodrigues’s review of statistical approaches to handling “unobserved” heterogeneity they conclude that:

“The methods proposed to correct for unobservable heterogeneity deliver less than is commonly assumed, particularly because of an inherent non-identifiability involved when the analyst must rely on observables to assess goodness-of-fit.” (Pg. 129)

Incorporating genetic information into large demographic surveys would reduce the amount of heterogeneity that is unobserved. This would reduce the importance of the mathematical assumptions about the distribution of unobserved frailty.

The next section discusses the effects of unobserved heterogeneity in a different context. The conclusions from that discussion apply here as well. In summary, it is not likely that we will be able to use standard regression techniques to control for many genotypes. Therefore, the contribution of genetic information to controlling for heterogeneity will depend on the discovery of a few genes associated with a large fraction of the unobserved heterogeneity.

Genetic research also might provide evidence of the functional form for the distribution of risks (Weiss, 1990). This would improve our controls for unobserved heterogeneity even if demographic surveys did not collect genetic information for individuals. This possibility was suggested by several of the authors in *Convergent Issues in Genetics and Demography* (Adams et al., 1990). However, several notes of caution are in order. First, genetics is only one of the sources of unobserved heterogeneity. Unobserved differences in behavior (e.g., attitudes toward health care, disease prevention behaviors) and in personal history (previous illnesses, accidents, etc.) have a significant impact on the health of the elderly. Therefore, firm estimates of the distribution of genetic risk factors would not completely solve the problem of choosing a

functional form for unobserved heterogeneity.

Second, genetic epidemiology is focused on gene finding. Their study designs and research agendas may not lead to useful information on the overall distribution of genetic risk. If demographers want to use genetic information to derive functional forms for the distribution of genetic frailty, we may have to tackle that question ourselves. Estimating the distribution of genetic frailty would require combining information on the co-occurrence of risky genotypes, the relative risks of each genotype, and information on how those risks are combined to form total risk. Data from large-scale demographic surveys could provide data for examining the distribution of risk associated with both genetic and behavioral variables.

Genes and Demographic Models of the Age Pattern of Mortality.

The rate of increase in mortality at the oldest ages can be thought of as the result of two factors: 1) the increase in risk with age for individuals and 2) variation in risk among individuals. We only observe the way mortality increases with age in a population, not the risks for individuals or the distribution of risks across individuals. However, we want to understand how individuals age. To do this, we either have to explain most of the variation among individuals or model the effects of unobserved variation on the rates for a population.

This problem can be illustrated using a simple simulation. We assume that the age pattern of mortality for white males in the U.S. results from averaging two hypothetical subgroups: one with low risk and one with elevated risk. We assume that at birth half of the population has low risk and half has high risk. We assume that the high risk subgroup has a mortality rate three times that of the low risk subgroup at every age. At age 50, the population composition is essentially the same as at birth and the risk for the population is close to the average of the risks for the subgroups. However, only about 0.2% of the high risk subgroup survives to age 100 compared to 12% of the low risk subgroup. Therefore, at age 100 the average risk in the population is very close to the risk in the low risk subgroup.

If we ignored the existence of two subgroups, we would guess that the risk for an individual increases by a factor of 7.5 between ages 70 and 90. However, this increase is based on the overall population which is 39% high risk at age 70 and 3.4% high risk at age 90. In fact, within each subgroup the risk increases by a factor of 9.5. Thus by ignoring heterogeneity we get a distorted picture of the effects of age and aging on individuals. Vaupel has stated that this is the “fundamental problem... for analyses of age-trajectories of mortality.” (1997: 25)

We rarely know anything about the number of subgroups that comprise the total population or about the variation in mortality among them. We can estimate the age pattern of risk in individuals from population data if we assume functional forms for the age pattern for individuals and for the distribution of risks in the population (Manton, Stallard, Vaupel, 1986). However, since the risks for individuals are unobservable and the distribution of risks across individuals is unknown, it is not possible to test the plausibility of our assumptions. We can improve our understanding of aging by reducing the amount of heterogeneity that is unobserved. Much of the unobserved variation in mortality risks is due to genetic variation. If a few genes were responsible for a large part of the variation in the risk of death, we could reduce the problem of unobserved heterogeneity by modeling mortality separately for subgroups defined by their genotype.

The model described above of a population composed of two subgroups provides insight into the types of genes that will be useful for reducing unobserved heterogeneity in mortality. The changing composition of the population is a result of the differences in the survival rates.

The proportion surviving to age x in a subgroup is equal to:

$$e^{-\int m_x dx}$$

The relative size of two groups at age y with mortality rates ${}^1\mu_x$ and ${}^2\mu_x$ is:

$$\frac{B_2 e^{-\int {}^2m_x dx}}{B_1 e^{-\int {}^1m_x dx}} = \frac{B_2}{B_1} e^{-\int ({}^2m_x - {}^1m_x) dx}$$

where the B_i are the numbers born into the two groups and the integration is over ages 0 to y . We see from this that the change in the relative size of the two groups depends on the absolute difference between the age-specific mortality rates in the two subgroups. Therefore, heterogeneity does not have much effect on the hazard rate unless the subgroups have very different risks.¹³

For genetic information to explain much heterogeneity, the relative risk associated with the risky genotypes must be large enough to cause a sufficient difference in the absolute risks. In addition, the risky genotypes must be sufficiently common to have a noticeable effect on overall mortality. Therefore to be useful for understanding heterogeneity, a gene would have to have common polymorphisms associated with relatively large differences in risk.

This leads to three conclusions about the potential gain from genetic research for understanding the effects of heterogeneity. First, it is not likely that any single genotype will explain much of the heterogeneity in mortality under age 80. Before age 80 the total mortality rates are low enough that it would take exceedingly large differences in relative mortality between genotypes to cause noticeable heterogeneity. For example, with equal size subgroups of U.S. white males at birth, it takes a relative risk of 1.5 to cause a noticeable change in population composition by age 80.¹⁴ A less equal split between subgroups would require a much larger relative risk. To put this into perspective, I estimate that the APOE ϵ_4/ϵ_4 genotype is associated with a relative risk of death at age 80 of about 2 relative to the most common genotype, ϵ_3/ϵ_3 . Less than 5% of the population has the ϵ_4/ϵ_4 genotype.

Second, even at the oldest ages only very common genotypes associated with large differences in risk are apt to be useful in demographic modeling of mortality. Third, since single genes are not apt to have a big enough impact, explaining a substantial fraction of the effects of heterogeneity will probably require the use of complex genotypes that combine information on several genes.

The situation is similar if we look at individual causes of death or chronic diseases instead of total mortality rates. Although there are probably genes with larger relative risks for specific diseases, the incidence rates are much lower than the overall risk of death. Therefore, to get sufficiently large absolute differences in rates would require even larger relative risks. For example, we can simulate two equal sized subgroups at birth which differ only in their mortality

¹³ This ignores changes in functional form associated with averaging hazards functions for the two subgroups. For example, the weighted average of two Weibull functions for the risk of repeatable events is a Weibull function only if the two subpopulations have the same exponent on age.

¹⁴ A noticeable change in population composition was taken to be a drop in the percent in the high risk subgroup from 50% at birth to 40%.

to ischemic heart disease (the most common cause of death). To get noticeable signs of heterogeneity by age 70 requires that the high risk subgroup have IHD mortality rates 3 times the low risk subgroup. Common genotypes for common conditions are not apt to have relative risks anywhere near this high. A relative risk for IHD mortality of 1.5 doesn't lead to a noticeable effect of heterogeneity until after age 90. Therefore, reducing the amount of unobserved heterogeneity will usually require several common risk factors that are each associated with large relative risks.

Studies of Differences among Populations

Research on differences among populations always confront the possibility that the differences are partly due to genetics. Without controls for genetic differences it is difficult to estimate the relative importance of various social, economic and cultural factors.¹⁵ Gene information can only help explain variation across populations if a) there are significant differences in allele frequencies across populations, or b) there are important gene-environment interactions. These two criteria are often related. Large differences in allele frequencies are often the result of current or historical differences in environment. Polymorphisms can arise in situations where there are advantages of different alleles. For example, the sickle cell mutation provides some protection against falciparum malaria to those who are heterozygous (i.e., have only one copy of the mutation). However, those who are homozygous suffer life-threatening anemia (Weiss, 1993). In populations that historically lived in malarious areas, the frequency of the mutation was determined by the balance between the survival advantage of being heterozygous and the disadvantage of being homozygous. In populations that emerged in areas free of malaria, the sickle cell mutation is very rare.

Genetic variation within populations does not guarantee differences among populations, so genes that are useful for explaining variation within populations may not explain variation across populations. This has one important implication for identifying genes that might be useful for demographic research. To this point, I have not differentiated between genes that have a few very common polymorphisms and genes that have numerous rare alleles. If there are many rare mutations that are associated with increased risk, it is not likely that those mutations will cluster in the same populations.¹⁶ An example of this is the genes for breast cancer which are discussed below.

For comparisons of populations, the relevant index is the sum of the frequencies of all of the risky mutations.¹⁷ If there are numerous risky alleles, a surplus of one risky allele in one

¹⁵ Sokal et al., (1997) have performed an interesting analysis of cancer mortality rates in Europe without examining the effects of individual alleles. They found that cancer mortality rates are more closely correlated with a measure of ethnohistorical distance between populations than with a measure of genetic distances. This was true of rates for many cancer sites as well as overall cancer mortality. It was also true in both Western and Central Europe. This suggests that cultural practices may be more important than genetic differences in determining cancer mortality. This research strategy is an interesting alternative to research using allele frequencies for individual genes.

¹⁶ This applies to naturally occurring populations, not populations defined by risk factors for an accelerated rate of mutation, such as exposure to radiation or occupational hazards that might increase the rate of mutations in germ cells.

¹⁷ For simplicity, I assume that risky alleles are associated with the same amount of excess risk.

population might be counter-balanced by a surplus of a second allele elsewhere. It is reasonable to assume that mutations occur with similar frequency in all populations and selection keeps each risky allele rare in all populations. In that case, there might be differences in the frequencies of individual risky alleles, but little variation in the total frequency of risky alleles.

This is not the case if there are one or more common polymorphisms associated with substantial risk. If one or two mutations become common in at least one population (say at least 5% of all alleles), there is no reason to believe that the same mutation will be equally common in all populations. For example, differences in the frequency of common alleles can be due to founder effects¹⁸ or differences in environment. Therefore, polymorphisms will be more useful for explaining variation across populations than will differences in the frequencies of numerous rare mutations.

Gene-environment interactions (including interactions with behavior) might be very important for understanding differences among populations. This is especially true since large differences in gene frequencies may reflect current or historical differences in environment. Therefore, differences in gene frequencies between populations may indicate that gene-environment interactions are potentially important.

Gene-gene interactions may also be important. The effects of a mutation of one gene might be counterbalanced by the effects of mutations of other genes. Therefore, an allele may not have the same effect in all populations even given the same current environment. This complicates studies of the contribution of genetics to race/ethnic differences in health. For example, we cannot assume that the effect of APOE genotype on serum cholesterol levels is exactly the same in all ethnic groups since numerous genes affect serum lipid levels. The effects must be documented in different ethnic groups (i.e., genetic environments) as well as in different social/behavioral environments.

Genes and Behavior.

Genes associated with basic personality traits or susceptibility to addictions could help to explain differences in behaviors like risk-taking, diet, and use of health services. The potential importance of behavioral genetics (which combines genetics and psychology) is suggested by the fact that an estimated 30% of all human genes are expressed primarily in the brain. Much of the brief summary that follows is based on recent reviews by Gilger (2000) and Merikangas and Swendsen (1997).

Twin and family studies provide intriguing insights into the possible role of genetics in personality and behavior. Numerous behavioral traits of interest to demographers have often been shown to have heritability in the range of 45% to 50% (Gilger, 2000). This includes variables associated with personality (e.g., risk-taking behavior, harm avoidance, self-control), cognition (IQ, memory, speed of processing information) and social status variables associated with achievement (occupation and years of education). For example, one study of female twins suggested that about half of the variance in several measures of perceived social support from friends, family, and confidants is attributable to genetics (Kendler, 1997).

Although these results are intriguing, genetic information might not provide much insight into behaviors of interest to demographers in the next ten to twenty years. First, even narrowly defined aspects of behavior are polygenic so it might be necessary to control for dozens of genes

¹⁸ Founder effects occur when migration leads to a population that is descended from a small number of ancestors. Weiss (1993) provides an excellent discussion of the factors that lead to genetic heterogeneity within and between populations.

to add much to our understanding of differences in complex behaviors (like savings decisions). However, it may also turn out that some genes have wide-spread effects on numerous aspects of behavior. For example, research in behavioral genetics suggests there may be some genes that affect multiple dimensions of cognition (Gilger, 2000).

The second issue is that it will be exceedingly difficult to identify the specific genes that affect behavior even with QTL models and dense maps of the human genome. This is even true for severe psychiatric disorders such as schizophrenia, major mood disorders, and panic disorder which have very high estimates of heritability. The evidence suggests that the genetic causes of these diseases are very complex and often involve gene-environment interactions. The use of linkage and association studies have produced a number of candidate genes for these disorders, but there have been problems of consistency and replicability of results (Merikangas and Swendsen, 1997).

One area of interest to demographers studying health and behavior is genetic research on addictive behaviors, including smoking and alcohol consumption. Despite newspaper headlines, little is known for sure about specific genetic markers for addiction. Merikangas and Swendsen concluded that “although several investigations have replicated significant associations between alcoholism and [several genetic] markers, the majority of investigations are either preliminary, nonconfirmatory, or have revealed potential sampling biases that may independently explain observed associations” (1997: 153).

Thus far, the most successful case in behavioral genetics is the study of dyslexia. Even there, the specific genes and alleles are uncertain. Only the identification of a region of chromosome 6 is definite (Gilger, 2000). Therefore, despite the HGP and numerous developments in genetic epidemiology, it may be decades before much is known about specific genes that affect behavior.

A third issue is that associations between genes and behavior might differ substantially by age, sex, ethnic groups and social environment. For example, numerous genes on the X chromosome have alleles associated with very low intelligence. It is possible that other alleles have less dramatic effects. Since women have two copies of the X chromosome and men only have one, the effects of genes associated with intelligence on the X chromosome might differ substantially by sex. This might explain the fact that men are more likely to be at the extremes of intelligence than women (Gécz and Mulley, 2000). The effects of genes on intelligence also vary by age. Estimates of heritability of IQ increase from about 15% among young children, to about 40% among adolescence and about 80% among older adults (Gilger, 2000). Similarly, the relative role of genetics in antisocial and criminal behavior is probably very different among teenagers than among adults (Lyons et al., 1995). Social and ethnic differences in gene action attributable to gene-environment and gene-gene interactions are especially difficult problems for the use of genetics in nationally representative samples. Behaviors or health conditions that vary by ethnic group (for example, between northern and southern European heritage) could be falsely linked to polymorphisms that also differ by ethnicity.

When behavioral genetics does discover specific genes associated with behaviors, demographic surveys might prove invaluable for putting the results into a social context. Associations between genes and personality traits are especially prone to exaggeration and misunderstanding in public discussions. Individual genes probably explain little of the variation in complex behaviors. However, preliminary reports of genetic factors affecting behavior are so intriguing that they invite speculation that often goes far beyond the actual research findings. It is possible to combine case-control data and gene frequencies to estimate how much variation is

attributable to a particular genotype. However, this will always be less convincing than a head-to-head comparison of genetic effects and social, behavioral, and economic effects using a single data set. Large-scale demographic surveys that measure numerous complex behaviors like saving rates, family caregiving, and health practices could provide invaluable tests of the relative importance of genetics for common behaviors.

Longevity Genes - A Special Case.

Demographers are fascinated by the possibility that one or more genes might determine the rate of decline in multiple organ systems. Several such genes have been identified in other species (Vaupel et al., 1998). These genes are sometimes called *gerontogenes* or *longevity genes*. The discovery of one or more genes that act as aging “clocks” in humans would be a major breakthrough for genetics. However, the mere existence of such genes would not have a major effect on demographic research. For example, a mutation in a longevity gene that was present in 0.1% of the population would still be rare (probably less than 1%) among centenarians.¹⁹ Such a genotype would not explain much about survival to the oldest ages. Therefore, in order to be important for demographic research, there would have to be common polymorphisms associated with large differences in survival.

Any discovery about the biological determinants of the rate of aging raises the possibility of therapies to slow aging. Therefore the discovery of a gerontogene with even very rare mutations that increased longevity would cause speculation about future trends in mortality. However, the discovery of such a gene would only be relevant to long-term (and, therefore, very speculative) projections.

Prospective Epidemiologic Surveys that Include Genetic Information

Some epidemiologic cohort studies of populations have collected genetic information that could be used for demographic research. It is instructive to examine a few examples of data on the APOE gene collected in population-based epidemiologic studies. One recent example is the Helsinki Ageing Study, a prospective study of a sample of individuals born in 1904, 1909, and 1914. The study began in 1989 and included blood samples that were tested for APOE. Tilvis, Strandberg, and Juva (1998) present five-year survival rates by the presence or absence of an ϵ_4 allele. Carriers of the ϵ_4 allele had a mortality rate between ages 75 and 80 that was 1.85 times that of the rest of the population. Between ages 80 and 85, the risk ratio was 1.52. There was no evidence of excess mortality between 85 and 90 (risk ratio of 0.98). A Cox regression controlling for age and sex showed a risk ratio of 1.61 associated with the ϵ_4 allele. The authors do not present results controlling for any other variables, so we don't know whether controlling for APOE changes the estimates of the effects of social and economic variables that are of interest to demographers.

This study is typical of epidemiologic studies that include APOE genotyping. Other studies in the U.S. that provide similar data include: a subsample of the MRFIT study (Eichner et al., 1993), the Framingham study (Myers et al., 1996), the Framingham Offspring Study (Schaefer et al., 1994; Wilson et al., 1994), the NHANES III (National Center for Health Statistics, 2000), the Iowa PESE study (Ferrucci et al. 1997), several epidemiologic studies of AD (e.g., Evans et al., 1997), and studies of other conditions (e.g., The Study of Osteoporotic

¹⁹ If the rare genotype was associated with a ten-fold increase in the chance of surviving to age 100, the gene frequency at age 100 would be slightly less than 10 times the frequency at birth.

Fractures, Vogt, Cauley and Kuller, 1997). Population-based European studies include the Kungsholmen Study in Sweden (Corder et al., 1996) and the Rotterdam Study (Slooter et al., 1998), etc.). The research from these studies is generally limited to the association of APOE genotypes with one outcome. Most of them control only for age and sex, although they rarely provide data by sex unless the differences are statistically significant.²⁰ Many of these studies could be used to study the relative importance of APOE genotype and other risk factors in determining mortality risks.

Identifying Genes That Might be Important for Demography

Genes that might be of interest to demographers can be termed *demogenes*. Their defining characteristic is that they have a noticeable effect at the population level. This simple criteria excludes virtually all genes that have been identified to date. However, this could change in coming years as a result of the rapid developments in genetic epidemiology, especially the HGP and progress in methods for identifying QTLs. Given the rapid pace of developments, it is useful to have criteria for identifying the genes that are most apt to be useful for demographic research.

The preceding discussion suggests several criteria for demogenes. The first is that they must be associated with one or more common conditions or behaviors. This screens out most known genotypes since they affect characteristics that are rare. Most of the diseases for which genetic causes or risk factors have been identified are not major causes of death or disability. The second criteria for demogenes is that many individuals carry alleles that are associated with substantial variations in risk. In other words, it must be a major gene or oligogenes with common polymorphisms. As a rule, we should look for risky (or protective) genotypes that have frequencies of at least 5%.

These two criteria for demogenes determine the *attributable fraction*, the proportion of cases of a disease that are associated with a given risk factor. For example, about 20-25% of individuals with Alzheimer's disease have at least one copy of the APOE ϵ_4 allele.²¹ Demogenes must be associated with a large attributable fraction of deaths, cases of disease, or variation in other variables of interest to demographers.

Each of these criteria eliminates a large number of genes identified to date, but the combination of the two criteria eliminates almost all known genes. Alleles associated with large effects on major causes of death are generally very rare due to natural selection. Most of the major diseases are complex traits whose heritability results from the effects of numerous, relatively rare mutations (the polygenes in segregation analysis). Therefore, only a very small

²⁰ This presents problems for meta-analytic studies since few studies have adequate sample sizes to detect moderate sized differences by sex (or many other variables). Consistent differences between the sexes observed in several studies could be lost when the studies don't even report the direction of differences by sex if the differences are not significant.

²¹ Attributable risk is often confused with the concept of causation. This is especially true of genetic risk factors. Since many cases of disease are associated with multiple risk factors, the simple association of a risk factor with a case does not mean that that risk factor caused the case. For example, it is not appropriate to state that 20-25% of cases of Alzheimer's disease are caused by APOE ϵ_4 .

proportion of genes are likely to meet these first two criteria for demogenes.

Research on differences between populations leads to a third criteria for demogenes: the frequency of genotypes should vary substantially across populations. Founder effects and differences in environment can lead to very large differences in the frequency of polymorphisms. On the other hand, it is not likely that large numbers of rare mutations of a single gene will cumulate in specific populations without becoming common in any population. Therefore, genes associated with common polymorphisms are much more apt to be useful for explaining variation across populations.

Fourth, interactions with social or behavioral variables of interest to demographers enhance the value of genes for demographic research. Research on gene-environment interactions is not as advanced as the search for single genes. This is an area of research that might benefit from collaboration between demographers and genetic epidemiologists.

Finally, many of the mutations recently linked to diseases are *somatic mutations*, i.e., mutations that occur in a single cell of the body and are not inherited. This is particularly true of much of the genetic research on cancer. It is not feasible to screen for somatic mutations in population surveys since they are often localized in individual organs or certain cell types. Although somatic mutations might play a role in demographic models of aging and disease in individuals (Manton and Stallard, 1979), *germinal* (i.e., inherited) mutations are apt to be more useful for demographic research based on large surveys.

An Illustrative Comparison: BRCA genes and APOE

A comparison of three genes illustrates these criteria. The Apolipoprotein-E (APOE) gene is associated with the risk of both ischemic heart disease (IHD) and Alzheimer's disease. BRCA-1 and BRCA-2 both are associated with the risk of breast and ovarian cancers. All three of these genes may prove to be very important to biomedical research and may play a role in future demographic work. However, APOE is much better suited to incorporation into demographic research.

The six common polymorphisms of the APOE gene (,2/2, ,2/3, ,2/4, ,3/3, ,3/4, or ,4/4) are described above. The ,3 allele is the most common form in all populations, but ,2 and ,4 are common polymorphisms. More than 200 mutations of BRCA-1 and more than 100 mutations of BRCA-2 have been associated with cancer susceptibility (Rahman and Stratton, 1998). None of these BRCA-1 and BRCA-2 mutations is sufficiently common to be termed a polymorphism. We can define a BRCA genotype in terms of the presence of one or two mutations of either BRCA-1 or BRCA-2 that are associated with increased risk.

We can apply the four criteria for demogenes to compare the potential usefulness of these genes to demographic research.

APOE is associated with more common causes of death and disability than the BRCA genes.

Table 1 shows the associations between polymorphisms of APOE and the risk of IHD mortality and the incidence of AD. IHD is the leading cause of death and AD is the third or fourth leading cause. (Ewbank, 1999) The range of risks for AD is especially large with the ,4/4 genotype having an odds ratio of more than 20 relative to the low risk ,2/3 genotype (i.e., 7.87 ± 0.33 or 23.8). When the risks of IHD and AD are combined, white males in the U.S. with the ,3/4 genotype have a risk of death at age 80 that is about 20-25% higher than those with the ,3/3 genotype. The ,4/4 have a risk of death almost twice that of the ,3/3 (Ewbank, unpublished results).

Table 1: Odds Ratios by APOE Genotype for the Risk of Death to Ischemic Heart Disease and The Incidence of Alzheimer's Disease

	Odds Ratio by APOE Genotype			
	,3/3	,3/4	,4/4	,2/3
IHD mortality (at age 65)	1.0	1.71	2.82	0.97
AD incidence (at age 85)	1.0	3.34	7.87	0.33

Although cancer is the second leading cause of death, the BRCA genes are only associated with the risks of breast and ovarian cancers which account for 21.5% of cancer deaths in women (Hoyert, Kochenek and Murphy,1999). For the total population, breast and ovarian cancers are only responsible for about 9% as many deaths as IHD and AD combined. For women, this rises to about 17%. It is more difficult to compare the morbidity burden of AD with that of breast and ovarian cancers. However, all three are associated with substantial disability and the relative disease burdens probably are not very different then the relative mortality burden.

It would be easier to control for BRCA genotype in demographic analyses if all of the known mutations were associated with similar amounts of excess risk. We could then simply include a binary variable indicating women with one or more BRCA mutations. However, the mutations are so rare that it will never be possible to get accurate estimates of relative risks for each allele. Almost all of the known mutations of BRCA1 are protein-truncating (i.e., mutations that prevent the gene from producing copies of the complete protein molecule) and, therefore, they are associated with substantial increases in risk (Rahman and Stratton, 1998).

Polymorphisms and Rare Alleles. Table 2 shows the APOE gene frequencies in several populations. In these populations at least 11% of the population carries one or more copies of the ,4 allele and at least 5% carries one or more copies of the ,2 allele.

Table 2: APOE Gene Frequencies, Five Populations.

	APOE Genotype			
	,3/3 or 2/4	,3/4	,4/4	,2/3 or 2/2
U.S. Whites	65%	21%	2%	13%
Italy	71%	17%	1%	11%
Finland	59%	31%	4%	6%
Chinese	72%	11%	0.4%	16%
Nigeria	47%	40%	9%	5%

The frequency of the ,4 allele is correlated with IHD in high income countries (Luc et al., 1994). For example, the rate among men aged 65-69 is more than twice as high in Finland as in Italy

(UN Demographic Yearbook, 1993). The differences in APOE gene frequencies explain some of the variation in IHD mortality at ages 65-69 and in total mortality at ages 80-84 (Ewbank, 1999; Stengård, Weiss, and Sing, 1998). These large differences in APOE gene frequencies also may be associated with differences in the prevalence of AD within Europe. However, variations in the prevalence of AD are not well documented.

The frequencies of mutations of BRCA-1 and BRCA-2 have not been determined in many populations. It is more difficult to determine BRCA frequencies since hundreds of mutations of BRCA-1 and BRCA-2 have been identified. An indirect estimate for Britain suggests that only about 0.6% to 0.12% carry a BRCA-1 or BRCA-2 mutation (Rahman and Stratton, 1998). The previous discussion suggests that the variations across populations are probably not as large as the variation in the frequency of the APOE ϵ_4 allele. The proportions carrying BRCA-1 or BRCA-2 mutations are very high in Icelandics and Ashkenazim, about 0.5% and 2.5%. These high rates are probably due to founder effects and small population sizes (Rahman and Stratton). Mutations of BRCA-1 or BRCA-2 will not explain much of the variation in breast and ovarian cancer mortality rates across populations.

The APOE ϵ_4 allele is associated with larger attributable risks than the BRCA genes.

The high frequency of the APOE ϵ_4 allele combined with high relative risks for IHD and AD leads to a high attributable risk of death. The ϵ_4 allele is associated with about 20-25% of cases of AD among U.S. whites (Evans et al., 1997) and I estimate that it is associated with about 10-11% of IHD deaths among white males in the U.S. The association of APOE ϵ_4 with IHD and with mortality are probably smaller among women (Wilson et al., 1994; Vogt et al., 1997). However, AD is slightly more common among women than men. The combined effects of ϵ_4 on IHD and AD lead to the large variations in mortality documented in several studies (e.g., Tilvis, Strandberg, and Juva, 1998; Corder et al. 1996).

In contrast, BRCA-1 and BRCA-2 are only associated with about 6 to 7% of cases of breast cancer and probably a smaller proportion of ovarian cancers (Rahman and Stratton, 1998). This proportion changes with age. About 30-35% of breast cancer cases at 20-29 are associated with BRCA mutations as opposed to about 2% over age 70 (Rahman and Stratton, 1998). In addition breast and ovarian cancers are only responsible for 5% of deaths in women in the U.S. (Hoyert et al. 1999). Therefore, the fraction of total deaths in women that is attributable to BRCA-1 and BRCA-2 is less than 1%.

The risk associated with APOE ϵ_4 probably differs across environments.

There is some evidence of gene-environment interactions which increase the importance of APOE for demographic research. In particular, APOE appears to interact with dietary consumption of fats to determine the levels of serum lipids (Lopez-Miranda et al., 1994; Lehtimäki, et al, 1995; however, see Lefevre et al. 1997). Therefore, the ϵ_4 allele may carry less risk of IHD in populations that consume low levels of fats. This is consistent with the finding that IHD does not appear to be as significant a cause of death in Nigeria and other areas of Africa where APOE ϵ_4 is very common. There may be a similar interaction at work with AD since it appears that the prevalence of AD is low in Nigeria but high in African-Americans (Hendrie et al., 1995). Similarly, a study of AD in Japanese-Americans living in Hawaii found a higher prevalence of AD than prevalence studies in Japan (White et al. 1996). All of these findings are consistent with an important interaction between environmental (probably including dietary lipids) and the effect of the ϵ_4 allele. This increases the potential value of the ϵ_4 allele to demographic research. There may also be interactions between environment and the BRCA genes. However, there is as yet little research on gene-environment interactions involving

BRCA genes.

How many demogenes are there and how soon will we find them?

This is the million dollar question for demographers interested in genetics. If there are apt to be a dozen demogenes discovered in the next ten years, then we have to begin planning for the collection of genetic information in demographic surveys. If APOE is apt to be the only true demogene for the next ten years, then genetic information collected in demographic surveys will not have much impact on demographic research in the next ten to twenty years.

It is useful to consider the evidence for likely demogenes from genetic epidemiology. The previous discussion of behavioral genetics provided a summary of the current state of knowledge about genes affecting behavior. The following paragraphs review the evidence for four other traits of interest to demographers. Body mass index (BMI) and blood pressure are biomarkers that would be relatively easy to add to demographic surveys. They are important risk factors for death and disability and have been the focus of social, economic and behavioral research. Osteoporosis is a major chronic disease which plays an important role in disability at the older ages and coronary artery disease is a major cause of death. There is substantial evidence that all four of these traits have high rates of heritability.

Body Mass Index There is substantial evidence from segregation analyses for major gene effects on BMI in numerous populations (Ginsburg et al, 1998; Colilla et al., 2000). Although the estimates vary across studies, it is common to find estimates that there is a single gene with alleles that have a frequency of 20-30% that explain 15-50% of the variation in BMI. A few notes of caution are important. First, even if BMI is largely controlled by a single major gene in each population, it is theoretically possible that it is not the same gene or the same alleles in each population. Second, segregation studies use significance tests to select among models with different assumptions about the number of major genes. It is often difficult to determine whether heritability is due only to numerous polygenes or whether there are one or two major genes. This is one reason for different conclusions from different studies. Third, BMI is especially difficult because of gene-environment interactions and the possibility of assortative (i.e., non-random) mating, although segregation analyses attempt to control for these complications.

Siffert et al. (1999) present evidence for men in two populations linking the 825T allele of the G protein $\beta 3$ subunit (GNB3) with the risk of obesity as defined by BMI.²² They also show that the frequency of the 825T allele varies substantially across populations with values of about 30% in much of Europe and 85% in Africa (Siffert et al., 1999). If their findings are replicated by other labs in additional populations, this allele might be valuable for demographic research on body weight and obesity.

Hypertension Twin studies and pedigree studies suggest that the heritability of blood pressure is probably in the range of 25% to 50% (Williams et al., 1994). However, it is still not clear whether there are one or more major genes for blood pressure. Livshits, Ginsburg, and Kobylansky (1999) reviewed previous segregation analyses and genetic studies of blood pressure before presenting their own results. They find only two conclusions that can be drawn unequivocally. First, blood pressure is closely tied to other variables (especially body mass) so

²² They present significant findings for German men and Chinese men. Results for Zimbabwean men were similar but not significant, possibly due to the small number of obese men in the sample and because of strong urban/rural differences probably due to environmental differences.

the genetic effects on blood pressure may be only indirect. Second, genetics does affect blood pressure, but it is not clear whether this includes a major gene with a direct effect.

One possibility is that the genetic influences on blood pressure are dominated by different genes in different populations. Schork (1997) provides an excellent discussion of the possible effects of gene interactions, migration, population size and subdivision, inbreeding, and stochastic factors on the genetics of blood pressure. He concludes that

“[s]ince there are so many physiological and biochemical pathways that mediate blood pressure regulation and the human species is relatively old... there are quite likely to be, on a worldwide scale, different mutations and gene combinations contributing to [hypertensive cardiovascular disease].” (1997:??)

Coronary Artery Disease. There is a clear genetic component to the risk of death to coronary artery disease (Peyster, 1997), the major cause of death in the U.S. However, this is the result of numerous genes which affect many of the common risk factors for coronary artery disease. For example, heritability estimates from a study of Mexican-Americans in San Antonio Texas (Peyster, 1997) ranged between 18% and 69% for ten major risk factors for coronary artery disease (including total cholesterol, systolic blood pressure and body mass index). None of these risk factors appears to be controlled by a single gene, although segregation analyses suggest that there are major unidentified genes for several of these traits. Most of these genes have not been identified. Some reports of alleles associated with heart disease have not yet been adequately replicated. Those genes that have been identified for measures of lipids, lipoproteins, and apolipoproteins generally explain less than 10% of the observed variation (Peyster, 1997). APOE is the only gene that has been consistently associated with coronary artery disease (Peyster, 1997) and even its association with mortality is not clear in females (e.g., Vogt, Cauley, and Kuller, 1997). In summary, variation in the risk of coronary artery disease is the result of numerous risk factors which are themselves complex traits for which few genes have been identified.

Osteoporosis Reduced bone mass and fracture are major health problems in the elderly. Genetics appears to explain about 75% to 85% of the variation in bone mass (Ralston, 1997). However, segregation analyses suggest that there are several genes with relatively small effects rather than one or two major genes with common alleles (Ralston, 1997; Zmuda, 1999). In addition, genetics may explain little of the risk for fractures in the elderly (Kannus et al., 1999) which is only partially determined by bone density.

The main candidate for a major gene associated with osteoporosis is the vitamin D receptor gene (VDR). Several studies of alleles at the BsmI site of VDR have found significant differences in bone mass, however, these findings have not been universal (Ralston, 1997; Zmuda et al., 1999) There are numerous other candidate genes. For example, Zmuda et al., (1999) discuss research on six candidate genes (including APOE), mention several more candidates, and refer to several linkage studies that have identified promising regions. There are also alternative candidate alleles for some of the genes that have been studied. The pace of research on this topic is suggested by the fact that Zmuda et al. (1999) reference 12 important association studies published in 1997 alone.

Summary of the Prospects for Demogenes. These four conditions are indicative of the state of research on the genetics of complex conditions. Segregation analyses sometimes suggest major genes that would be potential candidates for demogenes. However, locating the genes, characterizing the alleles, and replicating the results in different populations will require

extensive research. Documenting allele frequencies for different populations is relatively easy, but generally requires work by numerous researchers. Unraveling gene-environment interactions and documenting differences in results across populations will require additional years of research. However, the rate of progress on complex diseases is increasing steadily. More labs have genetic samples appropriate for research on specific conditions and are geared up for rapid assessments of candidate genes. Some of the associations that have already been found in single studies will prove to be important and the consensus of opinion on individual genes could change rapidly. In addition, the rate of discovery will continue to increase as the HGP completes the sequencing of the human genome and begins to catalogue SNPs.

Summary and Recommendations

Social and behavioral researchers have always been intrigued by genetics, but the exciting developments in the study of rare genetic diseases have found little application in demography. This could change as genetic epidemiology discovers genes associated with common conditions. However, in the short-run the flood of new genetic research has led to a more complex view of the genetics of complex traits. Genetic epidemiology has demonstrated that genetic variation in humans does not fit the simple Mendelian model of diseases associated with a small number of genes each with a few alleles (Weiss, 1996). The amount of genetic variation in human populations is much greater than many experts expected twenty years ago. That diversity resulted from historical differences in environment, population migrations of small groups of related individuals, and numerous random events (Weiss, 1993). The resulting diversity is staggering. For example, research on cystic fibrosis, a “simple genetic” disease, has uncovered more than 800 mutations of the cystic fibrosis gene that are associated with the disease (Kaprio, 2000).

This review has examined how future research in the genetic epidemiology of complex diseases might affect demographic research. This provides insight into the characteristics of the types of genes that are most apt to be useful to demographers. Demogenes are those that:

- 1) are associated with the most common diseases, causes of death or other variables of interest to demographers,
- 2) have common polymorphisms associated with substantial variation in risk,
- 3) have large variations in allele frequencies across populations, and
- 4) interact with environmental or behavioral characteristics being studied by demographers.

APOE is the only gene that has been proven to meet all of these criteria. The impact of genetic research on demography during the next ten to fifteen years will depend on how many additional demogenes are discovered. Demographic research on mortality would be significantly altered if genetic epidemiologists discovered two or three additional genes with impact on overall mortality as large as APOE. Research on functional disability would benefit greatly from four to six genes associated with large attributable fractions for the most common chronic conditions.

It is hard to predict how many demogenes there are and how soon they will be discovered. One expert recently predicted that during the next five to ten years “genetic dissection of complex traits will continue to yield specific genes, each accounting for only a relatively small fraction of cases” (Kaprio, 2000). This seems to be the general consensus. On the other hand, Peyser predicted in 1997 that “within a few years, most of the risk factors [for

coronary artery disease], both established and proposed, will be found to be associated with specific measured genes.” (1997: 86) Predicting the future is complicated by the recent acceleration of research. The Human Genome Project is completing the sequencing of the human genome and is just starting its search for SNPs and there are still disagreements over the relative advantages of association studies and linkage methods for studying complex traits.

The potential role of large-scale demographic surveys in genetic research

We will almost certainly decide to add the collection of genetic material to large-scale demographic surveys. The question is should we begin designing supplements to current surveys or is it premature to collect genetic material before genetic epidemiology has identified more demogenes. There are a few conclusions that follow from the preceding review,

1) Surveys of diverse populations are not useful for identifying the genes associated with specific conditions. Spurious correlations associated with population diversity would completely overwhelm genome-wide scans in nationally representative samples.

2) Large-scale surveys may have a role to play in replicating studies of the effect of previously identified genes. In this case the diversity in nationally representative samples could be an advantage if the data were analyzed properly. However, demographic surveys would only be useful for replicating findings for phenotypes that are carefully measured in those surveys. Studies of a few measures, like body mass index, might rely on self-reported measures. Others, like blood pressure and insulin levels, might be added to large surveys. However, errors in defining phenotypes severely complicate replication in genetic research.

3) Perhaps the biggest potential for demographic surveys in the next ten years is putting genetic research into a social or public health context. This will certainly be true of genetic effects on behavior. In many cases this might involve demonstrating that individual genes contribute little to understanding complex behaviors. However, given the likelihood of misperceptions about the generalizability of findings in behavioral genetics to everyday life, negative findings could be very important.

4) There is a long time lag between the first plans for a major survey and the availability of useful data. For example, if genetic material was collected in 2002, two more biannual rounds of data collection would provide reasonable follow-up data by 2006. Most of the data analysis would occur after 2007 and would probably have to rely on genes first identified by 2005. If data collection continues until 2020, most of the research would involve genes discovered by 2015. It will take several years to plan the collection and analysis of nationally-representative samples and determine the best approach to testing them for a wide range of candidate genes. Given the speed of progress in genetic epidemiology, it would be wise to begin work on the design of appropriate strategies for demographic surveys.

5) Currently available data from epidemiologic surveys could be used to develop models for incorporating genetic information into demographic research. For example, a number of data sets already include APOE genotype, basic social and economic indicators, and prospective data on survival. Some of these surveys probably have additional data on health such as functional health and nursing home placement.

6) It is useful to remember that even if genetic information does not become important for demography until after 2010, this is well within the time horizon of our current graduate students. Ten years from now they will be the young Associate Professors who will determine how genetics ultimately affects demography.

The preceding discussion started from the perspective of demography by asking what will genetic information add to our current research and what new demographic research might it stimulate. An alternative approach is to ask what demographic questions will be raised by the accelerated pace of discoveries in genetics. Will public discussions of genetics pose demographic questions that we are not currently able to address?

Clinicians are already feeling pressure from the public for further information. Patients whose parents had Alzheimer's disease are asking whether they should get tested for APOE or for rare mutations of the Presenilin genes (Mayeux et al., 1998). Women with familial risk of breast cancer are interested in testing for BRCA-1 and BRCA-2 mutations (Coughlin, Khoury and Steinberg, 1999). Those who carry these mutations are looking for appropriate prevention strategies. Therefore, while biomedical researchers study how individual mutations cause disease, clinical researchers are struggling to understand what these findings already mean for their patients.

As the steady stream of new genetic findings cumulates into a flood, social scientists will be under pressure to figure out what this all means for society. We will face new questions that need to be answered and common perceptions that need to be tested. For example, does the higher prevalence of APOE ϵ_4 in African-Americans explain much of the differential in mortality by race? How important can tobacco advertising be if it turns out there is a gene associated with addiction to nicotine? If there is a gene for caution or risk-taking, does it explain differences in savings or income? How much of the relationship between poverty and poor health is "simply" due to bad genes? What would be the implications of genes associated with "intelligence?"

Newly discovered genes will lead to new thinking about who we are as individuals and what we are as a society. It is almost certain that speculation will outpace evidence. Popular perceptions will stray beyond what has been demonstrated by scientists. The old aphorism about "seeing the forest for the trees" may be replaced by "seeing the person (or the ethnic group) for the genes." We will be faced with the problem of putting the flood of genetic information into a social and demographic context.

Epidemiology will respond to some of these challenges. However, the perspective of most epidemiologists is a desire to understand the causes of specific disease. Demographers, including social demographers and economic demographers, have always concentrated on the bigger picture - the social, demographic issues. To demographers, the social and economic context of health is more than mechanisms complicating disease rates. We are also interested in non-health behaviors such as retirement decisions, savings behavior, and care-giving.

Adding genetic and bioindicator data to large demographic surveys may be useful to epidemiologists. However, these data will be crucial to demographers if we are to put genes into a wider social context. It is difficult to predict where genetic research will lead us as a society in the next 20 years. For that reason it is difficult to predict what social, economic, and demographic questions will arise and what new avenues of research will open up. However, if the current promises of new genetic discoveries are even partially realized, they could change what questions demographers study and how we study them.