

# Models of human genetic disease: how biased are the standard formulae?

IAN M. HASTINGS\*

*Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK*

*(Received 5 January 1999 and in revised form 4 March 1999)*

## Summary

Standard theory provides a simple prediction for the frequency of a recessive lethal allele conferring heterozygous protection against an infectious disease (the best-known example being sickle cell protection against malaria). This relationship allows historic disease mortality rates to be estimated. There are, however, hidden biases in this approach. Reproductively active human females in archaic societies normally produce children at intervals of around 4 years. If death of the fetus or young infant (less than around 3 years of age) occurs, then the mother re-enters oestrus and produces another child. This ‘reproductive compensation’ reduces selection against the agent causing early mortality (the recessive allele or infective agent) and biases our estimates of historic mortality rates. The magnitude of these biases is investigated. Re-conception also constitutes a demographic selective pressure acting alongside natural selection: lethal genetic diseases (or tightly linked loci) will be selected to become ever more virulent, killing at ever decreasing ages, to allow the mother to re-enter oestrus and re-conceive a (hopefully unaffected) sibling; this effect also invalidates statistical tests using the number of alleles to distinguish overdominance from drift as explanations for high allele frequency. The same bias affects calculations of mutation/selection balance: for any given mutation rate, syndromes which kill early in life will reach much higher frequencies than those killing at later ages. An intriguing consequence is that lethal recessive disorders in humans will increase in frequency by up to 45% as a consequence of the recent demographic transition to planned family size.

## 1. Introduction

There are two main models invoked to explain the persistence of genetic disease in human populations. The first explanation is that the alleles which cause disease when homozygous, confer some benefit when heterozygous: this is the overdominance model. The second explanation is that the alleles are present at a balance caused by their input by mutation and their removal by natural selection: this is the mutation/selection model. Infectious diseases are one of the major selective forces affecting human evolution (Haldane, 1949) and are known to have caused genetic responses in human populations. The best-known example is probably that of the sickle cell gene. The sickle cell homozygote suffers from lethal

anaemia, the heterozygote is phenotypically normal and apparently protected from fatal malaria, while the homozygote wild-type is phenotypically normal but at risk from developing fatal malaria. The sickle cell allele is therefore maintained in a population at a frequency determined by the certainty (in the absence of medical care) of the sickle cell homozygote dying from anaemia, balanced against the risk of homozygous wild-types contracting fatal malaria.

This is the classic example of overdominance cited in almost all introductory textbooks. However, there are two important implications of this genetic system besides its use as a paradigm. First, its frequency can be used to infer historic malaria mortality rates (e.g. Jones, 1997). Calculations suggest that in many populations 10–15% of human mortality can be ascribed to a single cause: malaria. This is a huge selection pressure, far exceeding those directly observed in extant natural populations (Endler, 1986).

Current address: Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA. Tel: +44 (0)151-708-9393.

A second, more practical use is that estimates of past malaria mortality can serve as a baseline when assessing the efficacy of current anti-malaria public health programmes (e.g. Saul, 1997).

Given the importance of these calculations, it is worthwhile briefly re-considering the assumptions upon which they are made. The calculations are based on a model of discrete, non-overlapping generations: adults contribute gametes to a gene pool and then immediately die; the gametes fuse at random to produce zygotes which, after allowing for mortality due to the genetic disease and infectious agents, develop to form the individuals of the next generation. A cursory consideration of the human reproductive life-cycle reveals biases inherent in this model. Reproduction and subsequent fate of offspring in human populations are not the independent events implicitly assumed in the models. Females in archaic populations generally produce children approximately every 4 years (e.g. Konner & Worthman, 1980; Short, 1994); this occurs because immediately after the birth of a child the mother enters a state of amenorrhoea (suspension of reproductive functions) which lasts for around 3.25 years. Importantly, if the fetus or child less than about age 3.25 years dies, the mother comes back into oestrus and reproduces again; this effect is known as 'reproductive compensation' (e.g. Charlesworth, 1994; Crow & Kimura, 1970, p. 291; Cavalli-Sforza & Bodmer, 1971, p. 191). Put simply, it is not so 'costly' for a female to produce an offspring that dies very young, because she reproduces again shortly afterwards and replaces the dead child.

Assuming sickle cell children die early in life, reproductive compensation reduces selection against the sickle allele so that, for any given malaria mortality rate, its equilibrium frequency will be higher than predicted. Subsequent application of the standard formula will overestimate malaria mortality rates. The converse also applies. If malaria kills early in childhood then the mother re-enters oestrus and a replacement child is conceived. In this scenario the effects of disease mortality on the population are reduced, the frequency of the protective allele falls and we underestimate the mortality rate from allele frequencies. These effects mean we have to consider both the epidemiology of the infectious disease and the pathology of the genetic syndrome when inferring mortality rates. The standard methods are therefore biased, but the magnitude of this bias is unclear. The following discussion explicitly considers the sickle cell allele, but there are several other alleles thought to provide overdominant protection against infectious diseases (see later discussion) and the discussion is equally applicable to them.

It is informative to consider the other main genetic model for the occurrence of recessive lethal disorders in human populations: that they are invariably

deleterious and present in the population at a balance between their input (by mutation) and elimination (by selection). Standard theory describes a relatively simple relationship between equilibrium allele frequency and the forces which determine this frequency, i.e. mutation rate, dominance coefficient and strength of natural selection (see, e.g., Falconer & Mackay, 1996). The latter two factors are usually difficult to quantify but there is one class of allele, recessive lethals, where they are known by definition. This leads to a simple relationship between mutation rate and equilibrium frequency. However, the same bias occurs as in consideration of overdominance: if lethality occurs early in gestation or childhood then reproductive compensation reduces the effective selection pressure against the mutation compared with one acting later in life. Interestingly, this makes the prediction that for equivalent mutation rates, genes whose lethality occurs early in embryonic life (such as cell cycle genes) will have higher frequency of lethals than those expressed later in development. As before, the magnitude of this bias is unknown and will be calculated. In addition, these demographic effects have implications for the evolution of virulence of human genetic disorders and for statistical techniques designed to test whether the alleles may protect against infective disease.

## 2. Methods

### (i) Overdominant, recessive lethal alleles

The standard case, where reproductive compensation is absent, can be described as follows. Represent the protective allele as  $a$ , the wild-type as  $+$ , and death rate due to the infective agent as  $s$ ; thus the fitnesses of genotypes  $aa$ ,  $a+$  and  $++$  are 0, 1 and  $1-s$ , respectively. The fates of genotypes are followed, rather than individual alleles. Following the terminology of Falconer & Mackay (1996),  $H$  is the frequency of  $a+$  adults and  $P$  the frequency of  $++$  adults; note that  $P = 1 - H$  because  $aa$  genotypes die prior to reproduction. Assuming no sex-related effects, the frequencies of genotypes next generation are

$$H' = \frac{H^2(1/2) + 2HP(1/2)}{\bar{W}}, \quad (1a)$$

$$P' = \frac{[H^2(1/4) + 2HP(1/2) + P^2](1-s)}{\bar{W}}, \quad (1b)$$

where  $\bar{W}$ , the mean population fitness, is the sum of numerators. Solving for  $H' = H$  gives  $\hat{H}$ , the equilibrium frequency of the  $a+$  genotype, and  $f$ , the equilibrium frequency of the  $a$  allele, as  $\hat{H}/2$ . This gives the standard solution that

$$f = \frac{s}{1+s}, \quad (2)$$

or alternately that

$$s = \frac{f}{1-f} \tag{3}$$

(Falconer & Mackay, 1996, eqn 2.19; Hartl & Clark 1989, eqn 4.5).

It is possible to investigate one limiting case, where death from the genetic syndrome occurs *in utero* while death from disease occurs after the opportunity for reproductive compensation has passed, as follows. Assume each *aa* zygote produced in an *a+*/*a+* mating immediately dies *in utero* from the genetic syndrome and is immediately replaced by an offspring of another genotype, this new fetus having the same father. Thus we revise the expected output from this type of mating from the predicted (1/4):(1/2):(1/4) proportions of *aa*:*a+*:*++* to (2/3):(1/3) proportions of *a+*:*++* and modify (1) to

$$H' = \frac{H^2(2/3) + 2HP(1/2)}{\bar{W}}, \tag{4a}$$

$$P' = \frac{[H^2(1/3) + 2HP(1/2) + P^2](1-s)}{\bar{W}}, \tag{4b}$$

allowing equilibrium frequencies of the allele to be calculated as

$$f = \frac{1 + 3s - \sqrt{(1 + 6s - 3s^2)}}{4s} \tag{5}$$

or, by re-arrangement, as

$$s = \frac{2f}{3 - 6f + 4f^2}. \tag{6}$$

In the other limiting case, assume that individuals die *in utero* from the infective agent while mortality from the genetic syndrome occurs at later ages when reproductive compensation is not possible. As in (4) we need to recalculate the ratios of offspring from each type of mating. A proportion *s* of the homozygous *++* zygotes die *in utero* of infective disease and are immediately replaced. In *a+* by *a+* matings, the frequency of *++* homozygotes produced is 1/4 so the proportion of zygotes dying is (1/4)*s* = *s*/4. These will immediately be replaced, but of these replacements a further *s*/4 will die, and so on. The total number of offspring produced is therefore increased by a factor of

$$1 + \left(\frac{s}{4}\right) + \left(\frac{s}{4}\right)^2 + \left(\frac{s}{4}\right)^3 + \left(\frac{s}{4}\right)^4 \dots = 4/(4-s).$$

The zygotes at conception consists of 1/2 genotype *a+*, 1/4 of *++*, and 1/4 of the lethal *aa* genotypes, which can be ignored. However, the output at birth is (1/2) *a+* and (1/4)(1-*s*) *++*, because a proportion

(1-*s*) of the latter die *in utero*. Similarly, in *a+* by *aa* matings the total is increased by

$$1 + \left(\frac{s}{2}\right) + \left(\frac{s}{2}\right)^2 + \left(\frac{s}{2}\right)^3 + \left(\frac{s}{2}\right)^4 \dots = 2/(2-s),$$

of which (1/2) offspring are *a+* and (1/2)(1-*s*) are *++*. These values can be substituted into a modified (1) as

$$H' = \frac{H^2[4/(4-s)](1/2) + 2HP[2/(2-s)](1/2)}{\bar{W}},$$

*P'* =

$$\frac{H^2[4/(4-s)](1/4)(1-s) + 2HP[2/(2-s)](1/2)(1-s) + P^2}{\bar{W}}$$

and solved for *H'* = *H* to obtain the equilibrium allele frequency as

$$f = \frac{2 - \sqrt{(4 - 8s + 6s^2)}}{2(2-s)} \tag{7}$$

(dropping a negligible factor in *s*<sup>3</sup>).

(ii) *Mutation/selection balance*

As before, *H* is the frequency of the heterozygote and *P* that of the wild-type homozygote. In the conventional case we obtain

$$H' = \frac{H^2(1/2) + 2HP(1/2)}{\bar{W}}$$

and

$$P' = \frac{H^2(1/4) + 2HP(1/2) + P^2}{\bar{W}}.$$

Mutation occurs at rate *μ* from wild-type to mutant while back-mutation to wild-type is assumed to be negligible. The relative frequencies after mutation are

$$H'' = \frac{H'(1-\mu) + P'2\mu(1-\mu)}{\bar{W}}$$

and

$$P'' = \frac{P'(1-\mu)^2}{\bar{W}}.$$

Solving for *H''* = *H* gives

$$\hat{f} = \frac{\sqrt{\mu - \mu^2}}{1 - \mu}. \tag{8}$$

Assuming *μ* is sufficiently small that *μ* ≪ 1 and *μ* ≪ √*μ* gives the standard result for a fully recessive allele that *f* = √*μ*, or alternately,

$$\mu = \hat{f}^2. \tag{9}$$

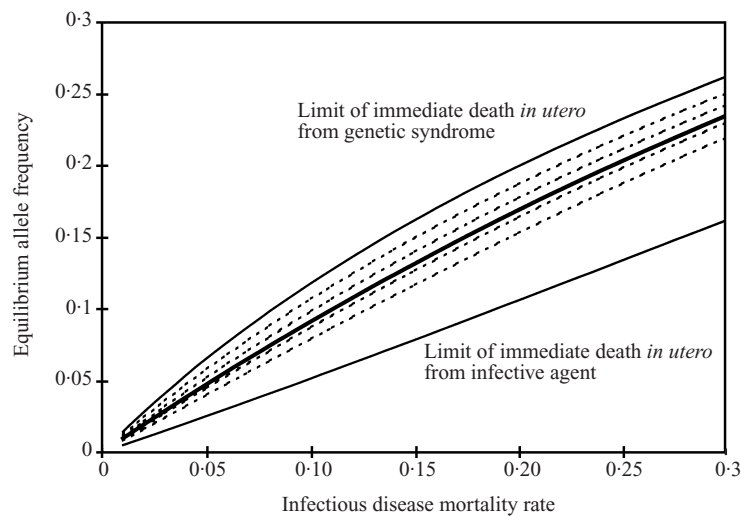


Fig. 1. The effects of human demography and parasite epidemiology on the equilibrium frequencies of an allele providing complete heterozygote protection against an infective disease; the protective allele is assumed to be recessive lethal. The  $x$ -axis represents the disease mortality which would occur in the absence of this genetic protection. The central continuous line is the equilibrium frequency predicted by standard theory (equation 2). The upper and lower continuous lines represent the limits obtained assuming death occurs immediately following conception and that immediate reproductive compensation occurs (see text for explanation). The upper continuous line assumes that death at conception is due to the genetic syndrome and death due to infective disease occurs at an age after the opportunity for reproductive compensation (equation 5). The lower continuous line assumes that death at conception is caused by the infective agent and death from the genetic syndrome occurs after the opportunity for reproductive compensation (equation 7). Computer simulations were used to generate the more plausible predictions represented by the four dotted lines. From top to bottom these are: (i) The ‘ovalocytosis’ model, which assumes the fetus dies as a result of the genetic syndrome during the first month *in utero* while deaths from infection by the pathogen occur after age 3.25 years. (ii) The ‘sickle cell’ model, in which death from the genetic syndrome occurs at age 1.3 years and deaths from infection occurs after age 3.25 years. (iii) A model in which death from the infective agent occurs at age 2 while death from the genetic syndrome occurs after age 3.25 years. (iv) A model in which death from the infective agent occurs during the first year of life and death from the genetic syndrome occurs after age 3.25 years.

In the case of full reproductive compensation we obtain

$$H' = \frac{H^2(2/3) + 2HP(1/2)}{\bar{W}}, \quad (10a)$$

$$P' = \frac{H^2(1/3) + 2HP(1/2) + P^2}{\bar{W}}, \quad (10b)$$

as before (equation 4; noting that  $s = 0$  as infectious disease is absent). Mutations then occur and  $f$  is obtained as a function of  $\mu$ . The exact solution for  $\hat{f}$  is somewhat complex but can easily be solved numerically. Alternately, assuming  $\mu^2 \ll 1$  and  $f^2 \ll 1$  gives a good approximation as

$$\hat{f} = \frac{\sqrt{24\mu - 9u}}{4} \quad (11)$$

or

$$\mu = \frac{2\hat{f}^2}{3(1 - 3\hat{f})}. \quad (12)$$

### (iii) Computer simulations

Equations (5) and (7) provide limits for the magnitude of the bias inherent in the standard analysis. More

realistic biological systems were investigated by simulation. The salient points of these simulations are: individuals reach puberty and enter the breeding population at age 20 years; individuals die (or females reach menopause) at a rate 1/20 per year giving a mean reproductive lifespan of 20 years. Females re-enter oestrus 4 years after conception of a surviving baby (9 months gestation and 3.25 years of nursing) and take 0.3 years to re-conceive: hence children are spaced at intervals of 4.3 years. If the child of a gestating/nursing female dies, the mother re-enters the breeding population and takes the normal 0.3 years to re-conceive. The simulation was tested thoroughly and, in particular, when run to equilibrium: (i) gives the same results as those obtained analytically at the demographic limits of no, or complete, reproductive compensation (equations 2 and 5); and (ii) gives the standard results (equation 2) when malarial and genetic disorder deaths both occur after weaning, i.e. after age 3.25 years or when both occur at exactly the same age and so have the same demographic effect.

Two models will be investigated, based loosely on alleles conferring protection against malaria. The first is a model where children die at age 1.25 years. For convenience, we will refer to this as the ‘sickle cell’

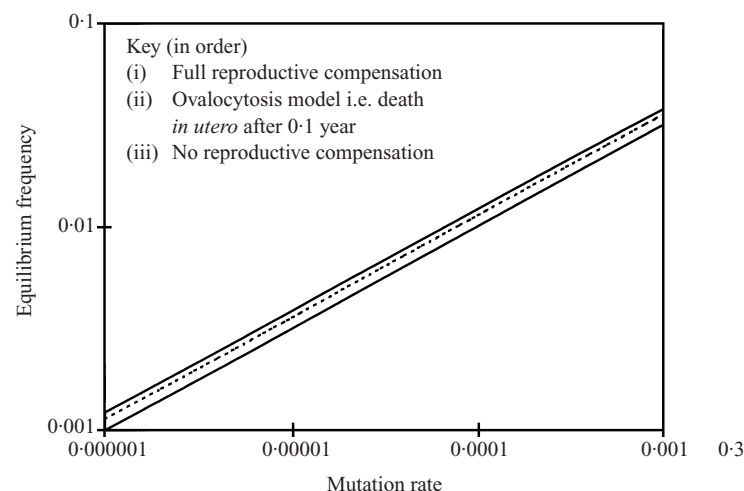


Fig. 2. Equilibrium frequencies of recessive lethal alleles under mutation/selection balance. The lower continuous line is that predicted from standard methods (equation 8), and the upper continuous line is that predicted assuming complete reproductive compensation (equation 11). The dotted line is that obtained from an ovalocytosis model (see caption to Fig. 1).

model as it is based on a syndrome where children are fatally affected during the first year of life (although in practice the onset and prognosis of sickle cell anaemia are much more complex (Weatherall *et al.*, 1989), at least in modern health care conditions). The second is an ‘ovalocytosis’ model where affected offspring die *in utero* after 0.1 years (i.e. 5 weeks). Malaria deaths in both models were assumed to occur in children after age 3.25 years. This may not always be the case in areas of high transmission, where most mortality may occur in young children. The effects of pathogen epidemiology can be investigated by allowing death from the infective agent to occur at variable ages while death from the genetic syndrome occurs after weaning.

### 3. Results

The effects of demographic factors on equilibrium allele frequency under a model of overdominance are shown in Fig. 1. Continuous lines represent the analytical results and those obtained by simulation are represented by dotted lines. The estimates for allele frequency under mutation/selection balance are shown in Fig. 2. The lower line is the balance predicted by the standard model (equation 8), the upper is that predicted by full reproductive compensation (equation 11), while the dotted line shows results obtained under an ovalocytosis model (in which malaria is absent so ovalocytosis is a recessive lethal syndrome maintained purely by mutation/selection).

### 4. Discussion

Most overdominant genes have been identified by their deleterious effects when homozygous, notably those responsible for blood pathologies such as sickle

cell anaemia and thalassaemias, which protect against malaria when heterozygous (e.g. Weatherall *et al.*, 1989; Weatherall, 1996). Among Jewish populations there are five syndromes of lysosomal storage disorders which are thought to protect against respiratory infections: Tay–Sachs, Gaucher, Niemann–Pick, mucopolipidosis type IV, Hunters (Rotter & Diamond, 1987; Bach *et al.*, 1992). These are easily detected through their deleterious effects in children or adults, but other homozygous pathologies which kill *in utero* cannot, by definition, be observed directly and remain undetected in the population. Progress in understanding the molecular basis of infection has allowed us to identify such protective alleles and also to infer their lethality through a lack of homozygous neonates. One such example is ovalocytosis, which appears to protect against severe malaria: the lack of observed homozygotes leads to the conclusion that it is a recessive lethal, death occurring *in utero* (Liu *et al.*, 1994; Genton *et al.*, 1995; Mgone *et al.*, 1996). As molecular technologies become more sophisticated, it seems likely that other such alleles will be identified as giving heterozygote protection against infectious diseases while being homozygous lethal *in utero*. The development of effective genome scanning techniques (Hill, 1996; Hill & Motulsky, 1998) may go some way to rectifying this ascertainment bias. Indeed the question as to how much human genetic variation is explicable by overdominant protection against disease (infective or uninjective) will be one of the more interesting questions addressed by the application of this technology.

The demographic factors used for the simulations are based on human data obtained from populations of !Kung (Kalahari bushmen) where mean interbirth interval is around 4 years, weaning of surviving children occurring at around age 3.5 years while the

mother gestates a sibling (Konner & Worthman, 1980; Short, 1994; references therein). This population is generally taken as a paradigm of an archaic hunter-gatherer society, so their demography appears to provide the most appropriate parameters for the simulation. Other parameter values have been investigated (re-entry into oestrus when the current offspring is 2 or 5 years old) but this results in only minor quantitative differences (results not shown). The simulations suggest that moderately large differences in estimated disease mortality rates occur between different models. An allele frequency of 15% (a rough figure for sickle allele frequency in highly endemic areas; Weatherall *et al.*, 1989) corresponds to a standard estimate of an 18% mortality rate (equation 3), while under an ovalocytosis model (rapid death *in utero*) it falls to 15% and under a model of immediate post-natal mortality from infection (Fig. 1) it rises to 20%. For an allele frequency of 8% (a rough figure for the sickle allele in west Africa (Jones, 1997) and for the ovalocytosis allele in Papua New Guinea (Genton *et al.*, 1995)), the equivalent figures are 8.7%, 7.1% and 10.2%, while for allele frequencies of 2% (as in the allele responsible for Tay–Sachs disease in Ashkenazi Jews; Motulsky, 1995) the equivalent figures are 2.0%, 1.6% and 2.5%. Thus the biases appear to be potentially large, at least in mathematical terms. In practice, the errors associated with field estimates of allele frequencies and assumptions such as their frequencies being at equilibrium and that the human population is large and randomly mating, means that the standard estimates at least give a good approximation of the likely mortality rates in historic times.

The results presented in Fig. 1 allow correction factors to be estimated for prevailing epidemiology, and equations (5) and (7) enable the estimates to be given within algebraic limits. Inspection of these equations shows that the bias will be at a maximum of one-third at low values of allele frequency, the magnitude decreasing as allele frequency increases. The results also emphasise that demographic and epidemiological data (e.g. mean age of death from malaria) need to be presented when drawing inferences about past mortality rates. The effects of parasite epidemiology can cause problems when attempting to infer mortality rates in different geographic regions or in social conditions which differ in their intensity of disease transmission. In areas of intense transmission, first infection, and presumably death from the infective agent, occurs at a lower age than in areas of less intense transmission. Such differences can be observed in extant populations (e.g. Anderson & May, 1992, table 3.3) and may bias our estimates of mortality rates (Fig. 1). Similarly, the gross epidemiology of the disease may also affect our estimates of mortality. In endemic regions the diseases may predominantly infect

and kill infants, while the same disease occurring as epidemics may infect all age groups and so predominantly kill non-infants.

The models of immediate death *in utero* followed by immediate re-conception are, of course, physiologically impossible. They were developed to provide simple algebraic expressions for the maximum bias attributable to reproductive compensation. There are, however, important biological situations in which these models exactly describe natural systems. Females of many species produce eggs or broods within a patch of resource capable of sustaining only a proportion of them, the ‘surplus’ dying from causes such as starvation or cannibalism. A hypothetical example would be a *Drosophila* laying 200 eggs in a location capable of supporting only 100 through to adult stage. More concrete examples are bird species which lay more eggs than can be fed, the younger and/or weaker chicks progressively starving to death. In these situations the individuals in the brood dying of the genetic syndrome or infectious disease are essentially replaced by siblings who would otherwise have died; this is exactly the process described by the models used to investigate immediate death and replacement *in utero*. Care must be taken in estimating disease mortality from overdominant loci in such species. For example, Fig. 1 reveals that if the genetic disorder kills in the nest while infectious disease kills after fledging then a 5% allele frequency corresponds to a 4% disease mortality rate, while the converse (infectious death in the nest, post-fledging death by genetic disorder) corresponds to a 10% disease mortality rate.

Exactly this effect may occur in humans. Many human societies have recently started to limit overall family size either as a consequence of governmental pressure (such as China’s policy of limiting family size) or through economic considerations (Western societies). This corresponds exactly to the paradigm of immediate replacement *in utero* as dead offspring are replaced by living siblings. The correspondence becomes even closer in situations where genetic technology allows affected fetuses to be identified *in utero* and the pregnancy terminated (see, e.g. Milunsky, 1998). There is a curious consequence of this demographic transition for recessive lethal genetic disorders maintained by mutation/selection balance. Historically, those disorders causing death after early childhood were unaffected by reproductive compensation, and have allele frequencies predicted by the standard equation. The demographic transition to fixed family size means that reproductive compensation does affect these syndromes, selection against them effectively falls, and their allele frequencies will start to rise. The results given by equations (8) and (11) (illustrated in Fig. 2) suggest that a 20% increase in frequency will occur. The syndrome is recessive so

a 20% increase in allele frequency results in a 44% increase in the frequency of the genetic disorder. The speed at which the deleterious alleles increase depends on the underlying mutation rate.

As a concrete example, consider spinal muscular atrophy type 1 (Werdnig–Hoffmann disease). The frequency of the syndrome is about 1 in 10000, allele frequency is about 1%, prenatal diagnosis is possible and death normally occurs in early infancy (Conner & Ferguson-Smith, 1997); these figures imply an underlying mutation rate of around  $10^{-4}$ . Application of equation (10) reveals that the rate of increase in Werdnig–Hoffmann disease immediately following the human demographic transition is likely to be in the order of 0.6% per generation. Syndromes caused by genes with lower mutation rates will show correspondingly lower rates of increase (0.2% per generation for a mutation rate of  $10^{-5}$ ), although they will eventually still reach an equilibrium 44% higher than original.

Falconer & Mackay (1996, p. 37) and Crow (1997) have recently speculated that decreasing selection pressures in modern societies will allow the frequency of genetic disorders to rise. The intriguing suggestion here is that even if selection has not fallen (the disorders remain lethal), then their frequency may still rise substantially purely as a result of demographic changes in human societies. There was a period of interest in reproductive compensation in the 1940s and 1950s as a possible factor in the dynamics of the newly discovered Rhesus blood factor. A review of this work can be found in Cavalli-Sforza & Bodmer (1971, p. 191); interestingly R. A. Fisher was aware of the problem (cited in Race, 1944), although it appears not to have been a factor in his later works, which is somewhat ironic given his interest in human society and eugenics (Fisher, 1958).

The results also have a bearing on considerations of mutational load. Each lethal mutation that enters a population must ultimately be eliminated through the death of an individual bearing it; the proportion of individuals dying for this reason constitutes a mutational load. It is usually implicitly assumed that these ‘individuals’ are fully independent (i.e. post-weaning) children or adults. However, it appears that up to 50% of all successful human fertilisations are lost within 9 weeks (Sadler, 1995). Assuming these losses are due to genetic (as well as congenic) reasons then the load attributable to lethal mutations may be greatly reduced. If children are spaced at 4 year (46 month) intervals in archaic human societies, and a lost foetus is replaced within 5 months (2 month gestation for lethality to occur, 3 months for the mother to re-enter oestrus and become pregnant) then the mutation is eliminated at a cost of 5/46 or approximately 0.1 of an ‘individual’. This is a time cost; energetic costs of developing then resorbing an

embryo are presumably much less and may even be negligible. In current Western human societies with planned family sizes, each loss is replaced so even the time loss is essentially zero. This argument applies only to mutations which are lethal in early development (presumably including genes encoding most cell cycle and housekeeping enzymes), those acting later in life are manifest as genetic diseases resulting in the death of an individual. Most human differentiation is completed within 10 weeks of gestation (all major organs are present and functional) and the consequences of this early development as a ‘test’ for functional tissue-specific genes will have constituted a potent selection pressure for rapid differentiation of human embryos. This selection pressure certainly exists, but whether it has exerted a significant effect given all the other constraints and selection pressures acting on development, remains a matter of conjecture.

The Ashkenazi Jewish population appears to have high frequencies of several recessive deleterious alleles, one of the better known being that responsible for Tay–Sachs disease (Motulsky, 1995). These alleles also appear to occur within the same populations; four of the lysosomal storage disorders appear within the same populations of the Ashkenazim (Motulsky, 1995). There has been much speculation on the underlying reason for these high frequencies. One hypothesis is that these alleles provided protection against infectious disease; in the case of Tay–Sachs disease this may have been tuberculosis (Rotter & Diamond, 1987). An alternative hypothesis is that they reached high frequencies as a result of chance founder effects during the establishment and subsequent population bottlenecks experienced by the Ashkenazi population. One objective method of distinguishing between these hypotheses is to identify the number of allelic variants present in the population. If there are a large number of variants of independent origin it suggests that they have accumulated in the population as a consequence of their ability to protect against infective disease. Alternatively, if there is predominantly a single allelic variant it is indicative of founder effects (see discussion in Motulsky, 1995).

These calculations are based on the premise that all allelic variants provide the same level of protection against disease and have the same effect (lethality when homozygous) in the human; thus they are selectively equivalent and their exact number is a balance between input by mutation and loss by random genetic drift. The results shown above indicate that it is not enough to assume that these variants are selectively equivalent but that they must also be *demographically* equivalent. The reason for this assertion is that demographic considerations mean that the most virulent (i.e. the one causing death at the earliest age) will come to dominate the population.

This occurs because, in many genetic syndromes, a homozygous fetus/child is certain to die anyway so the faster the allele kills the fetus/child, the faster the mother will come back into oestrus, re-conceive and produce a sibling of the fetus/child. Since the sibling may carry a copy of the allelic variant in the heterozygous state, this provides the selective/demographic pressure ensuring its spread through the population (the process is a variant of kin selection theory and a similar argument was invoked by Charlesworth (1994) to explain the lethality of the mouse *t* haplotype). We would therefore expect the protective allele to evolve to kill earlier and earlier in gestation; in which case, as argued above, it no longer gives rise to an obvious infant phenotype and hence its presence is unlikely to be detected. Even if it cannot kill directly *in utero*, it may still evolve to do so if it can become tightly linked (perhaps through an inversion) to a recessive lethal that does act early in gestation. This contrasts with the more conventional view that alleles (and modifiers at other loci) would evolve to reduce the damage done to the individual. Another unfortunate conclusion from this line of argument is that we cannot formally cite the predominance of a single allelic variant as evidence that high allele frequencies are due to population bottlenecks (see above); this may indeed be the case but it may also have reached high frequencies by displacing other, less virulent, variants.

I thank Bruce Wedgwood-Oppenheim and Mark Kirkpatrick for comments on an earlier version of the manuscript, and two anonymous referees for comments on the submitted version. This work was supported by the Medical Research Council.

## References

- Anderson, R. M. & May, R. M. (1992). *Infectious Diseases of Humans*. Oxford: Oxford University Press.
- Bach, G., Zlotogora, J. & Ziegler, M. (1992). Lysosomal storage disorders among Jews. In *Genetic Diversity Among Jews: Diseases and Markers at the DNA Level* (ed. B. Bonne-Tamir & A. Adam), pp. 301–304. Oxford: Oxford University Press.
- Cavalli-Sforza, L. L. & Bodmer, W. F. (1971). *The Genetics of Human Populations*. San Francisco: W.H. Freeman.
- Charlesworth, B. (1994). The evolution of lethals in the *t*-haplotype system of the mouse. *Proceedings of the Royal Society of London, Series B* **258**, 101–107.
- Conner, M. & Ferguson-Smith, M. (1997). *Essential Medical Genetics*, 5th edn. Oxford: Blackwell Science.
- Crow, J. F. (1997). The high spontaneous mutation rate: Is it a health risk? *Proceedings of the National Academy of Sciences of the USA* **94**, 8380–8386.
- Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper & Row.

- Endler, J. A. (1986). *Natural Selection in the Wild*. New Jersey: Princeton University Press.
- Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th edn. Harlow: Longman.
- Fisher, R. A. (1958). *The Genetical Theory of Natural Selection*, 2nd edn. New York: Dover Publications.
- Genton, B., Alyaman, F., Mgone, C. S., Alexander, N., Paniu, M. M., Alpers, M. P. & Mokela, D. (1995). Ovalocytosis and cerebral malaria. *Nature* **378**, 564–565.
- Haldane, J. B. S. (1949). Disease and evolution. *La Ricerca Scientifica* **19** (Supplement), 68–76.
- Hartl, D. L. & Clarke, A. G. (1989). *Principles of Population Genetics*, 2nd edn. Sunderland, MA: Sinauer Associates.
- Hill, A. V. S. (1996). Genetic susceptibility to malaria and other infectious diseases: from the MHC to the whole genome. *Parasitology* **112**, S75–S84.
- Hill, A. V. S. & Motulsky, A. G. (1998). Natural selection for disease susceptibility and resistance genes: examples and prospects. In *Evolution in Health and Disease* (ed. S. C. Stearns). Oxford: Oxford University Press.
- Jones, T. R. (1997). Quantitative aspects of the relationship between the sickle-cell gene and malaria. *Parasitology Today* **13**, 107–111.
- Konner, M. & Worthman, C. (1980). Nursing frequency, gonadal function and birth spacing among !Kung hunter-gatherers. *Science* **207**, 788–791.
- Liu, S. C., Jarolim, P., Rubin, H. L., Palek, J., Amato, D., Hassan, K. & Zaik, M. (1994). The homozygous state for the band-3 protein mutation in southeast-asian ovalocytosis may be lethal. *Blood* **84**, 3590–3591.
- Mgone, C. S., Koki, G., Paniu, M. M., Kono, J., Bhatia, K. K., Genton, B., Alexander, N. & Alpers, M. P. (1996). Occurrence of the erythrocyte band-3 (AE1) gene deletion in relation to malaria endemicity in Papua New Guinea. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **90**, 228–231.
- Milunsky, A. (1998). *Genetic Disorders of the Fetus: Diagnosis, Prevention and Treatment*, 4th edn. Baltimore: Johns Hopkins University Press.
- Motulsky, A. G. (1995). Jewish diseases and origins. *Nature Genetics* **9**, 99–101.
- Race, R. R. (1944). Some recent observations on the inheritance of blood groups. *British Medical Bulletin* **2**, 165.
- Rotter, J. I. & Diamond, J. M. (1987). What maintains the frequencies of human genetic diseases? *Nature* **329**, 289–290.
- Sadler, T. W. (1995). *Longman's Medical Embryology*, 7th edn. Baltimore: Williams & Wilkins.
- Saul, A. (1997). Malaria mortality and transmission rates in Africa. *Parasitology Today* **13**, 318–319.
- Short, R. V. (1994). Human reproduction in an evolutionary context. *Annals of the New York Academy of Sciences* **709**, 416–425.
- Weatherall, D. J. (1996). Host genetics and infectious disease. *Parasitology* **112**, S23–S29.
- Weatherall, D. J., Clegg, J. B., Higgs, D. R. & Wood, W. G. (1989). The hemoglobinopathies. In *The Metabolic Basis of Inherited Disease* (ed. C. R. Scriver, A. L. Beaudet, W. S. Sly & D. Valle), pp. 2281–2339. New York: McGraw-Hill.