A late response to the first William Allan memorial award: seven years of human genetics

As the first recipient of this award in 1962, I was preoccupied by directing a research group in Brazil and accepting a professorship at the University of Hawaii. Separated from the publications that followed my Wisconsin Ph.D. in 1955, I therefore failed to attend my own award ceremony and the second recipient was equally silent. James Neel and later award winners were more mature, responsive to the William Allan memorial, and communicative to the American Society of Human Genetics and its journal. In remorse, I describe herein major research during my Wisconsin years, greatly influenced by colleagues there and in Hawaii, Japan, and Brazil. For each contribution the impact of subsequent developments is briefly mentioned.

As a child in New Haven, Connecticut I learned from a cousin to collect butterflies, but lost enthusiasm for entomology as an undergraduate in Swarthmore College. There I encountered Dobzhansky's book "Genetics and the Origin of Species", leading to a lifelong fascination with genetics. After two years I transferred to the University of Hawaii, where Gordon Mainland expanded my interests and recommended his colleague James Crow at the University of Wisconsin as his successor after I graduated early in 1951. This association was initially expressed in Drosophila, leading to my M.S. the following year. Simultaneously Crow greatly stimulated my interest in human genetics and encouraged me to spend 1952-53 in Japan with James Neel and colleagues. The focus was the effect of exposure of parents to the atomic bombs on the first generation offspring in Hiroshima and Nagasaki. Among by-products of these data were papers on non-randomness in consanguineous marriage (1) and inheritance of human birth weight (2). Of greater stimulus were two papers on linkage of blood groups with diseases of simple dominant inheritance (3,4). The samples were small and nonsignificant, but they stimulated interest that led two years later to my Ph.D. in Wisconsin and first major research contribution.

1. Linkage

Analysis of human linkage began with Bernstein (1931) and methodology developed rapidly from a few enthusiasts and their mathematical stimuli. I was influenced equally by Haldane and Smith (1947) who applied LODs to linkage and Wald (1947) who did not consider linkage but developed sequential analysis whereby two hypotheses were discriminated by the smallest number of observations. His book was soon a subject of great interest for statisticians in Wisconsin and my fascination with linkage. Among the problems to be addressed was the frequency of random autosomal linkage. From rough estimates in three lower organisms, the frequency in humans was taken to be approximately .05, sufficient to require stronger linkage evidence than had previously been demanded. Sequential analysis was shown to require less than 1/3 as many observations as the immediate predecessors of sequential tests. Whether or not sequential analysis was used, published tables of LOD scores (5) made detection and estimation of linkage easy and exact for single loci with complete penetrance and reliable frequency.

The first application was to linkage between elliptocytoses and the Rh blood type (6). Early studies of 14 pedigrees gave contradictory evidence. Linkage was highly significant ($\chi^2_1 = 34.2$), but so was heterogeneity among pedigrees ($\chi^2_{13} = 36.4$). Omitting 7 small pedigrees with fewer than six double backcross progeny, the remaining 7 pedigrees fell unequivocally into two groups. The first group of 4 showed close linkage to the Rh locus, with estimated recombination frequency .033 and 95% confidence interval .012-.078. The other 3 samples showed no suggestion of linkage. Heterogeneity within each class was far from significant. At the time such examples were familiar only to Drosophila enthusiasts. The elliptocytosis locus closely linked to Rh was identified 25 years later as protein band 4.1. Subsequently several dominant alleles at different loci unlinked to each other were identified, including α spectrin (SPTA1) and β spectrin (SPTB), initially coded EL1, EL2, EL3. A similar code was later applied to other diseases of high penetrance at more than one locus. The present century of research on innumerable genes

of low penetrance made codes like EL1 much less practical than more informative titles derived from human molecular genetics.

Long before that development the few people pursuing human linkage were reanalysing such data as was then available. Separation of the sexes of parents had been developed for linkage in man (Smith, 1954). Using that method for cystic fibrosis and the MNS locus, Steinberg et al. (1956) had estimated high significance (P<.0006) for one subsample, but this was not supported by the rest of the data. This led them to conclude that Smith's scores "may lead to spuriously significant results when applied to samples of practical size". Using the LOD tests Steinberg and Morton (7) found no significant evidence for that subsample or overall.

Finally (8) I reviewed evidence of human linkage published before 1957. Traits were limited to complete dominance or recessivity. Markers were limited to colour blindness on the X chromosome and a few blood groups on the autosomes. Earlier evidence by Haldane of linkage on the X chromosome between colour blindness, G6PD and haemophilia was supported by Haldane and Smith (1947). Beginning in 1951 Jan Mohr discovered the first autosomal linkage, later recognised as the Lutheran blood group LU and the secretor locus SE (Sanger and Race, 1958). The next two linkages were reported by the Galton Laboratory, who recognised the association between elliptocytosis and the Rh locus on chromosome 1 (Lawler, 1954). As noted above (6), compelling evidence was soon detected of unlinked loci that if unrecognised cause highly significant heterogeneity and overestimate linked recombination. Renwick and Lawler (1955) demonstrated highly significant linkage on chromosome 9 between ABO and the nail-patella locus, which showed no evidence of unlinked loci. Many early claims of linkage were not confirmed because they accepted a P-value less than .001, neglected or could not determine ascertainment, used other methods that underestimated P-values, or accepted traits with multiple loci or penetrance below unity.

For the next half century it was generally agreed that routine linkage tests would be profitable in extensive family studies of rare genes. This led from a few linkages detected by a small number of researchers and markers to

hundreds of researchers and many markers, using techniques that added power to the linkage tests, although as this complexity increased sequential tests were seldom exploited. Successes approaching 45 per cent per annum stimulated the Human Genome Project conceived in the mid-1980s and launched in 1990. In the next century this led to common diseases and multiple loci. The uncertainty that once challenged linkage studies with high penetrance has been replaced by genome-wide association studies (GWAS), mostly with small effect and therefore resistant to linkage analysis. The number of significant loci is high for several common diseases, but heritability explains only a minority of the genetic variants. Many alternatives have been proposed and are beginning to be explored. The number of researchers is huge and progress in DNA, RNA, potentially environmental risk factors, and other complications is impressive. It took half a century for linkage of genes with high penetrance to be explored. The much more difficult problem of genes with low penetrance may take longer.

2. Mutation

During the last century the young science of population genetics was confronted with efforts to assess hazards of ionizing radiation and other mutagenic agents. Deleterious mutations were known from Drosophila and the mouse to have a large recessive component, but human evidence was not available until Muller (1948) examined mortality data of one set of unrelated and consanguineous marriages and concluded that “every person on the average contains heterozygously at least one lethal gene or groups of genes which [homozygously would] … kill an individual ... between birth and maturity”. Subsequent analysis showed that this is an underestimate based on gametes rather than persons. Our best estimate including two other studies was .03-.05 mutations per gamete per generation (9). Since early embryonic deaths and detrimental effects after maturity were not detected, we assumed that the total mutation rate for genes contributing to consanguinity with high penetrance in a homozygote (~1) but low penetrance in a heterozygote (~.02) is 2-3 times as high as our estimate of .06-.15 mutations per gamete per generation. With 10^4 loci per gamete, this corresponds to 6-15x10^{-6} mutations per locus per generation. At mutational equilibrium and

heterozygous penetrance of .02 per locus, the estimated number of heterozygous genes that would be lethal in homozygotes is 6-15 per individual but only .0006-.0015 per locus.

Since the evidence came from consanguineous marriage unfamiliar to most human geneticists, it was disputed in several ways. The epidemiological argument is that the number of lethal equivalents depends on past inbreeding, definition and levels of mortality and morbidity, accuracy of diagnosis, and possible ascertainment bias or confounding with environmental variables that differ among populations or over time. If S is the fraction of survivors, then we expect $-\log_e S = A+BF$, where $0 \leq F \leq 1$ is a set of inbreeding values, B is a corresponding effect of inbreeding, and A measures the amount of expressed damage when F = 0. An estimate of B/A with an unusually high value would be suspect if the residual variance were significant, but that was not found. Covariance of S with environment cannot be estimated reliably and has not been provided in these data. The genetic argument that $B/A >> 1$ is evidence of overdominance, is more easily addressed. Among the many thousand human polymorphisms, the few known examples of heterozygote advantage are functionally diallelic and give $B/A < 1$, since A depends on environmental effects but B depends only on inbreeding. The estimate of B/A as 11 or more indicates that overdominant loci are making no substantial contribution to B.

In support of these results substantial underestimates of multiple loci were estimated for three diseases (10), with correspondingly higher morbidity of B/A (11). However, mortality risks for recent Japanese and U.S. samples were less. The mean value of A in earlier studies was estimated as about 3 times as much as recent rates. After more than a century susceptibility and deaths from infectious diseases are becoming much rarer, and inbreeding has decreased. It will be many generations before A increases near to equilibrium with disease and inbreeding. Until then analysis of past consanguineous marriages remains the best estimate of mutational damage.

However, a persistent problem remains with multiple genes of low penetrance and therefore a z value that may approach 1. Consanguineous marriage has

become rare in many countries, unfriendly in some others, and of little relevance to low penetrance in any case. Contribution to the mutational load remains unknown, but there is a general belief that current risks are not the most dangerous hazard to humans in a world of global warming, population increase, and wars increasingly supported by lethal weapons. Now we are cautious about mutation and enjoying decreased disease and inbreeding (though they are a loss for detecting rare recessives). Currently there is no indication that human mutation will be pursued beyond its present evidence.

3. Segregation analysis of major genes

For about 25 years the U.S. had greater access than other countries to mathematical statistics by computers, which encouraged human geneticists to apply Mendelian genetics to linkage, mutation, and finally segregation analysis (12). Although complexity gradually developed, common genes with low penetrance still defy complete analysis, but otherwise analysis is provided largely by maximum likelihood scores for 4 parameters (p is deleterious gene frequency, π is the probability that an affected person is a proband, x is the probability of sporadic cases in the population, and h is the probability that a parent of a dominant phenotype be homozygous.) Other parameters are q=1-p, t≥1 is the number of times a proband may be independently ascertained, a is the number of probands in a sibship with r affected, and $1 \leq a \leq r$. Further problems are given for p if onset is delayed and incomplete, or for x if there is a deleterious sex-linked recessive trait. Equations derived under any of these assumptions give a more powerful analysis than was previously available.

More information about genetic tests under incomplete ascertainment was soon provided (13). It was first applied to congenital deaf mutism in Northern Ireland (14). Autosomal recessives from normal parents were estimated to have complete penetrance and at least 36 ± 12 unidentified genes, supported in 2009 by at least 46 identified ones. The major types of muscular dystrophy from Wisconsin and other populations were identified by the evidence available (15) and formal genetics applied as accurately as then possible (16). The first group (facioscapulohumeral) is due to an autosomal dominant gene

or genes with complete penetrance in individuals of both sexes who survived to the age of onset. The fertility is nearly normal and the incidence is 4 persons who will develop the trait per million births. The second group (limb-girdle) is composed of autosomal recessive cases with penetrance in both sexes. The fertility is 25% of normal, with an estimated incidence of 38 persons per million births. A third group is also limb-girdle, composed of truly sporadic cases with unknown aetiology. Fertility and other characteristics are the same as for recessive cases. These sporadic cases are not due to dominant mutations, since all of 110 children were normal. The number of people who will develop the disease is estimated as 27 per million births. The fourth group (Duchenne) consists of sex-linked classes with a prevalence estimated as 66 living cases per million male births. Their fertility is less than 4% of normal.

The splitting of a group of patients with a given disease into smaller but genetically more uniform subgroups has been a topic of research in medical genetics since 1960, but progress has been slow. Until genes are identified that may or may not correspond with these subgroups, segregation analysis will be inconclusive.

4. Genetics of interracial crosses

During 1958-1959 we collected data on nearly 180,000 live births and late fetal deaths from 1948-1958 to answer two main questions about genetic effects of outcrossing. Hawaii was the best place for such a study, first because of the relative equality and short outbreeding history of many racial groups and secondly because colleagues there were helpful in providing research space with anonymised data. The two questions that prompted this investigation have been answered unequivocally (17 and later):

1. First generation hybrids between major and minor races in man are intermediate in size, mortality, and morbidity between the parental groups. A slight advantage of intercrossing is suggested but not significant.
2. At the present time, human populations do not represent coadopted genetic combinations that are disrupted by outcrossing. It remains an

untested hypothesis that outcrossing effects might be observed in a more vigorous environment.

This study was published later in much greater detail. Surprisingly, attempts have been made recently to replace race by a more obscure term like group, population, affinity, ancestry, nationality, or ethnicity. These ambiguities are often used by anthropologists to explore groups too small to encourage genetics or medicine. As Curt Stern and many others have argued, race covers major and minor groups of individuals while ethnicity does not, but the outcome is not settled. Meanwhile interest of some individuals has focused on major racial issues, assuming that the current environment does not reflect selection that in the past differentiated these races. On that assumption an advantage of crossing is possible, but no data comparable to interracial crosses in Hawaii has been examined. Could a study of greater magnitude and complexity than Hawaii provided half a century ago be performed today?

On return to Wisconsin and preliminary analysis of Hawaii data, I became interested in high-mortality populations and chose emigrants from northeastern Brazil passing through the Hospedaria de Imigrantes in São Paulo where we studied more than 1,000 large nuclear families comprising almost 7,000 tested people. Our research in 1962 extended beyond my Allan Award and preceded return with Brazilian and other colleagues to the University of Hawaii. There we confirmed the endemic goitre-PTC association, strongly negative results for the alleged smallpox-ABO association, no convincing evidence of selection operating on the markers we studied, but interesting results from inbreeding, racial admixture, and population structure. The wide variety of later research in Hawaii and elsewhere lies after 1962, beginning my last century of human genetics.

REFERENCES

Some research by Morton et al. before 1963

1. Morton NE (1955). Non-randomness in consanguineous marriage. *Ann Hum Genet* 20:116-124.
2. Morton NE (1955). The inheritance of human birth weight. *Ann Hum Genet* 20:125-134.

3. Morton NE, Moloney WC, Fujii T (1954). Linkage in man. Pelger's nuclear anomaly, taste, and blood groups. *Am J Hum Genet* 6:38-43.
4. Fujii T, Moloney WC, Morton NE (1955). Data on linkage of ovalocytosis and blood groups. *Am J Hum Genet* 7:72-75.
5. Morton NE (1955). Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277-318.
6. Morton NE (1956). The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am J Hum Genet* 8:80-96.
7. Steinberg AG, Morton NE (1956). Sequential test for linkage between cystic fibrosis of the pancreas and the MNS locus. *Am J Hum Genet* 8:177-189.
8. Morton NE (1957). Further scoring types in sequential linkage tests, with a critical review of autosomal and partial sex linkage in man. *Am J Hum Genet* 9:55-75.
9. Morton NE, Crow JF, Muller HJ (1956). An estimate of the mutational damage in man from data on consanguineous marriage. *Proc Natl Acad Sci USA* 42:855-863.
10 Morton NE (1960). The mutational load due to detrimental genes in man. *Am J Hum Genet* 12:348-364.
11. Morton NE (1961). Morbidity of children from consanguineous marriages. In *Progress in Medical Genetics,* edited by Steinberg AG, Grune and Stratton Inc, pp.261-291.
12. Morton NE (1958). Segregation analysis in human genetics. *Science* 127:79-80.
13. Morton NE (1959). Genetic tests under incomplete ascertainment. *Am J Hum Genet* 11:1-16.
14. Chung CS, Robinson OW, Morton NE (1959). A note on deaf mutism. *Ann Hum Genet* 23:357-366.
15. Chung CS, Morton NE (1959). Discrimination of genetic entities in muscular dystrophy. *Am J Hum Genet.* 11:339-359.
16. Morton NE, Chung CS (1959). Formal genetics of muscular dystrophy. *Am J Hum Genet* 11:360-379.

17. Morton NE (1962). Genetics of interracial crosses in Hawaii. *Eugenics Quarterly* 9:23-24.

References by others cited here

Bernstein F (1931). Zur Grundlegung der Vererbung bein Menschen. *Z indukt Abstummungs Vererbungsl* 57:113-138.

Haldane JBS, Smith CAB (1947). A new estimate of the linkage between the genes for haemophilia and colour-blindness in man. *Ann Eugen* 6:26-65.

Wald A (1947). *Sequential Analysis*. Wiley, New York.

Mohr J (1951). A search for linkage between the Lutheran blood group and other hereditary characters. *Acta Pathol Microbiol Scand* 28:207-210.

Lawler SD (1954). Family studies showing linkage between elliptocytosis and the Rhesus blood group system. *Proc Int Congr Genet IX, Carologia Suppl*. p1199.

Smith CAB (1954). The separation of the sexes of parents in the detection of human linkage. *Ann Eugen* 18:278-301.

Renwick JH, Lawler SD (1955). Genetical linkage between the ABO and nail-patella loci. *Ann Hum Genet* 19:312-331.

Steinberg AG, Shwachman FH, Allen Jr, Dooley RR (1956). Linkage studies with cystic fibrosis of the pancreas. *Am J Hum Genet* 8:162-172.

Sanger R, Race RR (1958). The Lutheran-secretor linkage in man. *Heredity* 12:513-520.