APPLICATIONS—VARIOUS DISEASES AND DISORDERS

Population Studies of Allele Frequencies in Single Gene Disorders: Methodological and Policy Considerations

Linda L. McCabe1 and Edward R. B. McCabe2

INTRODUCTION

The ascertainment of the frequency of single gene disorders and the various alleles responsible for each of these disorders, is an area of considerable interest. These frequencies influence the strategies for disease and carrier detection, the prioritization of screening programs for one disease over another, and the methodologies that will be employed to detect individuals with these mutant alleles. The number of disorders amenable to molecular genetic epidemiology is tremendous and growing daily. There are encyclopedic resources to provide the latest information on disease and allele frequencies (e.g., Scriver et al. (1) and Online Mendelian Inheritance in Man (OMIM) (2)).

In this review we will describe the frequency ascertainment of several classic single gene disorders, including their allele frequencies, and discuss the reasons why these frequencies may be important to investigators, clinicians, epidemiologists, and policy makers. We will then discuss methods that are being developed to reduce the bias in allele frequency detection. Finally, we will briefly review the controversy regarding the use of samples for genetic testing, describing the latest positions of professional organizations regarding this issue.

Received for publication October 8, 1996, and accepted for publication June 4, 1997.

Abbreviations: ACMG, American College of Medical Genetics; AS, carrier status; ASHG, American Society of Human Genetics; CDC, Centers for Disease Control and Prevention; FMR1, fragile X syndrome gene; FS, affected status; G6PD, glucose-6-phosphate dehydrogenase deficiency; HEXA, hexosaminidase A; mRNA, messenger RNA; NCHGR, National Center for Human Genome Research; S/β-thal, compound heterozygote for the S and β thalassemia alleles.

1Department of Psychiatry, University of California, Los Angeles, School of Medicine, Los Angeles, CA.
2Department of Pediatrics, and Program In Human Genetics, University of California, Los Angeles, School of Medicine, Los Angeles, CA.

Reprint requests to Dr. Edward R. B. McCabe, Department of Pediatrics, MDCC 22–412, UCLA School of Medicine, 10833 Le Conte Avenue, Los Angeles, CA 90095–1752.

ALLELLE FREQUENCIES IN VARIOUS POPULATIONS

Groups that are isolated either physically or culturally are frequently genetically isolated as well. Breeding within the group results in mutational alleles and allele frequencies that differ from other groups.

The Ashkenazi Jews represent a classic example of a population with an increased frequency of certain diseases. Disorders that will be discussed for the Ashkenazim include Tay-Sachs disease and Gaucher disease, as well as other rarer disorders.

Screening of Jewish communities in North America for Tay-Sachs disease heterozygotes began in the 1970s, and expanded throughout much of the world (3). More than 950,000 individuals have been screened, and 36,000 heterozygotes have been detected. The goal of these targeted screening programs has been to identify couples in which both partners are carriers, and nearly 1,100 couples at-risk have been identified, none of whom previously had a child with Tay-Sachs disease. Such screening programs provide large databases from which one can determine allele frequencies. The frequency of mutations in the gene for hexosaminidase A (HEXA), the enzyme which is responsible for Tay-Sachs disease, is 0.033 among Ashkenazi Jews in Israel and North America, but is much lower in the non-Jewish, general population among whom the frequency is estimated to be 0.006.

Three mutations in HEXA account for 95–98 percent of the Jewish heterozygotes. Two mutations are responsible for the infantile form of Tay-Sachs disease: a four base-pair insertion in exon 11 (+TATC1278) representing 75–80 percent and a mutation in intron 12 (+1 IVS-12 G→C) representing 15 percent. The third mutation, an alteration in exon 7 (Gly269Ser) is found in 3 percent of Jewish heterozygotes and is responsible for late-onset (adult or chronic) Tay-Sachs disease.

The mutation pattern for Tay-Sachs disease is very different among non-Jewish heterozygotes (3); approximately 20 percent have the exon 11 insertion and...
virtually none have the intron 12 mutation for infantile Tay-Sachs disease. In the United States, 15 percent have an abnormal splice-site mutation at +1 IVS-9 (G→A). Approximately 35 percent of non-Jewish heterozygotes have one of two pseudodeficiency mutations, Arg247Trp or Arg249Trp. About 5 percent carry the exon 7 mutation for late-onset Tay-Sachs disease.

Tay-Sachs disease is also observed in geographically isolated, non-Jewish populations, including locales in eastern Quebec, Japan, and Switzerland, and two groups in the United States, one among the Pennsylvania Dutch and another in the Louisiana Cajuns (3). In non-Jewish French Canadians, at least two novel mutations (a 7.6-kb deletion at the 5' end and a G-to-A alteration at +1 IVS-7) have been discovered. The most common mutation in Ashkenazi Jews, the exon 11 insertion, was also observed in the Quebec sample. In a study of French Canadians living in New England, the same three mutations were noted, among others (4). The main mutation in the Japanese is a nucleotide transversion at an acceptor splice site, −1 IVS-5 G→T. Among the Pennsylvania Dutch, the +1 IVS-9 G→A and a pseudodeficiency mutation, Arg247Trp, are the most common. Both of these mutations also are observed in other non-Jewish heterozygotes. In the Cajuns, the +1 IVS-9 G→A and the +TATC 1278 have been detected. The first is common among non-Jewish heterozygotes, while the second is the most common mutation among Ashkenazi carriers, and, as expected, both are observed in a related population, the Acadians living in New England.

In addition to Tay-Sachs disease, there are a number of other disorders which occur predominantly among Ashkenazi Jews, including Gaucher disease. In this population, the Gaucher mutation frequency is 0.0343 (5). There are three types of Gaucher disease: type 1 is the most common and does not involve the central nervous system; type 2 is the acute neuropathic form with an early onset, severe central nervous system involvement, and death by age 2 years; and type 3 is the subacute neuropathic form with neurologic symptoms, a later onset, and more chronic course than type 2. All genotypes show a variety of phenotypes, making genetic counseling difficult. Using the five most common mutations, 97 percent of those in Ashkenazi Jews can be detected, but only 75 percent of the mutations in the non-Jewish groups can be found. Based on the frequency of Gaucher disease among non-Jews in Holland, the gene frequency is estimated to be 0.005, and many of the affected Dutch individuals have the 1226G mutation common among the Ashkenazi Jews. This mutation has not been seen in Japan, where the overall gene frequency is probably lower than that for Holland.

While families with multiple affected members have different clinical presentations, there can be some general conclusions regarding the relations between genotype and phenotype (6). Individuals homozygous for the 1226G mutation have only type 1 disease. Those who are compound heterozygotes with one 1226G allele have more severe type 1 disease than the 1226G homozygotes. In the Swedish Norrbottian population, those individuals homozygous for the 1448C mutation usually present with type 3 disease, while those with the same mutation in Japan have type 1 disease. Thus, genotype/phenotype correlations vary with genetic background.

In addition to Tay-Sachs disease and Gaucher disease, the Ashkenazim have a higher frequency of mutations for other, rarer disorders. There are at least four complementation groups accounting for Fanconi anemia (7). For complementation group C, the most common mutation in Ashkenazim is IVS4 + 4A→T, and this allele was found in 83 percent of the Ashkenazi patients tested. The carrier frequency for this allele was 2 of 314 among Ashkenazi Jews tested and it was not detected in 130 non-Jewish individuals. Another disorder, the late infantile form, metachromatic leukodystrophy, has an incidence of 1 per 40,000 live births in most populations (8, 9). However, among the Habbanite Jews living in Israel, the frequency is 1 per 75 live births. All Habbanite Jews tested have the P377L mutation in combination with pseudodeficiency alleles, and the high frequency of this disorder among the Habbanite Jews has been attributed to a founder effect (8). Moslem Arab patients from Jerusalem are homozygous for a splice donor site mutation exon/intron 2, which has been found in other population groups as well (8). In another study (9), Arab patients from Jerusalem had the same mutation found in the previous study (8), while those from Galilee had five different mutations. Three of the mutations found in Galilee patients were in Moslem Arabs and two were in Christian Arabs. Given the variety of mutations in such a small geographic area, it is difficult to attribute these mutations to a founder effect (8).

Other disorders of interest among Ashkenazim are myopathic glycogen storage disease, idopathic torsion dystonia, and factor XI deficiency. Yet another disorder which occurs with a higher frequency among the Ashkenazim is the myopathic glycogen storage disease, type VII, phosphofructokinase deficiency, or Tarai disease. Three mutations accounted for this disorder in 18 Ashkenazi Jewish families (10): 11 of 18 had a g→a mutation at the splice donor site of intron 5; 6 of 18 had a frameshift mutation in exon 22; and a nonconservative amino acid substitution in exon 4.
occurred in 1 of 18. These are different mutations from the point mutation in intron 15 in a Japanese family. Careful study of Ashkenazi Jewish families with idiopathic torsion dystonia (11) showed that more than 90 percent of the early-onset cases were due to a single founder mutation presumed to appear approximately 350 years ago in Lithuania and Byelorussia. A mutation for factor XI deficiency in Ashkenazi Jews has also been observed in four unrelated Iraqi-Jewish families, suggesting that this particular mutation was present in Jews 2,500 years ago (12).

While the overall frequency of cystic fibrosis is lower among the Ashkenazim than among others of northern European descent, molecular genetic testing can be much more focused in this population because of the more limited number of mutations among the Jews. The W1282X mutation is present in 60 percent of the Ashkenazim with the disorder (13) and the ΔF508 mutation accounts for an additional 23 percent (14), with six mutations encompassing 95 percent of the mutant Ashkenazim alleles. While there is an ongoing discussion regarding the utility of carrier screening and newborn screening for cystic fibrosis, the ability to detect 95 percent of mutant alleles among Ashkenazi Jews suggests that DNA-based screening for cystic fibrosis is definitely possible within this group. At least 350 mutations have been identified in total for cystic fibrosis, but less than 10 have a frequency greater than 0.5 percent. The ΔF508 mutation accounts for 70 percent of cystic fibrosis worldwide, with an increased incidence in northern Europe and a decreased incidence in southern Europe. The worldwide frequency of cystic fibrosis among all Caucasians is estimated to be one in 2,500. The relation between genotype and phenotype is substantial for pancreatic function, with two severe mutations accounting for 85 percent of the cases of pancreatic insufficiency. Mutation analysis is highly reliable for prenatal diagnosis and carrier detection. The states of Colorado and Wisconsin, and countries such as Australia, Belgium, Canada, France, Italy, and the United Kingdom, include cystic fibrosis as part of their newborn screening program (15). In all of these programs, DNA diagnosis is used as a follow-up for suspect initial immunoreactive trypsinogen results. If one only screened for the ΔF508 mutation (with an observed frequency of 75.8 percent in one US study (16), one would expect that true positive individuals following the initial positive immunoreactive trypsinogen results would be 57 percent homozygous for ΔF508, 37 percent would be compound heterozygotes with one ΔF508, and only 6 percent would not have a ΔF508 allele.

Unlike cystic fibrosis, there is no single common mutation for phenylketonuria (17). Over 100 mutations have been associated with phenylketonuria, with some mutations being associated at higher frequency in certain populations. Because there are so many different mutations, most patients with phenylketonuria are compound heterozygotes. Further complicating our understanding of the molecular genetics of phenylketonuria are the eight restriction fragment length polymorphisms at the locus for phenylalanine hydroxylase, the enzyme with reduced activity in phenylketonuria. More than 70 different haplotypes have been defined by combination of these eight restriction fragment length polymorphisms, and these have been observed at differing frequencies in different populations. Therefore, there are wide geographic and ethnic variations in the incidence of phenylketonuria, ranging from five cases (Ashkenazi Jews) to 385 cases (Turks) per million births (17). The incidence of phenylketonuria in the United States ranges from one in 10,000 to one in 25,000 (18), and among Chinese the incidence is one in 16,500 (17). Several examples of founder effects and genetic drift have been seen in specific populations. A single deletion mutation is seen in all phenylketonuria cases among Yemenite Jews currently living in Israel. The origin of this mutation has been traced to a single member of the Jewish community in the capital of Yemen during the early 18th century. A recent study of the molecular basis of phenylketonuria in Spain (19) showed a predominance of haplotypes 1, 6, and 9, which are also frequently found in other Mediterranean populations. In Spain, 41 percent of the mutations were represented by IVS10, I65T, E280K, and P281L. Not found were R408W and IVS12 (prevalent in northern Europe), and R252W, R261Q, and L249F (present in other southern European populations). Molecular analysis of phenylketonuria patients in Ireland (20) showed R408W in 42 percent of the patients, and this mutation was associated with haplotype 1. In eastern Europe, R408W was associated with haplotype 2. These observations suggest two independent founding events for the R408W mutation, one Balto-Slavic and another Irish/Celtic. One group (21) attributes these two mutation/haplotype combinations to recurrence, since the CpG dinucleotide involved has a mutation rate 42-fold greater than the mutation rate for non-CpG dinucleotides. The presence of the R408W mutation with haplotype 1 (Irish/Celtic) in Norway can be explained by the assimilation of the Irish into the Norwegian population when the Norwegian Vikings were in Ireland and northwest Scotland from the 8th into the 12th centuries. Thus, DNA studies can be used to track ancient human migrations.

Epidemiol Rev Vol. 19, No. 1, 1997
METHODS TO REDUCE THE BIAS IN ALLELE FREQUENCY DETECTION

The 1987 demonstration that DNA can be extracted from dried blood specimens on filter paper blotters used in newborn screening provides a means for the reduction of bias in allele frequency detection (22). However, the quantity of extracted DNA limited these initial attempts. The polymerase chain reaction allowed for amplification of the DNA extracted from filter paper blotters in quantities sufficient for the diagnosis of sickle cell disease (23). Newborn screening for sickle cell disease was recommended when it was demonstrated that penicillin prophylaxis significantly reduced the morbidity and mortality from infection (24). The goal was to have affected infants on penicillin prophylaxis by 4 months of age. Primary newborn screening involves the detection of the hemoglobin protein phenotype. Unfortunately, many children required repeat screening at 2-4 months of age when the proportion of hemoglobin F would be diminished and the adult β-globin phenotype could be more readily determined. Molecular genetic diagnosis using DNA extracted from the original newborn screening blotter would permit immediate diagnosis of patients with sickle cell disease using the original newborn screening specimen, allowing more efficient diagnosis and more efficacious management.

Comparison of molecular genetic diagnosis with the hemoglobin phenotype determined by electrophoresis in the Texas Department of Health Newborn Screening Laboratory (Austin, Texas) was carried out in a double-blind study (25). Each sample was tested using four different molecular genetic methods. Of the 75 samples in the study, there were disagreements between the DNA diagnosis and the hemoglobin phenotype in five samples. In two of these samples, there was disagreement between the original hemoglobin electrophoretic screen and all four molecular genetic methods the first time the DNA sample was tested. When molecular genetic testing was repeated, there was complete agreement among the DNA tests and between the DNA tests and the hemoglobin electrophoresis. This suggested contamination during the first DNA test. For a third sample, three of the four DNA tests agreed with the hemoglobin phenotype the first time, and the second time all four DNA tests agreed. Contamination for the one DNA test was suspected on the initial test. Because of these problems with contamination, procedures were instituted that reduced such problems dramatically. The fourth sample showed very poor amplification the first time, and no amplification the second time, and it was considered to be an inadequate sample. The fifth sample consistently showed differences between the DNA tests, which showed AS (i.e., carrier status), and the hemoglobin phenotype, which showed FS (i.e., affected status). Eventually the clinical diagnosis of this infant resolved the discrepancy, when it was determined that the infant was a compound heterozygote for the S and β-thalassemia alleles (S/β-thal). In order to deal with such samples, a method of extraction of messenger RNA (mRNA) from newborn screening blotters was developed which was quite successful in identifying individuals with S/β-thalassemia (26).

Application of these methods as follow-up to a positive newborn screening result for sickle cell disease at the Texas Department of Health Newborn Screening Laboratory reduced the age at diagnosis from an average of 4 months to an average of 2 months (27), thus removing laboratory barriers that previously prevented initiation of penicillin prophylaxis before the recommended 4 months of age. Of the more than 500 samples studied, 13 percent were probable S/β-thalassemia by DNA and RNA testing. In addition to the manual methods, an automated method was demonstrated using microtiter plate technology, which would significantly reduce labor intensity and costs and increase sample throughput. Some unexpected results were informative during this study. DNA testing will only identify the alleles targeted by the specific detection methods: two infants with FS screening results had AA genotypes, but had screening phenotypes that were actually F, other, having one or two abnormal hemoglobins migrating at the same position as the S. The E allele was demonstrated using molecular techniques in seven infants when it had not been observed in the original hemoglobin screen, and one time it was seen in the original screen but not detected by DNA methods. In one infant, the C allele was shown with molecular genetic methods, but not with hemoglobin electrophoresis. These studies demonstrate that the advantages of dried blood specimens for newborn screening are also advantages for the determination of allele frequency, including ease of sample collection, ease of transport to the laboratory, and stability of the analyte.

Studies of β-thalassemia in India (28) showed differences in the most common mutations in Punjabi and Maharashtrian, and between these groups and Indian immigrants in the United States and the United Kingdom. In the Punjabis, the IVS-1 nt 1 (G-T) mutation accounted for nearly 25 percent of the β-thalassemia mutations, but represented only 5 percent or less of the mutations in each of the other groups. The IVS-1 nt 5 allele accounted for 60 percent of the β-thalassemia mutations in the Maharashtrans, 35 percent among the immigrants to the United States, and 23 percent among the Punjabis.
Regional differences in mutations for β-thalassemia also have been demonstrated in Thailand (29). In the northeast, the codon 71/72 frameshift accounts for 13 percent of the mutations, but is absent in the central portion of the country and represents only 1 percent of the mutations in the south of Thailand. In the central region of Thailand, the codon 41/42 frameshift, the IVS-2 nt 654 splicing, and the codon 17 amber mutations together account for 82 percent of the alterations. The IVS-2 nt 654 allele is also common in China. In southern Thailand, the codon 19 splicing and the IVS-1 nt 5 splicing mutations are prevalent. The latter is also common in Indian Maharastrians. Eight alleles represent 98 percent of the β-thalassemia mutations in Thailand.

The frequency of β-thalassemia mutations also shows regional differences in Italy. The IVSII-745 mutation is found in four of 11 patients of Latium origin in Rome and was localized to the southeastern province of Frosinone (30). The frequency of this mutation was less than 10 percent in patients from Sicily (31). Use of molecular screening has made prenatal diagnosis widely available and possible for almost all cases. In Greece, prenatal diagnosis targeting the four most common mutations identifies 80 percent of the affected fetuses, and using eight alleles in the testing panel increases the success rate to 95 percent (32).

In addition to the hemoglobinopathies, molecular genetic diagnosis from newborn screening blood blotters has also been effective for Duchenne muscular dystrophy (33). Multiplex polymerase chain reaction targeting regions of the Duchenne muscular dystrophy genome with the highest frequency of deletions can detect between 80 and 90 percent of the deletions in Duchenne muscular dystrophy patients using nine sets of polymerase chain reaction primers. This method would ascertain 50 percent of all patients with Duchenne muscular dystrophy using newborn screening blotters. DNA analysis is being used as a second tier test, following an initial screening that identifies an elevated circulating level of creatine kinase (34).

The advantage of DNA extraction from newborn screening blotters in the ascertainment of mutant allele frequencies in a particular study population is that one does not need to evaluate a large number of homozygous individuals. A much smaller sample can be used since identification of the frequency of a specific allele in the heterozygotes gives the frequency of homozygotes, compound heterozygotes, and true heterozygotes. These methods can also be adapted to use chromogenic probes rather than radioactive probes (35). Using discarded newborn screening specimens from the Texas Department of Health (Austin, Texas), our prospective study determined for the first time the unexpectedly high frequency of S/β-thalassemia among African-Americans (36). Previous studies of allele frequency relied on clinical ascertainment in adults and older children, and greatly underestimated the frequency of S/β-thalassemia. Individuals with AS genotypes need to be tested for S/β-thalassemia in order to initiate appropriate management in a timely fashion.

Another red cell disorder, glucose-6-phosphate dehydrogenase deficiency (G6PD) is common in the Mediterranean region, on the Indian subcontinent, and in southeast Asia. More than 50 different mutations have been described. A study of males living in northwest Pakistan showed a frequency of this disease of 10 percent (37). Among those with glucose-6-phosphate dehydrogenase deficiency in northwest Pakistan, 76 percent have a C→T substitution at nt 563. Unlike the European and Middle Eastern populations, this is not in linkage disequilibrium with the C→T substitution at nt 1311. Glucose-6-phosphate dehydrogenase deficiency also occurs among the Chinese, and studies of allele frequency can be useful in determining migration and intermarriage patterns of the aboriginal tribes to Taiwan (38). The frequency of this disorder varies among the tribes: Saisiat, 9 percent; Ami, 6 percent; and Yami, 0 percent. The mutations are also different for the Saisiat and Ami. The most prevalent mutant allele among the Saisiat is the nt 493 mutation (G6PD Taipei) at 67 percent, while nt 592 mutation predominated in the Ami (67 percent). It is unclear whether the high rate of the G6PD Taipei mutation among the Saisiat indicates original migration from mainland China, intermarriage with mainland Chinese or Taiwanese, or both. The predominant mutation in the Ami has not yet been found in the Chinese population in Taiwan, but has been identified in two unrelated individuals from southern Europe and one person from mainland China. The absence of glucose-6-phosphate dehydrogenase deficiency in the Yami may be due to their more extreme isolation on Orchid Island.

Newborn screening blood specimens were used in Manitoba to determine the frequencies of permutations of the FMR1 gene responsible for fragile X syndrome, the most common form of familial mental retardation (39). The CGG allele ranged from nine to 106 repeats in their population. The most frequent allele was a repeat of 28, and 97 percent of the alleles had fewer than 40 repeats. About 2.3 percent of alleles had between 40 to 49 repeats, and 0.37 percent had repeats ranging from 50 to 59. Alleles with greater than 60 repeats occurred in 0.37 percent of this population. Alleles between 40 and 60 repeats can be either stable
or unstable, depending upon the number of AGG trinucleotides among the CGG repeats.

In a study of families with fragile X syndrome in Finland, one DXS548 allele was present in 92 percent of the affected individuals and in 17 percent of the normals (40). One specific haplotype was found in 73 percent of the patients and in only one of 34 normals examined. This particular haplotype was found throughout Finland, which suggests that it was present in the first settlers in that country, and the investigators concluded that it may have been silently carried through 100 generations.

Another disorder involving expansion of trinucleotide (CAG) repeats is Huntington disease. This disorder occurs with variable prevalence rates in different populations. Among western Europeans, the prevalence rate ranges between five and 10 per 100,000. Individuals in China, Finland, and Japan, and Africans indigenous to a number of countries, have a rate that is one-tenth that of the western Europeans. Predisposition for Huntington disease depends not only on the number of CAG repeats, but also on the number of adjacent CCG repeats and a GAG polymorphism at residue 2642 (41). Unaffected western European individuals typically have seven CCG repeats and higher numbers of CAG repeats. African blacks, Chinese, and Japanese unaffecteds have 10 CCG repeats, and fewer CAG repeats. The GAG polymorphism is absent in African blacks, Chinese, and Japanese controls, but present in 10 percent of western European normal individuals.

Another disorder involving CAG repeats is Machado-Joseph disease. Between 68 and 84 repeats are observed in affected individuals, with the normal range represented by 14–37 repeats (42). The expansion of the repeats associated with Machado-Joseph disease shows intergenerational instability, particularly during male meioses. The size of the expanded allele is inversely related to the age at onset of Machado-Joseph disease and directly related to the frequency of additional clinical symptoms. Caucasian and Japanese patients affected with Machado-Joseph disease share haplotypes at several markers surrounding the gene for Machado-Joseph disease which are uncommon in unaffected Caucasian and Japanese individuals. These observations suggest either common founders in these populations or certain features of these chromosomes that make them susceptible to pathologic expansion of the CAG repeat in the gene for Machado-Joseph disease.

The use of dried blood specimens facilitates collection and transport of samples, permitting the investigation of allele frequencies among groups that may be culturally or geographically isolated. In addition, the possibility of examining a sample that is broadly representative of a population, such as all newborns, allows evaluation of allele frequencies that are unbiased by targeted collection strategies and do not require initial detection by clinical phenotype.

**POLICY DEVELOPMENT REGARDING THE USE OF SAMPLES FOR GENETIC TESTING**

Allele frequency studies, like those cited above, are improving our understanding of genetic disease. As the knowledge base of genetics expands, and the technology of molecular biology improves to further accelerate the pace of growth, genetics professionals are considering new questions regarding the study of samples and the implications of the results of genetic research for the individuals providing these samples. Since genetic information is available in a wide array of biologic materials collected from an individual, a number of organizations are developing policies on the storage and use of genetic material and on the procedures for informed consent when obtaining such material.

One of the first attempts to deal with these issues came as the result of a working group convened by the Centers for Disease Control and Prevention (CDC) and the National Center for Human Genome Research (NCHGR) (43). This group recommended that if samples that have been collected in the past and still retain identifiers, and if a new project has the same intention as the original project or if it meets current guidelines exempting the project from obtaining informed consent, then the appropriate Institutional Review Board should review the new project and determine whether new informed consent would be required for the use of these samples. For research that would remove identifiers from such samples, this group suggests that the Institutional Review Board should consider the previous consent process, the scientific basis and importance of the study, the difficulty of recontacting participants, the finite nature of the samples, and the availability of effective medical treatment for the disorder under study. If previously collected samples are anonymous, the group agreed that it would be unnecessary to obtain consent, but that the Institutional Review Board could still review the protocol’s scientific basis and significance, as well as the possibility of conducting a similar study that would require informed consent. The group’s recommendation for the collection of samples in the future for research use would be to ask the potential subjects whether they wanted their sample to be identified or to remain anonymous. For identified samples, the participants should be informed about the risks and benefits of participation, issues of confidentiality, the circum-

---

*Epidemiol Rev* Vol. 19, No. 1, 1997
stances under which they would be contacted again by the researcher, and their ability to withdraw from the study. For anonymous samples, the subjects should be instructed that they would be unable to receive specific information regarding the results from their specimens. Additional information that all participants should receive includes potential commercial profits from the sample, possible sharing of their sample with other researchers, and the possibility of limiting the study of their sample to certain disorders.

Pathologists reacted strongly against the tone of the preamble to the CDC/NCHGR recommendations, which emphasized the harm to research subjects over the benefits of participation (44). The concern of the pathologists was that the policies outlined in the CDC/NCHGR document would severely limit access to archived clinical specimens for a variety of purposes, including molecular genetic research. Permanent storage of these samples is required by law, and these samples are also DNA databases providing the opportunity for studying the relations between genotype and morphology, pathogenesis, and prognosis. The pathologists interpreted the CDC/NCHGR document as forbidding the accepted practice of using the remains from a clinical laboratory test for research, if the sample might be linked to some clinical information, but would be otherwise anonymous. The pathologists objected to the suggestion of the CDC/NCHGR working group that informed consent would need to be obtained before these stored samples could be used for research. The pathologists were also concerned that the working group was suggesting that consent forms for medical procedures would be required to contain an extensive, burdensome, and confusing consent form to accommodate all of the possible research uses of the samples. These two issues would preclude any meaningful epidemiologic research due to the cost of obtaining informed consent and the impossibility of securing a representative sample with such a consent form, even if the samples were anonymous.

The American College of Medical Genetics (ACMG) was also concerned about the issues of storage and use of genetic material (45). The ACMG committee recommended that the following issues should be clarified when obtaining clinical samples: purpose, limitations, and possible outcomes of the test; methods for communicating and maintaining confidentiality of results; additional use of sample or its destruction; scope of permission to use samples or results in counseling and testing relatives; potential for future test development and application to current specimen; permission to use anonymized samples for research; duration of storage; future access by patient or designee; option to withdraw or destroy a sample at any time; and possibility of inadvertent sample loss. The ACMG considered the following issues for samples for research: purpose, limitations, and possible outcomes of the research; methods for communicating results and maintaining confidentiality of the results; potential for future test development and application to current specimen; possible commercial value of the sample; permission to remove identifiers and use the sample in other research; provision for future contact regarding additional research; duration of storage and plans for destruction; and adherence of Institutional Review Board procedures. In dealing with the use of stored genetic material which had been previously collected, the ACMG suggested consideration of recontacting the donor, anonymizing the sample, prior consent forms, and risks and benefits. They argued that contacting donors regarding new diagnostic tests would be the same as obtaining samples for clinical tests, and that contacting donors regarding new research would be the same as obtaining samples for research.

The American Society of Human Genetics (ASHG) formulated a statement on informed consent for genetic research (46). The ASHG group attempted to strike a balance between the goals of expanding knowledge through research and respecting the interests of research subjects. For research using prospectively collected samples, the ASHG encouraged that: informed consent should be obtained except for anonymous sample collection; subjects’ confidentiality should be maintained; information should be given regarding the purpose, limitations, possible outcomes, and communication of results; subjects should be informed concerning storage of samples and possible storage failure; and they should be given options concerning the scope of other investigations and the possibility of sharing their samples with other researchers. The ASHG suggested that by anonymizing existing samples, the need for consent was eliminated. However, they cautioned that researchers should consider the appropriateness of anonymization when there was the possibility of medical intervention. For research with identifiable, existing samples, researchers should recontact the participants to obtain consent for new studies unless the researcher obtains a waiver based on a decision that the research would provide no more than minimal risk, and there would be no adverse effect on subjects’ rights and welfare. An additional issue that would influence this decision would be the impracticality of obtaining consent. The ASHG group noted the need to communicate pertinent information after the study when appropriate.

Each of these documents represents an attempt to deal with a variety of complicated issues regarding the
use of biologic materials in research studies. Since the existing collections of these specimens are DNA databases, they contain sensitive information. Information obtained from the study of this material is sufficiently complex so as to require trained genetics professionals to present this information to the subjects, including identification of medical risk, carrier status, or risk to offspring affected by genetic disease. There is a general consensus regarding certain issues. Results should be presented to the subject and/or the subject’s physician. This information should not be shared with relatives, employers, insurance companies, or other parties.

There is a delicate balance to be maintained that will ensure that participants in genetic research, including those in epidemiologic studies, receive the maximal benefits from the study while minimizing any risks or harmful effects. Stored samples and prospective studies have great potential for improving our understanding of the genetic basis of human disease. As the field of genetics develops and matures, new issues will arise which need to be considered involving the protection of human subjects. Professionals from many disciplines will need to work together to solve these problems as they arise and to try to anticipate them so they may continue to pursue studies such as those concerned with the unbiased ascertainment of the frequency of mutant alleles.

REFERENCES


